technische universität
dortmund

Measurements of Decays in the $B \to D\overline{D}$ Family at the LHCb Experiment

Dissertation zur Erlangung des akademischen Grades
Dr. rer. nat.

vorgelegt von
Antje Mödden

Fakultät Physik
Technische Universität Dortmund

Juni 2022

Der Fakultät Physik der Technischen Universität Dortmund zur Erlangung des akademischen Grades eines Dr. rer. nat. vorgelegte Dissertation.

Betreuer:	Prof. Dr. Bernhard Spaan
1. Gutachter:	Prof. Dr. Johannes Albrecht
2. Gutachter:	Prof. Dr. Kevin Kröninger

Abstract

Precision measurements of CKM-matrix elements are used to test the Standard Model of particle physics. Decays of B_s^0 mesons that proceed via $b \to c\bar{c}s$ transitions are used to measure the *CP*-violating weak mixing phase ϕ_s that is directly related to the CKM angle β_s and has a precise theoretical prediction. The mixing phase ϕ_s can be measured using $B_s^0 \to D_s^+ D_s^-$ decays, where higher-order Standard Model effects can be assessed by exploiting results from analyses in additional decays of the $B \to D\overline{D}$ family. These analyses comprise *CP*-violation and branching-ratio measurements. Two different analyses of $B \to D\overline{D}$ decays are presented, which utilise data collected by the LHCb experiment.

A search for the $B_s^0 \to D^{*\pm} D^{\mp}$ decay is presented using data corresponding to an integrated luminosity of $9\,\text{fb}^{-1}$ collected at centre-of-mass energies of 7, 8 and 13 TeV. The $B_s^0 \to D^{*\pm} D^{\mp}$ decay is observed and its branching fraction is measured for the first time.

The ongoing measurement of *CP* violation in $B_s^0 \to D_s^+ D_s^-$ decays uses data corresponding to $6\,\text{fb}^{-1}$ collected at 13 TeV. The measurement of the *CP* parameters is still kept blind, but the sensitivities to the *CP* parameters are evaluated and indicate that this will be the most precise measurement in this decay channel.

Kurzfassung

Präzisionsmessungen von Parametern der CKM Matrix werden durchgeführt, um das Standardmodell der Teilchenphysik zu untersuchen. Zerfälle von B_s^0-Mesonen mit $b \to c\bar{c}s$ Übergängen werden verwendet, um die *CP*-verletzende schwache Mischungsphase ϕ_s zu bestimmen, welche direkt mit dem CKM Winkel β_s zusammenhängt und eine präzise theoretische Vorhersage hat. Die schwache Mischungsphase ϕ_s kann in $B_s^0 \to D_s^+ D_s^-$ Zerfällen gemessen werden, wobei Effekte höherer Ordnung durch Ergebnisse von Analysen zusätzlicher Zerfälle der $B \to D\overline{D}$ Familie abgeschätzt werden können. Diese Analysen umfassen Messungen von *CP*-Verletzung und Verzweigungsverhältnissen. Es werden zwei unterschiedliche Analysen solcher Zerfälle vorgestellt, welche Daten des LHCb-Experiments nutzen.

Die Analyse von $B_s^0 \to D^{*\pm} D^{\mp}$ Zerfällen verwendet Daten, die einer integrierten Luminosität von $9\,\text{fb}^{-1}$ entsprechen und bei Schwerpunktsenergien von 7, 8 und 13 TeV aufgenommen wurden. Der $B_s^0 \to D^{*\pm} D^{\mp}$ Zerfall wird erstmals beobachtet und das Verzweigungsverhältnis gemessen.

Die Messung von *CP* Verletzung in $B_s^0 \to D_s^+ D_s^-$ Zerfällen ist noch nicht abgeschlossen. Es werden Daten verwendet, die $6\,\text{fb}^{-1}$ entsprechen und bei einer Schwerpunktsenergie von 13 TeV aufgenommen wurden. Die Messung der *CP*-Parameter ist noch blind, aber die Sensitivitäten können evaluiert werden. Es wird die präziseste Messung von *CP*-Verletzung in diesem Zerfallskanal erwartet.

Table of contents

1 Introduction

The Standard Model of particle physics (SM) is a theoretical model developed in the 1960s that describes all fundamental particles and the interactions between them [1–3]. Over the last decades, the SM has been intensively tested through many experiments and has been proven to be a successful theory on the smallest scales. It provided predictions of several particles that all have been confirmed. The Higgs boson was already postulated in 1964 [4,5] and with its experimental observation in 2012 [6,7] the last missing particle of the SM was found.

Despite this success, the SM is known to be incomplete as it lacks explanations for several open questions. For example, the SM does not include gravity. Gravity is described by the general theory of relativity, which is difficult to integrate with the SM to find a unified theoretical description of all fundamental interactions. Furthermore, neutrinos are massless in the SM, but the observation of neutrino oscillations [8,9] implies that neutrinos in fact have a non-zero mass. Moreover, all matter that is described by the SM only accounts for 5% of the energy density of the universe [10] and no explanation for the remaining components — dark matter and dark energy — is provided by the SM. Another open question is the observed matter-antimatter asymmetry in the universe. Although matter and antimatter must have been produced in equal amounts in the Big Bang, today's universe is dominated by matter. Three necessary conditions to explain the matter-antimatter asymmetry were postulated in 1967 [11]: a deviation from the thermal equilibrium at some point of the evolution of the universe, violation of baryon number conservation, and violation of the charge symmetry C and the combination of charge and parity symmetry CP. In the SM, CP violation is manifested in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which describes the probability of quark-flavour transitions. However, the amount of CP violation in the SM is not large enough to account for the matter-antimatter asymmetry. This indicates the existence of physics beyond the SM, also referred to as New Physics.

The Large Hadron Collider (LHC), the world's most powerful particle accelerator, is designed to perform searches for New Physics. The so-called direct searches aim to measure unknown particles produced in the collisions. These kind of searches are limited by the centre-of-mass energy of the collider. So far, no New Physics particles have been found, making the so-called indirect searches all the more important. Here, precision measurements of SM observables are performed and compared with the SM predictions. Deviations can be caused by New Physics particles that appear in loop processes and are not limited by the available energy scale. Thus, far higher energy scales can be probed than which are accessible in direct searches. The LHCb experiment at the LHC is specialised for these indirect searches. It is designed to perform precision measurements of *e.g.* CP violation and branching fractions in decays of b and c hadrons. Data collected by the LHCb experiment is utilised in this thesis.

One promising probe of the SM is the measurement of the weak mixing phase ϕ_s, which is directly related to the CKM parameter β_s and has a very precise SM prediction. It is predicted to be small in the SM and measurements of deviating values can be an indicator of New Physics effects. This measurement is commonly performed in B decays to charmonium final states with the most precise measurements being performed in $B_s^0 \to J/\psi\,\phi$, $B_s^0 \to J/\psi\,KK$ and $B_s^0 \to J/\psi\,\pi\pi$ decays [12–15]. The measurement can also be performed using $B_s^0 \to D_s^+ D_s^-$ decays [16]. As experimental precision improves, pollution from higher-order SM contributions will cause the leading uncertainty in the measurements of ϕ_s. It will be mandatory to control these contributions and to exploit a wide range of decay modes that give access to ϕ_s. Additionally, it is beneficial to perform the measurement in decay modes which are sensitive to different higher-order contributions to assess these effects. While contributions entering through the electroweak penguin sector can affect the measurements in the $B_s^0 \to J/\psi\,\phi$ decay mode, they are colour-suppressed in the $B_s^0 \to D_s^+ D_s^-$ decay mode [17]. This difference in the electroweak penguin sector makes the measurement of ϕ_s in $B_s^0 \to D_s^+ D_s^-$ decays attractive, though the decay suffers from lower selection efficiencies and more backgrounds. Also, higher-order contributions in $B_s^0 \to D_s^+ D_s^-$ decays can be constrained utilising additional $B \to D\overline{D}$ decays. CP-violation measurements in the family of $B \to D\overline{D}$ decays provide information to fundamental SM parameters and access to New Physics effects [17–20]. These measurements can be used to relate different CP-violating $B \to D\overline{D}$ decays and constrain hadronic effects [21, 22]. Also, branching ratio measurements of $B \to D\overline{D}$ decays are useful to quantify these hadronic effects to obtain precise results [21, 22].

In this thesis, two measurements in the field of $B \to D\overline{D}$ decays are presented. First, a search for the $B_s^0 \to D^{*\pm} D^{\mp}$ decay is presented. The decay is experimentally observed and its branching fraction is measured for the first time in the scope of this thesis. It confirms a theory prediction that assumes prominent contributions from rescattering from *e.g.* $D_s^{*\pm} D_s^{\mp}$ states [23], and refutes a prediction using a perturbative QCD approach [24]. This result can be related to other $B \to D\overline{D}$ decays to assess subleading contributions in $B_s^0 \to D_s^+ D_s^-$ decays. The analysis is published in the Journal of High Energy Physics in Ref [25]. The second analysis aims to measure CP violation in $B_s^0 \to D_s^+ D_s^-$ decays and is an update of a previous measurement of ϕ_s in $B_s^0 \to D_s^+ D_s^-$ decays [16]. The measurement is ongoing but in an advanced state with systematic uncertainties being currently evaluated. The measured statistical sensitivities to the CP parameters are reported.

The thesis is structured as follows: Chap. 2 introduces the SM and gives an overview of $B \to D\overline{D}$ decays. In detail, the formalism for the ϕ_s measurement in

$B_s^0 \rightarrow D_s^+ D_s^-$ decays and the influence of higher-order SM effects are discussed. Chap. 3 gives an overview of the LHCb detector. The tools and techniques used in the analyses are shortly described in Chap. 4. Afterwards, the analyses of the $B_s^0 \rightarrow D^{*\pm} D^{\mp}$ and $B_s^0 \rightarrow D_s^+ D_s^-$ decays are presented in Chap. 5 and Chap. 6, respectively. Finally, Chap. 7 discusses the results of the measurements and gives an outlook for further measurements.

2 The Standard Model of Particle Physics

This chapter presents the theoretical concepts in regards to measurements of branching ratios and decay-time-dependent *CP* violation in $B \to D\overline{D}$ decays. First, a short introduction of the Standard Model of particle physics (SM) [1–3] is given in Sec. 2.1. The SM is a relativistic quantum field theory and gives the currently best description of all elementary particles and the interactions between them. Afterwards, the quark mixing and the origin of *CP* violation are described in Sec. 2.2, followed by a general introduction of the time-evolution of neutral mesons in Sec. 2.3, which is necessary to describe decay-time-dependent *CP* violation. The different types of *CP* violation are introduced in Sec. 2.4. Finally, $B \to D\overline{D}$ decays in the SM are discussed in Sec. 2.5, showing that these can be used to measure the weak mixing phases of B mesons and to constrain higher-order effects with branching ratio measurements.

2.1 Fundamental Particles and Interactions

The information given in this section is based on Refs. [26, 27]. A graphical overview of the constituents of the SM and their properties is given in Fig. 2.1. In the SM, all matter is made out of twelve fermions with spin $^1/_2$ divided into six quarks, q, and six leptons, ℓ. For each particle, there is an antiparticle with the same properties but inverted charges. The fermions are categorised into three generations in ascending order of mass. Each generation contains an up-type quark (u, c, t) with the electric charge $+2/3\,e$ and an down-type quark (d, s, b) with the electric charge $-1/3\,e$, with e being the elementary charge. Due to so-called *confinement* [28] quarks cannot propagate freely and only appear in bound states, called hadrons. While each quark carries a colour charge, hadrons are colour neutral. Hadrons are further classified as mesons ($q\bar{q}$), baryons (qqq), tetraquarks ($qq\overline{q}\overline{q}$) and pentaquarks ($qqqq\bar{q}$). While tetraquarks have first been observed by the Belle collaboration [29], the LHCb collaboration has recently observed additional tetraquark states [30–32] and also discovered pentaquarks [33]. The leptons comprise the electron, e, muon, μ, or tauon, τ, with charge $-1\,e$ and their corresponding uncharged neutrino (ν_e, ν_μ, ν_τ). Leptons are colour neutral.

The SM describes the electromagnetic, weak and strong interaction of fermions. The interactions are described by the exchange of spin-1 gauge bosons, particles with integer spin. The mediators of the strong force are the eight gluons, g, that couple to the colour charge of particles. Gluons carry a colour and an anticolour and can therefore interact with each other. The mediator of the electromagnetic force is the neutral and massless photon, γ, which couples to the electric charge of particles. The mediators of the weak interaction couple to all fermions. They are the massive $W^{\pm}$ and Z bosons. While the strong and weak interaction have a limited range, the range of the electromagnetic force is limitless. The SM is completed by the Higgs boson, H^0, a scalar particle. It was predicted in 1964 [4,5] and discovered in 2012 by the ATLAS and CMS collaborations [6,7]. It is a consequence of the Higgs-mechanism through which the gauge bosons obtain their masses. The fermions obtain their masses through Yukawa interactions with the Higgs field.

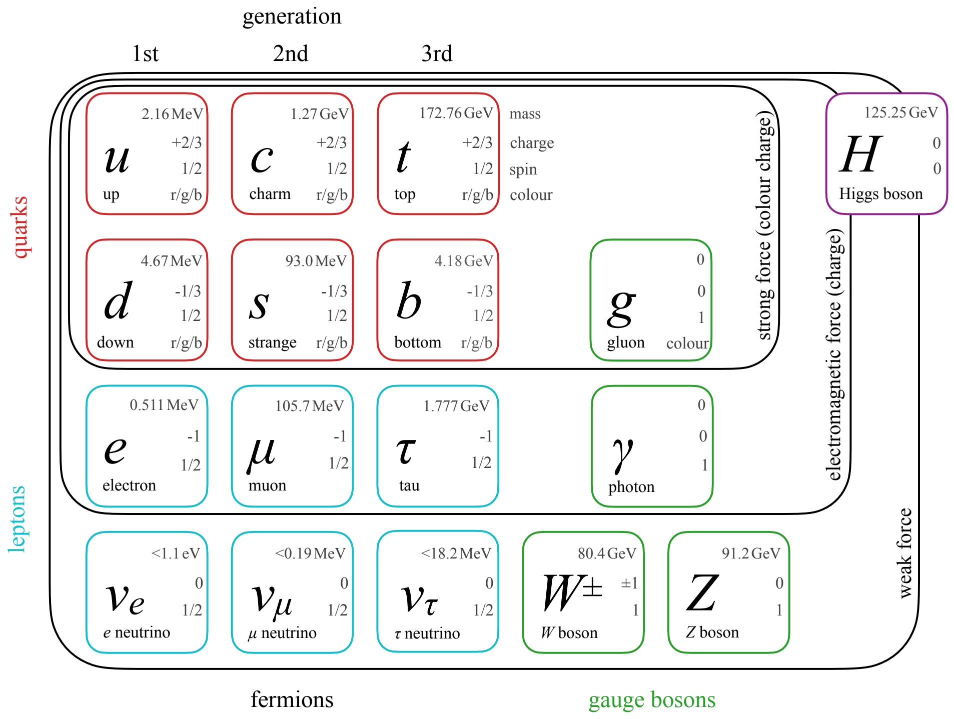

Figure 2.1: Overview of the fundamental particles and forces of the SM, and their properties. All numerical values are taken from [34].

2.2 The CKM Mechanism

The description in this section follows Refs. [34–36]. The fermion masses originate through the Yukawa interaction between left-handed and right-handed fermion fields and because of the non-zero vacuum expectation value of the Higgs field. Considering multiple generations, the coupling constants become arbitrary matrices and can be diagonalised to transform fermion fields from the weak-eigenstate basis into the measurable mass eigenstates. In case of leptons, the fact that neutrinos are massless in the SM allows the diagonalisation of the matrices to be performed in a way that the weak eigenstate and mass eigenstate of charged leptons are equal. In the quark sector, this cannot be done for both up-type and down-type quarks at the same time since none of them are massless. Instead, the matrices for up-type and down-type quarks have to be diagonalised separately and are then merged to one quark mixing matrix of the form

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

By convention, the so-called Cabibbo-Kobayashi-Maskawa (CKM) matrix, V_{CKM}, describes the mixing between the mass eigenstates (d, s, b) and the weak eigenstates (d', s', b') of the down-type quarks. The squared matrix elements, $|V_{ij}|^2$, are proportional to the transition probability of the up-type quark i to the down-type quark j. The quark transitions are possible via weak charged currents whereas transitions between the generations are allowed due to the non-zero off-diagonal matrix elements. The CKM matrix is a 3×3 complex and unitary matrix and can be fully described by four free parameters: three real angles and one complex phase, which is responsible for *CP* violation in the SM. A standard parametrisation [37] is given by

$$V_{\mathrm{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\mathrm{e}^{-\mathrm{i}\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & c_{23}c_{13} \end{pmatrix} , \tag{2.2}$$

where s_{ij} (c_{ij}) is the sine (cosine) of the three Euler angles θ_{12}, θ_{23} and θ_{13}, and δ the *CP*-violating phase. Another common representation of the CKM matrix is the Wolfenstein parametrisation using the relations

$$s_{12} = \lambda\,, \quad s_{23} = A\lambda^2\,, \quad s_{13}\mathrm{e}^{\mathrm{i}\delta} = A\lambda^3\left(\rho + \mathrm{i}\eta\right)\,, \tag{2.3}$$

with the three real parameters $\lambda \approx 0.23$, $A \approx 0.82$, $\rho \approx 0.14$ and one complex phase $\eta \approx 0.36$ [34]. Making use of the experimental finding of the hierarchy $|V_{ub}|^2 \ll |V_{cb}|^2 \ll |V_{us}|^2 \ll 1$, the Wolfenstein parametrisation is commonly expressed through a series expansion in power of the parameter λ resulting in the approximation [38]

$$V_{\mathrm{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & \lambda^3 A(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & \lambda^2 A \\ \lambda^3 A(1 - \rho - i\eta) & -\lambda^2 A & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)\,. \tag{2.4}$$

The magnitudes of the CKM matrix elements are the highest on the main diagonal and are of order $\mathcal{O}(1)$, *i.e.* quark transitions within a generation are most likely. The magnitudes decrease to the off-diagonal: quark transitions from the

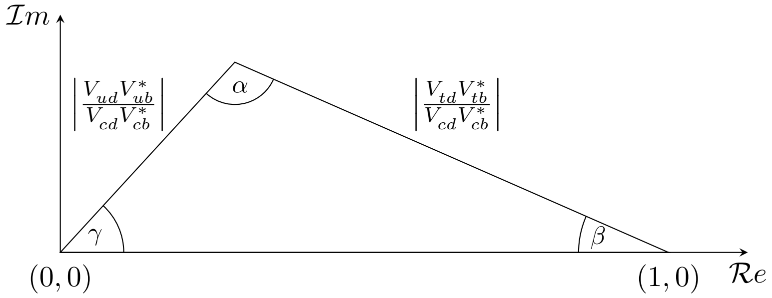

Figure 2.2: Schematic representation of the CKM triangle in the complex plane.

first to the second generation are of order $\mathcal{O}(\lambda)$, transitions from the second to third generation are of order $\mathcal{O}(\lambda^2)$ and from the first to third generation the magnitudes are of order $\mathcal{O}(\lambda^3)$.

