Department of Physics and Astronomy
Heidelberg University

Bachelor Thesis in Physics
submitted by

Hannes Trescher

born in Freiburg (Germany)

2024

Measurement of the B_s^0 oscillation frequency Δm_s in the $B_s^0 \to D_s^- \pi^+$ decay with 2024 data

This Bachelor Thesis has been carried out by Hannes Trescher at the
Physikalisches Institut in Heidelberg
under the supervision of
Prof. Dr. Stephanie Hansmann-Menzemer

Abstract

This thesis provides an analysis of the oscillation frequency Δm_s of the B_s^0-$\overline{B}_s^0$ system. A dataset collected in 2024 using proton-proton collisions at the LHCb detector with an integrated luminosity equivalent to 1.73 fb^{-1} is used. The analysis is performed using the $B_s^0 \to D_s^-\,(K^+K^-\pi^-)\pi^+$ decay channel. Flavour tagging algorithms are used to distinguish between mixed and unmixed signal candidates. The oscillation frequency Δm_s is determined with a decay time fit accounting for effects of the decay time resolution and the calibrated mistag probability. The oscillation frequency of the B_s^0-$\overline{B}_s^0$ system is measured to a value of $\Delta m_s = (17.792 \pm 0.011)\,\mathrm{ps}^{-1}$. This value agrees with the current world average within its uncertainties.

Kurzfassung

Diese Bachelorarbeit umfasst die Messung der Oszillationsfrequenz Δm_s des B_s^0-$\overline{B}_s^0$ Systems anhand des $B_s^0 \to D_s^-\,(K^+K^-\pi^-)\pi^+$ Zerfallskanals. Dafür wurde ein aktueller Datensatz von Proton-Proton Kollisionen am LHCb Detektor aus 2024 aufgenommen, welcher einer integrierten Luminosität von 1,73 fb^{-1} entspricht. Um oszillierte und nicht-oszillierte Mesonen zu unterscheiden werden Algorithmen eingesetzt, die den anfänglichen Zustand der B_s^0 Mesonen bestimmen. Die Oszillationsfrequenz wird durch das Anpassen einer Wahrscheinlichkeitsdichtefunktion an die Verteilung der Zerfallszeit des B_s^0 mesons gemessen. Dabei werden Effekte der Zerfallszeitauflösung und die Wahrscheinlichkeit für eine falsche Bestimmung des anfänglichen Zustandes berücksichtigt. Die Oszillationsfrequenz ergibt sich zu einem Wert von $\Delta m_s = (17,792 \pm 0,011)\,\mathrm{ps}^{-1}$, welcher innerhalb der Fehlergrenzen mit dem weltweiten Durchschnittswert übereinstimmt.

Contents

1 Introduction

The Standard Model of particle physics is able to describe the fundamental particles of our universe and their interaction via three out of the four fundamental forces [1]. It is the most complete theory to date combining the electromagnetic, the weak and the strong interaction. But there are limits of the Standard Model in terms of observations seen in nature that cannot be described by this theory. An example for this is the particle anti-particle asymmetry in our observed universe. This asymmetry is deeply connected to CP-violation and although the Standard Model incorporates CP-violation to some extend through the Cabbibo-Kobayashi-Maskawa mechanism [2,3], it is not able account for the extend of the asymmetry in matter. This motivates the search for physics beyond the Standard Model. CP-violation was fist observed in the decay of neutral K mesons [4] in 1964, while modern particle physics searches for CP-violation primarily in the decays of the B mesons. The LHCb experiment is a dedicated b- and c-physics experiment probing the Standard Model with measurements of CP-violation using particles containing b-quarks and c-quarks. An important observable in terms of CP-violation is the neutral B_s^0 meson oscillation frequency Δm_s. This quantity is needed in measurements of CP-violation to reduce systematic uncertainties in the B_s^0-$\overline{B}_s^0$ system. Furthermore the B_s^0 oscillation frequency Δm_s combined with the B^0 oscillation frequency Δm_d provides a constraint on the Cabbibo-Kobayashi-Maskawa [2] matrix. Measurements of the oscillation frequency Δm_s have been carried out for many years. The first measurement was executed with the collider detector at Fermilab in 2006 [5] and were followed by Run1 and Run2 measurements of the LHCb detector [6–10]. A recent software and hardware upgrade of the LHCb detector gives opportunity to push the precision frontier in this measurement. This thesis aims to be a proof of concept working with the newly commissioned LHCb detector. The analysis presented in this thesis uses this newly gathered data and performs a measurement of the oscillation frequency Δm_s.

2 Theory

2.1 The Standard Model of particle physics

The Standard Model of particle physics (SM) is the most complete and well tested theory describing the fundamental particles of our universe as well as the fundamental interactions between these particles. In the SM [1] three of the four known fundamental forces are described, namely the electromagnetic force, the weak and the strong force.

2.1.1 Particles of the Standard Model

The particles of the Standard Model can be split up into groups based on specific properties of the particles. The particles that make up the matter in our universe are called fermions. Fermions are particles having a half-integer spin and can be further distinguished by looking at the interaction they are sensitive to.
Only six of the twelve fermions can interact through the strong force, they are called quarks. There are two types of quarks called up-type quarks with a electrical charge of $+\frac{2}{3}e$ and down type quarks with a $-\frac{1}{3}e$ charge, with e being the elementary charge. Quarks can not be observed as individual particles, since they are confined to be bound in colourless states by the strong force. The colourless states of particles are called hadrons and they are typically combinations of a quark with an anti-quark ($q\bar{q}$) called mesons or combinations of three quarks (qqq) or three anti-quarks ($\bar{q}\bar{q}\bar{q}$) called baryons or anti-baryons respectively.
The remaining six fermions are called leptons. Leptons have either integer electrical charge or are electrically neutral. The neutral particles are called neutrinos and do not interact with the electromagnetic force, which makes them only sensitive to weak interactions. The remaining three leptons, the electron, the muon and the tauon are able to interact with the electromagnetic as well as the weak force.
The fermions can be grouped into three generations, with the first generation containing the lightest and most stable fermions, which make up the ordinary matter in our universe. Each generation of quarks contains an up-type and a down-type quark. Each lepton generation contains a neutrino as well as a charged lepton. Each particle is associated to an anti-particle with the same physical properties except for an inverse the electrical charge. An overview of the particles is shown in Figure 1.

2.1.2 Fundamental forces of the Standard Model

In the SM the interaction of the fundamental forces happens as an exchange of particles called bosons, which are integer spin particles. Each fundamental interaction is described by a quantum field theory. The electromagnetic force is described by the Quantum Electrodynamics (QED) with the electrically neutral and massless photon γ as the mediator. The strong force, described by the Quantum Chromodynamics (QCD), is mediated through the massless gluon g. The weak force is mediated by the charged massive $W^{\pm}$ boson and the massive neutral Z^{0} boson. The Higgs boson is a quite special boson, since it is the only particle with a spin of zero. The particles of the SM get their masses by interacting with the Higgs boson.
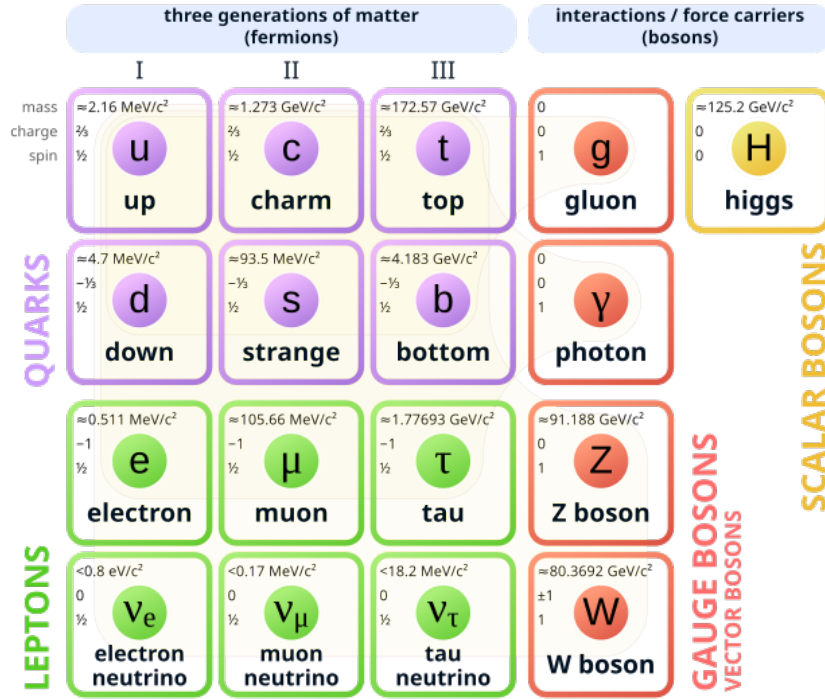
Figure 1: Summary of fundamental particles of the Standard Model of particle physics. Taken from [11].

2.1.3 Weak charged-currents

All fermions participate in the weak interaction. In the weak interaction it is possible to change the flavour of the fermion, where flavour refers to the specific type of the fermion. Only processes involving the electrically charged $W^{\pm}$ gauge bosons allow the change of flavour. Up-type quarks can change into down-type quarks and vice versa, since the quark types differ by one elementary charge. The process of changing the flavour is important for this analysis, because it forms the basis of the mixing process of neutral B_s^0 mesons, described by the Feynman box diagrams in Figure 2. The strength of the interaction coupling is described by the components of the unitary 3×3 Cabbibo-Kobayashi-Maskawa (CKM) matrix.

$$\mathrm{CKM} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \tag{1}$$

The CKM matrix plays an important role in the SM, since it incorporates charge-parity-violation (CP-violation) into the SM. The charge symmetry describes the invariance of a process after applying charge conjugation, which turns a particle into its anti-particle. Parity describes the process of reversing spacial variables. The strong and electromagnetic interaction conserve the CP-symmetry, whereas processes in the weak interaction violate this symmetry. Given the unitary of the CKM matrix it can be parameterised by using four parameters, which are three mixing angles and a complex phase. This complex phase accounts for CP-violation in the CKM mechanism. The measurement of Δm_s puts a
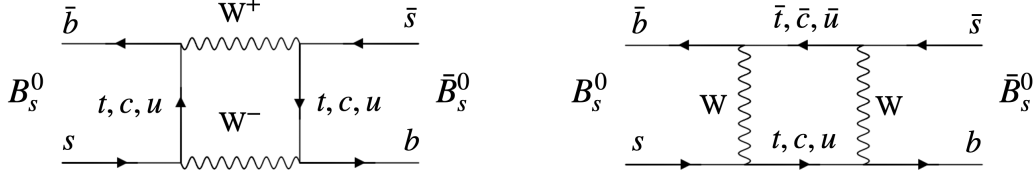
Figure 2: Dominant box diagrams describing the mixing process.

constraint on the amount of CP-violation described with CKM mechanism.

2.2 Mixing theory of the B_s^0 meson

In the absence of mixing in the B_s^0- $\overline{B}_s^0$ system, the time evolution of the neutral B_s^0-meson, with the constraint of $|B_s^0(t=0)\rangle = |B_s^0\rangle$ is described by the following equation:

$$|B_s^0(t)\rangle = e^{-\Gamma t/2}e^{-imt}|B_s^0\rangle. \tag{2}$$

with the total decay with Γ, which accounts for the exponential decay of the wave function due to the decay of the unstable B_s^0 meson. Equation 2 is a solution of the *effective* Schrödinger equation:

$$i\frac{\partial}{\partial t}|B_s^0(t)\rangle = (m - \frac{i}{2}\Gamma)|B_s^0(t)\rangle = \mathcal{H}|B_s^0(t)\rangle. \tag{3}$$

Considering the B_s^0-$\overline{B}_s^0$ mixing system, the *effective* Hamiltonian becomes:

$$\mathcal{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} - \frac{i}{2}\begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}. \tag{4}$$

The diagonal elements of the mass matrix $\mathbf{M}$ are the masses of the flavour eigenstates. The off-diagonal elements represent the contributions of the mixing process.

Assuming CPT invariance of the SM, the diagonal elements of both the mass matrix $\mathbf{M}$ and the decay matrix $\mathbf{\Gamma}$ need to be equal, hence $\Gamma_{11} = \Gamma_{22} = \Gamma$ and $M_{11} = M_{22} = M$. The Hamiltonian therefore can be written as [12]:

$$\mathcal{H} = \begin{bmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M - \frac{i}{2}\Gamma \end{bmatrix}. \tag{5}$$

The diagonalisation of the Hamiltonian yields the following equation:

$$i\frac{\partial}{\partial t}\begin{bmatrix} |B_L\rangle \\ |B_H\rangle \end{bmatrix} = \begin{bmatrix} \lambda_L & 0 \\ 0 & \lambda_H \end{bmatrix}\begin{bmatrix} |B_L\rangle \\ |B_H\rangle \end{bmatrix}, \tag{6}$$

with the eigenvalues $\lambda_{H,L} = m_{H,L} - \frac{i}{2}\Gamma_{H,L}$ and the eigenstates:

$$|B_H\rangle = p|B_s^0\rangle - q|\overline{B}_s^0\rangle \quad \text{and} \quad |B_L\rangle = p|B_s^0\rangle + q|\overline{B}_s^0\rangle, \tag{7}$$

which can be interpreted as the heavy and light mass eigenstates of the B_s^0-$\overline{B}_s^0$ system. The amplitudes p and q fulfill the normalisation $|p|^2 + |q|^2 = 1$. The eigenvalues are composed of the masses:

$$m_{H,L} = M \pm \text{Re}\left[\sqrt{\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\left(M_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}\right] \tag{8}$$

4

and decay widths:

$$\Gamma_{H,L} = \Gamma \mp \mathrm{Im}\left[\sqrt{\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\left(M_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}\right]. \tag{9}$$

Assuming CP invariance of the mixing process $|q/p| = 1$, the time evolution of pure $|B_s^0\rangle$ and $|\bar{B}_s^0\rangle$ states at $t = 0$ can be obtained using a basis transformation which yields the following result:

$$|B_s^0(t)\rangle = g_+(t)|B_s^0\rangle + \frac{p}{q}g_-(t)|\bar{B}_s^0\rangle \qquad |\bar{B}_s^0(t)\rangle = g_+(t)|\bar{B}_s^0\rangle + \frac{p}{q}g_-(t)|B_s^0\rangle, \tag{10}$$

with the following equation for the time dependent amplitude:

$$|g_\pm(t)|^2 = \frac{e^{-\Gamma_s t}}{2}\left[\cosh\left(\frac{\Delta\Gamma_s}{2}t\right) \pm \cos(\Delta m_s t)\right]. \tag{11}$$

The time evolution of the flavour eigenstates is a superposition of both the particle and the anti-particle state. Over time the B_s^0 meson can oscillate into the $\bar{B}_s^0$ meson state and vice versa. This oscillation is described by the time dependent amplitudes $g_\pm(t)$. The above stated equation for the amplitude uses the following definition:

$$\Delta m_s = m_H - m_L \qquad \Delta\Gamma_s = \Gamma_L - \Gamma_H, \tag{12}$$

where Δm_s is the mass difference between the heavy and the light mass eigenstates and $\Delta\Gamma_s$ is the difference in decay width of the two eigenstates. The amplitudes $g_\pm(t)$ are used in this analysis for the multidimensional fit to measure the Δm_s oscillation frequency.

3 The LHCb experiment

The LHCb experiment is one of the four large experiments at the Large Hadron Collider LHC. The LHC is located near Geneva at the border of Switzerland and France and is part of CERN, the European Organisation for Nuclear Research. The LHC is between 50 meters and 170 meters below the ground and has a circumference of 26.7 kilometers. Inside of the LHC the two counter-rotating hadron beam are accelerated and made collide at four interaction points, at which the four large experiments ATLAS, CMS, ALICE and LHCb are located. At the LHC a center-of-mass energy of $\sqrt{s} = 14$ TeV for proton-proton collisions (pp collisions) can be achieved.

3.1 The LHCb detector

The LHCb experiment set its focus on the decay of the heavy b- and c-quarks. These quarks, produced at the pp collision point, have high momentum in the directions along the beam line. This characteristic reflects in the design of the LHCb detector. Unlike the other LHC experiments, the LHCb detector is a single-arm forward spectrometer, meaning that the sub-detectors are placed one after another along the beam pipe, in contrast to a design with multiple layers around the collision similar to an onion. It is called a 'single-arm'-detector, because it covers only one direction along the beam pipe. The LHCb detector operation periods are called Runs. The detector is currently in Run3 and had software and hardware upgrades made during the last shutdown period. The upgraded LHCb detector is able to detect particles from every pp-collision happening with a frequency of 40 MHz. It is designed to reach a nominal instantaneous luminosity of $\mathcal{L} = 2 \times 10^{33}$ cm^{-2} s^{-1} [13].

The following sections describe the subdetectors of the upgraded LHCb detectors. For further description of the LHCb detector, a right-handed coordinate system is used with its y-axis pointing towards the surface and its z-axis along the beam direction. The origin of the coordinate system is set to be the interaction point of the pp-collision. This choice of coordinate system can be seen in Figure 3.

3.1.1 Magnet

A key component of the LHCb detector is the dipole magnet. The magnetic field acts on charged particles through the Lorentzian force and bends the trajectory of the particle based on its momentum and charge. With the knowledge of the magnetic field strength along a particle track, provided my a field map, it is possible to reconstruct the momentum of the track. The magnetic field is oriented in the vertical direction and has a bending power of approximately 4 Tm. The magnet consists of two saddle-shaped aluminium coils, with increasing distance to match the geometry requirements of the detector. In the data taking process, the polarity of the magnet is reversed regularly.