The unitarity of the CKM matrix, $V_{\text{CKM}}V_{\text{CKM}}^\dagger = \mathbb{1}$, leads to six vanishing relations that can be represented as triangles in the complex plane. All of these unitarity triangles have an area equal to half of the Jarlskog parameter J [39], a measure of the amount of CP violation in the SM given by

$$J = \pm\mathcal{I}m\left(V_{ij}V_{kl}V_{il}^*V_{kj}^*\right) \quad \text{for } i \neq j, k \neq l\,. \tag{2.5}$$

It is measured to be $J \approx 3 \times 10^{-5}$ [34]. However, the ratios of side length of the triangles differ, with triangles showing side lenghts of different order of λ being nearly degenerate. One triangle where all side length are of $\mathcal{O}(\lambda^3)$ relates to the condition

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0\,, \tag{2.6}$$

with terms including transitions of b- and d-quarks. It is therefore referred to as the bd triangle or the unitarity triangle. By normalising Eq. (2.6) with the experimentally best known term $V_{cd}V_{cb}^*$ the triangle shown in Fig. 2.2 is obtained.

The three angles are given by

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right)\,, \quad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right)\,, \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right)\,. \tag{2.7}$$

An important goal of experimental flavour physics is to overconstrain the unitarity triangle by performing multiple measurements of the angles and side lengths providing a strong test of the unitarity of the CKM matrix. Measurements of a non-closing triangle would indicate physics beyond the SM. Currently, all measurements of the parameters are in good agreement with the SM and the uncertainty of the apex position of the triangle is relatively small [40].

The nearly degenerate bs triangle is obtained by the condition

$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^*\,, \tag{2.8}$$

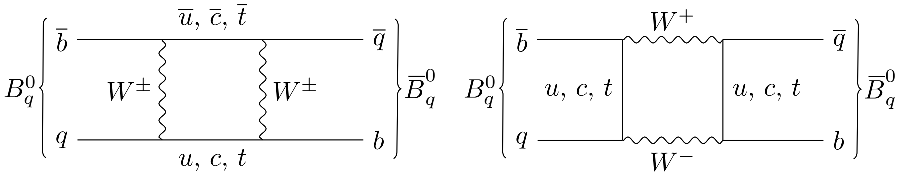

Figure 2.3: Dominant Box diagrams describing the $B_q^0 - \overline{B}_q^0$ oscillation.

with the sides being of order $\mathcal{O}(\lambda^4)$, $\mathcal{O}(\lambda^2)$ and $\mathcal{O}(\lambda^4)$, respectively. In particular, the small angle β_s between the two longer sides, defined by

$$\beta_s = \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) , \tag{2.9}$$

is an interesting parameter for *CP* violation measurements in the B_s^0 meson system. It is directly related to the *CP*-violating mixing phase ϕ_s of B_s^0 mesons via $\phi_s = -2\beta_s$.

2.3 Neutral Meson System and Time Evolution

The description in this section follows Ref. [35, 41]. Flavoured neutral mesons can oscillate between particle and antiparticle before they decay via the weak interaction. Such mixing processes provide crucial information about *CP* violation and the CKM matrix. In the lowest order, the process is described by box diagrams shown in Fig. 2.3. In the following, the description of the mixing formalism is outlined for the neutral B^0-meson system, whereby B^0 denotes both B_d^0 and B_s^0 mesons. However, the general formalism is also accurate for the K^0- and D^0-meson system.

Mixing needs to be included in the time evolution of a initial B^0 ($\overline{B}^0$) meson. The time-dependent wave function is given by a superposition of the flavour eigenstates $|B^0\rangle = |\bar{b}q\rangle$ and $|\overline{B}^0\rangle = |b\overline{q}\rangle$, with $q = d(s)$ for a B^0 (B_s^0) meson:

$$|\Psi(t)\rangle = \Psi_1(t)\,|B^0\rangle + \Psi_2(t)\,|\overline{B}^0\rangle \ . \tag{2.10}$$

The time evolution is then expressed by the time-dependent Schrödinger equation

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\Psi_1\\ \Psi_2\end{pmatrix} = \left(\mathbf{M} - \frac{\mathrm{i}}{2}\mathbf{\Gamma}\right)\begin{pmatrix}\Psi_1\\ \Psi_2\end{pmatrix} , \tag{2.11}$$

with the hermitian 2×2 mass and decay matrices $\mathbf{M}$ and $\mathbf{\Gamma}$, respectively. The non-zero off-diagonal elements allow for transitions from B^0 to $\overline{B}^0$ and vice versa. The elements on the main diagonal describe flavour-conserving transitions. In consequence of CPT invariance [42–44], particles and antiparticles have the same masses and decay widths and the diagonal elements are equal ($M_{11} = M_{22}$, $\Gamma_{11} = \Gamma_{22}$).

In order to develop the time evolution of the initial states B^0 and $\overline{B}^0$, the hamiltonian ($\mathbf{M} - \frac{\mathrm{i}}{2}\mathbf{\Gamma}$) needs to be diagonalised. This results into two eigenstates that can be distinguished by their well defined masses, a light mass eigenstate $|B_L\rangle$ and the heavy mass eigenstate $|B_H\rangle$, given by the superposition of the flavour eigenstates

$$\begin{aligned} |B_L\rangle &= p\,|B^0\rangle + q\,|\overline{B}^0\rangle \ , \\ |B_H\rangle &= p\,|B^0\rangle - q\,|\overline{B}^0\rangle \ . \end{aligned} \tag{2.12}$$

Here, the complex coefficients p and q that fulfill the normalisation $|q|^2 + |p|^2 = 1$. The ratio of p and q is related to matrix elements of $\mathbf{M}$ and $\mathbf{\Gamma}$ via

$$\begin{aligned} \frac{q}{p} &= \sqrt{\frac{2M_{12}^* - \mathrm{i}2\Gamma_{12}^*}{2M_{12} - \mathrm{i}2\Gamma_{12}}} \\ &\approx \sqrt{\frac{M_{12}^*}{M_{12}}} = \frac{V_{tq}V_{tb}^*}{V_{tb}V_{tq}^*} \ , \end{aligned} \tag{2.13}$$

where the fact that contributions from top quarks are dominant in the box diagrams [45] is exploited in the approximation. The eigenstates have defined masses and decay widths and thus can be characterised by the mass difference and decay width difference

$$\Delta m = m_H - m_L \, , \quad \Delta\Gamma = \Gamma_H - \Gamma_L \, . \tag{2.14}$$

The time evolution of the initial mass eigenstates is given by

$$|B_{L,H}(t)\rangle = \exp^{-(\mathrm{i}m_{L,H} + \Gamma_{L,H}/2)t}\,|B_{L,H}\rangle \ . \tag{2.15}$$

Inserting Eq. (2.12) into Eq. (2.15) gives the time evolution of the flavour eigenstates

$$\begin{aligned} |B^0(t)\rangle &= g_+(t)\,|B^0\rangle + \frac{p}{q} g_-\,|\overline{B}^0\rangle \ , \\ |\overline{B}^0(t)\rangle &= \frac{p}{q} g_-(t)\,|B^0\rangle + g_+\,|\overline{B}^0\rangle \ , \end{aligned} \tag{2.16}$$

with the time-dependent coefficients

$$\begin{aligned} g_+(t) &= \exp^{-\mathrm{i}mt}\exp^{-\Gamma t/2}\left[\cosh\frac{\Delta\Gamma t}{4}\cos\frac{\Delta m t}{4} - \mathrm{i}\sinh\frac{\Delta\Gamma t}{4}\sin\frac{\Delta m}{4}\right] \ , \\ g_-(t) &= \exp^{-\mathrm{i}mt}\exp^{-\Gamma t/2}\left[\sinh\frac{\Delta\Gamma t}{4}\cos\frac{\Delta m t}{4} + \mathrm{i}\cosh\frac{\Delta\Gamma t}{4}\sin\frac{\Delta m}{4}\right] \ . \end{aligned} \tag{2.17}$$

Experimentally, the time-dependent decay rates of the flavour eigenstates B^0 and $\overline{B}^0$ into final state f are investigated

$$\begin{aligned} \Gamma(B^0(t) \to f) &= \mathcal{N}_f |\,\langle f|B^0(t)\rangle\,|^2 \, , \\ \Gamma(\overline{B}^0(t) \to f) &= \mathcal{N}_f |\,\langle f|\overline{B}^0(t)\rangle\,|^2 \, , \end{aligned} \tag{2.18}$$

where $\mathcal{N}_f$ is a time-dependent normalisation factor arising from kinematics. Further, the amplitudes

$$A_f = \langle f|B^0\rangle \tag{2.19}$$

$$\overline{A}_f = \langle f|\overline{B}^0\rangle \tag{2.20}$$

and the quantity describing the relative phase between the p/q ratio and the decay amplitudes

$$\lambda_f = \frac{q}{p}\frac{\overline{A}_f}{A_f} \tag{2.21}$$

are useful. In the calculation of the decay rates into final state f, the parameters $|A_f|$ and $|\lambda_f|$ are factored out and the *CP* observables

$$A_f^{\Delta\Gamma} = -\frac{2\mathcal{R}e(\lambda_f)}{1+|\lambda_f|^2}\,, \quad C_f = \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}\,, \quad S_f = \frac{2\mathcal{I}m(\lambda_f)}{1+|\lambda_f|^2}\,, \tag{2.22}$$

are defined to simplify the formulas. With this, the decay rates are given as

$$\begin{aligned}\Gamma(B^0(t)\to f) &= \frac{1}{2}\mathrm{e}^{-\Gamma t}\mathcal{N}_f|A_f|^2(1+|\lambda_f|^2)\\ &\quad\left[\cosh\left(\frac{\Delta\Gamma}{2}t\right) + A_f^{\Delta\Gamma}\sinh\left(\frac{\Delta\Gamma}{2}t\right) + C_f\cos(\Delta m t) - S_f\sin(\Delta m t)\right],\\ \Gamma(\overline{B}^0(t)\to f) &= \frac{1}{2}\mathrm{e}^{-\Gamma t}\mathcal{N}_f|A_f|^2(1+|\lambda_f|^2)|\frac{p}{q}|^2\\ &\quad\left[\cosh\left(\frac{\Delta\Gamma}{2}t\right) + A_f^{\Delta\Gamma}\sinh\left(\frac{\Delta\Gamma}{2}t\right) - C_f\cos(\Delta m t) + S_f\sin(\Delta m t)\right].\end{aligned} \tag{2.23}$$

The *CP* observables satisfy the condition

$$(A_f^{\Delta\Gamma})^2 + (C_f)^2 + (S_f)^2 = 1\,. \tag{2.24}$$

If B^0 and $\overline{B}^0$ can decay into the *CP*-conjugate final state $\overline{f}$ that differs from final state f, the corresponding decay rates can be obtained by replacing f with $\overline{f}$.

2.4 *CP* Violation in the Standard Model

A fundamental concept of the SM is the validity of the CPT theorem [42–44], which states that all interactions are invariant under the successive application of the discrete transformations charge conjugation C, parity operation P and time reversal T in any order. The operation of C inverts the signs of the charges of a particle, hence it transforms particles into their antiparticles. The operator

P changes the sign of all spatial coordinates, $\vec{x} = -\vec{x}$, and thus reverses the handedness of space. It corresponds to a transformation of left-handed fermions into right-handed fermions. By applying the transformation T, the time coordinate t is changed into $-t$ while leaving $\vec{x}$ unchanged. The SM allows for violation of the individual transformations in the weak interaction, which was experimentally observed by Wu and Garwin in 1956 [46]. The combination of C and P is conserved in most weak interactions. However, CP violation has been discovered 1964 in the neutral kaon system [47], 2001 in the B system [48] and just recently, in the year 2019, in the D system [49].

Three classes of CP violating effects can be distinguished in weak decays of neutral and charged hadrons and are described based on [41].

***CP* violation in the decay:** Direct CP violation, also called CP violation in the decay, is defined by

$$\left|\frac{\overline{A}_{\overline{f}}}{A_f}\right| \neq 1\,, \tag{2.25}$$

i.e. the decay and its CP conjugated decay have different amplitudes. This is the only type of CP violation that can occur in decays of charged mesons or baryons since it does not rely on mixing. It has been first observed in the kaon system [50, 51], is now also established in the B^0 and B_s^0 system [52] and has recently been observed in the D^0 decays by the LHCb experiment [49].

***CP* violation in the mixing:** Indirect CP violation occurs due to a non-vanishing relative phase between M_{12} and Γ_{12} in Eq. (2.13) leading to

$$\left|\frac{q}{p}\right|^2 = \left|\frac{2M_{12}^* - \mathrm{i}2\Gamma_{12}^*}{2M_{12} - \mathrm{i}2\Gamma_{12}}\right| \neq 1\,. \tag{2.26}$$

As this is a result from the mass eigenstates being different from the CP eigenstates, it is also refered to as CP violation in the mixing. This corresponds to a different transition probability of a particle into its antiparticle and vice versa. Measurements in the neutral B^0 and B_s^0 systems are in agreement with the small SM prediction and show that indirect CP violation is negligible here [53].

***CP* violation in the interference of direct decay and decay after mixing:** If neutral mesons decay into a final state that is common to both initial flavour states, CP violation can occur in the interference between the direct decay ($B^0 \rightarrow$f) and the decay with mixing ($B^0 \rightarrow \overline{B}^0 \rightarrow f$), also refered to as interference CP violation. It can occur even if direct and indirect CP violation are absent, *i.e.* $|q/p| = 1$ and $|\overline{A}/A_f| = 1$. In this case there is a relative non-vanishing phase between $|q/p|$ and $|\overline{A}/A_f|$. If the final state is a CP eigenstate,

f_{CP}, Eq. (2.21) can be rewritten as

$$\lambda_f = \eta_{f_{CP}} \frac{q}{p} \frac{\overline{A}_f}{A_f} \,, \tag{2.27}$$

where $\eta_{f_{CP}} = \pm 1$ is the *CP* eigenvalue of f_{CP}. In this case, interference *CP* violation is defined by

$$\mathrm{Im}(\lambda_f) \neq 0 \,. \tag{2.28}$$

It leads to a non-vanishing time-dependent *CP* asymmetry that can be measured experimentally

$$\mathcal{A} = \frac{\Gamma(\overline{B}^0(t) \to f_{CP}) - \Gamma(B^0(t) \to f_{CP})}{\Gamma(\overline{B}^0(t) \to f_{CP}) + \Gamma(B^0(t) \to f_{CP})} = \frac{S_{f_{CP}} \sin(\Delta m t) - C_{f_{CP}} \cos(\Delta m t)}{\cosh(\Delta\Gamma t/2) + A^{\Delta\Gamma}_{f_{CP}} \sinh(\Delta\Gamma t/2)} \,. \tag{2.29}$$

The $B_s^0 \to D_s^+ D_s^-$ decay analysed in the scope of this thesis is subject to this type of *CP* violation.

2.5 $B \to D\overline{D}$ Decays in the Standard Model

This section discusses the family of B-meson decays into two mesons containing one c quark each, abbreviated as $B \to D\overline{D}$ decays. These decays are sensitive to elements of the CKM matrix. *CP*-violating observables in $B \to D\overline{D}$ decays provide valuable information about fundamental parameters of the SM and access to New Physics [21,22]. Decays of a B^0 meson into a $D^{(*)+}D^{(*)-}$ pair provide access to the B^0-$\overline{B}^0$ mixing phase ϕ_d [54–59], whereas $B_s^0 \to D_s^{+(*)+} D_s^{+(*)-}$ decays can be used to measure the B_s^0-$\overline{B}_s^0$ mixing phase ϕ_s [60]. Within the $B \to D\overline{D}$ family, loop and non-factorisable contributions can be notably prominent [22]. These contributions can be constrained with the help of additional $B \to D\overline{D}$ decays. Tests of the SM have been performed by relating *CP* asymmetries and branching ratio measurements of various $B \to D\overline{D}$ decay modes [21,22].

Sec. 2.5.1 discusses *CP* violation in the $B_s^0 \to D_s^+ D_s^-$ decay. Sec. 2.5.2 gives a short overview of the $B_s^0 \to D^{*\pm} D^{\mp}$ decay, whose branching fraction is measured in the scope of this thesis.

2.5.1 Measuring ϕ_s in $B_s^0 \to D_s^+ D_s^-$ decays

In decays involving $b \to c\bar{c}s$ transitions, the *CP*-violating mixing phase ϕ_s can be measured in the interference between direct decay and decay after mixing of B_s^0 mesons. In the SM and assuming negligible sub-leading penguin contributions, the mixing phase ϕ_s is directly related to the doubly Cabibbo suppressed CKM angle β_s with $\phi_s = -2\beta_s$. The mixing phase ϕ_s can be predicted using a global fit of the unitarity triangle [40]

$$\phi_s^{\mathrm{SM}} = -2\beta_s = -0.0368^{+0.0009}_{-0.0006}\,. \tag{2.30}$$

This precisely-predicted small value makes the measurement of β_s via the mixing phase ϕ_s sensitive to possible New Physics effects contributing to the B_s^0-$\overline{B}_s^0$ mixing and thus an excellent probe of the SM.

The mixing phase ϕ_s can be precisely measured in the "golden" channel $B_s^0 \to J/\psi\,\phi$. Measurements in this channel are subject to clean theoretical interpretation and thus have relatively small theoretical uncertainties. From an experimental point of view, these measurements profit from clean signatures in the detector and high signal yield due to high branching fractions. As the decay has two vector final states, an angular analysis is necessary to disentangle the interfering *CP*-even and *CP*-odd components. Measurements have been performed using $B_s^0 \to J/\psi\,\phi$, $B_s^0 \to J/\psi\,KK$ and $B_s^0 \to J/\psi\,\pi\pi$ decays.

The world average value [53]

$$\phi_s = -0.050 \pm 0.019 \tag{2.31}$$

is consistent with the SM prediction but has larger uncertainties. With improving experimental precision, the pollution from higher order SM contributions will become the source of the leading uncertainty. It will be mandatory to have good control over possible hadronic effects that could mimic New Physics signal. Exploiting the wide range of $b \to c\bar{c}s$ decay modes accessible at LHCb will be essential to achieve this. Multiple independent measurements improve the precision of the average value and provide a powerful consistency test of the SM.