3.1.2 Particle tracking

Right around the interaction region of the colliding proton beams, the **Ve**rtex **Lo**cator (VELO) is placed. It consists of pixelated hybrid silicon detectors arranged in modules to detect tracks of ionising particles. Its location makes it possible to reconstruct the interaction point of the colliding proton beams, referred to as the primary vertex (PV),
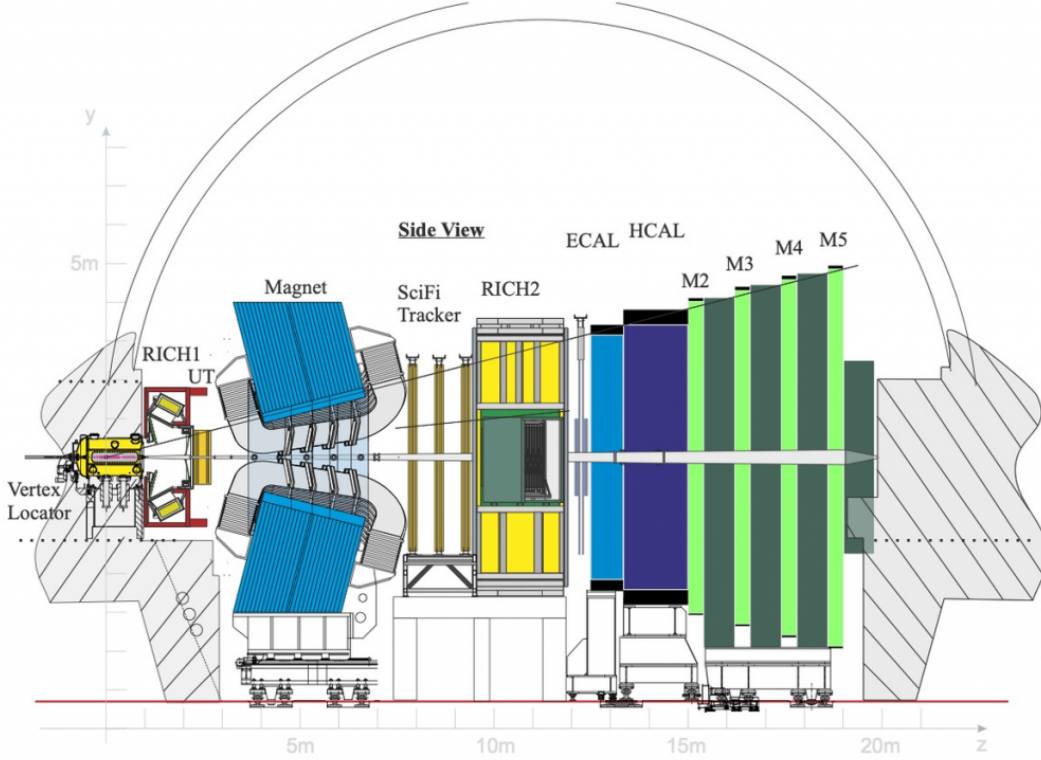
Figure 3: Schematic layout of the LHCb detector. Taken from [14].

as well as the so called secondary vertex (SV) of decaying particles, which is spatially separated from the PV. By calculating the distance between these vertices, the decay time of unstable particles can be calculated using the momentum of the particle. The decay time is a crucial quantity for this analysis and the precise measurement of the decay time available at LHCb is a key factor in determining the fast B_s^0-$\overline{B}_s^0$ oscillation. The VELO information is used by the reconstruction algorithm of the LHCb and additionally provides discriminatory information for event selection.

The **U**pstream **T**racker (UT) is located in front (upstream with respect to the particle track) of the dipole magnet. The UT consists of four planes of silicon strip detectors. Combined with the VELO information and the magnetic field in the region between the VELO and the UT, a first estimate of the particle momentum can be made. This also reduces the rate of ghost tracks. A ghost track describes hits in the detector from more than one particle that are randomly combined to form a track. The two middle detector planes of the UT are tilted with an angle of $\pm 5°$,with respect to the y-axis, to get information about the vertical position of the tracks.

The **Sci**ntillating **Fi**bre tracker (SciFi) is the only tracking detector located downstream of the dipole magnet. The SciFi consists of multilayered mats of scintillating fibers. There are tree stations of this tracker named T_1, T_2 and T_3. Each station consist of four layers. Similarly to the UT, the two middle SciFi mats are tilted $\pm 5°$ with respect to y-axis, to gather information about the vertical position of the track. To cover the acceptance range of the LHCb detector, the SciFi mats are 5 meters high and 6 meters wide. The readout of the scintillating signal is done by silicon photomultipliers (SiPMs) at the top and bottom

of the fiber mats, with a mirror placed in the middle of the scintillating fibre to direct the signal to the SiPMs. The SciFi provides excellent momentum resolution. This is especially useful for calculations of the decay time and the invariant mass of particles.

3.1.3 Particle identification

The **R**ing **I**maging **Ch**erenkov detectors 1 and 2 (RICH1 and RICH2 respectively) provide information for the identification of charged particles. RICH1 is located upstream of the magnet and provides hadron separation in a momentum range of 2.6 - 60 GeV[1] whereas the RICH2 detector is located downstream of the SciFi and provides particle identification in a momentum range of 15-100 GeV. The different momentum regions are realised with different diffraction indices of the gasses used in RICH1 and RICH2. Charged particles travel through the fluorocarbon gaseous radiators of the RICH detectors and produce Cerenkov light. The Cerenkov photon trajectories follow a cone shape with opening angles dependent on the trigger particle mass and momentum. The Cerenkov light is reflected with mirrors outside of the acceptance of the detector and read out by multi-anode photomultiplier tubes (MaPMTs). The RICH detectors are crucial for flavour tagging of neutral B_s^0 mesons as they provide good charged kaon identification from the $b \to c \to s$ decay chain.

The **E**lectromagnetic **Cal**orimeter (ECAL) consists of alternating scintillator and lead layers and is located behind the RICH2 detector. It has a thickness of 25 radiation lengths. This provides a high probability that the electromagnetic shower of electrons or photons is completely covered by the calorimeter. The complete coverage of the electromagnetic shower enables a high energy resolution. To measure the energy with low uncertainties is important for the precisely calculating the invariant mass of a particle. A charged particle travelling through the calorimeter produces a photon through bremsstrahlung. This high energy photon produces an electron-positron pair and the process is repeated until the critical energy is reached. The light produced prom this process in the scintillators is transmitted to photomultiplier tubes.

The **H**adronic **Cal**orimeter is a sampling tile calorimeter located downstream of the ECAL. It consist of alternating layers of iron and plastic scintillator tiles. With only 5.6 radiation lengths, due to space limitations, it is thinner than the ECAL. The light of the scintillators is transmitted to photomultiplier tubes similarly to the ECAL. In contrast to the ECAL, the HCAL has a higher granularity, motivated through the physical properties of hadronic showers, which are typically more spread out than the electromagnetic showers.

The four **M**oun chambers (M2-M5) are located downstream of the HCAL and are the most upstream sub-detectors of the LHCb experiment with respect to the particle path. They consist of layers of multi-wire-proportional chambers (MWPCs) with layers of 80cm thick iron in between. The iron absorbers provide absorption for low energy particles. The MWPCs consists of multiple anode wires surrounded by gas between cathode plates. Ionizing particles will produce ion pairs in the gas that are accelerated to the wires or the plates depending on their charge. The resulting current in the wire is proportional to the energy of the ionizing particle. Due to logical combinations of readouts of multiple wires, positional information about the particle is efficiently gained.

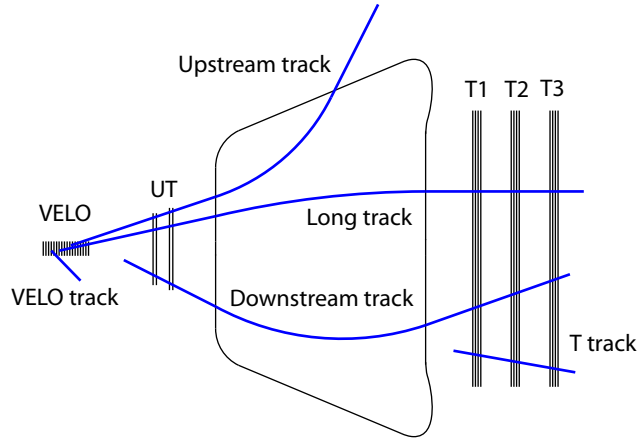[1]Natural units c=$\hbar$=1 are used throughout this analysis

Figure 4: Track types used in the reconstruction process. Taken from [16].

3.1.4 Trigger and reconstruction

The reconstruction process uses the information of the LHCb subdetectors for each (pp collision) to form particles and infer their physical properties. The reconstruction is done in real time for each pp-collision using the combination of the two software based triggers, namely the **H**igh **L**evel **T**rigger 1 and 2 (HLT1 and HLT2). The HLT1 is the first trigger stage and provides a partial event reconstruction and a first selection of events. This information is then fed into the HLT2, amongst real time alignment and calibration data of the detector, to perform a full event reconstruction and a final selection. The two trigger stages make sure that only the relevant information of events for the analysis is stored to reduce processing time and storage space.

The HLT1 trigger sequence starts with the track reconstruction in the VELO to locate the PV through a straight line fit. The track is then extrapolated to the UT and SciFi, to form a so called long track shown in Figure 4. This information is used calculate the particles displacement from the PV. Furthermore, tracks are classified as muons or non-muons based on the muon system information and combined to two-body displaced vertex candidates. This simplified reconstruction of events is performed to identify events matching the LHCb physics program. This selection is done to reduce the amount of computation needed in the more accurate reconstruction of the HLT2. The reconstructed and selected tracks of the HLT1 are then fed into the HLT2.

The HLT2 takes into account real time alignment and calibration information about the detector. In the reconstruction process tracks are combined with particle identification information to form individual charged particles. There is no magnetic field in the VELO, so information from the UT and the SciFi is needed for momentum measurement. A Kalman filter [15] provides uncertainties for the PV and SV of the particles and the best estimate of the particle momentum. Further information of the ECAL is used to identify electrons and photons.

After the two HLT phases, the data is stored and a more detailed reconstruction and selection can be done offline.

4 Analysis methods

This section provides a short description about the technical details and challenges of methods used in this analysis

4.1 Boosted Decision Trees

A boosted decision tree (BDT) is used in this analysis, to classify candidates in the data sample as either signal or background. A BDT is an ensemble of decision trees (DTs) combined in a process referred to as boosting. A DT gives an estimate $\hat{y}_i$ for the true class label y_i of candidate i, by testing given variables for different criteria. These criteria are derived by subsequently splitting samples into two subsamples, setting a cut on a single variable. The best variable for the cut as well as the cut itself are determined by minimising the loss function in the training process [17]:

$$L(y_i, \hat{y}_i) = \sum_i (y_i - \hat{y}_i)^2. \tag{13}$$

The samples are partitioned until no improvement of the loss is recorded or a given depth is reached. A split of the DT is called a node. Nodes are connected by branches and the nodes, where the partitioning process is stopped are called leafs.

Boosting is the process of building a model consisting of consecutive DTs. The individual DT is rather simple, but the combination of multiple simple DTs results in an advanced model. Before adding a new DT to the model, the previous model is evaluated. Based on the performance of the previous model, the data sample is weighted. Wrongly classified candidates in the sample are weighted more heavily. The newly added DT receives the weighted data sample and therefore sets its focus on the errors of the previous model. The finished model is called a boosted decision tree and is well suited to perform classification tasks.

4.1.1 Mathematical definition of a Boosted Decision Tree

A DT can be defined mathematically as

$$f_k(\vec{x}_i) = \vec{w}_{q(\vec{x}_i)}, \quad \vec{w} \in \mathbb{R}^T, \quad q : \mathbb{R}^m \to \{1, 2, ..., T\}. \tag{14}$$

The vector $\vec{w}$ contains all the scores of the T leafs and q is a function mapping each set of the m features of the i-th candidate $\vec{x}_i$ to the corresponding leaf.

A boosted decision tree contains multiple decision trees in series. The output of each DT is used in the training of the consecutive DT.

$$\hat{y}_i^{(0)} = 0$$
$$\hat{y}_i^{(1)} = y_i^{(0)} + f_1(\vec{x}_i)$$
$$\vdots \tag{15}$$
$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(\vec{x}_i) = \hat{y}_i^{(t-1)} + f_t(\vec{x}_i),$$

with i being the *i-th* candidate of the data sample. The shape of the DT f_t that is added in step t is determined by minimizing the objective function

$$obj^{(t)} = \sum_i L(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \omega(f_k). \tag{16}$$

The regularization term $\omega(f_k)$ is used to control the model complexity and is an arbitrary function. The regularization term of the *XGBoost* package [17] used in this analysis is defined as:

$$\omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{17}$$

To be in control of the model complexity is a key feature in BDT training, because it effects the so called overtraining of the BDT.

4.1.2 Overtraining

An overtrained BDT specialises on fluctuations of the training sample instead of the physical differences between signal and background distributions of the training variables. The overtrained BDT therefore performs bad on unseen data, so overtraining needs to be avoided since this is the intended use of the BDT. Overtraining can be observed in the training process, when the BDT performance is improving on the training sample but not improving on the test sample. A way to observe and avoid overtraining is the use of cross validation. In this analysis the K-folding cross validation method is used. K-folding is the process of shuffling the training sample and splitting it into k subsamples. The BDT is trained on k-1 samples and evaluated on the remaining subsample. The evaluation data is stored and the model gets reset. The training is repeated in a way that every subsample was used for the evaluation of the BDT once. This process yields no final model and is intended to check the model complexity. With the stored evaluation data from each fold, the logarithmic loss (log-loss) curve of the training and evaluation sample is plotted. The log-loss curve is the logarithm of the loss function Eq. 13 plotted over the number of trees added to the BDT. By studying the log-loss curve of the training sample and the evaluation sample, it is possible to identify overtraining. In this case the log-loss curve of the evaluation sample is rising or reaching a plateau while the log-loss curve of the training sample is still declining.

To counteract on this behaviour a method called early stopping is used. In the early stopping method a threshold value on the maximum number of trees added in the boosting process without reduction of the log-loss curve is set. If the threshold number of trees is added to the model, but the log-loss curve showed no improvement, the training is stopped. This reduces the model complexity and the time used for training the BDT, since fewer DTs are added to the model.

Another way to control overtraining are hyperparameters [17]. These parameters set constraints for the DT building process. The *minimum child weight* for example sets a lower boundary for the sum of the weights in a child leaf. If the minimum is not reached, a further split of the leaf is restricted. The γ-*factor* defines a minimum for the reduction of the loss function required in each step to further partition a leaf. Both of these hyperparameters control the model complexity which is crucial to avoid overtraining.

For each fold of the K-folding process, the receiver-operating-characteristic curve (ROC-curve) is plotted. The ROC-curve is the result of plotting the true positive rate (TPR) against the false negative rate (FPR) defined as:

$$TPR = \frac{S_{\text{true}}}{S_{\text{true}} + B_{\text{false}}} \quad \text{and} \quad FPR = \frac{S_{\text{false}}}{S_{\text{false}} + B_{\text{true}}}, \tag{18}$$

with S being events classified as signal by the BDT and B events classified as background. The labels true and false correspond to the truth of the classification, received by comparing the label of the event with the prediction of the model.

A perfect BDT would have a TPR of one and a FPR of zero. This would correspond to an area under the ROC-curve of one. A randomly classifying BDT, on the other hand, would have an area of 0.5. Therefore the area under the ROC-curve is a good quantisation for the performance of the BDT and is referred to as the ROC-score. To further investigate overtraining of the BDT, the following approach is used: the overtraining is quantised by comparing the classification distribution of the BDT for the true signal and true background events of the test sample with the true signal and true background events of the training sample. A non-overtrained BDT would produce similar distributions for the test and the training sample, whereas an overtrained BDT would show differences in the distributions of the two samples.

4.2 Unbinned Maximum Likelihood Fit

To retrieve physical quantities from data, an unbinned maximum likelihood method is used. For a set of measurements $X = \{\vec{x}^{(1)}, \vec{x}^{(2)}, \ldots\}$ an estimate for the parameters $\vec{\theta} = (\theta_1, \theta_2, \ldots)$ of a probability density function (PDF) is generated. The estimated parameters $\hat{\vec{\theta}}$ are the ones, for which the given set of measurements is most likely if it was described by the PDF. These parameters are determined by maximising the likelihood function:

$$L(\vec{\theta}) = \prod_i f(\vec{x}^{(i)}; \vec{\theta}) \tag{19}$$

The likelihood function is a function of the parameters $\vec{\theta}$ evaluated on the fixed set of measurements X. Instead of maximizing the likelihood function, it is computationally more convenient to minimize the negative logarithmic likelihood (NLL) function. Intuitively speaking it is easier to add together the values of the Likelihood function than to multiply them.

$$\text{NLL}(\theta) = -\ln(L(\vec{\theta})) = -\sum_i \ln(f(\vec{x}^{(i)}; \vec{\theta}). \tag{20}$$

To find the parameters for which the given data is most likely one needs to minimize this NLL function. This minimisation is performed with the *iminuit* package [18].
The uncertainty of the estimated parameters are calculated using the Hessian matrix, which consists of second derivatives of the logarithmic likelihood function.

$$\left. \frac{\partial^2 \ln L(\vec{\theta})}{\partial \vec{\theta}^2} \right|_{\vec{\theta} = \hat{\vec{\theta}}} \tag{21}$$

An approximation of the Hessian matrix is calculated during the minimisation process. By inverting the Hessian matrix, the uncertainties of the parameters can be determined

using the Rao-Cramer-Frechet minimum variance bound [19]:

$$V[\hat{\theta}] \geq \frac{1}{E\left[\frac{\partial^2 \ln L(\vec{\theta})}{\partial \vec{\theta}^2}\Big|_{\vec{\theta}=\hat{\vec{\theta}}}\right]}, \tag{22}$$

which can be further approximated in the limit of a large sample to:

$$V[\hat{\theta}] \geq \frac{1}{\left[\frac{\partial^2 \ln L(\vec{\theta})}{\partial \vec{\theta}^2}\Big|_{\vec{\theta}=\hat{\vec{\theta}}}\right]}. \tag{23}$$

4.3   Flavour tagging at LHCb

A key information in determining the B_s^0-$\overline{B}_s^0$ oscillation frequency Δm_s is the particle or antiparticle state of the B_s^0-meson at the time of production (initial) and at the time of the decay (final). The information about the initial and final state of the B_s^0-meson is crucial to separate the data into mixed and unmixed candidates: the unmixed B_s^0 candidates have the same initial and final flavour state, whereas the mixed B_s^0 candidates have different initial and final flavour states.

Since the $B_s^0 \rightarrow D_s^- \pi^+$ is a flavour-specific decay, the final flavour of the B_s^0-meson can be determined with the charge of the decay products.

To retrieve information about the initial flavour state of the B_s^0 meson, machine learning algorithms, also called taggers, are used in a procedure referred to as flavour tagging.
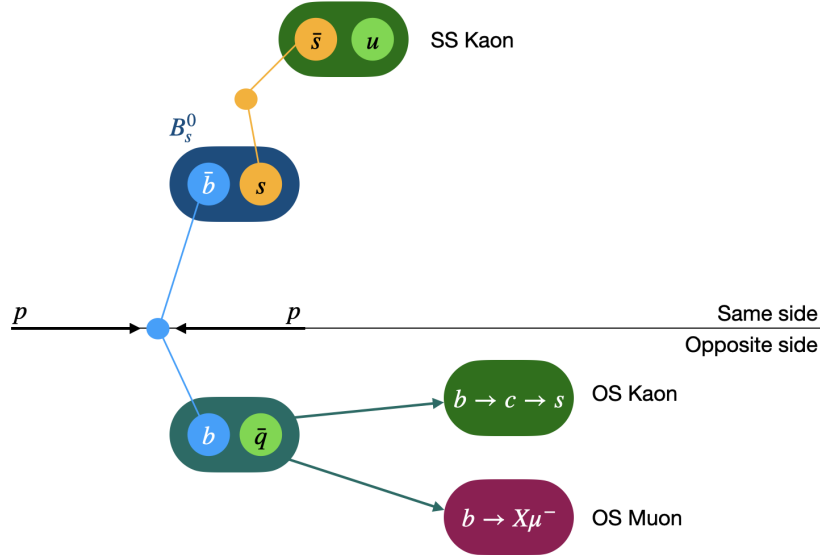

Figure 5: Schematic view of the tagging particles used for the specific flavour tagging algorithms.

In this analysis two distinct categories of taggers are used, namely the same side taggers (SS) and the opposite side taggers (OS).

The SS taggers use information from the hadronisation process of the B_s^0 meson. This is

shown in the top half of Figure 5. The $\bar{b}$-quark produced in the pp-collision needs an s-quark to form a B_s^0-meson. The s-quark is produced in a $s\bar{s}$-quark pair via the strong interaction. The flavour of the $\bar{s}$-quark is therefore directly related to the initial flavour of the B_s^0-meson. In the case of the SSKaon tagger, the associated $\bar{s}$-quark forms a kaon and the initial state of the B_s^0 meson is directly related to the charge of this kaon. In this case, the kaon is referred to as a tagging particle.

The OS taggers exploit the fact that the b-quarks are predominantly produced in $b\bar{b}$-quark pairs in a pp-collision. Therefore additionally to the signal B_s^0 meson, there is another b-particle in the event. The initial flavour of the opposite side b-particle is directly related to the initial flavour of the signal B_s^0-meson. The OS taggers follow along the decay chain of the b-quark.

The OSMuon tagger focuses on semileptonic decays of the b-quark containing a muon. The tagging decision is made based on the charge of the muon in the weak decay $b \to X\mu^-$.

The OSKaon tagger follows along the $b \to c \to s$ decay chain resulting in a charged kaon, for example the $B^- \to D^0 \, (K^-\pi^+)\pi^-$ decay. The tagging decision is based on the charge of the kaon.

Preselection requirements on particle identification, momentum and distance from the PV are applied to select tagging particles. The selected tagging particles provide tagging decisions assigned to the B_s^0 candidate. A tagging decision of $d = +1$ (-1) corresponds to an initial flavour of B_s^0 $(\bar{B}_s^0)$ of the signal candidate. If the tagger is not able to give information about the initial flavour, the tagging decision is $d = 0$.

A neural network (NN) is trained on the selected tagging particles with information about their kinematic properties and the topology of the decay, to estimate the probability of wrongly tagging the B_s^0-meson. This estimated mistag probability $\eta \in [0.0, 0.5]$ is 0.0 for a perfect tagging decision and 0.5 for a random guess. If the estimated mistag probability is greater than 0.5, the value $1 - \eta$ is used and the tagging decision is reversed.

The performance of the flavour tagging algorithms are evaluated by defining the following quantities.

The tagging efficiency:

$$\epsilon_{\text{tag}} = \frac{N_{\text{tag}}}{N_{\text{tag}} + N_{\text{untag}}}, \tag{24}$$

with $N_{tag}(N_{untag})$ being the amounts of tagged (untagged) events respectively, is simply the fraction of tagged events and contains no information about the quality of the tags.

The mistag probability ω, describing the fraction of wrongly tagged events:

$$\omega = \frac{N_{\text{wrong}}}{N_{\text{wrong}} + N_{\text{right}}}, \tag{25}$$

with $N_{\text{wrong}}(N_{\text{right}})$ being the amounts of wrongly (rightly) tagged events.

An important quantity for this analysis is the dilution factor:

$$\mathcal{D} = (1 - 2\omega). \tag{26}$$

It is a factor accounting for imperfect tagging by diluting the oscillation amplitude. This is further discussed in Section 8.1.

A combined measurement of the tagging efficiency and the mistag probability is called the tagging power or the effective tagging efficiency. It combines the efficiency of the flavour tagging algorithm with the quality of the tags:

$$\epsilon_{\text{eff}} = \epsilon_{\text{tag}}\mathcal{D}^2. \tag{27}$$

This quantity is a measurement of the statistical power of the flavour tagged sample with N events, since the uncertainty of the measured mixing amplitudes scales with $\sigma_A \sim 1/\sqrt{N \cdot \varepsilon_{\text{eff}}}$.

5 Offline event selection

This analysis uses an early 2024 data sample of $B_s^0 \to D_s^-$ $(K^+K^-\pi^-)\pi^+$ decay[2] taken in Run3 of the LHCb experiment. Samples with both magnet polarities are combined. The data sample for this analysis corresponds to an integrated luminosity of about $1.73\,\mathrm{fb}^{-1}$. It was taken without information from the UT subdetector Sec. 3.1. As mentioned in Section 3.1, the detector was upgraded during the last shutdown period. The alignment of the new detector has not reached an optimal state yet. The imperfect alignment induces a systematic loss in track reconstruction efficiency. Furthermore it leads to uncertainties in the calculation of the particle momenta. This impacts all variables calculated based on the particle momenta. A Monte Carlo (MC) sample of the $B_s^0 \to D_s^-$ $(K^+K^-\pi^-)\pi^+$ decay is created. The created signal particles are passed through a simulation of the detector to match the conditions of the real data sample [20]. No information from the UT is used and samples with both magnet polarities are combined.

5.1 Signal selection

For this analysis the $B_s^0 \to D_s^-$ $(K^+K^-\pi^-)\pi^+$ decay is studied, to determine the B_s^0 oscillation frequency. A signal selection is important to efficiently store events from the decay of interest while keeping background contributions from other decays small. To select the B_s^0 signal candidates, the first step is to use the reconstructed particle trajectory, discussed in Sec. 3.1.4. The reconstruction of the B_s^0 meson starts with the stable charged particles and subsequently combines them to composite particles. To reconstruct the signal B_s^0 candidates, tracks with high probability to be a K^+ as well as tracks with high probability to be a K^- or a π^-, which can be associated to a common vertex are searched. By taking into account the four momenta of the three stable particles, the four-momentum of the D_s^- can be reconstructed by varying the track properties within their associated uncertainties. This process is done by the decay tree fitter (DTF). This fit additionally provides a χ^2 value as measure of the fit quality. The resulting D_s^- four-momentum is then combined with a π^+ track in the same manner to form a B_s^0 signal candidate.

Occasionally random tracks that happen to be spatially close to each other are combined to a B_s^0 candidate. If it fulfills the criteria for a signal candidate, it ends up in the data sample as combinatorial background. This combinatorial background events do not represent signal decays and for this analysis they need to be removed from the sample.

Another source of background in the reconstruction process are misidentified particles. In the particle identification process, a probability like variable is assigned to each stable particle, which determines the likelihood of the particle being a specific hadron or lepton. Wrongly identified particles can be used to form a signal candidate, which ends up in the data sample. This background is also suppressed in this analysis using particle identification cuts in the preselection as well as vetoes for specific decays with a similar topology to the signal decay, such as the $B^0 \to D^-\pi^+$ decay.

In the process of creating the data sample, a set of sprucing cuts is applied to the selected events. These sprucing cuts are summarised in Table 1.

---

[2]The charge conjugated mode is always implied

Particle	Sprucing cut
π^+	$PID_K < 0$
	$p > 5\,\text{GeV}$
	$p_T > 500\,\text{MeV}$
π^-	$PID_K < 5$
	$p > 2\,\text{GeV}$
	$p_T > 250\,\text{MeV}$
K^-,K^+	$PID_K > -5$
	$p > 2\,\text{GeV}$
	$p_T > 250\,\text{MeV}$
D_s^-	$1830\,\text{MeV} < m(D_s^-) < 1910\,\text{MeV}$
	$\text{Sum}_{p_T} > 1800\,\text{MeV}$
	$\text{Vtx}_{\chi^2_{red}} < 10$
	$\text{BPV}_{\chi^2} > 36$
	$\text{BPV IP}_{\chi^2} > 0$
	$\text{BPV DIRA}_{\chi^2} > 0$
	$\text{DOCA}_{12} > 0.2\,\text{mm}$
	$\text{DOCA}_{13} > 0.2\,\text{mm}$
	$\text{DOCA}_{23} > 0.2\,\text{mm}$
B_s^0	$5000\,\text{MeV} < m(B_s^0) < 7000\,\text{MeV}$
	$\text{Sum}_{p_T} > 5\,\text{GeV}$
	$\text{Vtx}_{\chi^2_{red}} < 20$
	$\text{BPV IP}_{\chi^2} < 25$
	$\text{BPV DIRA} > 0.999$
	$\text{BPV lifetime} > 0.2\,\text{ps}$

Table 1: Sprucing cuts on the data sample.

5.2   Preselection

To further reduce background candidates in the data sample, a preselection is performed. This selection contains cuts on the quality of the reconstruction and particle identification criteria as well as veto cuts for specific background decays and requires at least one tagging decision Sec. 4.3. The applied preselection cuts are summarised in Table 2. The preselection cuts are loosely based on a previous analysis of the B^0 oscillation frequency Δm in $B^0 \rightarrow J/\psi K^{*0}$ and $B^0 \rightarrow D^-\pi^+$ [21].

The reduced χ^2 variable, χ^2_{red}, describes how well the primary vertex is fitted with respect to the degrees of freedom of this fit. A cut reduces the amount of combinatorial background, as combinatorial background candidates usually have a higher χ^2_{red} value than signal candidates. This is the case, because combinatorial background results from the combination of random particles. As the particles tracks are uncorrelated, they are more inclined to have a worse fit than those that actually belong to a single vertex.

The GhostProb is a measurement of the quality of the track reconstruction. A cut on this variable reduces combinatorial background, because the GhostProb should be per definition higher for randomly reconstructed particles.

The particle identification (PID) variables are combined differential logarithmic likelihoods

rectangular cuts	
K^+, K^-, π^+, π^-	$\chi^2_{red} < 3.5$
	$GhostProb < 0.5$
D_s^-	$1940\,\mathrm{MeV} < m(D_s^-) < 1990\,\mathrm{MeV}$
	$\chi^2_{red} < 9$
π^+	$PID_p < 10$
K^+, K^-	$PID_K > 0$
veto cuts	
$B_s^0 \rightarrow D_s^- (K^+\pi^-\pi^-)\pi^+$	$K^- : PID_K > 10$
	or not $(1840\,\mathrm{MeV} < m(D_s^-) < 1890\,\mathrm{MeV})$
$\bar{\Lambda}_b^0 \rightarrow \bar{\Lambda}_c^- (K^+\bar{p}\pi^-)\pi^+$	$K^- : PID_K - PID_p > 0$
	or not $(2255\,\mathrm{MeV} < \text{reconstructed } m(\bar{\Lambda}_c) < 2315\,\mathrm{MeV})$
tagging decision cut	
B_s^0	not $(d = 0 \,\forall$ SS- and OS-Taggers$)$

Table 2: Preselection cuts on the data sample.

(DLL) [22]. They are a measure of how likely a specific mass hypothesis for a given track is relative to the π mass hypothesis. For this the likelihoods from each particle identification subdetector is combined to a single likelihood $\mathcal{L} = \mathcal{L}_{RICH} \cdot \mathcal{L}_{ECAL/HCAL} \cdot \mathcal{L}_{MUON}$. For example the PID_K variable is defined as follows:

$$PID_K = DLL(K) = ln(\mathcal{L}(K)) - ln(\mathcal{L}(\pi)) \tag{28}$$

Additionally, a cut on the D_s^- mass is performed. The veto cuts are set to remove contributions from the $B^+ \rightarrow D^- (K^+\pi^-\pi^+) \pi^+$ decay as well as the $\bar{\Lambda}_b^0 \rightarrow \bar{\Lambda}_c^- (K^+ \bar{p} \pi^-) \pi^+$ decay. These background decays appear in the data sample due to misidentification of particles. For the $B^+ \rightarrow D^- (K^+\pi^-\pi^+) \pi^+$ decay, one of the π^- is wrongly identified to be a K^-. To suppress this background contribution, a tighter particle identity cut is used for the reconstructed K^-. This tighter cut is applied if the reconstructed D_s mass, using the four vectors of the reconstructed particles and a π^- mass hypothesis for the K^-, lies in a mass region of 25 MeV around the nominal D_s^+ mass. Similarly, the $\bar{\Lambda}_b^0 \rightarrow \bar{\Lambda}_c^- (K^+\bar{p}\pi^-)\pi^+$ decay is suppressed by setting a tight particle identity cut. Candidates where the D_s^+ mass, reconstructed under a proton mass hypothesis, lies in a 30 MeV range around the $\bar{\Lambda}_c^-$ mass need to pass this tighter cut. The results of these veto cuts on a subsample of the data is shown in Figures 6b and 7. Looking at the shape of the removed candidates it is visible that the veto cuts are effective. The peaks of the removed candidate distributions do not lie inside of the signal area, which is clearly visible in the distribution of the kept candidates. Looking at the mass peaks of the kept candidates, it is visible that the peak gets more narrow on the side where the misidentified candidates are removed. Also the peaks of the reconstructed masses under the two hypotheses lie right in the mass region that is selected for the veto.

The tagging decision cut excludes all candidates that have no tagging decision d from all available flavour tagging algorithms. These candidates are excluded, as they do not contribute to the determination of the B_s^0 oscillation frequency Δm_s.