The $B_s^0 \to D_s^+ D_s^-$ decay is another channel involving $b \to c\bar{c}s$ transitions giving access to ϕ_s. The decay channel is as theoretically clean as the $B_s^0 \to J/\psi\,\phi$ channel. However, from an experimental point of view, the $B_s^0 \to D_s^+ D_s^-$ channel is disadvantageous. Though no angular analysis is needed, measurements are challenging due to more backgrounds and lower trigger efficiencies. Still, measurements in the $B_s^0 \to D_s^+ D_s^-$ decay are of particular interest because the decay is sensitive to contributions from different penguin amplitudes. The measurement of ϕ_s in $B_s^0 \to J/\psi\,\phi$ can be affected by contributions entering through the electroweak penguin sector, whereas the measurement in the $B_s^0 \to D_s^+ D_s^-$ decay mode is almost unaffected by this as electroweak penguins contribute only in colour-suppressed form [17]. With uncertainties coming from penguin contributions becoming more relevant, measurements in decay modes with different penguin amplitudes and different amounts of penguin pollution become important. Additionally, the penguin pollution in $B_s^0 \to D_s^+ D_s^-$ can be controlled using measurements of additional $B \to D\overline{D}$ modes [21, 22].

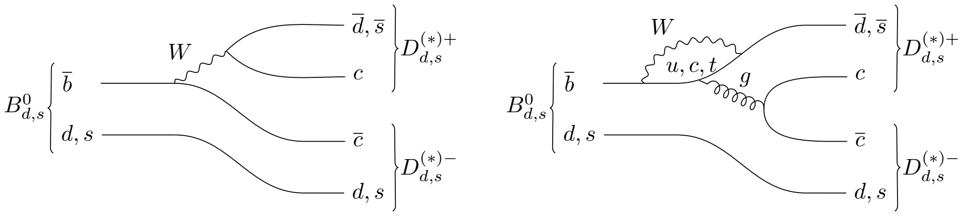

Figure 2.4: (Left) Tree-level and (right) penguin diagrams contributing to $B^0 \to D^{(*)\pm}_{(s)} D^{(*)\mp}$ and $B^0_s \to D^{(*)\pm}_{(s)} D^{(*)\mp}_s$ decays.

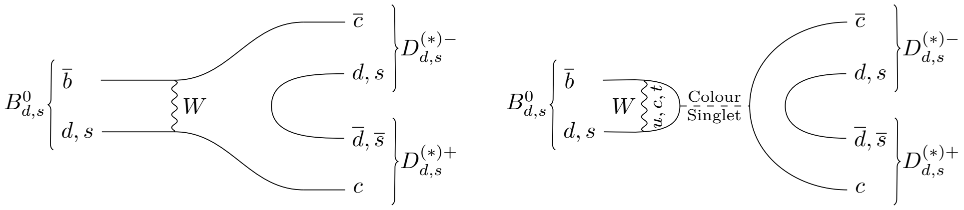

Figure 2.5: (Left) W-exchange and (right) penguin-annihilation diagrams contributing to $B^0_{(s)} \to D^{(*)+}_{(s)} D^{(*)-}_{(s)}$ decays.

Amplitudes and Observables

In this paragraph, the amplitudes and *CP* observables of the $B^0_s \to D^+_s D^-_s$ decay are summarised. The discussion closely follows Ref. [22]. The dominant diagrams of the $B^0_s \to D^+_s D^-_s$ decay are shown in Fig. 2.5. Contributing diagrams are the tree (T), penguin (P_q), exchange (E) and penguin annihilation (PA_q) diagrams, where the q subscript refers to the quark in the loop.

The amplitude of the $B^0_s \to D^+_s D^-_s$ decay is given as

$$\begin{aligned} A(B^0_s \to D^+_s D^-_s) &= V_{cs}V^*_{cb}(T + E + P_c + PA_c) \\ &\quad + V_{us}V^*_{ub}(P_u + PA_u) + V_{ts}V^*_{tb}(P_t + PA_t) \end{aligned} \tag{2.32}$$

The $V_{ts}V^*_{tb}$ term is eliminated using the CKM unitarity condition in Eq. (2.8)

$$\begin{aligned} A(B^0_s \to D^+_s D^-_s) = V_{cs}V^*_{cb}\underbrace{(T + E + P_c + PA_c - P_t - PA_t)}_{\mathcal{A}} \\ + V_{us}V^*_{ub}(P_u + PA_u - P_t - PA_t)\,. \end{aligned} \tag{2.33}$$

Factoring out the term $V_{cs}V^*_{cb}\mathcal{A}$ leads to

$$\begin{aligned} A(B^0_s \to D^+_s D^-_s) &= V_{cs}V^*_{cb}\mathcal{A}\big(1 + \frac{V_{us}V^*_{ub}}{V_{cs}V^*_{cb}}\frac{P_u + PA_u - P_t - PA_t}{\mathcal{A}}\big) \\ &= V_{cs}V^*_{cb}\mathcal{A}\big(1 + \frac{V_{us}V_{cd}}{V_{cs}V_{ud}}\frac{V_{ud}V^*_{ub}}{V_{cd}V^*_{cb}}\frac{P_u + PA_u - P_t - PA_t}{\mathcal{A}}\big)\,. \end{aligned} \tag{2.34}$$

Including the factor ${}^{V_{cd}V_{ud}}\!/\!{}_{V_{ud}V_{cd}}$ allows to utilise the the side length R_b of the bd triangle, which is enclosed by the angles α and γ

$$-R_b e^{\mathrm{i}\gamma} = \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \,. \tag{2.35}$$

The parameter

$$a e^{\mathrm{i}\theta} = R_b \left(\frac{P_u + PA_u - P_t - PA_t}{\mathcal{A}} \right) , \tag{2.36}$$

and

$$\varepsilon = -\frac{V_{us}V_{cd}}{V_{cs}V_{ud}} \approx -\frac{\lambda^2}{1-\lambda^2} \,, \tag{2.37}$$

simplify the amplitude equation to

$$A(B_s^0 \to D_s^+ D_s^-) = V_{cs}V_{cb}^* \mathcal{A} \left(1 + \varepsilon a e^{\mathrm{i}\theta} e^{\mathrm{i}\gamma} \right) \,. \tag{2.38}$$

Here, γ and the CKM elements V_{cs} and V_{cb}^* carry weak phases and the parameters a and θ are hadronic parameters with strong phases. The sign of the weak phases flips when applying the CP transformation while the signs of strong phases are unchanged. Therefore, the amplitude of the CP conjugate decay is given by

$$A(\overline{B}_s^0 \to D_s^+ D_s^-) = V_{cs}^* V_{cb} \mathcal{A} \left(1 + \varepsilon a e^{\mathrm{i}\theta} e^{-\mathrm{i}\gamma} \right) \,. \tag{2.39}$$

The parameter describing interference CP violation λ_f defined in Eq. (2.21) is obtained using these amplitudes together with the approximation of ${}^{q}\!/\!{}_{p}$ from Eq. (2.13) and the definition of the CKM angle β_s from Eq. (2.9):

$$\begin{aligned} \lambda_{D_s^+ D_s^-} &= \frac{V_{ts}V_{tb}^*}{V_{tb}V_{ts}^*} \frac{V_{cs}^*V_{cb}}{V_{cs}V_{cb}^*} \frac{1 + \varepsilon a e^{\mathrm{i}\theta} e^{-\mathrm{i}\gamma}}{1 + \varepsilon a e^{\mathrm{i}\theta} e^{\mathrm{i}\gamma}} \\ &= e^{2\mathrm{i}\beta_s} \frac{1 + \varepsilon a e^{\mathrm{i}\theta} e^{-\mathrm{i}\gamma}}{1 + \varepsilon a e^{\mathrm{i}\theta} e^{\mathrm{i}\gamma}} \,. \end{aligned} \tag{2.40}$$

The CP observables defined in Eq. (2.22) are then given by

$$\begin{aligned} C_{D_s^+ D_s^-} &= \frac{-2\varepsilon a \sin(\theta)\sin(\gamma)}{1 + 2\varepsilon a \cos(\theta)\cos(\gamma) + \varepsilon^2 a^2} \,, \\ S_{D_s^+ D_s^-} &= \frac{\sin(2\beta_s) + 2\varepsilon a \cos(\theta)\sin(2\beta_s + \gamma) + \varepsilon^2 a^2 \sin(2\beta_s + 2\gamma)}{1 + 2\varepsilon a \cos(\theta)\cos(\gamma) + \varepsilon^2 a^2} \,, \\ A^{\Delta\Gamma}_{D_s^+ D_s^-} &= -\frac{\cos(2\beta_s) + 2\varepsilon a \cos(2\beta_s + \gamma) + \varepsilon^2 a^2 \cos(2\beta_s + 2\gamma)}{1 + 2\varepsilon a \cos(\theta)\cos(\gamma) + \varepsilon^2 a^2} \,. \end{aligned} \tag{2.41}$$

Due to interferences between tree and penguin contributions, the CP parameter $C_{D_s^+ D_s^-}$ can differ from zero. The CP parameters $S_{D_s^+ D_s^-}$ and $A^{\Delta\Gamma}_{D_s^+ D_s^-}$ are caused in the interference between the direct decay and the decay after mixing and give access to the mixing phase $-2\beta_s = \phi_s$. However, only an effective phase ϕ_s^{eff} can be measured in $B_s^0 \to D_s^+ D_s^-$ due to the contributions of non-tree diagrams:

$$\phi_s^{\mathrm{eff}} = \phi_s + \Delta\phi_s \tag{2.42}$$

with

$$\sin(\phi_s^{\mathrm{eff}}) = -\frac{S_{D_s^+ D_s^-}}{\sqrt{1 - C^2_{D_s^+ D_s^-}}} \,. \tag{2.43}$$

The phase shift $\Delta\phi_s$ is defined by

$$\tan \Delta\phi_s^{D_s^+ D_s^-} = \frac{2\varepsilon a \cos\theta \sin\gamma + \varepsilon^2 a^2 \sin 2\gamma}{1 + 2\varepsilon a \cos\theta \cos\gamma + \varepsilon^2 a^2 \cos 2\gamma} \,. \tag{2.44}$$

2.5.2 Control of Higher-order Effects

Higher-order effects contributing to the $B_s^0 \to D_s^+ D_s^-$ decay can be accessed with the use of the $B^0 \to D^+ D^-$ decay. Here, the amplitude is given by

$$A(B^0 \to D^+ D^-) = V_{cd} V_{cb}^* \mathcal{A}' \left(1 - a' e^{\mathrm{i}\theta'} e^{\mathrm{i}\gamma}\right) \,, \tag{2.45}$$

where the $a' e^{\mathrm{i}\theta} e^{\mathrm{i}\gamma}$ term is not suppressed by ε. Due to the U-spin symmetry of the strong contributions, *i.e.* they are the same when exchanging all s with d quarks, the hadronic parameters are related by

$$\begin{aligned} a e^{\mathrm{i}\theta} &= a' e^{\mathrm{i}\theta'} \,, \\ \mathcal{A} &= \mathcal{A}' \,. \end{aligned} \tag{2.46}$$

The phase shift is defined by

$$\tan \Delta\phi_d^{D^+ D^-} = \frac{a'^2 \sin 2\gamma - 2a' \cos\theta' \sin\gamma}{1 - 2a' \cos\theta' \cos\gamma + a'^2 \cos 2\gamma} \tag{2.47}$$

and is not suppressed by ε. The hadronic parameters a' and θ' can be precisely measured and transferred to the phase shift of the $B_s^0 \to D_s^+ D_s^-$ decay.

Branching ratio measurements in the $B_s^0 \to D^{*\pm} D^{\mp}$ decay channel and all other $B_s^0 \to D^{(*)+} D^{(*)-}$ like decay channels can be used to assess contributions from the exchange and penguin-annihilation diagrams shown in Fig. 2.5. Decays via the tree or penguin diagrams from Fig. 2.4 are not possible. Thus, these measurements complement *CP*-violation measurements, to evaluate higher-order effects in $B \to D\bar{D}$ decays. The $B_s^0 \to D^{*\pm} D^{\mp}$ decay is first observed and its branching fraction is measured for the first time in the scope of this thesis. Using a perturbative QCD approach, the branching fraction is predicted to be $(3.6 \pm 0.6) \times 10^{-3}$ [24]. This is in disagreement with another prediction, which

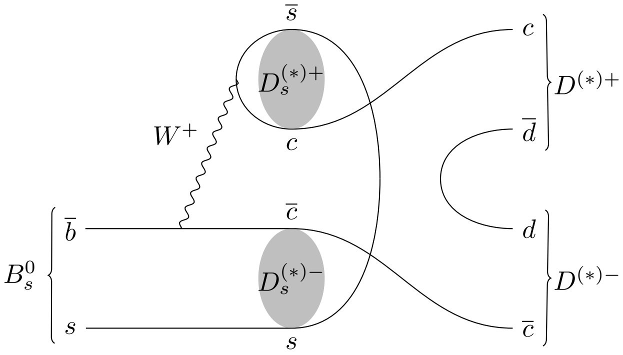

Figure 2.6: Rescattering diagram contributing to $B_s^0 \to D^{*\pm}D^{\mp}$ decays via the intermediate state $D_s^{*\pm}D_s^{\mp}$.

assumes prominent contributions from rescattering, *e.g.* from $D_s^{*\pm}D_s^{\mp}$ states, and predicts the value $(6.1 \pm 3.6) \times 10^{-5}$ [23]. A diagram showing such a rescattering process is shown in Fig. 2.6. The measurement of the branching fraction will show the importance of rescattering processes in $B \to D\overline{D}$ decays. It can be related to *e.g.* the branching fraction of the currently unobserved $B^0 \to D_s^+D_s^-$ decay to assess subleading contributions in $B_s^0 \to D_s^+D_s^-$ as outlined in [21, 22].

3 The LHCb Experiment at the LHC

The Large Hadron Collider beauty (LHCb) experiment is one of the four large experiments at the Large Hadron Collider (LHC) operated by CERN, the European Organization for Nuclear Research. A short description of the LHC is given in Sec. 3.1 followed by an overview of the LHCb detector in Sec. 3.2, based on [61] and [62], respectively. Sec. 3.3 covers the software and data processing at LHCb.

3.1 The LHC

The LHC is the world's most powerful particle accelerator located near Geneva. It is installed in a circular tunnel $50 - 175\,\mathrm{m}$ beneath the ground with a circumference of $26.7\,\mathrm{km}$, where two proton beams are accelerated in counter-rotating directions inside two separate rings. The beams are collided at four interaction points inside the four large LHC experiments ATLAS [63], ALICE [64], CMS [65] and LHCb [62]. Fig. 3.1 shows the accelerator complex. Before entering the LHC, the proton beams are pre-accelerated to an energy of $450\,\mathrm{GeV}$ in the linear accelerator LINAC 2, the Proton Synchrotron Booster (BOOSTER), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). Inside the LHC, the beams are further accelerated to the final centre-of-mass energy. A proton beam consist of up to 2808 bunches each comprising $1.15 \cdot 10^{11}$ protons. The LHC is designed to collide proton beams at a centre-of-mass energy of $14\,\mathrm{TeV}$ with a luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and a bunch spacing of $25\,\mathrm{ns}$. To keep the high-energetic beams on their circular path, superconducting dipole magnets are used. Further, quadrupole magnets focus the beams. During the first data taking period from 2011 to 2012, also referred to as Run 1, the centre-of-mass energies were $7\,\mathrm{TeV}$ in 2011 and $8\,\mathrm{TeV}$ in 2012. The second data taking period, Run 2, took place from 2015 to 2018 at a centre-of-mass energy of $13\,\mathrm{TeV}$.

The two largest detectors, ATLAS and CMS, are general-purpose detectors, whereas the LHCb and ALICE detectors have a more specialized focus. The LHC is also able to collide ion beams. The ALICE detector uses this to study strongly interacting matter at extreme energy densities, where a state of matter called quark-gluon plasma is created. The LHCb detector is designed to study decays of b and c hadrons with high precision and is described in detail in the next section.

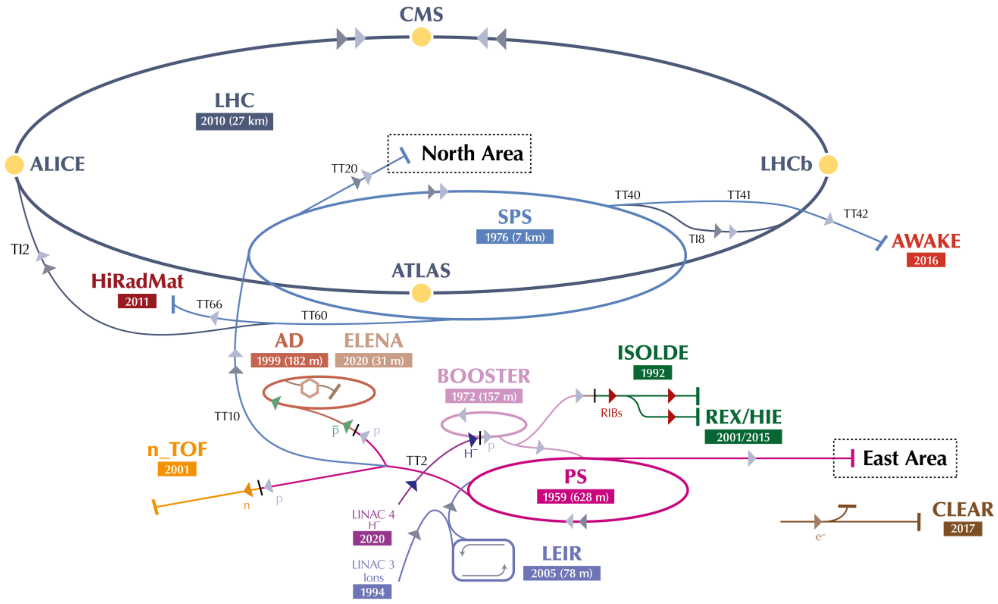

Figure 3.1: The CERN accelerator complex [66]. The beams first pass the LINAC2, BOOSTER, PS and SPS before being injected into the LHC. At the four collision points, the detectors of ATLAS, ALICE, CMS and LHCb are located.

3.2 The LHCb Detector

The LHCb experiment focuses on indirect searches for physics beyond the SM by performing precise measurements of *e.g.* *CP* violation and rare decays in the beauty and charm sector. The LHCb design fulfils fundamental aspects needed for precision measurements: high quality particle identification and accurate reconstruction of the measured hadrons production and decay vertices allowing for a precise determination of the decay times.

In contrast to the 4π detectors ATLAS and CMS, which cover almost the entire spatial angle range, the LHCb detector is a single-arm forward spectrometer with the detector components being sequentially ordered in forward direction with the interaction point in the beginning. This choice of design is justified by the production of the b quarks, which are mainly produced in $b\bar{b}$ pairs via gluon-gluon fusion. Due to the high collision energy it is likely that the two gluons have different momenta and as a result, the $b\bar{b}$ pair receives a strong boost in forward or backward direction along the beam axis. This is visualised in Fig. 3.2, where the production angles of simulated $b\bar{b}$ pairs at a centre-of-mass-energy of 14 TeV is shown. The LHCb detector covers a pseudorapidity range of $2 < \eta < 5$. The angular acceptance ranges from about $10-300$ mrad in the horizontal plane and from $10-250$ mrad in the vertical plane. About 25% of all produced $b\bar{b}$ pairs are inside the acceptance of the detector. The LHCb detector operates at a reduced

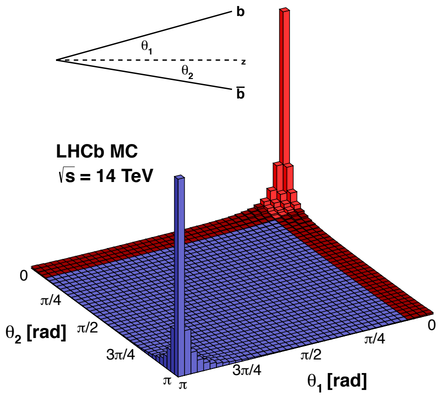

Figure 3.2: Production angle of simulated $b\bar{b}$ pairs as a function of the azimuthal angles θ_1 and θ_2 at a centre-of-mass energy of 14 TeV. Around 25% of the produced $b\bar{b}$ pairs are inside the acceptance of the LHCb detector represented by the red shaded region [67].

instantaneous luminosity of $4 \cdot 10^{23}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ by not colliding the beams directly but displaced to another. This way, the number of tracks is reduced, but the remaining tracks can be reconstructed with a higher efficiency.

A schematic view of the detector is given in Fig. 3.3. The individual components can be categorised into the track reconstruction system and the particle identification system and are summarised in the following.

3.2.1 Tracking System

In a hadronic environment like at the LHCb detector, high track multiplicity occurs. This means an excellent tracking system is required that reconstructs charged particle tracks and vertices. The tracking system at LHCb uses information from the Vertex Locator (VELO), the tracking stations (TT and T1–T3) and the dipole magnet. The T1–T3 stations are divided into Inner Tracker (IT) and Outer Tracker (OT).

Vertex Locator: The VELO is located around the pp interaction point, where the B hadrons are produced at the primary vertex (PV). After a flight distance of about 1 cm, B hadrons decay into secondary particles at the secondary vertex (SV). The VELO provides measurements of the PV and the displaced SV with high resolution. With these, decay times of the B hadrons can be precisely determined. The typical decay-time resolution is approximately 50 fs [68]. The VELO consists of two halves surrounding the beam pipe each comprising 21 silicon modules that can be moved up to 8 mm from the beam axis when the

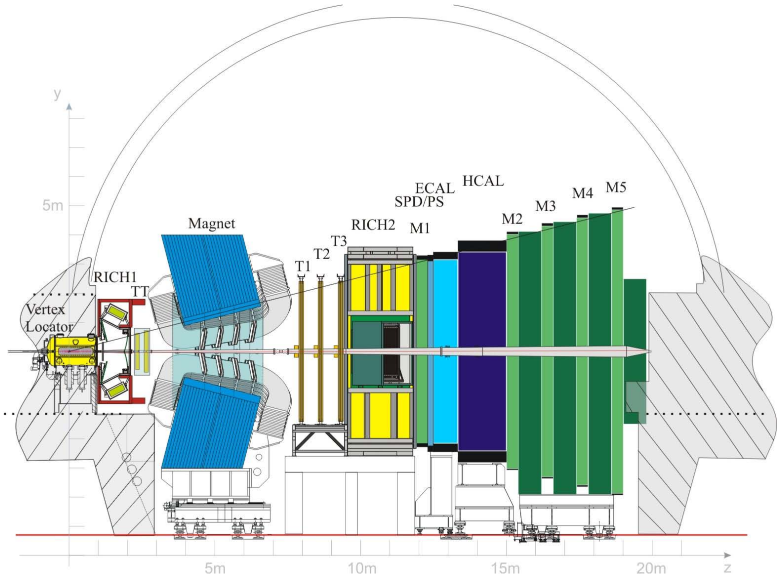

Figure 3.3: Schematic view of the LHCb detector. The beam pipe is installed parallel to the z-axis. On the left, the Vertex Locator (VELO) is located around the interaction point of the protons. Downstream of the VELO are the other components comprising two Ring Imaging Cherenkov (RICH1 and RICH2) detectors, the tracking system and a magnet, a calorimeter system and muon chambers [62].

beams are stable. A module comprises two silicon half discs placed perpendicular to the beam axis as shown in Fig. 3.4. One disc uses R-sensors and the other uses ϕ-sensors to measure the track position of a passing particle in cylindrical r and ϕ coordinates. Registered hits generated by a passing particle are required in at least three modules to reconstruct the track.

Tracking stations: The TT and the IT use silicon strip detectors to detect tracks. The TT is a planar tracking station with a width of 150 cm and a height of 130 cm covering the full detector acceptance. It is placed before the magnet so that low-momentum particles that are bend out of the detector acceptance after the magnet can still be detected. The IT covers a cross shaped region with a width of 120 cm and a height of 40 cm around the beam pipe. The OT is a drift detector consisting of straw tubes filled with a gas mixture (70% argon, 28.5% carbon dioxide and 1.5% oxygen). The TT, IT and OT stations consist of four overlapping layers. The first and last layer are aligned to the vertical axis, while the second and third layer are tilted by an angle of ±5°. This way both radial coordinates can be measured.

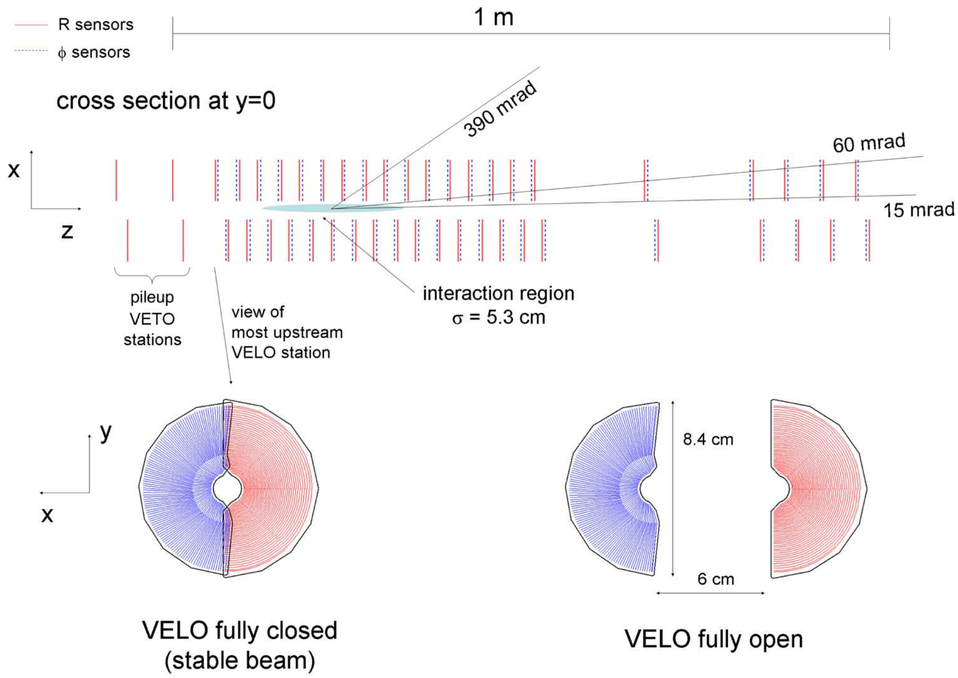

Figure 3.4: Cross-section of the VELO silicon sensors in the (x, z) plane with $y = 0$ in the fully closed position (top). The R sensors are shown in red and the ϕ sensors are shown in blue. Also the front view of the first modules is illustrated in the closed and open position (bottom) [62].

Magnet: A dipole magnet with a vertical magnet field (y-axis) is used to measure the momenta of charged particles. It is a non-superconducting magnet providing an integrated magnetic field of 4 Tm. The momentum of a passing charged particle is measured via the curvature of its bent track within the x-plane. The magnet is placed between the TT and the tracking stations T1–T3 downstream of the VELO.

3.2.2 Particle Identification System

An important task at LHCb is the identification of the particle type. In the presented analyses, distinguishing between charged hadrons (kaons, pions and protons) is essential for *CP*-violation measurements, background rejection and flavour tagging. This is achieved by two Ring Imaging Cherenkov (RICH1 and RICH2) detectors. A calorimeter system and muon chambers complete the particle identification (PID) system.

RICH detectors: The RICH detectors make use of the Cherenkov effect, where photons are emitted when a charged particle travels with a speed v greater than the speed of light c through a medium. The photons are emitted in a cone along the particle's flight direction. The opening angle of the cone, $\cos\theta = c/nv$, where n is the refractive index of the medium, gives access to the velocity of the particle. Combined with the momentum provided by the tracking system, the particle mass can be determined. The dependency of the Cherenkov angle on the momentum of the track is shown in Fig. 3.5, where clear separated pion, kaon and protons bands are visible.

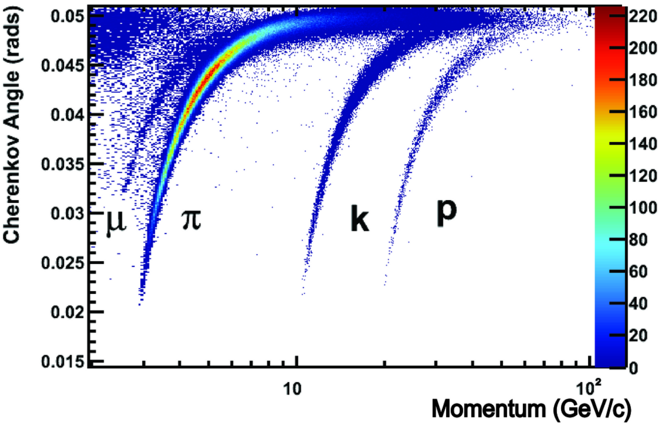

Figure 3.5: Cherenkov angle as a function of the momentum of the track using information of the RICH1 detector. The events are distributed into clear separated pion, kaon and proton bands. Although the role of the RICH detectors is the identification of hadrons, also a muon band is visible [69].

RICH1 is located between the VELO and the TT and uses C_4F_{10} as the medium to identify charged particles with low momenta ranging from 1 to 60 GeV that would leave the detector acceptance without passing RICH2. RICH2 is located directly after the tracking stations and identifies particles with momenta in the range 15 to 100 GeV with CF_4 gas as the medium. A mirror system outside the detector acceptance is used to focus the photons on photo-detectors.

Calorimeter system: The calorimeter system provides the identification of hadrons, electrons and photons, and measurements of their energies and positions for the L0 trigger. The calorimeter system is installed after the RICH2 detector and comprises four subdetectors. All calorimeters have an alternating order of absorber layer and scintillators. Passing charged particles can produce showers of secondary particles in the absorber, which then produce scintillation light in the scintillating material that is transmitted to photomultipliers. The first two detectors are the scintillator pad (SPD) and preshower detector (PS) that are able to separate between charged and uncharged particles, and electrons and charged hadrons, respectively. The following electromagnetic calorimeter (ECAL) and the hadron calorimeter (HCAL) allow for a distinction between electrons and hadrons with the ECAL being installed before the HCAL. The SPD, PS and ECAL utilise lead as absorber material, while the HCAL uses an iron absorber.

Muon chambers The five muon chambers M1–M5 serve as an identification system for muons. The M1 station is placed in front of the calorimeter system and is used to improve transverse-momentum measurements for the trigger. The stations M2–M5 are located downstream of the calorimeter system at the end of the detector. The muon stations are multi-wire proportional chambers, except for the inner region of M1, where gas electron multiplier are used. Between the stations, 80 cm thick iron layers are placed to absorb muons. Only muons with a minimum momentum of 6 GeV/c are able to pass the whole detector.

3.2.3 Trigger

Due to limited storage space and bandwidth, it is not possible to save the data of every collision. The bunch-crossing rate of 40 MHz is reduced by the trigger system to 3 kHz in 2011 [70], 5 kHz in 2012 [71] and 12.5 kHz in Run 2 [72]. The trigger is divided into two levels: the hardware-based level 0 (L0) trigger and the software-based High-Level Trigger (HLT). The L0 trigger operates fully synchronously with the initial bunch crossing rate and reduces the rate to 1 MHz, at which rate the entire detector can be read out. Information from the calorimeter and muon systems are utilised to search for hadrons, electrons and photons with high transverse energy, and muons and muon pairs with high transverse momentum. In addition, the total number of hits in the SPD is used to remove large multiplicity events, *i.e.* events with too many tracks. Data that passes the L0 trigger is processed by the HLT. As the data rate is still very high, the HLT is divided into two stages. First, the HLT1 reduces the rate with the help of a partial event reconstruction. Information from the VELO, tracking station and muon stations are utilised for the reconstruction. Finally, the data passing the HLT1 is sufficiently reduced that a complete event reconstruction can be performed in the HLT2.

3.3 LHCb Software

The recorded data is processed in several stages before it can be used for analysis. The processing is done with the software packages based on the GAUDI framework [73]. The MOORE package [74] contains the HLT software introduced in Sec. 3.2.3. Afterwards, the raw data passing the trigger is transformed from detector hits to tracks. This is done by the software BRUNEL [75] that combines PID and tracking information to so-called proto-particles. The resulting data is stored in Data Summary Tape files (DSTs) and contains the full event information. The DSTs are further processed for analysis in DaVinci [76], where proto-particles are combined to intermediate particles and the whole decay chain is reconstructed. In this step, a first preselection is performed. This preselection is performed centralised at LHCb and is organised in so-called *stripping lines* containing loose requirements to filter certain decay modes. Multiple similar decay modes can be part of the same line. The standard method to reconstruct the decay chain, referred to as `LoKiVertexFitter`, uses a bottom-to-top approach. It combines particles one after another starting with the final-state particles and calculates the vertices and kinematic properties until the whole decay chain is reconstructed. Another, more time-consuming approach, is reconstructing the whole decay chain in one single fit referred to as decay-tree fit (DTF) [77]. An advantage of this is that correlations and uncertainties on the parameters are correctly included. Furthermore, constraints on particles origin vertex and masses can be applied. After this processing step, the data is available in the nTuples format that can be used by the individual analyst.

In addition to recorded data, most analyses utilise data sets generated in Monte Carlo simulations to verify the analysis strategy, *e.g.* the presented analyses in this thesis use simulated samples for efficiency calculation, modelling the shape of distributions and training multivariate algorithms. The generation of simulated data is implemented the Gauss framework [78]. It comprises the packages Pythia [79, 80], which simulates the pp collison, EvtGen [81], which is responsible for the simulation of the decay of particles with the package Photos [82] to add final-state radiation, and Geant4 [83], which simulates the particle's propagation and interaction through and with the detector as described in Ref. [84]. The output of Gauss is converted with the Boole package [85] so that it equals the output of the detector. The remaining data processing is equivalent to the processing of real data explained above starting with the Moore package. An advantage of simulated data is the fact that the true information about the generation process is saved. Thus, quantities of the reconstructed particles can be compared with the originally generated ones. *e.g.* the truth information can be utilised to study resolution effects.

4 Data Analysis Tools and Techniques

This chapter introduces several data analysis tools and techniques used within the analyses of this thesis. At LHCb, the recorded data sample comprises candidates of various sources. In physics analyses, it is necessary to separate signal candidates, which one wants to analyse, from background candidates, which contaminate the data sample. Apart from specific cut-based requirements, machine learning algorithms are useful for classification of data. A multivariate classifier in the form of a boosted decision tree (BDT) is described in Sec. 4.1. To extract physics parameters, *e.g.* the number of signal candidates, the maximum likelihood method is used and explained in Sec. 4.2. To statistically separate signal from background candidates, the *sPlot* method is used, outlined in Sec. 4.3. In decay-time-dependent measurements of CP violation, the knowledge of the initial flavour of B^0 and B^0_s mesons is essential and is determined via the flavour tagging described in Sec. 4.4.

4.1 Boosted Decision Trees

In high energy physics, selection methods based on multivariate algorithms are a crucial ingredient to most analyses. They are used to classify data points by analysing multiple features simultaneously. In contrast to a cut-based selection that makes requirements on single variables, multivariate methods have the advantage that they can exploit correlations between the input features. A common usage of these methods is to distinguish between signal and background data. This is usually done with supervised learning where the input data for the training of a classifier is labelled, *i.e.* the classification for the training data is known. In the scope of this thesis, Boosted Decision Trees (BDT) are used. For the training of the BDT classifier a signal and a background sample proxy are needed.

BDTs comprise multiple binary decision trees [86], illustrated in Fig. 4.1. A decision tree starts at a root node, where a requirement is made on the input variable with the best separation power, splitting the candidates in the node into two branches. To determine the variable with the best separation power, the Gini index, $p \cdot (1 - p)$, with the signal purity p being the fraction of signal candidates to all candidates, is used. This process is repeated at each following sub-node until a leaf node is reached. This happens when a stop criterion, *e.g.* a maximum of number of sub-nodes or a minimal number of events in the node, is fulfilled. The class of the leaf is specified by the label of the majority of the

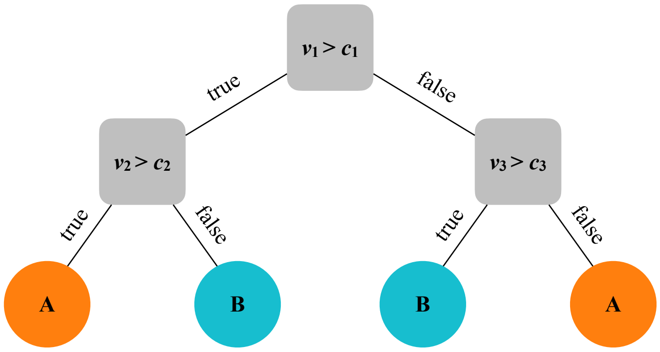

Figure 4.1: Illustration of a binary decision tree with a depth of two. Input candidates are divided into the classes A and B by requirements c_i on the variables v_i at each node.

remaining candidates.

An improvement of the performance of a decision tree is achieved by creating multiple decision trees consecutively and applying a boost weight to subsequent decision trees, obtaining a BDT. To configure the BDT, a set of variables with high separation power is needed and the hyperparameters of the utilised boosting algorithm have to be tuned. Hyperparameters are *e.g.* the number of trees, the maximum depth of the trees, which corresponds to the number of cuts that are made before a candidate is assigned to a class, or the minimal amount of candidates that must remain at each node. Different boosting algorithms exist. Two of them are the AdaBoost algorithm [87] and the Gradient Boosting algorithm [88], which are used in the scope of this thesis and are summarised in the following.