The MC sample contains more information about a particle than real data. A specific
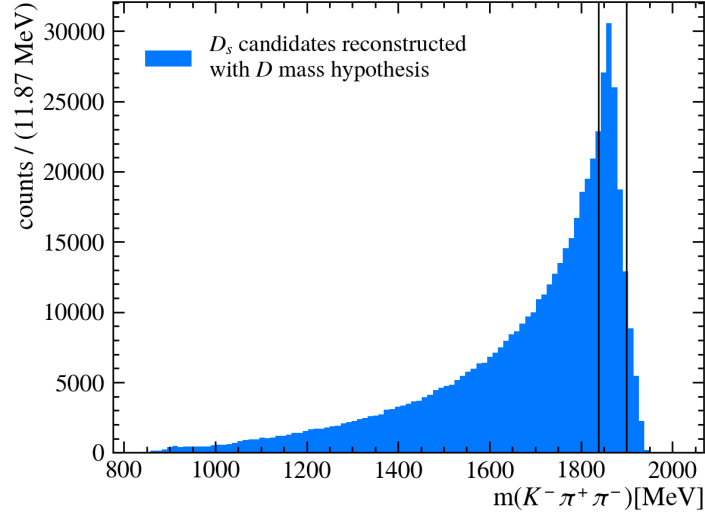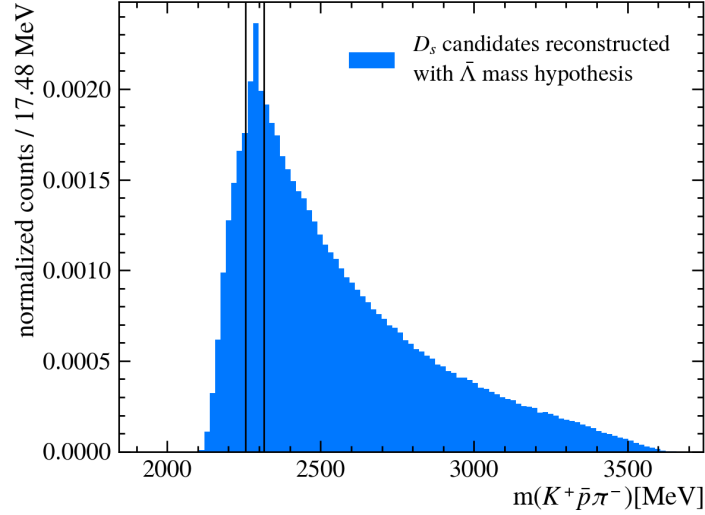
(a) Mass distribution of D_s^- candidates reconstructed with the D^- mass hypothesis.
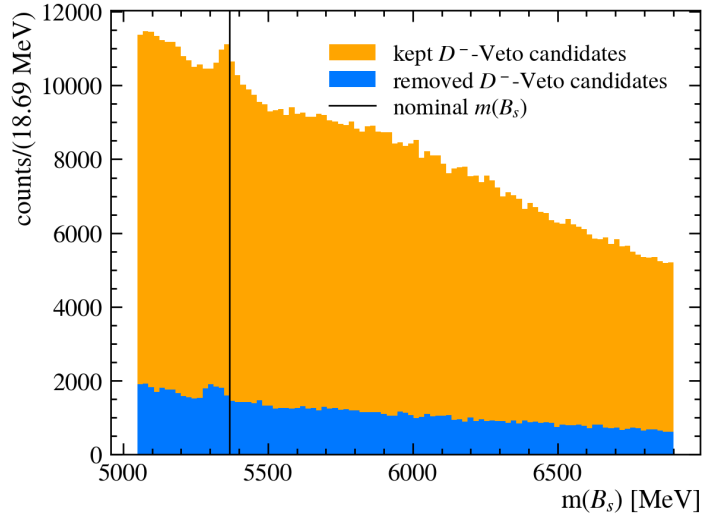
(b) B_s^0 mass distribution with and without the D^- veto cut.

Figure 6: Preselection $B^0 \rightarrow D^-(K^+\pi^-\pi^-)\pi^+$ veto cuts.

kind of variable that is only available in the MC sample is the particle *key*. This particle *key* is a unique number associated to every individual particle. The decay products of a decaying particle contains a variable called *mother key* that contains the particle *key* of the ancestor particle. With this information a process called truth matching is possible. In the process of truth matching, the key variables of all particles in the decay chain are compared to those of the signal candidates. Only candidates are kept if the key information of the mother particle and the daughter particles match. This ensures that the MC sample only contains particles from perfectly reconstructed signal decays. This truth matching is done for the MC sample in this analysis.

(a) Mass distribution of D_s^- candidates reconstructed with the $\bar{\Lambda}_c$ mass hypothesis.
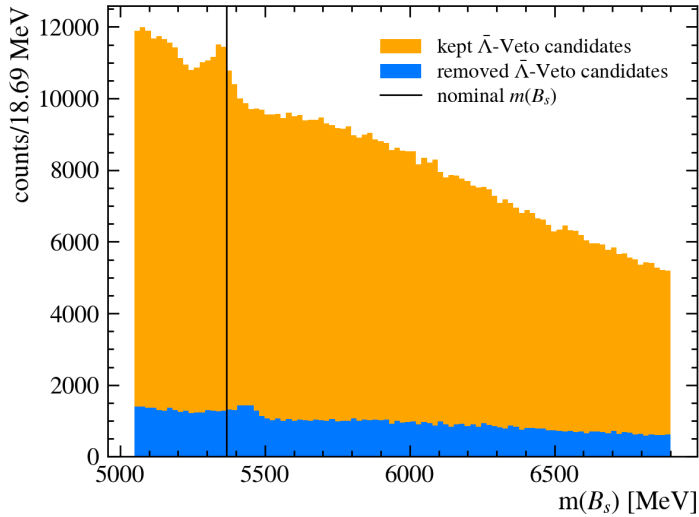
(b) Mass distribution with and without the $\bar{\Lambda}_b$ veto cut.

Figure 7: Preselection $\bar{\Lambda}_b^0 \to \bar{\Lambda}_c^-(K^+\bar{p}\pi^-)\pi^+$ veto cuts.

5.3 Boosted decision trees

After applying the sprucing and the preselection cuts, the amount of combinatorial background has to be removed further. In this analysis a boosted decision tree (BDT) is used. The BDT classifies every candidate in the data sample as signal or background. For this, the BDT needs to be trained with a set of training variables to make the right cuts on the data sample for classification. The first step is therefore to find a good set of training variables. Also the complexity of the trained BDT model needs to be controlled in order to avoid overtraining. And in the end, the performance of the trained BDT is to be evaluated to ensure efficient removal of combinatorial background while keeping most of the signal candidates in the process.
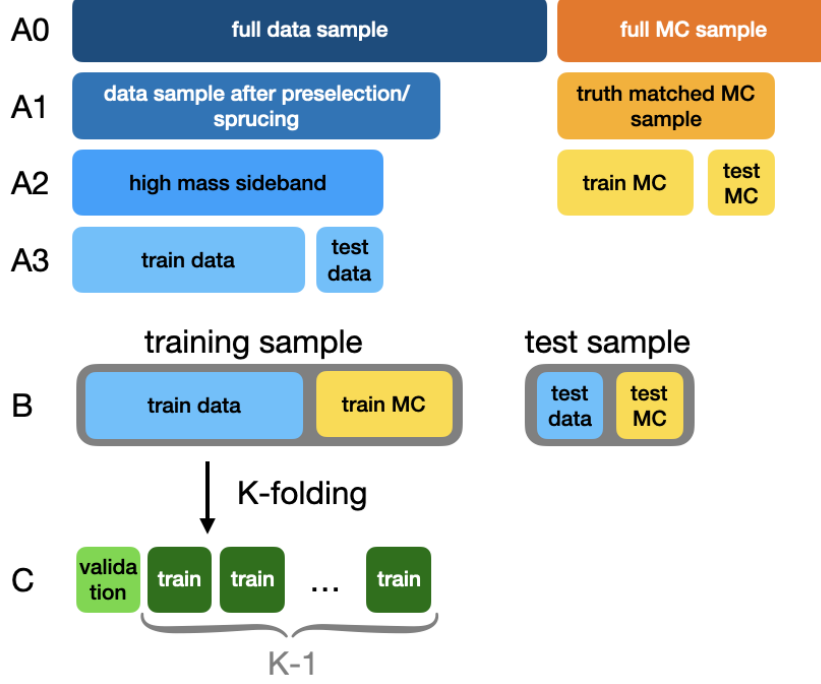
Figure 8: Flow chart of the creation of training samples.

5.3.1 Training variables

In the training process of the BDT, a set of training variables containing information about the kinematics of the particles and the goodness of the reconstruction process is used. Additionally a label is assigned to each event in the training sample, which classifies it as either signal (label = 1) or background (label=0). MC simulated events are used as signal. The background events in the training process are taken from the high mass sideband. This high mass sideband is defined as the mass interval $m(B_s^0) \in [5800,6800]$ MeV of the $B_s^0 \to D_s^- \pi^+$ data sample.

Figure 8 shows a flow chart of the creation of samples for the training and testing process. Up to this point we are in step A2. To find a set of suitable training variables a comparison between the MC distributions of signal candidates and the candidates in the high mass sideband distribution in data of various variables is performed. Variables with big differences in signal and background distributions are well suited for training a BDT. To quantise how different the distributions are, a Kolmogorov-Smirnov two sample test [23] is performed and variables with a small p-values are looked for.

Another criteria for the set of training variables were weak correlations to the mass and the lifetime of the neutral B_s^0 meson and as among themselves. Figure 9 shows the signal and background distributions of the resulting set of training variables. The separation in signal and background distribution is achieved using the sPlot technique. A negative logarithmic likelihood fit is performed on the mass distribution of the B_s^0 meson. From the resulting signal and background PDFs, weights for the data sample can be calculated separating signal and background distributions.

The variables $\eta_d(D_s^-)$ and $\eta_d(\pi^+)$ are the decorrelated versions of the variables $\eta(D_s^-)$ and

$\eta(\pi^+)$. The decorrelation is done using the following equation:

$$\eta_d = \eta_{\text{child}} - \eta_{\text{mother}}, \tag{29}$$

where child refers to D_s^- and π^+ and mother to the B_s^0 meson. The variable η itself is the pseudorapidity:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{30}$$

where θ is the angle between the the particles three-momentum and the beam axis (i.e. the z-axis). Particles travelling along the beam axis have the highest pseudorapidity whereas particles travelling orthogonal to the beam axis have a pseudorapidity of $\eta = 0$. The variables $p_T(\pi^+)$ and $p_T(D_s^-)$ describe the transverse momentum of the particle relative to the beam line.

The variable $\text{DIRA}(B_s^0)$ describes the angle between the vector from the primary vertex to the decay vertex of the B_s^0 and the reconstructed momentum of the B_s^0 meson.

The flight distance $\chi^2(D_s^-)$ describes the significance of determination of the flight distance of the D_s^- meson.

Figure 10 shows the correlation of the variables amongst themselves as well as the correlation to the mass and the lifetime of the neutral B_s^0 meson. There are minor correlations visible between the mass of the B_s^0 meson and its pseudorapidity as well as the decorrelated pseudorapidity of the D_s^- meson. The correlation of training variables to the mass or the decay time of the B_s^0 meson is considered problematic, because the BDT can alter the shape of the distributions. As for the mass variable, this is less important, because the shape of the mass distribution is only used for the calculation of signal weights. For the decay time variable this would be considered as a problem, since the Δm_s is determined using a fit to the decay time distribution. The correlations are below ± 0.3 for the mass variable and even lower for the decay time variable and therefore considered weak. No correction for the correlation is done. The variables amongst themselves have higher correlations. The correlation of training variables among themselves means less information is available for the BDT training, because correlated variables contain parts of the same information. For this correlation a higher tolerance is chosen, as it has less impact on the further analysis.

The final training sample is then created, using a sub sample of simulated candidates as well as a subsample of the B_s^0 candidates in the high mass sideband. The remaining candidates are used as a test sample to quantise the performance of the BDT and to check if the BDT is overtraining Sec. 4.1.1. This is shown in step B of Figure 8. Training the BDT on MC signal makes the BDT to perform less efficient, when applying it to data, due to differences between MC simulation and data. These differences are caused by imperfections in the simulation of the hadronisation process as well as the alignment and efficiency of the detector. This is further investigated by comparing the MC distribution, with the weighted signal distribution, for the training variables shown in Figure 9. A bias to the analysis introduced by the training on simulated events is not expected and a less efficient selection is accepted.

5.3.2 Training of the boosted decision tree

The BDT is trained with the set of selected training variables. The goal of the training is to get a well performing BDT that is not overtrained. To achieve this, several parameters
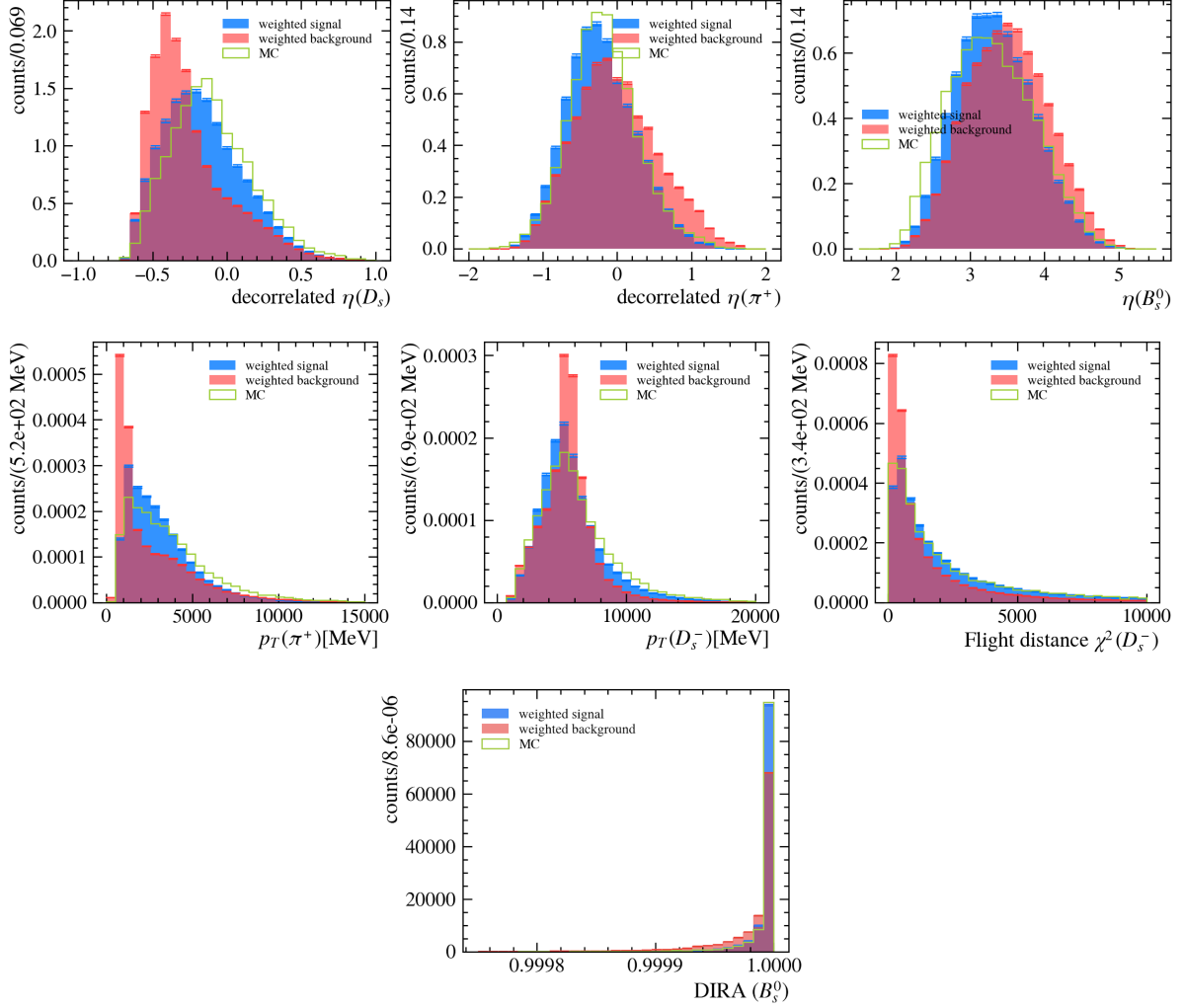
Figure 9: Distributions of training variables for sWeighted signal (blue) and background (red) candidates in data and MC candidates (green).

that control the complexity of the model, can be adjusted. The training is performed using a 6-fold cross validation. The training sample is split into six equally sized subsamples, which can be seen in step C of Figure 8. In the first step of the training, the BDT is trained on five of the six subsamples and the performance of the BDT is evaluated on the remaining subsample, called the validation set. In the next round the BDT is validated on a different validation set and trained on the remaining five subsamples. This is repeated until every subsample is once used as a validation set. The use of multiple validation sets enables a more generalized estimate of the performance of the BDT. It also uses the data efficiently and reduces the bias of training on just one sample. By training and evaluating the BDT model multiple times on different training and evaluation sets it is also possible to detect overtraining efficiently.

A first BDT is trained using the default parameter setting of the *XGBoost* package. Figure 11 shows the plot that tests if the BDT is overtraining. The trained BDT should perform equally well on the training sample and the testing sample. A comparison of the BDT classification for both of these samples separated in signal and background
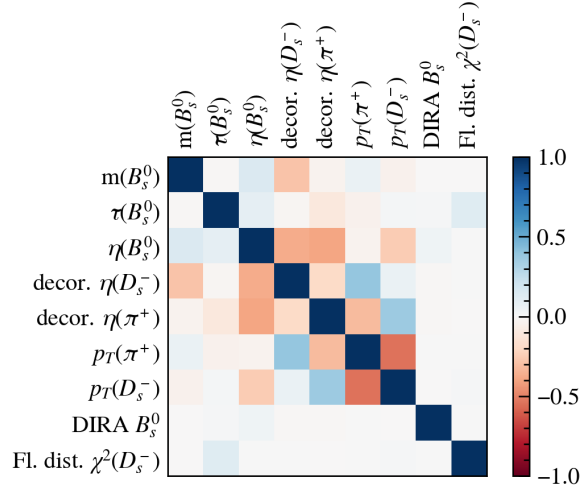
Figure 10: Correlation of training variables among themselves and to $m(B_s^0)$ and $\tau(B_s^0)$.
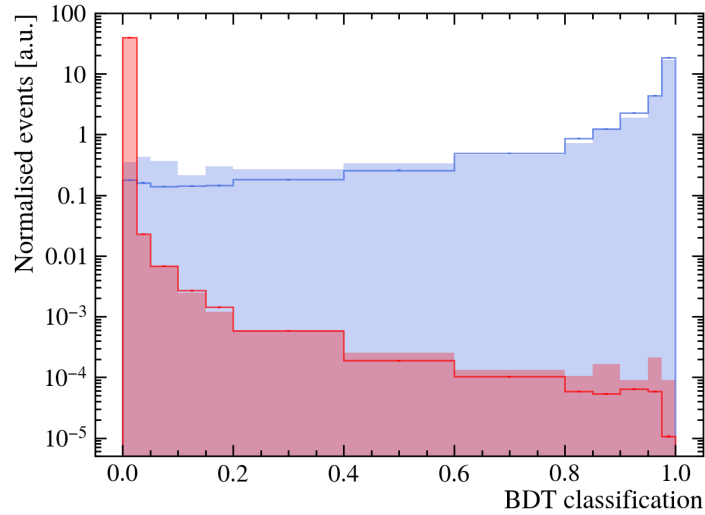

Figure 11: Overtraining test of the BDT model trained with default parameter settings. Comparison of the BDT classification on the signal test sample (blue filled) with the signal training sample (blue dots) and the background test sample (red filled) with the background training sample (red dots).

reveals discrepancies between the BDT performance on the different sample. The BDT trained with default settings is visibly overtrained on the signal sample. The biggest deviations between the distributions of the training and test sample can be seen at low BDT classifications. In the training sample, fewer signal candidates are classified with a low value than in the test sample. This classification is correct, because low BDT classification values correspond to background candidates, but it shows that the BDT is performing better on the training sample than on the test sample.