AdaBoost: The AdaBoost algorithm, short for adaptive boosting, gives incorrectly classified candidates a higher boost weight in the training of the following decision tree. Starting with initially unweighted candidates in the first decision tree training, for the following training steps the boost weight

$$\alpha^\beta = \left(\frac{1-\varepsilon}{\varepsilon}\right)^\beta , \tag{4.1}$$

is multiplied with the previous weights of the misclassified candidates. Here, ε is the misclassification rate and β is the learning rate. The weights are renormalised so that the sum of the weights is always constant. The output classifier of a BDT with N trees is given by

$$y_\text{boost} = \frac{1}{N}\sum_i^N \ln(\alpha_i^\beta) h_i , \tag{4.2}$$

with h_i being the output classifier of a single tree ranging from -1 to 1. Lower values indicate candidates classified as background, whereas higher values indicate candidates classified as signal.

Gradient Boosting: In the Gradient Boosting algorithm, the decision trees are trained additively, *i.e.* after each added tree, an objective function is minimised. The objective function is a measure of how well the prediction describes the training data and is defined as

$$\mathcal{L}_{\text{obj}} = \sum_i l(\hat{y}_i, y_i) + \sum_n \Omega(f_n) \quad \text{with} \quad \Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \,. \tag{4.3}$$

The first part of the objective function is a loss function that indicates the difference between the predicted label $\hat{y}_i$ and true label y_i and is commonly described by the mean squared error. The second part of the objective function is a regularization term that controls the complexity of a decision tree f_n considering the number of leaves in the tree, T, the weight that is assigned to each leaf, ω, and regularisation parameters γ and λ.

In general, BDTs can suffer from the effect of overtraining, *i.e.* the BDT gets trained on statistical fluctuations that are present in the training sample. This effect becomes visible through an overestimation of the separation performance and has to be avoided. Usually, overtraining can be identified by comparing the distribution of the output classifier of the training sample and of a validation sample that has not been used for the training.

A common technique to prevent overtraining is the k-fold technique. The data sample is randomly split into k subsamples of equal size. In total k BDTs are trained, whereby for each training $k-1$ subsamples are used and the remaining subsample is used as a validation sample to determine the performance of the trained classifier.

Another way to achieve a better stability of the classifier and avoid overtraining is the usage of the resampling technique *bagging*. The separate decision trees are trained using resampled training candidates in a subset of a specified size, whereby the same candidate is allowed to be picked several times. This way, each single tree of the BDT is trained with a different subsample. The combined classifier is more stable with respect to statistical fluctuations in the training sample as statistical fluctuations present in the original training data are not present in every subsample.

An additional approach to prevent overtraining is the *early stopping* method. After each added decision tree, this method checks if the performance has improved and aborts the training if it did not improve after n added trees. The parameter n needs to be specified.

4.2 Maximum Likelihood Method

A commonly used technique in particle physics to extract parameters of interest from a data sample is the maximum likelihood method [89]. This is a statistical method for an estimation of a set of parameters $\vec{\theta}$ by adapting an underlying model in form of a probability density function (PDF) $\mathcal{P}(\vec{x}, \vec{\theta})$ to a distribution observed on a data sample $\vec{x}$. Here, $\vec{x}$ contains n measured values of either a single observable or a vector of observables. The method uses the likelihood function, which is defined as the product of the likelihood over all measurements

$$\mathcal{L}(\vec{\theta}) = \prod_{i=1}^{n} \mathcal{P}(\vec{x}_i|\vec{\theta}) \,, \tag{4.4}$$

and indicates the likelihood to obtain the measured values $\vec{x}$ for given parameters $\vec{\theta}$. The best set of parameters is then obtained by maximizing $\mathcal{L}(\vec{\theta})$. In general, the PDF $\mathcal{P}$ can be composed of various PDFs to describe different components that need to be modelled. The likelihood function is then formed by the sum of the individual likelihoods, whereby also the number of measured values has to be considered. Assuming that the number of measured values is a random variable that follows a Poisson distribution, the Poisson term $\mathrm{e}^{-N}N^n/n!$ is introduced, which defines the probability to observe n candidates when N are expected. The so-called extended likelihood function is given by

$$\mathcal{L}(\vec{\theta}) = \frac{\mathrm{e}^{-N}N^n}{n!} \prod_{i=1}^{n} \mathcal{P}(\vec{x}_i|\vec{\theta}) \,. \tag{4.5}$$

When searching for the best set of estimators $\vec{\theta}$, it is necessary that the PDF is normalised while varying $\vec{\theta}$, which causes the most numerical effort. In practice, it is easier to instead minimise the negative logarithmic likelihood function. As the logarithm is a monotone function, the result stays the same. Furthermore, it is possible to include an uncertainty from an external parameter λ by constraining it via a Gaussian function. The likelihood is multiplied with a Gaussian function

$$\mathcal{G}(\lambda|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(\lambda-\mu)^2}{2\sigma^2}} \,, \tag{4.6}$$

with the mean μ and width σ set to the central value and uncertainty of the parameter λ, respectively. In the course of this thesis, the parameter estimation is performed using the ROOT framework [90], which utilises the minimisation package MINUIT [91]. Besides parameter estimations, also estimations for their uncertainties are provided.

4.3 Unfolding Data Distributions with the sPlot Method

After the whole selection has been applied, irreducible background candidates can remain in the data sample. When a maximum likelihood fit is performed to extract physics parameters, *e.g.* *CP* observables, their contribution needs to be parameterised. Finding a proper signal model is usually feasible as the model is often given by theoretical functions with modifications to include experimental effects. In contrast, finding a proper model for background components can face difficulties, *e.g.* for backgrounds from combinations of random tracks. Furthermore, large systematic uncertainties can arise from the definition of background models resulting in a poorer sensitivity of a measurement. An appropriate way to separate signal and background candidates is by statistically unfolding the distributions using the *sPlot* method [92]. In this technique, per-event signal weights, so-called sWeights, are calculated to project out single components from a data sample. The calculation uses the yields of an extended maximum likelihood fit to a discriminating observable, x, that needs to be uncorrelated to the actual observable of interest. The distribution of the discriminating observable has to differ between the components that need to be unfold. Moreover, it is required that the parametrisation of the individual components is known. A suitable observable is the reconstructed B^0 mass, where a proper model of the signal and background components can be found easily. In the extended maximum likelihood fit, all parameters except the yields of the components are fixed to the resulting values of a previous fit. The sWeights are then obtained as

$$ {}_s\mathcal{P}_n(x_i) = \frac{\sum_{j=1}^{N_s} \mathbf{V}_{nj} f_j(x_i)}{\sum_{k=1}^{N_s} N_k f_k(x_i)} \tag{4.7} $$

for each candidate x, where n denotes the component. The indices j and k sum over N_s, the number of components comprised in the PDF f. The estimate of the number of signal or background candidates is labelled as N_k and the correlation matrix of the yields is V_{nj}, which is given by the second derivative of the negative log likelihood function. The sum of all sWeights over one component returns the corresponding yield. When applying the sWeights to a variable uncorrelated to the discriminating variable, a statistically pure sample of the respective component is extracted. When applying the sWeights to candidates in a histogram the uncertainty on the bin content of bin i is given by

$$ \sigma(\mathrm{bin}) = \sqrt{\sum_{i \subset \mathrm{bin}} ({}_s\mathcal{P}_n)^2} \,. \tag{4.8} $$

4.4 Flavour Tagging

For decay-time-dependent measurements of *CP* violation, it is crucial to know the flavour state of the B meson (B or $\overline{B}$) at the time of its production, *i.e.* whether it contains a b or $\overline{b}$ quark. For charged B meson decays the production flavour can be inferred by the charges of the final-state particles. This is not possible for neutral B mesons as they are able to oscillate before the decay. At LHCb, the information about the production flavour is provided by the so-called flavour tagging. There are different flavour tagging algorithms, which can be divided into two groups: the same-side (SS) and the opposite-side (OS) taggers. The procedure of the flavour tagging and the flavour tagging algorithms for B_s^0 mesons are shown in Fig. 4.2.

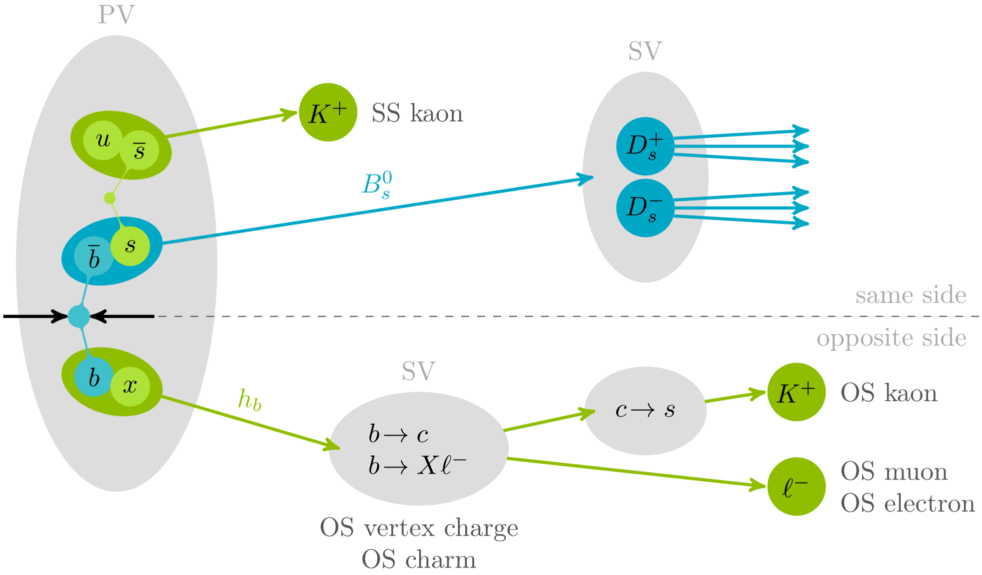

Figure 4.2: A schematic overview of the (green) flavour tagging algorithms in the analysis of (blue) $B_s^0 \rightarrow D_s^+ D_s^-$ decays. On the same side, the taggers search for particles associated to the hadronisation process of the signal b meson and on the opposite side, the taggers make use of decays of the accompanying b of opposite flavour produced in the pp collision.

Tagging algorithms: The SS taggers exploit remnants of the hadronisation process of the signal B meson. The hadronisation of a signal B meson comprising a $\overline{b}q$ quark pair leads to a $\overline{q}$ quark that can form another hadron called tagging particle, whose charge gives access to the production flavour. In case of signal B_s^0 mesons, the tagging particle is a kaon, which is selected by the SS kaon tagger.

The OS taggers make use of the decay of the hadron that comprises the non-signal b quark of the $b\overline{b}$ pair produced in the pp collision. There are several OS taggers, which search for different charged particles. The charge of this tagging particle gives information about the flavour of the non-signal b hadron and thereby allows to derive the production flavour of the signal B meson. The OS electron and OS muon taggers exploit the charge of the lepton of a semi-leptonic decay, the OS kaon tagger searches for kaons from decays with $b \to c \to s$ transitions and the OS charm tagger uses the decay products of a charm decay. The OS vertex charge tagger does not use a single particle but reconstructs a secondary decay vertex within the non-signal b hadron decay and calculates the average of the charges of all tracks belonging to this vertex weighted by the transverse momentum.

Output parameters and performance: All flavour tagging algorithms provide a tag decision $d \in \{-1, 0, 1\}$ for each reconstructed candidate and a mistag probability estimate $\eta \in [0, 0.5]$ for each tag , *i.e.* the probability that the tag decision is wrong. A tag of $d = +1(-1)$ corresponds to a B ($\overline{B}$), while a tag of $d = 0$ refers to an untagged candidate, for which the production flavour cannot be identified, *e.g.* when no tagging particle is selected. A mistag probability of $\eta = 0$ corresponds to a perfectly tagged candidate, while $\eta = 0.5$ equals an untagged candidate. The predictions of the tagging algorithms are the output of multivariate classifiers that exploit geometric and kinematic properties of the tagging particle and information on the event.

To parameterise the performance of a tagging algorithm, two quantities are introduced. The first quantity, the tagging efficiency, ε_{tag}, results from the fact that the tagging algorithms are not always able to make a prediction. It is defined as the ratio of tagged events over the number of all candidates

$$\varepsilon_{\text{tag}} = \frac{N_{\text{tagged}}}{N_{\text{tagged}} + N_{\text{untagged}}} \,. \tag{4.9}$$

The second quantity, the true mistag probability ω, is defined as the fraction of the number of incorrectly tagged candidates over all tagged candidates.

$$\omega = \frac{N_{\text{incorrect}}}{N_{\text{tagged}}} \,. \tag{4.10}$$

Incorrectly tagged candidates originate due to imperfect flavour tagging algorithms, *e.g.* an imperfect selection of the tagging particles. The true mistag probability is directly related to the dilution factor $D = 1 - 2\omega$ that describes how much the measured *CP* asymmetry is reduced by imperfect flavour tagging compared to a perfect flavour tagging. With these two quantities an effective tagging efficiency, $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}} D^2$, the so-called tagging power, can be formed. On a per-event basis i the tagging power is defined as

$$\varepsilon_{\text{eff}} = \frac{1}{N} \sum_{i}^{N} (1 - 2\omega)^2 \,, \tag{4.11}$$

summing over all candidates and with $\omega = 0.5$ and hence $D = 0$ for untagged candidates. The tagging power represents the effective statistical loss of a tagged analysis compared to a perfectly tagged sample and allows to compare the flavour tagging performance between different decay modes and also different experiments. In reality, a perfect flavour tagging cannot be achieved, and the number of candidates is reduced by the tagging power, $\varepsilon_{\text{eff}}N$, which reduces the statistical uncertainty on the measured *CP* asymmetry by

$$\sigma_{CP} \propto \frac{1}{\sqrt{\varepsilon_{\text{eff}}N}} \,. \tag{4.12}$$

Calibration and Combination: A multivariate classifier is used to determine the mistag probability and is trained on datasets of flavour specific decays, *i.e.* only $B \to f$ and $\overline{B} \to \bar{f}$ transitions are allowed. This datatset is different than the signal dataset and the distribution of the mistag probability estimate η might differ from the true mistag probability ω. Therefore the output of the tagging algorithms must be calibrated using a control sample that is similar with respect to the kinematic and decay topology of the signal channel. A calibration function is extracted that transforms the mistag probability estimate η into the true mistag probability $\omega(\eta)$ measured in the control sample. Oftentimes, a linear calibration function is sufficient

$$\omega(\eta) = p_0 + p_1(\eta - \langle\eta\rangle) \,, \tag{4.13}$$

to calculate the calibrated mistag probability $\omega(\eta)$. Here, $\langle\eta\rangle$ is the arithmetic mean of the estimated mistag probabilities, $\omega(\eta)$ is definied in Eq. (4.10) and p_0 and p_1 are the calibration parameters of the linear function. The introduction of $\langle\eta\rangle$ reduces the correlation between p_0 and p_1. A perfect calibrated tagger would result into $p_0 = \langle\eta\rangle$ and $p_1 = 1$.

Depending on the charge of the selected tagging particles, different interaction rates with the detector materials are possbile leading to different tagging efficiencies and true mistags for intial B and $\overline{B}$ mesons. The output of the tagging algorithms does not have to be independent of the initial b-meson flavour and two calibration functions are used to distinguish between B and $\overline{B}$ mesons:

$$\omega^{B}(\eta) = p_0^{B} + p_1^{B}(\eta - \langle\eta\rangle) \,, \quad \omega^{\overline{B}}(\eta) = p_0^{\overline{B}} + p_1^{\overline{B}}(\eta - \langle\eta\rangle) \,. \tag{4.14}$$

Here, the calibration parameters are defined as

$$p_i^{B} = p_i + \frac{\Delta p_i}{2} \,, \quad p_i^{\overline{B}} = p_i - \frac{\Delta p_i}{2} \quad \text{with } i = 0, 1 \,, \tag{4.15}$$

where p_i and Δp_i are the average and the difference of the calibration parameters of B and $\overline{B}$ mesons.

The tagging decisions and mistags of all taggers have to be combined. The combined probabilities that the initial B meson contained a b or $\bar{b}$ quark are calculated by

$$P(b) = \frac{p(b)}{p(b) + p(\bar{b})} \quad \text{and} \quad P(\bar{b}) = \frac{p(\bar{b})}{p(b) + p(\bar{b})} \,. \tag{4.16}$$

The propabilities $p(b)$ and $p(\bar{b})$ comprise tag d and mistag estimate η of a single tagger i that provides a non-zero tag decision and are given by

$$p(b) = \prod_i \left(\frac{1 + d_i}{2} - d_i(1 - \eta_i) \right) \,, \quad p(\bar{b}) = \prod_i \left(\frac{1 - d_i}{2} + d_i(1 - \eta_i) \right) \,. \tag{4.17}$$

The combined tag decisions and mistags are then defined as

$$(d, \eta) = \begin{cases} (+1, 1 - P(b)) & \text{if } P(b) > P(\bar{b}) \\ (-1, P(b)) & \text{if } P(b) < P(\bar{b}) \\ (0, 1/2) & \text{if } P(b) = P(\bar{b}) \end{cases} \,. \tag{4.18}$$

As correlations between the individual tagging algorithms are not considered but can be present, it is common to also calibrate the combinations again.

5 Measurement of $B_s^0 \to D^{*\pm}D^{\mp}$ Decays

This chapter describes the observation of the $B_s^0 \to D^{*\pm}D^{\mp}$ decay and the measurement of its branching fraction. The author performed all parts of the measurement. Advisory input but no direct contributions were given by Philipp Ibis and Margarete Schellenberg from the LHCb physics group from Dortmund. No other groups contributed. This analysis has been published in Ref. [25].

The branching fraction, $\mathcal{B}$, of a decay channel is the ratio of the number of *e.g.* B mesons decaying into a specific final state, $N(B \to f)$, with respect to all possible decay modes, $N(B)$. It is not possible to measure $N(B)$ directly, which corresponds to the total number of B meson candidates. Instead, it is calculated with the recorded integrated luminosity, $\mathcal{L}_{\text{int}}$, the $b\bar{b}$-production cross section, $\sigma_{b\bar{b}}$, the probability of a b quark hadronising with another quark q, f_q, and the total selection efficiency for the decay mode, ε. Accordingly, the branching fraction for the $B_s^0 \to D^{*\pm}D^{\mp}$ can be determined via

$$\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp}) = \frac{N(B_s^0 \to D^{*\pm}D^{\mp})}{2\sigma_{b\bar{b}}\mathcal{L}_{\text{int}} f_s \varepsilon_{B_s^0}} . \tag{5.1}$$

To cancel the large uncertainty on the $b\bar{b}$-production cross section and the uncertainty on the integrated luminosity, the ratio of the branching fractions of two decay modes can be measured. Dividing the branching fractions of the $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ decays leads to the ratio of branching fractions

$$\frac{\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})}{\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})} = \frac{N_{B_s^0}}{N_{B^0}} \frac{\varepsilon_{B^0}}{\varepsilon_{B_s^0}} \frac{f_d}{f_s} , \tag{5.2}$$

where $N_{B_s^0}$ and N_{B^0}, are the number of candidates reconstructed via specific D meson decays. By reconstructing both decays using the same D meson decays, the uncertainties on the branching ratios of the involved D meson decays cancel. The efficiencies in both decay channels are expected to be similar, so that the relative branching ratio in Eq. (5.2) is mostly defined by the ratio of the signal candidates, ${}^{N_{B_s^0}}/{}_{N_{B^0}}$, and the ratio of the hadronisation fractions, ${}^{f_d}/{}_{f_s}$. The hadronisation fraction is an external input with the value

$$\frac{f_s}{f_d} = 0.259 \pm 0.015 \tag{5.3}$$

for Run 1 [93]. The value for Run 2

$$\frac{f_s}{f_d} = 0.244 \pm 0.012 \tag{5.4}$$

is taken from [94], where $f_u = f_d$ is assumed.

The analysis is performed with data collected by the LHCb experiment in the years 2011 and 2012 at centre-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ (Run 1) as well as in the years 2015–2018 at $\sqrt{s} = 13\,\text{TeV}$ (Run 2). The whole dataset corresponds to an integrated luminosity of $9\,\text{fb}^{-1}$, where $3\,\text{fb}^{-1}$ was collected in Run 1 and $6\,\text{fb}^{-1}$ in Run 2. The recorded data sample contains $B^0 \to D^{*\pm}D^{\mp}$ as well as $B_s^0 \to D^{*\pm}D^{\mp}$ decays. Thus, the same data sample can be used to analyse both decay channels. Besides the recorded data samples, various simulation samples are utilised to develop the selection and evaluate efficiencies.

The reconstruction of the $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ candidates is explained in Sec. 5.1. Sec. 5.2 first discusses different types of background candidates that are present in the data sample and proceeds with detailing all steps of a selection chain developed to remove these candidates and to select the signal candidates. The evaluation of the selection efficiencies, ε_{B^0} and $\varepsilon_{B_s^0}$, is outlined in Sec. 5.3 and the extraction of the number of signal candidates, $N_{B_s^0}$ and N_{B^0}, via an extended maximum-likelihood fit is given in Sec. 5.4. Afterwards, the calculation of the significance of the measurement and the evaluation of systematic uncertainties are discussed in Sec. 5.5 and Sec. 5.6. Finally, with the measured number of signal candidates and the selection efficiencies, the relative branching ratio is calculated in Sec. 5.7 and the results are summarised in Sec. 5.8.

The measurement is performed blind to avoid an unconscious bias of the analyst during the optimisation of the selection. The blinding is done by cutting out the signal region of the invariant B_s^0 mass throughout the analysis. The control channel and simulation samples are used to optimise the selection and to calculate selection efficiencies, so that the information of the signal region is not needed at any point of the analysis. The cut-out signal region is only included after the selection is finalised and after systematic effects are fully understood.

5.1 Reconstruction

The B^0 and B_s^0 mesons are produced in the PV at the interaction point. They then pass through the detector with a flight distance of about $1\,\text{cm}$ and decay into $D^{*\pm}$ and $D^{\mp}$ mesons at the secondary vertex (SV). The $D^{\mp}$ meson is reconstructed through the decay $D^{\mp} \to K^{\pm}\pi^{\mp}\pi^{\mp}$, the most dominant decay with three charged hadrons in the final state. The $D^{\mp}$ meson traverses with a finite flight distance before its decay, which results into a distinct decay vertex of the $D^{\pm}$ meson downstream of the SV. The $D^{*\pm}$ meson is reconstructed through the decay $D^{*\pm} \to \overset{(-)}{D}{}^0\pi^{\pm}$, where the $\overset{(-)}{D}{}^0$ meson decays via $\overset{(-)}{D}{}^0 \to K^{\mp}\pi^{\pm}$. In the following chapters, charge conjugation is implied. The D^0 meson is not decaying promptly but downstream of the SV and the decay vertices of the $D^{*\pm}$ meson and D^0 meson can be separated from each other. Given the fact that the mass difference of the $D^{*\pm}$ meson and D^0 meson is barely larger than the pion mass, the pion that arises from the $D^{*\pm}$ decay has a low kinetic energy and a clean signature. This is why it is called *slow pion*. All particles that are analysed have to lie within the detector acceptance and be completely reconstructed.

5.2 Signal Selection

The data sample contains a large fraction of background candidates, which must be suppressed. Background candidates can be divided into combinatorial background and physical background. Physical background can be further divided into multiple classes. Candidates that mimic signal candidates resulting from misidentification of its final-state particles are classified as misidentified background. Partially-reconstructed background arises when particles are not reconstructed and the remaining tracks are wrongly combined to form a signal candidate. Additionally, non-resonant background candidate can arise, where the final state hadrons are the same as the ones from the signal decay, but do not stem from an intermediate meson and instead are formed directly in the B decay. These are referred to as single-charm background, where either no $D^{*\pm}$ or $D^{\mp}$ meson is present. Physical background appears as a peaking structure in the invariant mass distribution that has to be suppressed or otherwise be described by a PDF. Combinatorial background stems from unrelated particle tracks that do not necessarily come from the same vertex, but are combined and as a result mimic the signal candidate. In contrast to physical background, combinatorial background shows no peaking structures in the invariant mass distribution but has a flat, exponentially-decreasing shape. It cannot be removed completely and has to be described by a PDF. Due to the high track multiplicity at which the LHCb experiment operates, the data is dominated by combinatorial background.

To select signal candidates, a selection chain consisting of different selection steps is developed. At first, a preselection is applied (see Sec. 5.2.1). The selection proceeds with the use of vetoes to suppress physical backgrounds (see Sec. 5.2.2), the application of a multivariate classifier to suppress combinatorial background (see Sec. 5.2.3) and a suppression of the multiple candidates (see Sec. 5.2.4). The same selection is applied to the $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ decay channels. The overall selection strategy is the same for Run 1 and Run 2, but developed seperately for the Run 1, 2015–2016 and 2017–2018 data samples. Run 2 is split into 2015–2016 and 2017–2018 due to differences in the track reconstruction algorithms.

5.2.1 Preselection

The first part of this selection step is the centralised preselection, in which the stripping line `StrippingB02DstDBeauty2CharmLine` is used. Requirements on the properties of the final-state particles, intermediate particles and mother particle are applied as shown in Table A.1. Since minor differences in requirements on a few variables exist between different stripping versions for the different data-taking periods, these are aligned to have a better comparison of signal and background efficiencies. The variables are explained in detail in the following.

The number of long tracks, *i.e.* tracks that originate in the VELO and leave hits in all tracking stations, has to be less than 500 to restrict the track multiplicity and to enable a clean separation of the tracks. Either the trigger decision `Hlt2Topo`, which is triggered by particles decaying into two, three or four particles, or the trigger decision `Hlt2IncPhi`, which is an inclusive ϕ-trigger searching for the $\phi \to KK$ decay, is required.

A quantity that is used in several stripping requirements is the χ^2 variable, which is a measure for the quality of a fit. Oftentimes it is divided by the number of degrees of freedom, ndf. The final-state particles are required to have a low $\chi^2_{\text{track}}/\text{ndf}$ to ensure a good track reconstruction. Lower limits on their momenta, p, and the transverse momenta, p_{T}, guarantee the decay of a heavy meson. Final-state particles that do originate from the PV are excluded by requiring a high χ^2 of the impact parameter, IP, with respect to any PV, where IP is the shortest distance between the reconstructed track and the PV. The IP χ^2 is defined as the difference in the vertex-fit χ^2 of a given PV reconstructed with and without the B^0 being considered. The probability of a track being a ghost, which is a random combination of unrelated detector hits, is required to be small. Furthermore, pions and kaons have to be identified by the PID system through the difference of the log-likelihood between a pion and a kaon hypothesis, $\text{DLL}_{K\pi}$.

The shortest distance between pairs of final-state particles forming a vertex is required to be small by a cut on the distance of closest approach, DOCA, between all possible combinations of tracks. In addition, a cut on the $\chi^2_{\text{vtx}}/\text{ndf}$ ensures a correct vertex reconstruction. Moreover, a high χ^2 of the distance to any PV ensures a significantly displaced vertex from all PVs in the event. The direction angle, DIRA, is the cosine of the angle between the momentum direction of the respective particle and the flight direction from its production to its decay vertex. For all D mesons the DIRA has to be positive to ensure that they traverse into the detector plane. The sum of the p_{T} of all daughter particles of a D meson has to be larger than $1800\,\text{MeV}/c$ and the reconstructed masses of all D meson candidates are required to be around their nominal masses. Additionally, a cut on the mass difference of the $D^{*\pm}$ and D^0 mesons is used to reject misreconstructed candidates. A $D^{*\pm}$ and a $D^{\mp}$ candidate are combined to form a common vertex and reconstruct the B^0 or B_s^0 meson candidate. The vertex χ^2 has to have a good quality and the sum of the p_{T} of the D meson pair is required to be larger than $5\,\text{GeV}/c^2$. The IP χ^2 of the B^0 meson for the associated PV is required to be small. Furthermore, the decay time of the B^0 candidate is required to be larger than $0.2\,\text{ps}$ to reduce prompt background, where the reconstructed particles are produced directly in the pp collision in the PV and are wrongly combined to a B meson. Candidates are only kept if the invariant mass of the B meson lies inside a mass window covering the B^0 and

Table 5.1: Requirements of the additional preselection. Candidates that do not satisfy the requirements are rejected. The known values of the D-meson masses are taken from Ref. [34].

Requirement	Unit	Definition
$\lvert m_{K^{\pm}\pi^{\mp}} - m_{D^0,\mathrm{PDG}}\rvert < 50$	MeV/c^2	Invariant D^0 mass
$\lvert m_{K^{\pm}\pi^{\mp}\pi^{\mp}} - m_{D^{\mp},\mathrm{PDG}}\rvert < 50$	MeV/c^2	Invariant $D^{\mp}$ mass
$\lvert m_{(K^{\mp}\pi^{\pm})\pi^{\pm}} - m_{D^{*\pm},\mathrm{PDG}}\rvert < 50$	MeV/c^2	Invariant $D^{*\pm}$ mass
$m_{(K^{\mp}\pi^{\pm})\pi^{\pm}} - m_{K^{\pm}\pi^{\mp}} < 150$	MeV/c^2	$D^{*\pm}$-D^0 mass difference
$\tau/\sigma(D^{\pm}) > -1$		Decay time significance of $D^{\mp}$

B_s^0 signal candidates as well as background regions. In case there are multiple PVs in a collision, only the best PV is included, which is the PV for which the B^0 candidate has the lowest χ^2_{IP}.

After the centralised preselection, additional offline requirements are applied summarised in Table 5.1. While the centralised preselection uses variables determined by the `LoKiVertexFitter`, the variables used in the offline preselection are determined by the DTF. The three fits, in which either no constraints, a PV constraint or mass constraints of all D mesons are added, are required to have converged to be able to use the output variables of these fits in the rest of the analysis. These requirements retain almost all signal candidates. Requirements on all D mesons with high signal efficiencies are applied to suppress combinatorial background. Candidates lying inside a mass window of $\pm 50\,\mathrm{MeV}/c^2$ around the nominal masses of all D meson masses are selected. In addition, an efficient requirement on the mass difference of the $D^{*\pm}$ and D^0 meson is applied. The known mass difference is barely higher than the pion mass. By subtracting the invariant $K^-\pi^+\pi^+$ and $K^-\pi^+$ mass, mass resolution effects are reduced and correctly reconstructed $D^{*\pm} \to D^0\pi^{\pm}$ decays become visible as a sharp peak around the true mass difference. Candidates with larger mass differences are most likely caused by misreconstruction and are suppressed by limiting the mass difference to $150\,\mathrm{MeV}/c^2$. This requirement removes about 70% of the combinatorial background. Furthermore, the decay-time significance of the $D^{\mp}$ meson, defined as the decay time of the $D^{\mp}$ meson divided by its error, is required to be larger than -1 to suppress single-charm background with non-resonant $D^{\mp}$ mesons. The distributions of the variables used in the offline preselection are shown in Fig. 5.1 for the data-taking period 2017–2018. The total background rejection of the preselection is about 93% for all periods determined on the upper-mass sideband. The upper-mass sideband is defined as the region, where the invariant $D^{*\pm}D^{\mp}$ mass, $m_{D^{*\pm}D^{\mp}}$, is greater than $5400\,\mathrm{MeV}/c^2$. In this region no contributions from signal or physical backgrounds are expected. The mass distribution of the invariant $D^{*\pm}D^{\mp}$ mass after the preselection is shown in Fig. 5.2. A peak

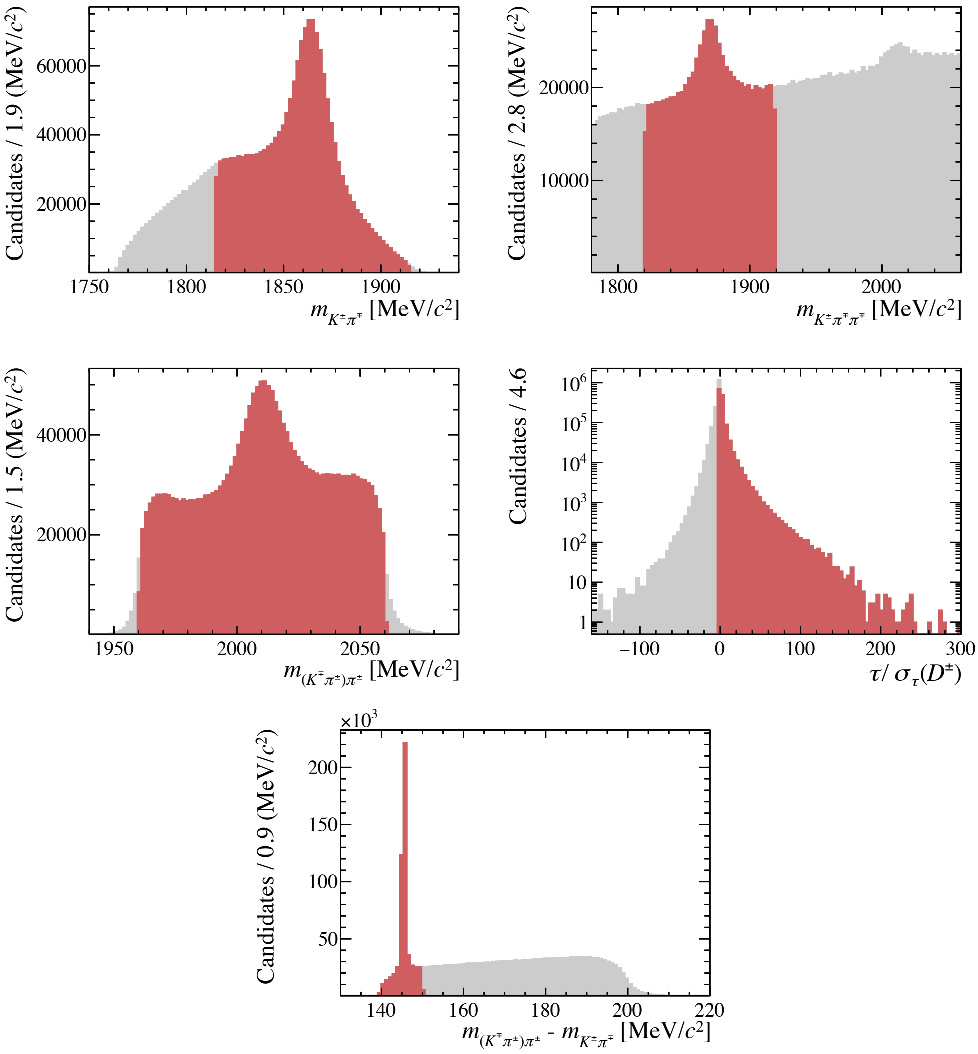

Figure 5.1: Distributions of the variables used in the preselection for the data sample of the data-taking period 2017-2018. The preselection comprises requirements on the invariant masses of the (top left) D^0, (top right) $D^\mp$ and (middle left) $D^{*\pm}$ mesons, (middle right) the decay time significance of the $D^\pm$ and (bottom) the difference of the invariant masses of the $D^{*\pm}$ and D^0 meson. By applying the respective requirement of the variable, candidates in the grey region are discarded and candidates in the red region are retained. The requirements are summarised in Table 5.1.

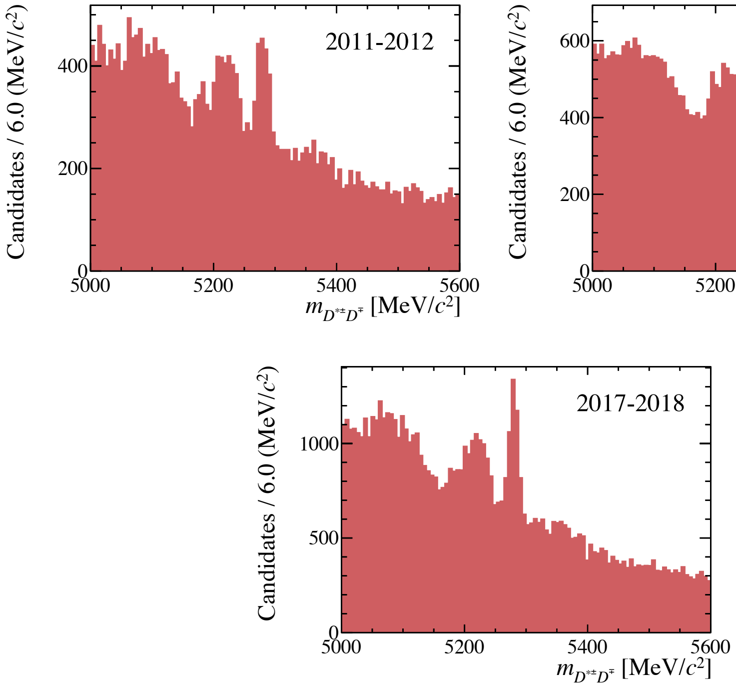

Figure 5.2: Invariant $D^{*\pm}D^{\mp}$ mass distribution after the preselection for the (top left) Run 1, (top right) 2015-2016 and (bottom) 2017-2018 data samples. A peak at the B^0 mass around $5280\,\mathrm{MeV}/c^2$ can be seen, while no clear peak at the B_s^0 mass around $5367\,\mathrm{MeV}/c^2$ is visible. At lower masses physical backgrounds are present. Combinatorial background constitutes the dominant background contribution.

at the B^0 mass is already visible, but the data sample is still dominated by combinatorial background and physical background at lower masses.