The main task of the BDT training is therefore to control the overtraining, while maintaining a well performing BDT. The model complexity and therefore the overtraining of the model is controlled by adjusting a set of hyper-parameters. Below the parameters

Parameter	Value
number of classifiers	50
learning rate	0.3
minimum child weight	50
γ	10
early stopping rounds	5

Table 3: Parameters of final BDT model.

used in this analysis are explained.

The total *number of classifiers* defines the upper limit of trees added in the boosting process. Setting a limit results in a less advanced BDT and saves computing time for the training process. A less advanced BDT is also less likely to be overtrained.

The *learning rate* is defined as the step size shrinkage. After each boosting steps, the resulting weights from the last tree are scaled by the learning rate. The weights are multiplied by the given learning rate. A lower learning rate slows the learning process down and makes the training process more conservative. Therefore this parameter is well suited to control overtraining.

The parameters *minimum child weight* and *γ-factor* are used to directly control the model complexity. The *minimum child weight* determines a lower limit for the sum of the weights in a child leaf node. If the leaf node does not reach the minimum child weight, further partitioning is restricted. This guarantees that the leafs contain enough data and therefore enough statistical information for the classification process. The next tree is trained using information from the leafs of the previous model. Prohibiting training in small data samples controls overtraining.

The *γ-factor* defines the minimum reduction of the loss function in order to further partition a leaf node. This is a useful parameter to keep the decision tree relatively simple. A higher value of the *γ factor*, makes the training process more conservative. This is useful to control the overtraining of the BDT.

The training is performed using early stopping. The value of the *early stopping rounds* parameter is fixed during the search for the best parameters.

The final BDT parameters and evaluation values are summarised in Table 3. A plot of the overtraining check and the ROC curve is included in Figure 12.

A reduction of overtraining is visible in Figure 12a compared to the BDT with default parameters. The distributions of the training samples compared to the test samples are in better agreement. There is still overtraining visible specifically for low BDT classification values in the signal samples and for high values in the background sample. Another difference between the two BDT versions can be seen in the overall shape of the classification value. The less overtraining model classifies more signal with low values and more background with high values than the overtrained BDT. This classification is incorrect, but only affects the efficiency of the BDT, without inducing a bias to the analysis. Therefore the weakly overtrained BDT is used for the further analysis.
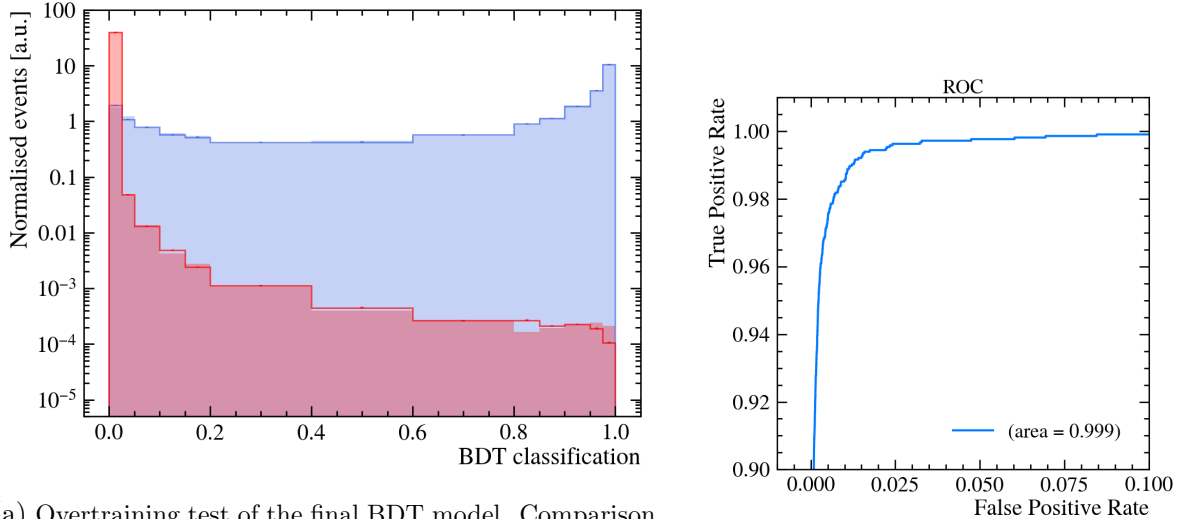
(a) Overtraining test of the final BDT model. Comparison of the BDT classification on the signal test sample (blue filled) with the signal training sample (blue dots) and the background test sample (red filled) with the background training sample (red dots).

(b) ROC curve of the final BDT model.

Figure 12: Training monitoring data of the final BDT model.

5.3.3 Evaluation of the BDT

The BDT output ranges from [0,1]. The optimal cut value needs to be determined. For this the following figure of merit (FOM) is used:

$$FOM = \frac{S}{\sqrt{S + B}}, \tag{31}$$

with S being the expected number of signal candidates remaining after a BDT cut and B the expected number of background candidates in the region of $\pm 2\sigma$ around the signal mass peak. The S is determined the following way:

$$S = N_{\text{data}} \cdot \frac{N_{\text{MC BDT}}}{N_{\text{MC}}}, \tag{32}$$

with N_{data} being the number of data in the signal region of $\mu \pm 2\sigma$ determined via a fit to the B_s^0 mass. The quantity $N_{\text{MC BDT}}$ is the number of events in the MC sample after applying a BDT cut and N_{MC} the total number of events in the MC sample. The quantity B is determined via a fit of the high mass sideband and extrapolating the PDF in a region $\pm 2\sigma$ around the central value of the signal fit and integrating in that interval. This metric is calculated for different cut values and the highest metric score is chosen for the optimal cut value of the BDT. The test sample is used for this evaluation. The result can be seen in Figure 13. The plot on the right shows the metric scores for BDT cut values up to 0.8. A broad peak at lower cut values is visible. This region is further investigated and the plot is shown on the right hand side of Figure 13. The best BDT cut value determined with the described method is 0.06. This value was used for the further analysis.

The BDT is trained with the goal to efficiently reject combinatorial background while
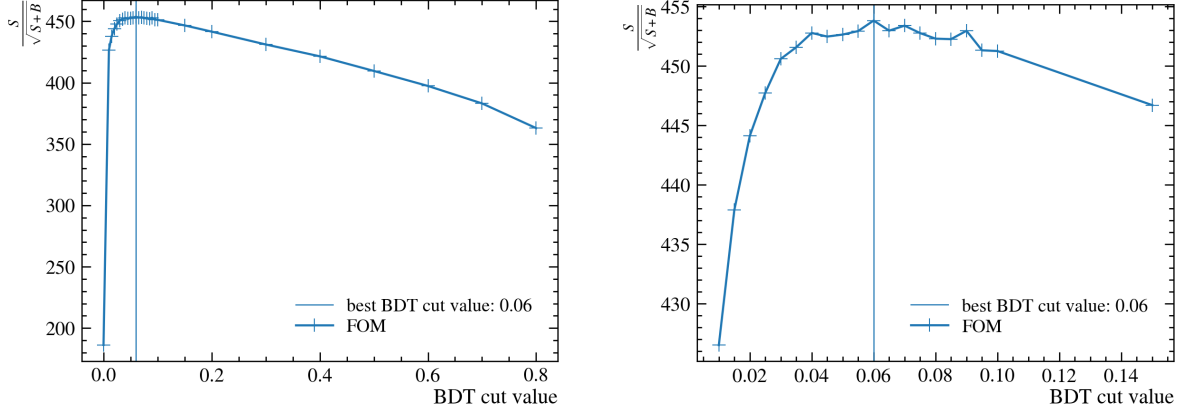
Figure 13: Determination of the best BDT cut value using the $S/\sqrt{S+B}$ metric. Interval up to a value of 0.8 (left) and more specific scan of the maximum region (right).

maintaining a high signal efficiency and avoiding overtraining. The rejection of the combinatorial background is quantised through the background rejection:

$$\varepsilon_{rej} = \frac{B_{tag}}{B}, \tag{33}$$

with B being the true background events of the test sample and B_{tag} the number of the former tagged as background by the BDT. In the same manner the signal efficiency is quantised through the signal efficiency

$$\varepsilon_{sig} = \frac{S_{tag}}{S}, \tag{34}$$

with S being the number of true signal candidates of the MC test sample and S_{tag} the number of true signal candidates tagged by the BDT as signal. The results can be seen in Figure 14.

The final BDT has a signal efficiency on MC signal between approximately 0.88 and 0.94. Considering the uncertainties of the calculated efficiency values, it can not be ruled out that the signal efficiency is described by a flat distribution over the signal region. The figure reveals minor deviations in bin to bin signal efficiency. The deviations show no clear trend over the signal region. Ideally the BDT would perform equally well throughout the signal region. This is not the case in this analysis, but the deviations are in an acceptable range. The background rejection shows a clear reduction of background rejection towards lower mass regions in the high mass sideband. This trend can be explained by the correlation of some training variables to the B_s^0 mass explained above. The signal and background events are now mixed and it is harder to separate them. By looking at the quantities, the background rejection is between approximately 0.996 and 0.9995. These values are close to the perfect background rejection score of 1 and the deviation between the highest and the lowest score is small. This leads to the conclusion that the BDT performs very well in terms of background rejection.

The BDT is now fully trained and evaluated. It can be applied to the data sample to efficiently remove combinatorial background. The effect the BDT classification cut has

27
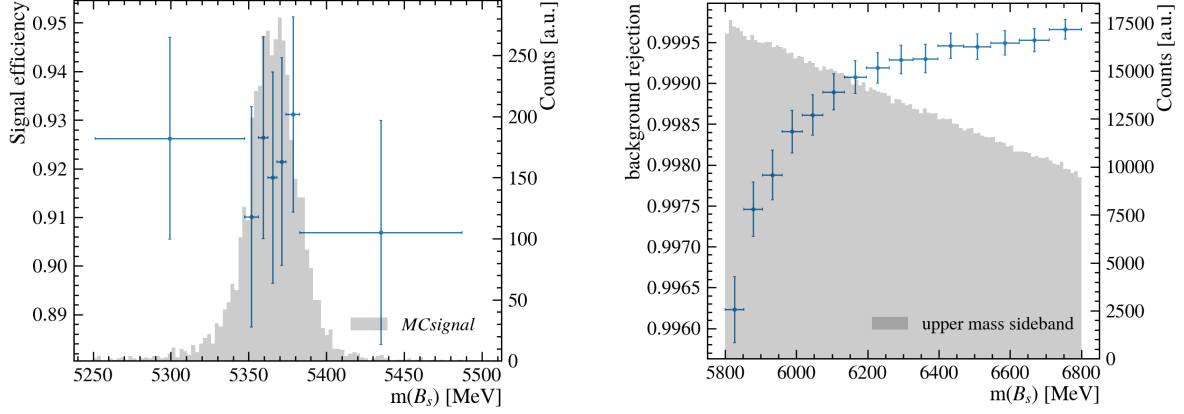
(a) Signal efficiency of the BDT.

(b) Background rejection of the BDT.

Figure 14: Evaluation of the BDT.

on the data sample is shown in Figure 15. The figure shows a rather simple distribution of data after applying the BDT cut. A visible reduction of background is present. The right hand plot has a logarithmic y-axis and reveals that the remaining background contributions can be approximated by an exponential function. This is especially useful for probability density function fits to the data distribution.
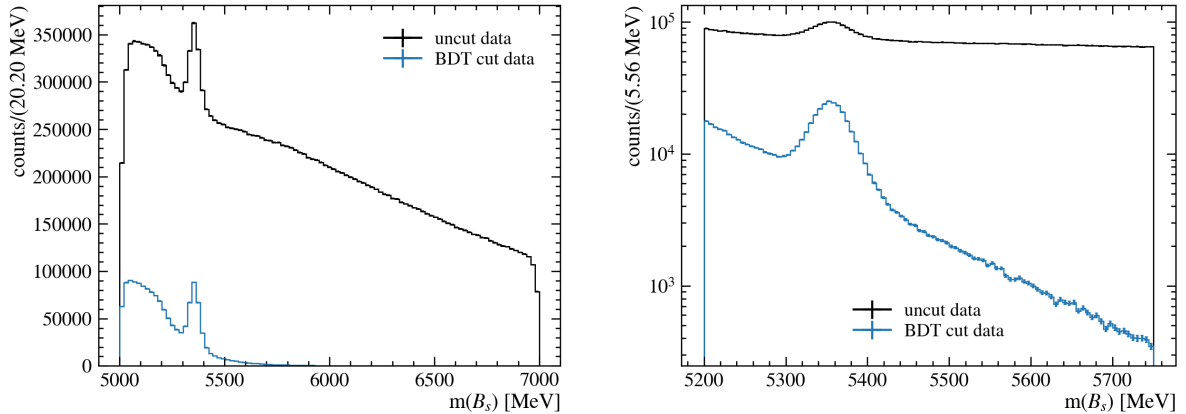

Figure 15: Effect of the BDT classification cut on data. Full mass range of data (left) and smaller selected interval plotted with logarithmic scale (right).

6 Fit to B_s^0 mass distribution

In order to calculate sWeights [24] for separating signal and background candidates in the distribution, a NLL fit to the B_s^0 mass distribution is performed. The following sections state the PDFs, which are used to model the B_s^0 mass distribution as well as the procedure of the fit. The *zFit* package [25] is used for the NLL fit, which provides a range of pre-defined PDFs.

6.1 Signal PDF

The signal PDF to describe the B_s^0 mass distribution is the sum of two PDFs. The first PDF is a Gaussian function, defined as follows:

$$\mathcal{G}(m; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(m-\mu)^2}{2\sigma^2}\right). \tag{35}$$

The second PDF is a double-sided Crystal Ball (DCB) function. The DCB function is a composite function of a Gaussian function and two exponential tails to each sides of the Gaussian function, defined as follows:

$$\mathcal{DCB}(m; \mu, \sigma, \alpha_L, n_L, \alpha_R, n_R) = \begin{cases} A_L \cdot \left(B_L - \frac{m-\mu}{\sigma}\right)^{-n_L} & ,\text{for } \frac{m-\mu}{\sigma} < -\alpha_L \\ \exp\left(-\frac{(m-\mu)^2}{2\sigma^2}\right) & ,\text{for} -\alpha_L \leq \frac{m-\mu}{\sigma} \leq \alpha_R \\ A_R \cdot \left(B_R - \frac{m-\mu}{\sigma}\right)^{-n_R} & ,\text{for } \frac{m-\mu}{\sigma} > -\alpha_R, \end{cases} \tag{36}$$

with

$$\begin{aligned} A_{L/R} &= \left(\frac{n_{L/R}}{|\alpha_{L/R}|}\right)^{n_{L/R}} \cdot \exp\left(-\frac{|\alpha_{L/R}|^2}{2}\right) \\ B_{L/R} &= \frac{n_{L/R}}{|\alpha_{L/R}|} - |\alpha_{L/R}| \end{aligned} \tag{37}$$

being scale parameters to ensure that the DCB is a continuous and differentiable function. A parameter f_m is introduced to scale the functions relative to each other in the process of adding them together. The parameters μ and σ are shared by the Gaussian and DCB PDF. The final PDF to describe the signal B_s^0 mass distribution is defined as:

$$\begin{aligned} \mathcal{PDF}_{m,\text{sig}}(m; \mu, \sigma, \alpha_L, n_L, \alpha_R, n_R, f_m) = f_m \cdot \mathcal{DCB}(m; \mu, \sigma, \alpha_L, n_L, \alpha_R, n_R) \\ + (1 - f_m) \cdot \mathcal{G}(m; \mu, \sigma). \end{aligned} \tag{38}$$

6.2 Background PDF

The background B_s^0 mass distribution in this analysis is modelled by an exponential function defined in the following way:

$$\mathcal{PDF}_{m,\text{bkg}}(m; \lambda) = A \cdot \exp(\lambda \cdot m), \tag{39}$$

with A as the normalisation factor depended on the mass range of the fit:

$$A = \int_{m_{min}}^{m_{max}} exp(\lambda \cdot m). \tag{40}$$

The final extended PDF to fit the B_s^0 mass distribution is the sum of the signal PDF $\mathcal{PDF}_{m,\text{sig}}$ and the background PDF $\mathcal{PDF}_{m,\text{bkg}}$ multiplied by their yields y_{sig} and y_{sig} respectively:

$$\mathcal{PDF}_m = y_{\text{sig}} \cdot \mathcal{PDF}_{m,\text{sig}} + y_{\text{bkg}} \cdot \mathcal{PDF}_{m,\text{bkg}}. \tag{41}$$

6.3 The fitting procedure

The NLL fit to the mass distribution of the B_s^0 meson is performed, by first fitting the parameters of the signal PDF $\mathcal{PDF}_{m,\text{sig}}$ to the MC sample. The resulting values of the parameters μ and σ are fixed to their pre-fitted values μ_{pre} and σ_{pre} respectively. Two new parameters are introduced to make differences between the pre-fit to the MC sample and the following fit to the data sample more accessible.

$$\mu' = \mu_{\text{pre}} + \text{shift} \qquad \sigma' = \sigma_{\text{pre}} \cdot \text{resolution factor}. \tag{42}$$

The parameters α_L, n_L, α_R and n_R of the double-sided Crystal Ball function for the final fit are Gaussian constraint within $\pm 3\sigma$ of the pre-fitted values. A Gaussian constrained is the likelihood of the estimated parameter value given the values set for the constraint (i.e. μ and σ). These Gaussian constraints are added to the loss function as penalty terms. The estimated parameters are assumed to follow a Gaussian distribution for this method. The results of the pre-fit are shown in Figure 16a and the parameter values are summarised in Table 9. The final fit to retrieve the sWeights from is shown in Figure 16b. The fit parameters are summarised in Table 10.