5.2.2 Vetoes

Physical background contributions can occur when kaons or protons stemming from hadronic decays of D_s^+ and Λ_c^+ hadrons are misidentified as pions and the decays are falsely reconstructed as a signal $D^+ \to K^-\pi^+\pi^+$ decay. The invariant mass of the background mother particle is shifted so that it still contributes to the invariant $D^{*-}D^+$ mass distribution. To identify these backgrounds, the mass of the $K^-\pi^+\pi^+$ system from the D^+ candidate is recalculated after applying the kaon or proton mass hypothesis to either of the pions. The background contributions are visible as peaks at the D_s^+ or Λ_c^+ mass in the recalculated invariant mass distribution. As a consequence of the preselection requirement on the mass difference between the D^{*-} and D^0 mesons, no physical background due to mis-reconstruction of the D^{*-} decay is present. Misidentified background contributions are reduced to a negligible level by applying vetoes using requirements on the mass of the background mother particle and PID variables. For the latter, ProbNN variables, which give a probability of the particle to be a kaon, pion, muon or electron provided by neural networks, are utilised.

The distributions of ProbNN variables show differences between simulation and data. These inaccuracies in simulation are corrected by a method, in which simulated ProbNN variables are transformed to match calibration samples of recorded data. This is based on a kernel density estimation in four dimensions: ProbNN value, transverse momentum and pseudorapidity of the track, and the event multiplicity [95]. An advantage of this method is that the correlations between the corrected ProbNN variables and other variables are preserved. A comparison of corrected and uncorrected simulated ProbNN distributions with sWeighted data shows an improvement after the correction for Run 1 samples. In the periods 2015–2016 and 2017–2018 most of the uncorrected simulated ProbNN distributions already describe the data well and a minor improvement is achieved with the correction. As an example of the correction method the pion ProbNN and kaon ProbNN distribution of a pion of the $D^{\mp}$ decay and the slow pion of the $D^{*\pm}$ decay, respectively, that are utilised in the selection, are shown in Fig. 5.3 for simulated $B^0 \to D^{*\pm} D^{\mp}$ decays and sWeighted data (see Sec. 5.4.2) of the period 2011–2012.

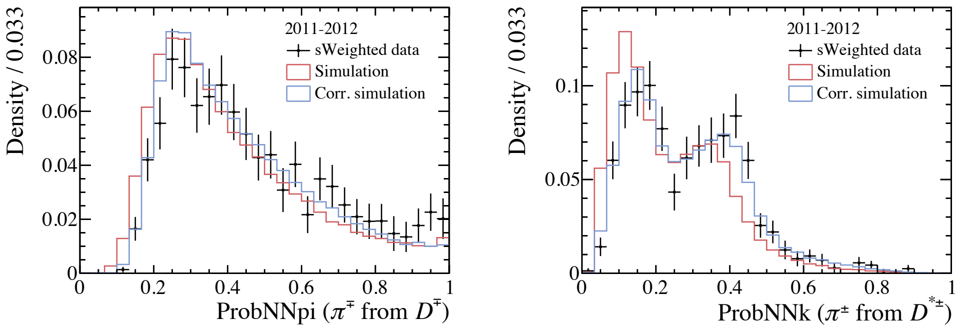

Figure 5.3: The (left) pion ProbNN distribution of a pion of the $D^{\mp}$ decay and the (right) kaon ProbNN distribution of the slow pion of the $D^{*\pm}$ decay. The corrected and uncorrected ProbNN distributions of simulated $B^0 \to D^{*\pm} D^{\mp}$ decays are illustrated in blue and red, respectively, and sWeighted data is shown in black.

An improved separation between two particle identifications is achieved with the ratio of the ProbNN variable for the new particle hypothesis and the sum of the ProbNN variables for the new particle hypothesis and a pion

$$\mathrm{ProbNN}_{h\pi} = \frac{\mathrm{ProbNN}_h}{\mathrm{ProbNN}_h + \mathrm{ProbNN}_\pi}\,. \tag{5.5}$$

If the pion is a different, heavier particle it most likely has a higher p_{T} than the real pion. Thus, a clear background peak is visible in the recalculated mass distribution when the different mass hypothesis is applied to the $\pi_{\mathrm{high}\,p_{\mathrm{T}}}$. When it is applied to the $\pi_{\mathrm{low}\,p_{\mathrm{T}}}$ a smaller background peak or even no resonant structure is present. The signal efficiency can be further increased by applying different requirements for $\pi_{\mathrm{high}\,p_{\mathrm{T}}}$ and $\pi_{\mathrm{low}\,p_{\mathrm{T}}}$.

In the following, all vetoes are explained. Each veto is optimised to have a high signal efficiency, while rejecting 95% of the background. For this, different combinations of mass and PID requirements are investigated. The background rejection is determined on simulated background samples in which the respective background decay is reconstructed as the signal decay.

D_s^+ and ϕ veto: Background contributions from $B^0 \to D^{*-} D_s^+$ decays with $D_s^+ \to K^- K^+ \pi^+$ occur when the K^+ meson is misidentified as a π^+ meson. The kaon mass hypothesis is assigned to either $\pi_{\mathrm{high}\,p_{\mathrm{T}}}$ or $\pi_{\mathrm{low}\,p_{\mathrm{T}}}$ when calculating the invariant masses of the daughter hadrons, which are labelled as $K^- K^+ \pi^+$ or $K^- \pi^+ K^+$, respectively. The candidate is rejected if the pion has a high probability to be identified as a kaon and the recalculated mass is compatible with the known D_s^+ mass [34]. The mass distributions of both combinations are shown in Fig. 5.4 for the 2017–2018 data sample. The D_s^+ decay can also proceed through the intermediate $D_s^+ \to \phi \pi^+$ decay with $\phi \to K^- K^+$. To further suppress this case, the mass of the $K^- \pi^+$ system is recalculated, where again the kaon mass hypothesis is applied to one of the two pions. Candidates are rejected if the pion has a high probability to be identified as a kaon and the mass of recalculated $K^- K^+$ pair is compatible with the known $\phi\,(1020)$ meson mass [34]. The distribution of the invariant $K^- K^+$ mass is also shown in Fig. 5.4, where a peak at the ϕ mass can be seen.

The D_s^+ and ϕ vetos have been optimised simultaneously and their requirements are reported in Table A.2 along with their signal efficiencies. The signal efficiencies of about 91% are relatively low. This comes from the fact that the D_s^+ mass is higher than the D^+ mass, but the kaon-pion misidentification results in a shift of the invariant mass to smaller values close to the mass of the D^+ meson. This can be seen in the left plot of Fig. 5.5, which shows the recalculated invariant $K^- K^+ \pi^+$ mass distribution for simulated $B^0 \to D^{*\pm} D^{\mp}$ decays that are used to calculate the efficiencies. A peak at roughly the nominal D_s^+ mass is present, which results in a rejection of a high percentage of simulated $B^0 \to D^{*\pm} D^{\mp}$ candidates with the D_s^+ veto.

Λ_c^+ veto: Background contributions from $\Lambda_b^0 \to \Lambda_c^+ D^{*-}$ with $\Lambda_c^+ \to K^- p \pi^+$ occur if the proton is mis-identified as a pion leading to a false reconstruction of the Λ_c^+ baryon as a D^+ meson. After assigning the proton mass to either $\pi_{\mathrm{high}\,p_{\mathrm{T}}}$ or $\pi_{\mathrm{low}\,p_{\mathrm{T}}}$ of the D^+ decay, the invariant masses of the $K^- p \pi^+$ or $K^- \pi^+ p$ combinations are calculated. In the invariant mass of the $K^- p \pi^+$ system, a small

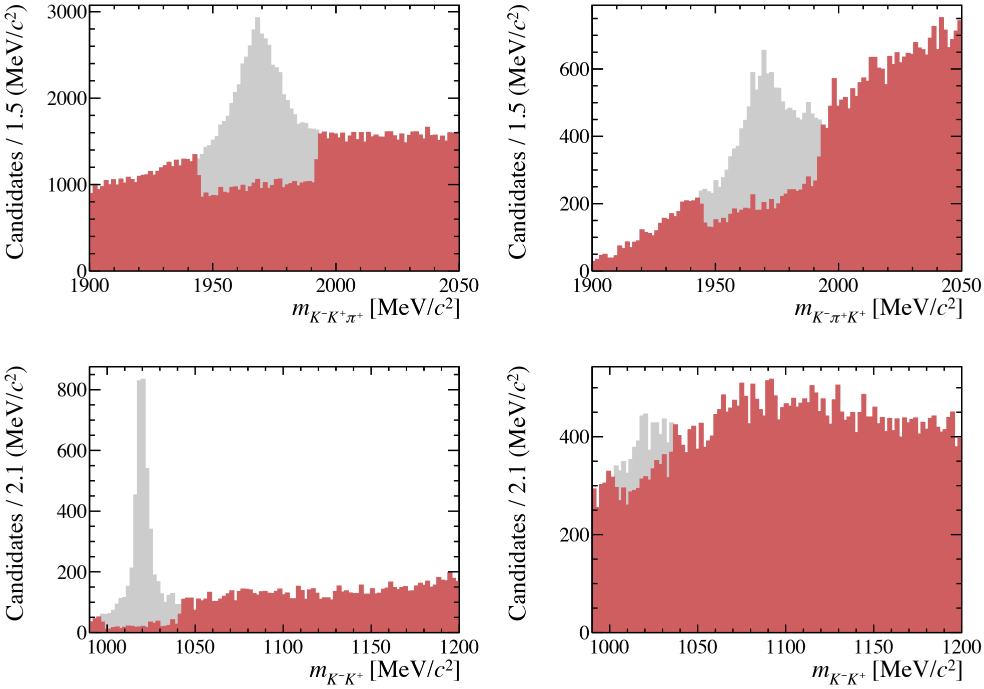

Figure 5.4: Effects of the (top) D_s^+ veto and (bottom) ϕ veto for 2017–2018 data. The kaon mass hypothesis is given to either the pion of the D^+ decay with the (left) higher p_T or with the (right) lower p_T to calculate the invariant masses of $K^-K^+\pi^+$ and $K^-\pi^+K^+$ for the D_s^+ veto or the invariant masses of the K^-K^+ pair for the ϕ veto. By applying the veto, candidates in the red area are kept while candidates in the grey area get rejected.

resonant structure is visible at the nominal Λ_c^+ mass [34]. In contrast, a resonant structure in the mass of the $K^-\pi^+p$ system is not as clearly visible. In both cases, candidates are rejected if the pion has a high probability to be identified as a proton and the recalculated mass is compatible with the known Λ_c^+ mass [34]. The mass distributions before and after the veto are shown in Fig. 5.6.

The requirements of this veto are reported in Table A.2 along with the respective signal efficiencies, which are about 98% for all data-taking periods. The recalculated invariant $K^-p\pi^+$ mass distribution using simulated $B^0 \to D^{*\pm}D^{\mp}$ decays is shown in the right plot of Fig. 5.5. In contrast to the D_s^+ veto, no resonant structure is present. Thus, only few simulated $B^0 \to D^{*\pm}D^{\mp}$ candidates are rejected with the Λ_c^+ veto.

Single-charm veto: Non-resonant background contributions from decays of the form $B_{(s)}^0 \to D^{*-}h^-h^+h^+$ can occur, where the three hadrons do not stem from a D^+ resonance. Dominant contributions of this type can arise from the

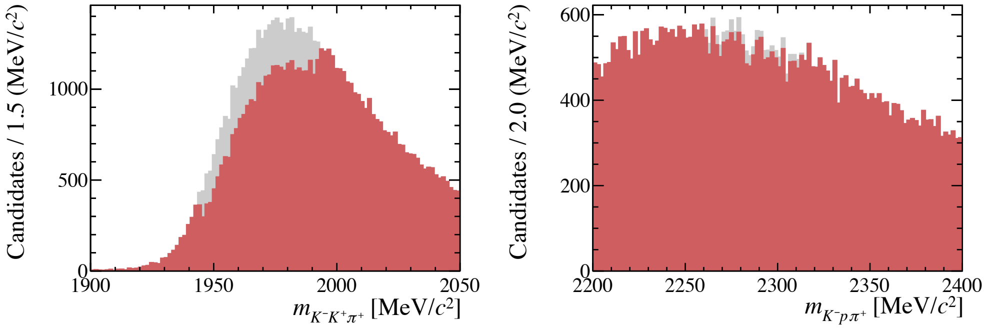

Figure 5.5: Effects of the (left) D_s^+ veto and (right) Λ_c^+ veto using simulated $B^0 \to D^{*\pm}D^{\mp}$ samples for the period 2017–2018. The kaon and proton mass hypotheses are given to the $\pi_{\mathrm{high}\,p_{\mathrm{T}}}$ of the D^+ decay to calculate the invariant masses of (left) $K^-K^+\pi^+$ and (right) $K^-p\pi^+$, respectively. In the $K^-K^+\pi^+$ mass distribution a peak at roughly the D_s^+ mass is present, whereas in the $K^-p\pi^+$ mass distribution no peak is visible. By applying the veto, candidates in the red area are kept while candidates in the grey area get rejected.

$B^0 \to D^{*-}\pi^-\pi^+\pi^+$ decay and possibly the $B_s^0 \to D^{*-}K^-\pi^+\pi^+$ decay. Other contributions are not considered as they are either Cabbibo suppressed or need multiple misidentifications. Such background candidates are rejected by a requirement on the χ^2 of the flight distance of the D^+ meson, $\chi^2_{D^\pm,\mathrm{FD}}$. This requirement ensures that the B^0 and D^+ decay vertices are well separated and the three hadrons do not stem from the SV. The $B_s^0 \to D^{*-}K^-\pi^+\pi^+$ decay has the same final-state particles as the signal $B_s^0 \to D^{*\pm}D^{\mp}$ decay and therefore its contribution should be located at the signal B_s^0 peak. Due to this fact, the $B^0 \to D^{*-}\pi^-\pi^+\pi^+$ decay is used as representative for both background contributions in the development of this veto. As the vertices of both background channels should have the same signatures, the veto should also suppress the $B_s^0 \to D^{*-}K^-\pi^+\pi^+$ contribution. In Fig. 5.7 the invariant $D^{*-}\pi^-\pi^+\pi^+$ mass is shown for the 2017–2018 data sample. A clear peak at the B^0 mass before applying the veto and no significant remaining background after applying the veto is visible. No background simulation is available, thus this veto is optimised using recorded data without the calculation of the background rejection. The requirements of the veto are reported in Table A.3 along with the signal efficiencies for all data-taking periods.

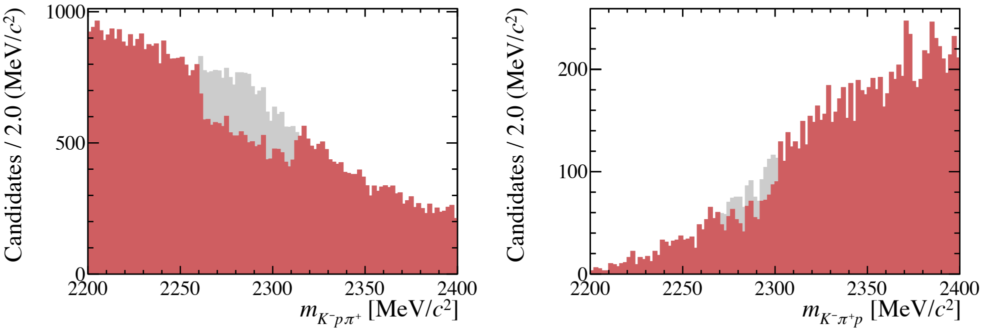

Figure 5.6: Effects of the Λ_c^+ veto for 2017–2018 data. The proton mass hypothesis is given to either (left) $\pi_{\mathrm{high}\,p_\mathrm{T}}$ or (right) $\pi_{\mathrm{low}\,p_\mathrm{T}}$ of the D^+ decay to calculate the invariant masses of $K^-p\pi^+$ and $K^-\pi^+p$. By applying the veto, candidates in the red area are kept while candidates in the grey area get rejected.

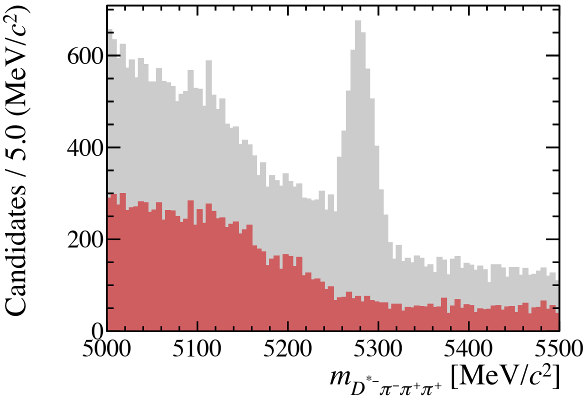

Figure 5.7: The invariant $D^{*-}\pi^-\pi^+\pi^+$ mass with constraint on the D^{*-} mass is shown for 2017–2018 data. By applying the veto against single-charm background candidates in the red area are kept and candidates in the grey area are rejected.

5.2.3 Multivariate Selection

After suppressing physical background contributions, a multivariate selection is developed to further reduce the amount of combinatorial background. In particular, a Boosted Decision Tree (BDT) [86, 87] implemented in the TMVA framework [96] is trained utilising the AdaBoost method. The training of the BDT and the optimisation of the BDT selection is described in the following.