Looking at the NLL fit to the B_s^0 mass, the model describes the data quite well. Judging
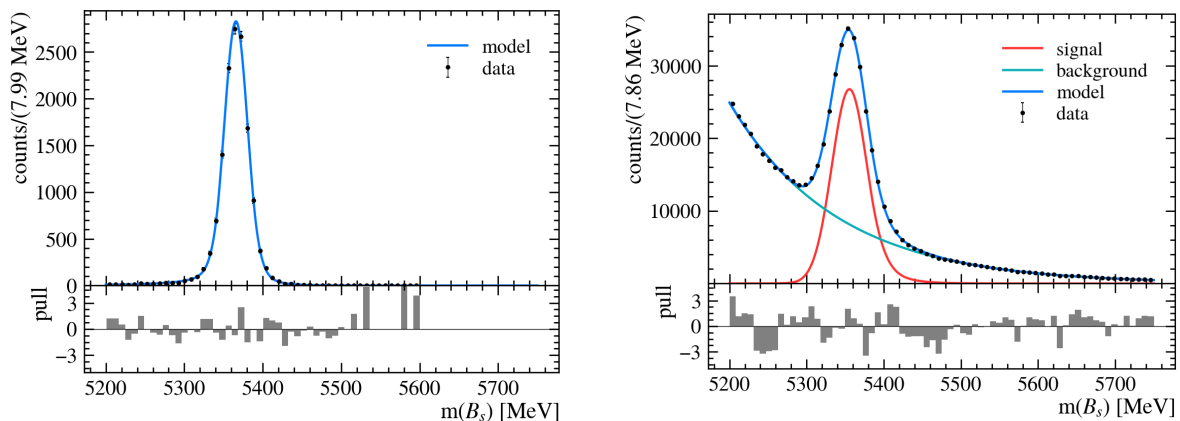
(a) Pre-fit of the signal PDF to the MC sample.

(b) NLL fit to the B_s^0 mass distribution.

Figure 16: Fits to the B_s^0 invariant mass distribution on MC sample and data sample.

from the pull distribution there are minor deviations between the PDF and the datapoints. The pull distribution of a perfect fit would show random deviations of around $\pm 1\sigma$ and

no trend of pulls in one direction. Regions where the datapoints are not well described by the PDF can be identified via a group of neighbouring pulls in the same direction. Such deviation are most visible in the transition of the Gaussian function into the exponential tail of the DCB function and at low mass values. The background in that region seems to be more complex than an exponential function. The main signal region and the upper mass background seems to be fitted well by the assumed PDF judging from the pull distribution in these mass regions. The fit is assumed to be valid for the calculation of sWeights.

Looking at the values for the newly added parameters defined in Equation 42 and summarised in Table 9 and 10, a shift of the mass peak of $\sim 10\,\mathrm{MeV}$ towards a lower mass can be observed and comparing the resolution of the B_s^0 mass, the uncertainty of the nominal mass increased by a factor of ~ 1.5 in the data fit.

These deviations can be explained by differences between the MC sample and the data sample caused by imperfections of the simulation process as well as imperfections in the detector efficiency and alignment. Especially the shift and the lower resolution of the B_s^0 mass are effects related to the flawed early alignment of the newly commissioned detector. These differences are assumed to introduce no bias into the measurement of the oscillation frequency Δm_s since the pre-fit to the MC sample is only used to get a more stable fit to the data sample and the further analysis is focused on the data sample.

7 Calibration and combination of flavour tagging algorithms

As explained in Section 4.3, flavour tagging is used to get the tagging decisions d and estimated mistag probability η of signal candidates. The flavour tagging algorithms used in this analysis are the SSKaon tagger, the OSKaon tagger and the OSMuon tagger, discussed in Section 4.3. To get a per-event mistag probability, the estimated mistag probability η needs to be calibrated on the specific data sample of this analysis. The calibration is needed for two reasons: firstly, the taggers are trained on MC simulated data samples and as already mentioned in Sec. 6, there are well-known discrepancies between the simulated sample and the data sample.

Secondly, the tagging algorithms are trained on the kinematics and topologies of specific decays. The OSKaon tagger for example is trained on the $B^+ \rightarrow J/\psi K^+$ decay. Applying the trained tagger to the $B_s^0 \rightarrow D_s^- \pi^+$ decay of this analysis might not give the same performance results compared to the training decay.

The *lhcb-ftcalib* package [26] is used for the calibration process, with the results shown in Figure 17. The x-axis represents the estimated mistag from the NN. On the y-axis, the measured mistag of the data sample is displayed. The measured mistag in a data sample can be determined by comparing the initial flavour of the B_s^0 meson to the tagging decision d. In the case of non-neutral B-meson decays, for example $B^+ \rightarrow J/\psi K^+$, the y-axis would be trivial, given the fact that the initial state of the B^+ meson can be determined by its decay products. The neutral B_s^0-meson on the other hand can oscillate and therefore the initial flavour needs to be determined using the mixing probability function:

$$\mathcal{P}(t) = \frac{1}{2}\left[1 - \frac{\cos(\Delta m_s t)}{\cosh(\frac{1}{2}\Delta\Gamma_s t)}\right], \tag{43}$$

together with the decay time information of the signal B_s^0 meson and information about its final flavour state. The mixing probability function is convoluted with the Gaussian resolution function, to model the decay time resolution:

$$\mathcal{P}(t) = \frac{1}{2}\left[1 - \frac{\cos(\Delta m_s t)}{\cosh(\frac{1}{2}\Delta\Gamma_s t)} \cdot \exp\left(-\frac{1}{2}\Delta m_s^2 \sigma_t^2\right)\right]. \tag{44}$$

Ideally the predicted mistag and the measured mistag would be distributed on the diagonal through the origin. Therefore the calibration uses a linear function to scale the mistag to the given data sample:

$$\omega(\eta) = p_0 + p_1(\eta - \langle\eta\rangle), \tag{45}$$

where p_0 and p_1 are floating fit parameters and the mean estimated mistag probability $\langle\eta\rangle$ is determined from the data sample. In the calibration process, fixed values for the oscillation frequency

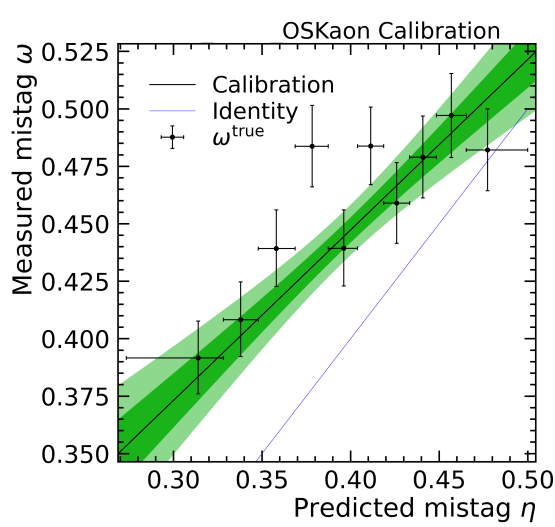$$\Delta m_s = 17.765 \pm 0.004 \pm 0.004 \, \text{ps}^{-1}$$

and the decay width

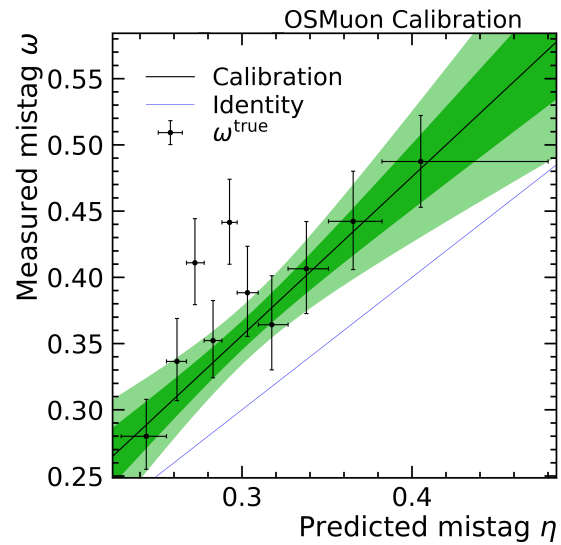$$\Delta\Gamma_s = 0.084 \pm 0.005 \, \text{ps}^{-1}$$

are used [27].

The calibration is performed on the sWeighted signal sample with the $B_s^0 \rightarrow D_s^- \pi^+$ decay,
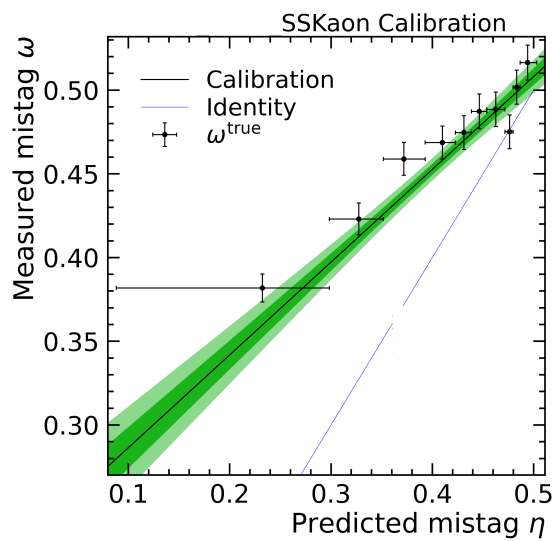
since no processed sample of a control decay is available in time for this thesis. This could induce a bias into the analysis due to the use of the fixed values for Δm_s and $\Delta \Gamma_s$.
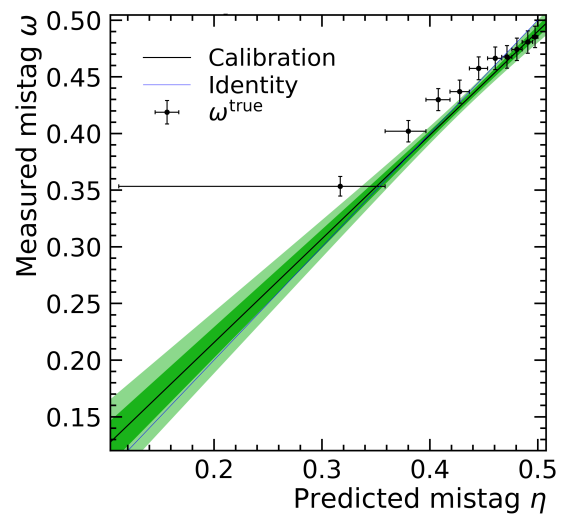

(a) Calibration of the OSKaon tagging algorithm.

(b) Calibration of the OSMuon tagging algorithm.

(c) Calibration of the SSKaon tagging algorithm.

(d) Calibration of the combined tagging algorithm.

Figure 17: Calibration of the flavour tagging algorithms.

The calibrated tagging performance of the individual taggers is summarised in Table 4. The individual calibrated OS taggers and SS taggers are combined to form a single

Tagger	Tagging efficiency [%]	Effective mistag [%]	Tagging power [%]
SSKaon	38.47 ± 0.14	43.4 ± 0.4	0.69 ± 0.08
OSKaon	14.91 ± 0.10	43.1 ± 0.6	0.28 ± 0.05
OSMuon	3.87 ± 0.05	34.9 ± 1.2	0.35 ± 0.06

Table 4: Calibrated tagging performances.

tagging decision with a corresponding mistag probability. To combine the flavour tagging decisions of each tagger i, the tagging algorithms are combined in the following manner:

$$
\begin{aligned}
p(d = -1) &= \prod_i \left(\frac{1 + d_i}{2} - d_i(1 - \eta_i) \right), \\
p(d = +1) &= \prod_i \left(\frac{1 - d_i}{2} + d_i(1 - \eta_i) \right).
\end{aligned}
\tag{46}
$$

To transform the quantities $p(d = \pm 1)$ into probabilities that the B_s^0 candidate contained a b-quark or an $\bar{b}$-quark at production, they need to be normalised:

$$
\begin{aligned}
\mathcal{P}(d = -1) &= \frac{p(d = -1)}{p(d = -1) + p(d = +1)}, \\
\mathcal{P}(d = +1) &= 1 - \mathcal{P}(d = -1).
\end{aligned}
\tag{47}
$$

The combined decision is made by comparing the probabilities of the B_s^0 meson to contain a b-quark or an $\bar{b}$-quark at production:

$$
d_{comb} = \begin{cases} +1 & \text{for} \quad \mathcal{P}(d = +1) > \mathcal{P}(d = -1) \\ -1 & \text{for} \quad \mathcal{P}(d = -1) > \mathcal{P}(d = +1). \end{cases}
\tag{48}
$$

The combined mistag probability η_{comb} is simply defined as:

$$
\eta_{comb} = 1 - \max \left[\mathcal{P}(d = +1), \mathcal{P}(d = -1) \right].
\tag{49}
$$

The combination of the flavour taggers is calibrated using the same linear calibration function as described above and the result is shown in Figure 17d.

The performance of the combined flavour tagger is summarised in Table 5. By combining

Tagging efficiency [%]	Effective mistag [%]	Tagging power [%]
44.77 ± 0.15	41.437 ± 0.021	1.31 ± 0.11

Table 5: Performance of the calibrated combination of taggers.

the single flavour taggers into one single tagger, an improvement in performance can be observed compared to the individual single taggers. The improvement in tagging efficiency is due to taking into account all events with at least one tagging decision, which results in

an increased tagging power and therefore a higher statistical power of the data sample. The tagging power observed is almost 4 times lower than in the published analysis using Run1 and Run2 data [9]. This big difference in performance can be explained by two reasons: firstly the published analysis used additional taggers, namely the OSCharm and the OSVertex taggers. The combination of only three taggers in this analysis is done due to the limited amount of available taggers for samples taken during the Run3 period. Secondly, the taggers of the previous analysis were trained on the same conditions as the data taking. The algorithms used in the analysis are trained with Run2 conditions and used for Run3 data. Further improvements on the final results are thus to be expected once the taggers trained on the Run3 upgraded conditions will become available.

8 Multidimensional fit

In order to measure the oscillation frequency of the B_s^0 meson Δm_s, a multidimensional fit to the decay time t of the B_s^0 meson as well as to the decay time resolution σ_t and to the calibrated mistag probability ω is performed. The following section provides information about the used PDFs and the fitting procedure.

8.1 PDFs for Multidimensional fit

The PDF for the multidimensional fit is made up of four contributions. The first contribution is the decay time PDF which accounts for the finite lifetime of the B_s^0 meson. The decay time distribution without considering effects of mixing can be described by an exponential function:

$$\mathcal{D}(t; \tau) \propto \exp\left(\frac{t}{\tau}\right). \tag{50}$$

The second contribution of the PDF, accounts for the mixing properties of the neutral B_s^0 meson described in Section 2. This gives rise to the following mixing decay PDF, which incorporates to the decay time PDF:

$$\mathcal{M}\mathrm{ix}(t, \omega; q, \tau, \Delta\Gamma_s, \Delta m_s) \propto \exp\left(\frac{t}{\tau}\right) \cdot \left(\cosh\left(\frac{\Delta\Gamma_s}{2}t\right) - q(1 - 2\omega)\cos(\Delta m_s t)\right). \tag{51}$$

The factor $(1 - 2\omega)$ is the dilution factor accounting for imperfect tagging. The influence of this dilution factor on the final PDF is discussed at the end of this section. The quantity q is used to account for the different mixing states of the B_s^0 meson. The case where $q = +1(-1)$ results in the mixing PDF for mixed (unmixed) candidates.

The third contribution to the multidimensional PDF is Gaussian PDF to model the decay time resolution of the LHCb detector. For this the mixing function is convoluted with the Gaussian PDF, which is equipped with a global parameter S_t to scale the per-event decay time resolution σ_t observed in data:

$$\mathcal{G}(t, \sigma_t; \mu, S_t) \propto \exp\left(-\frac{(t - \mu)^2}{2(S_t \cdot \sigma_t)^2}\right). \tag{52}$$

The convolution is done analytically before the fit. The PDF of the multidimensional fit up until now can be defined as:

$$\mathcal{P}(t, \omega, \sigma_t; q, \tau, \Delta\Gamma_s, \Delta m_s) \propto \exp\left(\frac{t}{\tau}\right) \cdot \left(\cosh\left(\frac{\Delta\Gamma_s}{2}t\right) - q(1 - 2\omega)\cos(\Delta m_s t)\right) \\ \otimes \mathcal{G}(t, \sigma_t; 0, S_t). \tag{53}$$

The last contribution is a decay time acceptance PDF, which accounts for time-biasing effects in the processing of the data. The sprucing cuts of the data sample introduce a bias on the decay time distribution, because cuts on variables like the impact parameter χ^2 (IP χ^2) have efficiencies dependent on the B_s^0 lifetime [28]. The IP χ^2 sprucing cut is more efficient for long lived B_s^0 mesons. This can be explained by the fact that the daughter particle tracks are more likely to point to the primary vertex for short lived B_s^0 mesons, so more short lived B_s^0 mesons pass this sprucing cut.

Also the reconstruction process of tracks has an efficiency depended on the decay time of

the B_s^0 meson. The reconstruction process of the VELO described in Section 3 assumes tracks with their origin close to the beam line. Displaced tracks from daughter particles of long lived B_s^0 mesons do not fulfill this assumption. The reconstruction is therefore slightly inefficient for long lived B_s^0 mesons.

The effects of the time-bias of the reconstruction and sprucing can be studied in detail and gives rise to a specific shape of the acceptance function. This is not done in this analysis. Instead an acceptance function with two floating parameters a_1 and a_2 is used. This decay time acceptance function is motivated based on the shape of the decay time distribution observed in data:

$$\mathcal{A}(t; a_1, a_2) = \arctan((a_1 \cdot t) + a_2). \tag{54}$$

The final mixing decay time PDF, used for the fit of the decay time distribution, is defined as:

$$\mathcal{P}(t, \omega, \sigma_t; q, \tau, \Delta\Gamma_s, \Delta m_s) \propto \exp\left(\frac{t}{\tau}\right) \cdot \left(\cosh\left(\frac{\Delta\Gamma_s}{2}t\right) - q(1 - 2\omega)\cos(\Delta m_s t)\right)$$
$$\otimes \mathcal{G}(t, \sigma_t; 0, S_t) \cdot \mathcal{A}(t; a_1, a_2) \cdot \Theta(0.5), \tag{55}$$

where the Heaviside function $\Theta(0.5)$ accounts for the lower limit of the decay time $t > 0.5\,\mathrm{ps}$ used in the fit.

To study the influence of the mistag probability ω and the decay time uncertainty σ_t on the mixing amplitude, a set of plots is created using the mixing decay time PDF with different conditions applied. No decay time acceptance function was used for the plots in Figure 18. Figure 18a shows the PDF with the assumption of perfect tagging ($\omega = 0$) and a perfect decay time resolution ($\sigma_t \approx 0$). A sharp separation between mixed and unmixed candidates is visible at each maxima and minima of the mixing PDF. The total number of candidates is decreasing as expected, due to the decay of the B_s^0 meson. Although this figure would be perfect to determine the oscillation frequency Δm_s, such data can not be recorded in a real world scenario.

Figure 18b shows the PDF under the assumption of a perfect decay time resolution and a realistic mistag probability ($\omega = 0.35$). The effect of the mistag probability on the amplitude can be studied. The increase in wrongly tagged events makes a sharp separation of mixed and unmixed candidates impossible. The dilution factor $(1 - 2\omega)$ decreases the amplitude of the mixing decay time PDF accordingly.
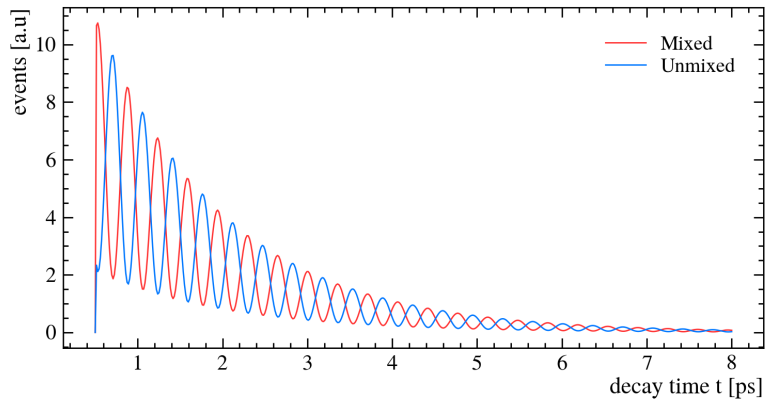
Figure 18c shows the influence of the decay time resolution on the mixing decay time PDF. Although a perfect mistag probability ($\omega=0$) was assumed, a clear separation between mixed and unmixed candidates is not observed, since the Gaussian convolution modelling the decay time uncertainty smears the decay time distribution. Mixed candidates can now smear into the distribution of unmixed candidates and vice versa. This results in the reduction of the mixing amplitude.
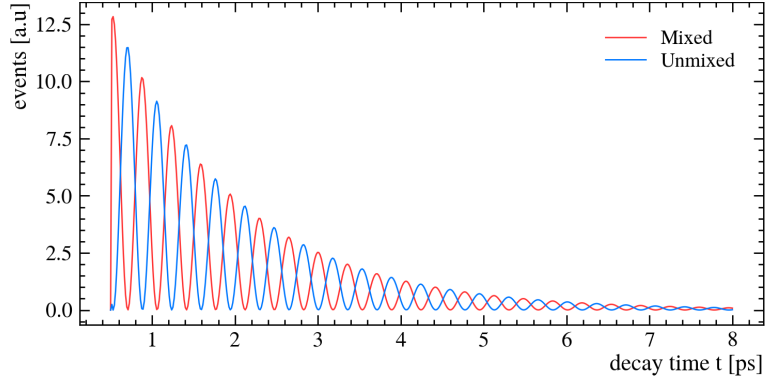
(a) Decay time distributions of mixed and unmixed signal candidates with perfect tagging and perfect decay time resolution.
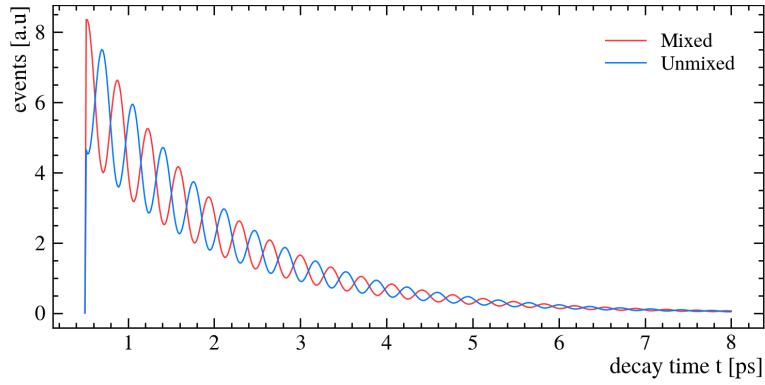
(b) Decay time distributions of mixed and unmixed signal candidates with realistic tagging and perfect decay time resolution.

(c) Decay time distributions of mixed and unmixed signal candidates with perfect tagging and realistic decay time resolution.

Figure 18: Influence of the mistag probability and the decay time uncertainty on the amplitude of the mixing decay time PDF.

8.2 The fitting procedure

To perform the multidimensional fit, the data sample needs to be split into a mixed and an unmixed sample based on the mixing state of the signal B_s^0 meson. For this a comparison between the tagging decision d representing the initial state and the sign of the particle

Observable	Limits
$m(B_s^0)$	$5200\,\mathrm{MeV} < m(B_s^0) < 5750\,\mathrm{MeV}$
t	$0.5\,\mathrm{ps} < t < 8.0\,\mathrm{ps}$
ω	$0.0 < \omega < 5.0$
σ_t	$0.0\,\mathrm{ps} < \sigma_t < 0.2\,\mathrm{ps}$

Table 6: Limits of the observables for the multidimensional fit.

identification (ID) number representing the final state is done. The tagging decision is $+1$, if the B_s^0 meson at production contained a $\bar{b}$-quark and -1 if the B_s^0 meson at production contained a b-quark. If the sign of the ID is $+1$, the B_s^0 meson contained a $\bar{b}$ and if the sign is -1, the B_s^0 meson contained a b-quark. The following two cases were treated as separate samples.

$$\begin{cases} d = \mathrm{ID} & \text{, for unmixed candidates} \\ d \neq \mathrm{ID} & \text{, for mixed candidates.} \end{cases} \tag{56}$$

Limits on the observables for the fit are set and summarised in Table 6. The fit consists of three parts: the first part is a fit to the B_s^0 mass distribution described in Section 6. With this fit, a set of sWeights [29] is calculated for the mixed and the unmixed sample separately. For the further fit only signal weighted candidates are used.

The second part of the fit is the determination of the acceptance function parameters a_1 and a_2. For this the acceptance function is multiplied with the decay PDF $\mathcal{D}$ and a Heaviside function:

$$f(t; a_1, a_2, \tau) = \mathcal{A}(t; a_1, a_2) \cdot \mathcal{D}(t; \tau) \cdot \Theta(0.5). \tag{57}$$

This function is fit to the combined mixed and unmixed distribution. Signal weights calculated for the combined sample are used. The resulting parameter estimates for a_1 and a_2 are fixed for the further fit.

The third part is the multidimensional fit in the observables t, ω and σ_t. The multidimensional fit is done simultaneously for the mixed and the unmixed sample. The PDF defined in Equation 55 is used with $q = +1$ for the mixed sample and $q = -1$ for the unmixed sample. The parameter $\Delta\Gamma_s$ is set the average value calculated by the Particle Data Group (PDG) [30].

$$\Delta\Gamma_s = (0.083 \pm 0.005)\,\mathrm{ps}^{-1}.$$

This is a combined value of multiple measurements. These measurements were made using different decay channels with higher branching fractions and also data with a much higher luminosity compared to this analysis. Therefore a more precise measurement of the $\Delta\Gamma_s$ value with the data of this analysis is not possible. The lifetime of the B_s^0 meson $\tau(B_s^0)$, the global scale factor for the decay time uncertainty S_t and the yields of the mixed and unmixed PDF as well as the oscillation frequency Δm_s are left floating in this fit.

9 Analysis results

The results of the multidimensional fit are summarised and discussed in this section. The oscillation frequency Δm_s of the B_s^0 meson is directly determined with the NLL fit. Figure 19 shows the mixed and unmixed PDFs plotted with the decay time distribution of the mixed and unmixed candidates. Since the fit treats the decay time resolution and mistag probability as conditional parameters, rather then fully modelled additional dimensions, the fitted parameters do not contain the full information of the fit. As such, the plot in Figure 19 is not a profile plot. Nonetheless, by averaging over the conditional variables as a function of the decay time, the figure can be interpreted similarly to a profile plot.
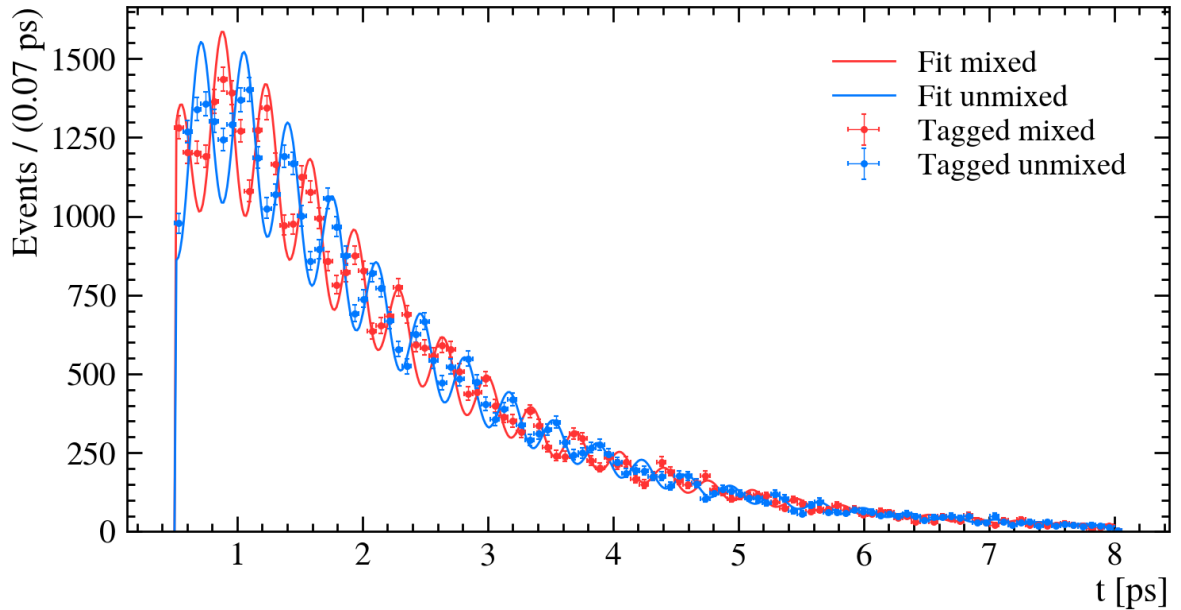

Figure 19: Fit of the multidimensional NLL fit to the mixed and unmixed sample to the decay time distribution.

A difference in amplitude between the PDFs and the data sample is especially visible for low decay time values. This deviation gets smaller towards higher decay times. Possible sources for this discrepancy of the amplitude are discussed in the following:
firstly, the decay time acceptance function is able to change the overall shape of the PDF. The effect of the acceptance PDF Eq. 54 on the deviation in amplitude between data and the fitted PDF is assumed to be small judging from the pre-fitted acceptance function. The fit is shown in Figure 20 and the acceptance PDF describes the decay time distribution fairly well. The acceptance function is pre-fitted to the combined mixed and unmixed sample and assumed to be the same for the mixed and unmixed sample separately.

Secondly, the discrepancy in amplitude can be a consequence of the decay time uncertainty model. The decay time uncertainty influences the PDF through the convolution of the mixing decay time PDF with the Gaussian function. This Gaussian function accounts for the decay time resolution of the LHCb detector and shapes the amplitude of the PDF as discussed in Section 8.1. A global scale factor S_t is multiplied to
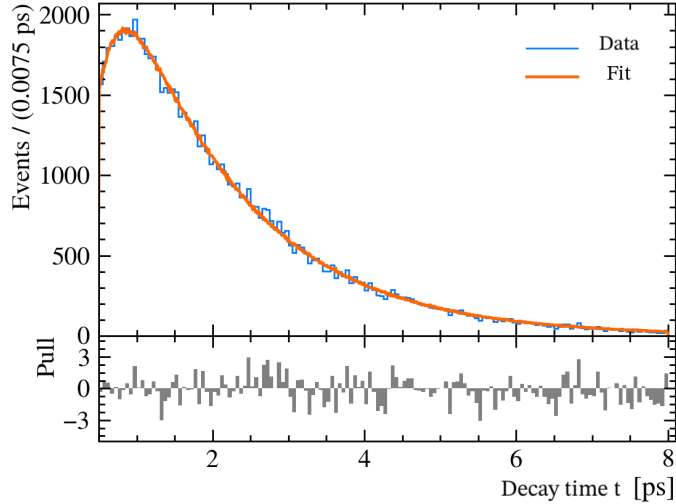
40

Figure 20: Pre-fit of the decay time acceptance function.

the per-event decay time uncertainty σ_t. The deviation in amplitude could be explained by assuming that the Gaussian convolution with a global scale factor S_t does not describe the decay time uncertainty of the detector correctly.

Thirdly, the difference in amplitude between the PDF and the data sample can be a described as a consequence of calibrating the flavour tagging algorithms. The mistag probability ω influences the amplitude of the mixing function through the dilution factor $\mathcal{D} = (1 - 2\omega)$. As already discussed in Section 7 the calibration of the flavour taggers is done on the $B_s^0 \rightarrow D_s^- \pi^+$ decay due to the the absence of an available processed control channel. Also the flavour tagging in this analysis is done using the taggers of Run2, which are essentially trained for a different detector. A systematic study of this effect is not the scope of this analysis. Both the use of the Run2 taggers and the calibration without a control channel effect the calibration of the mistag probability ω and therefore have an influence on the amplitude of the mixing decay time PDF as seen in Section 8.1.

Lastly, as this study required many of the components to be implemented in Python, a simple implementation error cannot be conclusively ruled out at this point.

Nonetheless influences on the amplitude are assumed to have no effect on the oscillation frequency Δm_s.

Figure 21 shows the correlation of the fit parameters among themselves. A weak positive correlation between the lifetime of the B_s^0 meson and the decay time uncertainty scale factor S_t is visible. This correlation is also assumed to not influence the determination of the Δm_s.
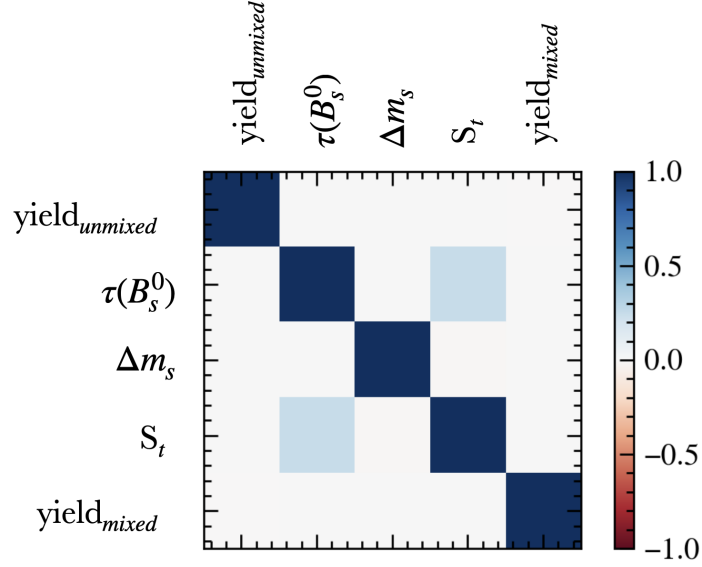

Figure 21: Correlation of the parameters of the multidimensional fit.

Looking at the oscillation frequency in Figure 19, the PDF is able to describe the frequency of the oscillation quite well. The resulting parameters of the fit are summarised in Table 7.

Parameter	Value
Δm_s	$(17.792 \pm 0.011)\ \mathrm{ps}^{-1}$
$\tau(B_s^0)$	$(1.576 \pm 0.006)\ \mathrm{ps}$
S_t	1.06 ± 0.04
y_{mixed}	42403 ± 210
y_{unmixed}	$41937 - 200 + 210$

Table 7: Summarised fit parameters of the multidimensional fit.

A comparison of the measured value for the oscillation frequency Δm_s and the world average value is performed. The current world average value is:

$$\Delta m_{s\,\mathbf{world\ average}} = (17.765 \pm 0.006)\,\mathrm{ps}^{-1}.$$

The sigma deviation between these two quantities is:

$$\sigma = \frac{|\Delta m_s - \Delta m_{s\textbf{world average}}|}{\sqrt{\sigma^2 + \sigma^2_{\textbf{world average}}}} = 2.15. \tag{58}$$

With a sigma value less than 3, the measured value is in reasonable agreement with the world average value within their uncertainties.

9.1 Validation of the estimated fit uncertainty

The following chapters provide measurements of the parameter estimate uncertainty. The measurements are performed using two different methods, namely the "bootstrapping" method and the "$\Delta \ln L = -1/2$" method. The uncertainties of both methods are compared to the uncertainties estimated during the NLL fit based on the Rao-Cramer-Frechet minimum variance bound method Eq. 22, to validate the estimated fit uncertainty.

9.1.1 Bootstrapping method

The statistical uncertainty of the last stage of the multidimensional fit is studied using the bootstrapping method. Test samples are produced with the mixing decay time PDF, given in Equation 55, using the world average values and a decay time uncertainty scale $S_t = 1$. The sample sizes are set corresponding to the yields of the final fit. The mixing decay time PDF is then fitted to the newly created samples with the same conditions of the multidimensional fit. This fit is performed 100 times to get 100 sets of estimated parameters. To evaluate the uncertainty of the fit, the pull distribution of the three variables Δm_s, $\tau(B_s^0)$ and S_t is studied. The pull of a parameter A in this context is defined as:

$$\text{Pull}(A) = \frac{A_{\text{fit}} - A_{\text{true}}}{\sigma_{\text{fit}}}, \tag{59}$$

with A_{true} being the value, the sample is created with and σ_A the uncertainty estimated from the *Minuit* package in the fitting process. The resulting plots are shown in Figure 22 and the fit parameters are summarised in Table 8. If the uncertainty is correctly

	Δm_s	$\tau(B_s^0)$	S_t
μ	-0.09 ± 0.10	-0.04 ± 0.09	-0.05 ± 0.10
σ	$1.02 - 0.07 + 0.09$	$0.91 - 0.06 + 0.07$	$0.94 - 0.07 + 0.08$

Table 8: Fit parameters of the pull fits.

estimated during the fit, the σ value of the pulls should be approximately one and the mean value of the pull distributions should be around zero. The pull distributions of all three parameters show a roughly Gaussian behaviour with a mean value of zero and a σ value of around one in the range of their uncertainties. However, a slight skewness is observed in the pull distributions for the lifetime parameter $\tau(B_s^0)$ and the oscillation frequency Δm_s, which could hint to asymmetric uncertainties of these parameters or that the estimated uncertainty intervals do not provide proper coverage. Nonetheless, the small deviations from a Gaussian distributions can be explained by the relatively small sample size of the bootstrapping method.
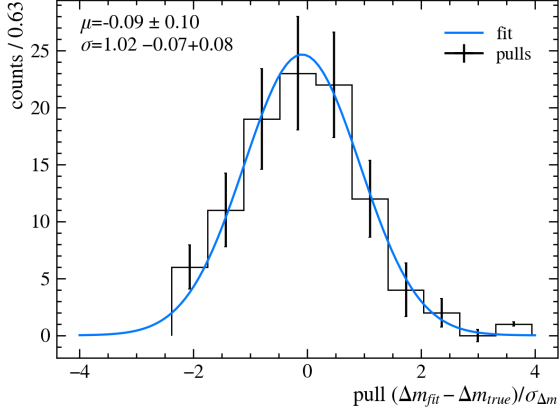
(a) Pull distribution of Δm_s.
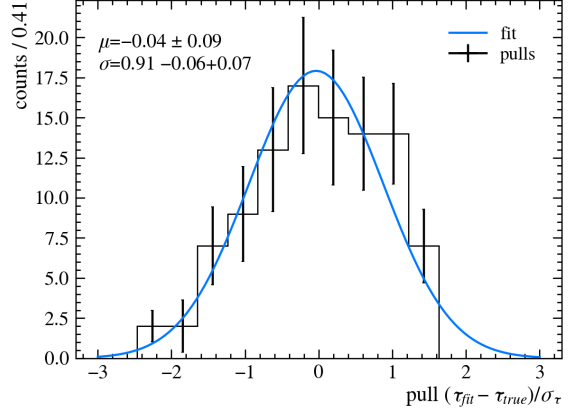
(b) Pull distribution of $\tau(B_s^0)$.
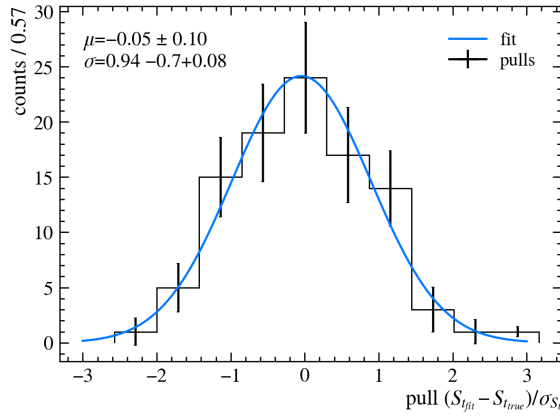
(c) Pull distribution of S_t.

Figure 22: Pull distributions of the bootstrapping method.

9.1.2 $\Delta \ln L = -1/2$ method

To further evaluate the estimated uncertainty of the NLL fit, the $\Delta \ln L = -1/2$ method is used. The NLL function is Taylor expanded around the function minimum. This yields the following formula:

$$- \ln L(\theta) \approx - \ln L(\hat{\theta}) - \frac{1}{2} \left[\frac{\partial^2 \ln L}{\partial^2 \theta} \right]_{\theta = \hat{\theta}} (\theta - \hat{\theta})^2 + \dots, \tag{60}$$

since the first derivative term is by definition zero in the minimum case where $\theta = \hat{\theta}$. Considering the minimum variance bound, described in Equation 22, the equation above can be written as:

$$- \ln L(\theta) \approx - \ln L_{\max} + \frac{(\theta - \hat{\theta})^2}{2\sigma_{\hat{\theta}^2}}. \tag{61}$$

Rewriting this above stated equation yields the rather simple correspondence between the NLL minimum and the uncertainty of the estimated parameter:

$$- \ln L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) \approx - \ln L_{\max} + \frac{1}{2} \tag{62}$$
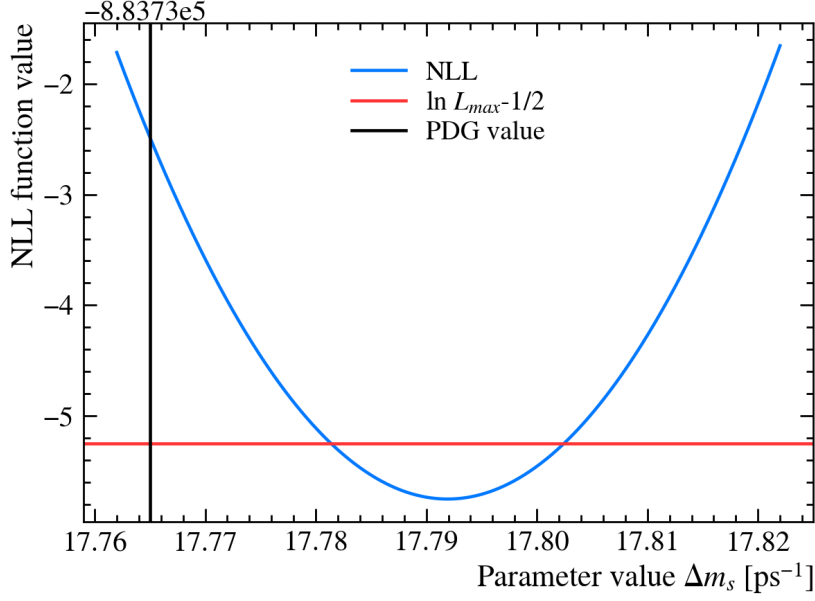
44

Figure 23: Determination of the Δm_s uncertainty using the $\Delta \ln L = -1/2$ method.

The uncertainty of the parameter can therefore be determined using the parameter values where the NLL minimum has increased by $\frac{1}{2}$. The resulting plot of this method can be seen in Figure 23.

The parameters, where the minimum of the NLL has increased by $\frac{1}{2}$ are

$$\Delta m_{s\text{lower}} = 17.7814 \, \text{ps}^{-1} \quad \text{and} \quad \Delta m_{s\text{upper}} = 17.8024 \, \text{ps}^{-1}.$$

The resulting uncertainties of the Δm_s values are

$$\sigma_{\Delta m_s, \text{lower}} = 0.0106 \, \text{ps}^{-1} \quad \text{and} \quad \sigma_{\Delta m_s, \text{upper}} = 0.0104 \, \text{ps}^{-1}.$$

These uncertainty estimates agree fairly well with the estimated uncertainty of $\pm 0.011 \, \text{ps}^{-1}$ calculated during the minimisation process. This process also reveals the symmetric nature of the NLL minimum.

10 Conclusion

The goal of this thesis is to measure the Δm_s oscillation frequency of the B_s^0-$\overline{B}_s^0$ system using the early 2024 data of the upgraded LHCb experiment.

A 2024 data sample with an integrated luminosity equivalent of $1.73\,\text{fb}^{-1}$ is used for this analysis. The oscillation frequency is studied on the $B_s^0 \to D_s^- (K^+K^-\pi^-)\pi^+$ decay channel. To reduce the background contributions in the data sample, a preselection is performed including veto cuts to effectively remove background decays, such as $B^0 \to D^- (K^+\pi^+\pi^-)\pi^+$, from the data sample. The removal of combinatorial background in the data sample is performed by a boosted decision tree (BDT). The BDT is trained on a sample containing MC simulated signal candidates and using candidates from the high mass sideband in data as background candidates. With the help of a Kolmogorow-Smirnow test, variables with big differences in signal and background distributions are determined and used as training variables for the BDT. The flavour tagging algorithms are calibrated on the data sample using sWeights and the $B_s^0 \to D_s^- \pi^+$ decay, since no processed sample of a control decay is available in time for this thesis. A multidimensional fit is implemented to measure the oscillation frequency Δm_s in the data sample. The first step of the fit is to calculate sWeights using a NLL fit to the B_s^0 mass distribution for mixed and unmixed candidates. A decay time acceptance function used in the mixing decay time PDF is pre-fitted to the decay time distribution of the B_s^0 signal candidates. A simultaneous fit to the decay time distribution of mixed and unmixed signal candidates is performed taking into account effect of the calibrated mistag probability ω and the decay time uncertainty σ_t. The oscillation of the B_s^0 meson is both visible in the mixed and unmixed data samples as well as in the fit of the mixing decay time PDF.

A comparison of the fitted mixing decay time PDF with the data points revealed differences in the mixing amplitudes. Possible sources of this discrepancy were investigated without finding a conclusive cause. An in depth analysis of the difference in the mixing amplitude needs to be done in the future.

The main goal of this analysis was to show that the oscillation of the B_s^0 meson can be resolved in the early 2024 data recorded with the upgraded LHCb detector. This goal was reached within this analysis.

Furthermore a good measurement of the oscillation frequency Δm_s could be performed, despite observing problems in modelling the mixing amplitude and using a limited amount of available tagging tools. The oscillation frequency was measured to be

$$\Delta m_s = (17.792 \pm 0.011)\,\text{ps}^{-1}.$$

The uncertainty of the fit parameter was validated using two different methods. With a value of less than 3σ, the measured value is in reasonable agreement within its uncertainty with the world average value of Δm_s.

Appendix

This appendix shows the parameters of the fit to the B_s^0 mass distributions, discussed in Section 6. Table 9 shows the estimated parameters of the pre-fit to the MC sample and Table 10 lists the parameters of the final fit to the B_s^0 mass distribution seen in data.

Parameter	Value
μ_pre	(5365.76 ± 0.16) MeV
σ_pre	(14.91 ± 0.19) MeV
α_L	2.3 ± 0.3
n_L	2.2 ± 0.6
α_R	1.19 ± 0.07
n_R	10.7 ± 1.5
f	0.53 ± 0.12

Table 9: Fit parameters for the pre-fit to the MC sample.

Parameter	Value
μ_pre	(5365.76 ± 0.16) MeV
shift	(-10.32 ± 0.09) MeV
σ_pre	(14.91 ± 0.19) MeV
resolution factor	1.531 ± 0.006
α_L	2.3 ± 0.3
n_L	2.2 ± 0.6
α_R	1.19 ± 0.07
n_R	10.7 ± 1.5
f	0.52 ± 0.16
y_sig	202200 ± 1100
λ_bkg	(-0.00719 ± 0.00015) MeV^{-1}
y_bkg	431300 ± 12000

Table 10: Fit parameters for the fit to the data sample with Gaussian constraints to calculate sWeights.

References

[1] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013.

[2] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652.

[3] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531.

[4] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13** (1964) 138.

[5] CDF collaboration, A. Abulencia *et al.*, *Observation of $B_s^0 - \bar{B}_s^0$ Oscillations*, Phys. Rev. Lett. **97** (2006) 242003, arXiv:hep-ex/0609040.

[6] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the B_s^0–$\overline{B}_s^0$ oscillation frequency in the decay $B_s^0 \to D_s^- \pi^+$*, New J. Phys. **15** (2013) 053021, arXiv:1304.4741.

[7] LHCb collaboration, R. Aaij *et al.*, *Observation of B_s^0–$\overline{B}_s^0$ mixing and measurement of mixing frequencies using semileptonic B decays*, Eur. Phys. J. **C73** (2013) 2655, arXiv:1308.1302.

[8] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle γ and B_s^0-$\overline{B}_s^0$ mixing frequency with $B_s^0 \to D_s^{\mp} h^{\pm} \pi^{\pm} \pi^{\mp}$ decays*, JHEP **03** (2021) 137, arXiv:2011.12041.

[9] LHCb collaboration, R. Aaij *et al.*, *Precise determination of the B_s^0-$\overline{B}_s^0$ oscillation frequency*, Nature Physics **18** (2022) 1, arXiv:2104.04421.

[10] LHCb collaboration, R. Aaij *et al.*, *Improved measurement of CP violation parameters in $B_s^0 \to J/\psi K^+ K^-$ decays in the vicinity of the $\phi(1020)$ resonance*, Phys. Rev. Lett. **132** (2024) 051802, arXiv:2308.01468.

[11] *File:Standard Model of Elementary Particles.svg - Wikipedia — en.wikipedia.org*, https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg#file. [Accessed 10-09-2024].

[12] M. Q. Führing, *Decay-time-dependent studies of strange beauty mesons*, 2023.

[13] LHCb collaboration, R. Aaij *et al.*, *The LHCb Upgrade I*, JINST **19** (2024) P05065, arXiv:2305.10515.

[14] R. Lindner, *LHCb layout$_2$. LHCb schema$_2$*, LHCb Collection., 2008.

[15] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.

[16] L. Calefice *et al.*, *Effect of the high-level trigger for detecting long-lived particles at lhcb*, Frontiers in Big Data **5** (2022) 1008737.

[17] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), 785–794, ACM, 2016.

[18] H. Dembinski and P. O. et al. *scikit-hep/iminuit*, doi: 10.5281/zenodo.3949207.

[19] C. Radhakrishna Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc. **37** (1945) 81.

[20] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

[21] T. Brambach *et al.*, *Measurement of the B_d^0 mixing frequency Δm_d in $B^0 \to J/\psi K^{*0}$ and $B^0 \to D^-\pi^+$*, .

[22] L. Anderlini *et al.*, *Computing strategy for PID calibration samples for LHCb Run 2* , CERN, Geneva, 2016.

[23] J. L. Hodges, *The significance probability of the smirnov two-sample test*, Arkiv för Matematik **3** (1958) 469.

[24] E. Rodrigues *et al.*, *The Scikit HEP Project – overview and prospects*, EPJ Web Conf. **245** (2020) 06028, arXiv:2007.03577.

[25] J. Eschle, A. N. Puig, R. Silva Coutinho, and N. Serra, *zfit: scalable pythonic fitting*, EPJ Web Conf. **245** (2020) 06025.

[26] Q. Führing and V. Jevtić, *lhcb-ftcalib: A software package for the calibration of flavour-tagged LHCb data*, 2024. doi: 10.5281/zenodo.12156328.

[27] Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021*, Phys. Rev. **D107** (2023) 052008, arXiv:2206.07501, updated results and plots available at https://hflav.web.cern.ch.

[28] F. Dordei, *Lifetime measurements of beauty hadrons at the LHCb experiment*, 2014. Presented 19 May 2015.

[29] F. Pedregosa *et al.*, *Scikit-learn: Machine learning in Python*, J. Machine Learning Res. **12** (2011) 2825, arXiv:1201.0490, and online at http://scikit-learn.org/stable/.

[30] Particle Data Group, N. S. *et al.*, *Review of particle physics*, to be published in Phys. Rev **D 110** (2024) 030001.

Erklärung

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

Heidelberg, den 23.09.2024,