BDT Training: Simulated $B^0 \to D^{*\pm} D^{\mp}$ decays are utilised as the signal proxy, because the sample of simulated $B_s^0 \to D^{*\pm} D^{\mp}$ decays of the data-taking period 2015–2016 has too few candidates for the BDT training. The background proxy is given by the upper-mass sideband of data with $m_{D^{*\pm}D^{\mp}} > 5450\,\text{MeV}/c^2$, in which neither signal nor physical backgrounds contribute and only combinatorial background candidates are present. All previous selection steps are applied to the samples used for the training. A k-fold technique with $k = 5$ is adopted. In order to select the input variables for the BDT training, multiple variables and their contribution to the BDT performance are evaluated based on the value of the area under the ROC curve (AUC). The receiver operating characteristic (ROC) curve describes the true positive rate (signal efficiency) as a function of the false positive rate (background efficiency). A ROC curve on the diagonal equals a random classification and has an AUC value of 0.5. In contrast, a perfect classification has an AUC value of one. Hence, a larger AUC value indicates a better BDT performance. Variables that do not contribute significantly to the BDT performance are discarded. Finally, 13 variables are used in the training listed in Table A.4. The variables comprise the mass difference between the $D^{*\pm}$ and D^0 meson, corrected ProbNN ratios (see Sec. 5.2.2), transverse momenta, p_{T}, the decay time significance, t/σ_t, of the $D^{\pm}$ meson, the impact parameter significances, χ^2_{IP}, and flight distance qualities, χ^2_{FD}. The hyperparameters used in the BDT training are the same for each of the three data-taking periods. To optimise these hyperparameters, the area under the ROC curve is utilised. Each BDT is build out of 300 trees with a maximum depth of two. At each node of the tree at least 2% of the training candidates must still be present. For each variable a scan at 20 points is performed to find the optimal cut point. The learning rate of the Adaboost method is set to $\beta = 0.8$. In addition, a bagging method is deployed, where the bagged sample has to be 80% of the size of the original sample. In Fig. 5.8 the BDT classifier distribution of the training and test samples as well as the corresponding ROC curve are shown for one fold and for the data-taking period 2017–2018 . The training and test samples are in good agreement for all data-taking periods. Hence, it can be assumed that no overtraining is present. The BDT classifier distributions of the five folds are compatible with each other. Thus, they are combined choosing a random BDT for candidates which were not used in the respective training, and afterwards treated as one.

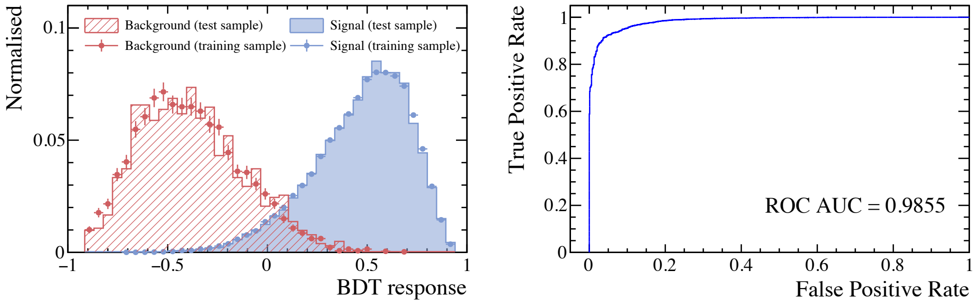

Figure 5.8: On the left, a comparison of the BDT classifier distribution of the training and testing samples of (red) background and (blue) signal for the data-taking period 2017-2018 is shown for one fold. The shaded areas illustrate the test sample and the data points describe the training sample. On the right, the ROC curve of the BDT is shown for the same fold. The ROC AUC value is close to one and indicates a good classification.

BDT requirement optimisation: To perform the multivariate selection, a requirement on the BDT response is applied. For this, a figure of merit (FOM) is deployed. The requirement is equal to the position of the maximum of the FOM. Here, a suitable FOM is one that is independent of expectations about the presence of signal candidates [97]:

$$\mathrm{FOM} = \frac{\varepsilon(\mathrm{cut})}{\frac{a}{2} + \sqrt{N_{\mathrm{B}}(\mathrm{cut})}}\,, \tag{5.6}$$

with the signal efficiency, ε, the number of background candidates in the signal region, N_{B}, and the target significance, a, in numbers of the standard deviation, which is set to three.

The efficiency of $B_s^0 \to D^{*\pm}D^{\mp}$ candidates cannot be determined on recorded $B_s^0 \to D^{*\pm}D^{\mp}$ data. Thus, it is investigated which way of determining the efficiency for different cuts on the BDT classifier matches real $B_s^0 \to D^{*\pm}D^{\mp}$ decays best. The efficiencies calculated with simulated $B^0 \to D^{*\pm}D^{\mp}$ decays and with $B^0 \to D^{*\pm}D^{\mp}$ candidates from recorded data, show small but significant deviations in the data-taking period 2011-2012 as well as 2015–2016. This implies differences between recorded data and simulation. Thus, it can be expected that the efficiency calculated with simulated $B_s^0 \to D^{*\pm}D^{\mp}$ decays is also not accurate. A comparison of the signal efficiencies calculated with simulated $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ decays shows that both distributions are compatible. It is assumed that differences between recorded data and simulation are similar in both decays and that the efficiencies using recorded $B^0 \to D^{*\pm}D^{\mp}$ data are compatible with the efficiency using recorded $B_s^0 \to D^{*\pm}D^{\mp}$ data. Therefore, the B^0 signal efficiency for a specific cut point is determined by maximum likelihood fits to the invariant $D^{*\pm}D^{\mp}$ mass on data around the known B^0 mass [34] before and after the application of this cut point. For the $B^0 \to D^{*\pm}D^{\mp}$ and combinatorial background components a model as discussed in Sec. 5.4 is used, where the shape parameters are obtained with simulated samples that passed the previous selection steps. Smaller deviations between $B^0 \to D^{*\pm}D^{\mp}$ and $B_s^0 \to D^{*\pm}D^{\mp}$ can lead to a slightly suboptimal BDT requirements but do not lead to a bias on the branching ratio.

The number of background candidates in the B_s^0 signal region is estimated with the upper-mass sideband ($5450\,\text{MeV}/c^2 < m_{D^{*\pm}D^{\mp}} < 5900\,\text{MeV}/c^2$). Due to the fact that the number of candidates in the sideband becomes very low at high BDT cut points, fits to the data become unstable and tend to not converge. Thus, another method is used to estimate N_B. The upper-mass sideband is split into two halves with the lower boundary, m_l, and higher boundary, m_h. An exponential function $f(m) = a\exp(-bm)$ is used to describe the number of candidates

$$
\begin{aligned}
N_l &= \int\limits_{m_l}^{m_l+\Delta m/2} a\exp(-bm)\mathrm{d}m = \frac{a}{b}\exp(-bm_l)[1-\exp(-b\frac{\Delta m}{2})]\,,\\
N_h &= \int\limits_{m_l+\Delta m/2}^{m_h} a\exp(-bm)\mathrm{d}m = \frac{a}{b}\exp(-b(m_l+\frac{\Delta m}{2}))[1-\exp(-b\frac{\Delta m}{2})]\,,
\end{aligned}
\tag{5.7}
$$

in the lower and higher half, with $\Delta m = m_h - m_l$. With Eq. (5.7) the parameters a and b of the exponential function are calculated as

$$
\begin{aligned}
b &= \frac{2}{\Delta m}\ln\frac{N_l}{N_h}\,,\\
a &= \frac{bN_l}{1-\exp\left(-b\Delta m/2\right)}\exp(bm_l)\,.
\end{aligned}
\tag{5.8}
$$

Using a and b from Eq. (5.8) allows to evaluate the number of background candidates in another mass range. The number of background candidates in the signal region around the known B_s^0 mass with the lower and higher boundary m_{l,B_s^0} and m_{h,B_s^0} is then given by

$$
N_\text{B} = \int\limits_{m_{l,B_s^0}}^{m_{h,B_s^0}} a\exp(-bm)\mathrm{d}m = \frac{a}{b}[\exp(-bm_{l,B_s^0}) - \exp(-bm_{h,B_s^0})]\,. \tag{5.9}
$$

The B_s^0 signal region is defined by the smallest interval that contains 98.76% (2.5σ interval) of the simulated candidates.

A three-dimensional scan of the requirement on the BDT response in the three data-taking periods is performed. The values of ε and N_B combined for all data-taking periods are determined for every combination of the three BDT requirements. All combinations of cuts for the three periods are scanned, where the cut points are varied in a range of 0.0 to 1.0 for each period. The optimal combination of the cut points is found to be 0.54, 0.42 and 0.48 for 2011–2012, 2015–2016 and 2017–2016, respectively. Fig. 5.9 shows the projection of the FOM for the period 2017–2018 with the requirements on the BDT of the other two periods being fixed to their optimal values, along with the signal and background efficiencies.

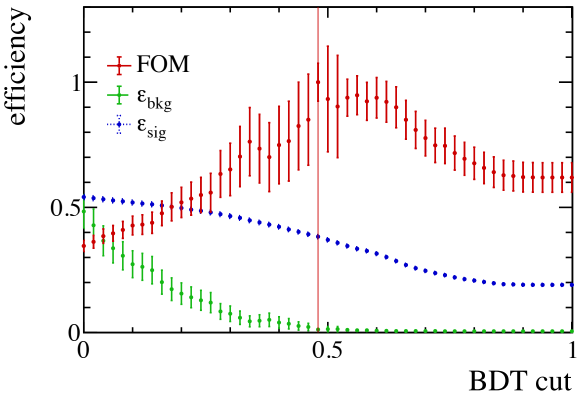

Figure 5.9: Projection of the FOM at different cuts on the BDTs for the period 2017–2018. The cuts on the BDT of the period 2011–2012 and 2015–2016 are fixed to their optimal values 0.54 and 0.42, respectively. The FOM, shown in red, is normalized to its highest value, indicated by the red vertical line. The maximum is at 0.48. The (green points) background efficiency is calculated using Eq. (5.9) and its error is propagated with the uncertainties of N_l and N_h from Eq. (5.7). The (blue points) signal efficiency is calculated with recorded data and its uncertainty comes from the yield of the fit to data after the cut on the BDT, where the uncertainty on the normalisation yield from the fit with no cut on the BDT is neglected. The uncertainties of the FOM are calculated with gaussian error propagation.

5.2.4 Multiple Candidates

It is possible that in an selected event more than one candidate remains. The fraction of events in data with such multiple candidates is less than 1% in Run 1 and about 1-2% for the 2015-2016 and 2017-2018 samples. The slightly higher fraction in Run 2 can be explained by the higher track multiplicity. Because of the low branching fraction of the $B^0 \to D^{*\pm}D^{\mp}$ decay mode, multiple signal decays per event are unlikely. As they are equally likely signal decays, one of the multiple candidates in an event is randomly selected and the others are exluded.

5.3 Efficiency Determination

The efficiencies of each selection step are needed as the total selection efficiency directly enters the calculation of the branching ratio (see Eq. (5.2)). The efficiency of a single selection step is determined with respect to the previous selection step. As a result, the total selection efficiency is the product of each single efficiency. Simulated $B^0 \to D^{*\pm}D^{\mp}$ and $B_s^0 \to D^{*\pm}D^{\mp}$ samples are used to obtain the efficiencies. This is feasible because differences between simulation and data are expected to cancel in the ratio of efficiencies of $B^0 \to D^{*\pm}D^{\mp}$ and $B_s^0 \to D^{*\pm}D^{\mp}$ decays. A systematic uncertainty is assigned due to any remaining differences coming mainly from the ProbNN variables (see Sec. 5.6). The individual efficiencies are described in the following. The selection efficiencies for $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ for each data-taking period are given in Table 5.2.

Table 5.2: The signal efficiencies of each selection step for the $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ decay channels for Run 1, 2015–2016 and 2017–2018, determined with simulation, as well as the total selection efficiencies.

		Signal efficiency [%]	
		$B_s^0 \to D^{*\pm}D^{\mp}$	
	Run I	2015–2016	2017–2018
Geometric	15.157 ± 0.020	16.19 ± 0.05	16.178 ± 0.027
Stripping	0.872 ± 0.003	1.238 ± 0.010	1.266 ± 0.005
Preselection	94.03 ± 0.08	93.68 ± 0.16	93.28 ± 0.10
Veto	79.31 ± 0.14	80.69 ± 0.26	80.56 ± 0.16
BDT	52.96 ± 0.20	69.25 ± 0.34	55.24 ± 0.23
$\varepsilon_{B_s^0}$ (total)	0.05221 ± 0.00030	0.10491 ± 0.00111	0.08502 ± 0.00053
		$B^0 \to D^{*\pm}D^{\mp}$	
	Run I	2015–2016	2017–2018
Geometric	15.006 ± 0.020	15.99 ± 0.05	16.034 ± 0.027
Stripping	0.847 ± 0.003	1.213 ± 0.005	1.213 ± 0.004
Preselection	93.69 ± 0.07	93.15 ± 0.10	93.11 ± 0.08
Veto	79.26 ± 0.12	80.52 ± 0.17	80.36 ± 0.15
BDT	52.02 ± 0.17	67.93 ± 0.22	54.41 ± 0.18
ε_{B^0} (total)	0.04910 ± 0.00026	0.09882 ± 0.00059	0.07918 ± 0.00043
$\varepsilon_{B_s^0}/\varepsilon_{B^0}$	1.063 ± 0.008	1.062 ± 0.013	1.074 ± 0.009

Geometric efficiencies: The signal decay chain has to be completely reconstructed and its final-state particles have to lie in the acceptance of the detector. When the simulated samples are generated, the so-called generator level cut rejects events with particles that are outside of the detector acceptance at an early stage of the generation. The LHCb simulation group provides the efficiency of this requirement.

Reconstruction, trigger and stripping efficiencies: The efficiency is calculated with simulated $B^0 \to D^{*\pm}D^{\mp}$ and $B_s^0 \to D^{*\pm}D^{\mp}$ samples. It comprises the reconstruction, trigger and stripping efficiencies, where the only trigger requirements are the ones made in the stripping. The stripping requirements are very loose, but signal candidates can still be rejected. In addition, it may happen that the reconstruction of the B^0 candidate fails even if the B^0 meson is in the detector acceptance. The significantly higher efficiency in Run 2 compared to Run 1 is caused by the trigger system being more efficient in the selection of beauty hadrons in Run 2 due to the new HLT design [72] and the higher centre-of-mass energy leading to higher momenta.

Preselection, Vetoes and BDT: The selection efficiencies of the preselection, vetoes and the BDT are calculated with simulated $B^0 \to D^{*\pm}D^{\mp}$ and $B_s^0 \to D^{*\pm}D^{\mp}$ samples. The vetoes and the BDT training utilise the corrected ProbNN variables. A comparison of the corrected variables between simulation and sWeighted data shows a good correspondence. Smaller, still persisting differences are expected to cancel when dividing the efficiencies of $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ decays.

5.3.1 Effect of the B_s^0 lifetime

In the course of this analysis the simulation used for the efficiency calculation is generated using the average B_s^0 lifetime for the $B_s^0 \to D^{*\pm}D^{\mp}$ decay mode, though the heavy and light eigenstates of the B_s^0 meson have significantly different lifetimes. Further, when measuring the relative branching ratio, time-integrated event yields are used without taking account of B_s^0-$\overline{B}_s^0$ mixing effects. As outlined in Ref. [98], the experimental branching ratio can be converted into a theoretical one that is independent of B_s^0-$\overline{B}_s^0$ mixing through

$$\mathcal{B}(B_s^0 \to f)_{\text{theory}} = \left(\frac{1 - (\Delta\Gamma_s/2\Gamma_s)^2}{1 + \mathcal{A}_f^{\Delta\Gamma}(\Delta\Gamma_s/2\Gamma_s)} \right) \mathcal{B}(B_s^0 \to f)_{\text{experiment}} \tag{5.10}$$

where $\mathcal{A}_f^{\Delta\Gamma}$ is the *CP* observable defined in Eq. (2.22). In particular, $\mathcal{A}_{\Delta\Gamma}^f = 0$ corresponds to an average B_s^0 lifetime, while $\mathcal{A}_{\Delta\Gamma}^f = \pm 1$ corresponds to either a purely heavy or purely light B_s^0 eigenstate. Since the selection efficieny depends on the B_s^0 lifetime, the effect of a changed B_s^0 lifetime on the efficiency, when considering either a purely heavy or a purely light B_s^0 eigenstate, is evaluated. Using selected simulated $B_s^0 \to D^{*\pm} D^{\mp}$ decays, each event is weighted with

$$w_{L,H} = \frac{\exp\left(-t/\tau_{L,H}\right)}{\tau_{L,H}} \bigg/ \frac{\exp\left(-t/\tau\right)}{\tau} \ , \tag{5.11}$$

which describes the normalised PDF of the light (heavy) eigenstate divided by the normalised PDF of the average, where $\tau_{L,H}$ are the life times of the light and heavy eigenstates. The correction factor for the efficiency is then produced by dividing the sum of the weights by the total number of events. It is found that the correction factors for all data-taking periods are compatible. Lastly, the resulting values of the integrated correction factors are 1.042 (0.949) for a purely heavy (light) B_s^0 eigenstate. The equivalent effect for the B^0 decay mode is not considered as it is assumed to be negligible due to the small value of $\Delta\Gamma_d$ [34].

5.4 Extraction of Signal Candidates

After all selection steps, an unbinned maximum likelihood fit to the invariant $D^{*\pm} D^{\mp}$ mass distribution, with constraints on the $D^{*\pm}$, D^0 and $D^{\mp}$ mass, is performed to determine the number of signal B^0 and B_s^0 candidates. This is done separately for each data-taking period. Background contributions are still present in the invariant mass distribution and need to be taken into account. Apart from combinatorial background, partially reconstructed backgrounds from $B^0 \to D^{*\pm} D^{*\mp}$ and $B_s^0 \to D^{*\pm} D^{*\mp}$ decays are present. In these, one of the $D^{*\pm}$ mesons decays into a charged D meson and an unreconstructed π^0 or photon. For the parametrisation of the mass distribution, a PDF, $\mathcal{P}$, consisting of components for $B^0 \to D^{*\pm} D^{\mp}$ and $B_s^0 \to D^{*\pm} D^{\mp}$ signal decays, combinatorial background, and partially reconstructed $B^0 \to D^{*\pm} D^{*\mp}$ and $B_s^0 \to D^{*\pm} D^{*\mp}$ decays, is utilised:

$$\begin{aligned} N\mathcal{P} = {} & N_{B^0 \to D^{*\pm} D^{\mp}} \mathcal{P}_{B^0 \to D^{*\pm} D^{\mp}} + N_{B_s^0 \to D^{*\pm} D^{\mp}} \mathcal{P}_{B_s^0 \to D^{*\pm} D^{\mp}} + N_{\text{Comb}} \mathcal{P}_{\text{Comb}} \\ & + N_{B^0 \to D^{*\pm} D^{*\mp}} \mathcal{P}_{B^0 \to D^{*\pm} D^{*\mp}} + N_{B_s^0 \to D^{*\pm} D^{*\mp}} \mathcal{P}_{B_s^0 \to D^{*\pm} D^{*\mp}} \ . \end{aligned} \tag{5.12}$$

Here, N_i denotes the number of candidates of the respective component.

5.4.1 Mass Model

There is no physical reason to use a specific PDF as a model and various PDFs can describe a component reasonably. To develop a proper model that describes the different components, simulation samples are used. Only the parametrisation of the combinatorial background is extracted in the fit to recorded data. In the following the parametrisations of each component are explained.

Signal: The $B^0 \to D^{*\pm}D^{\mp}$ component is modelled by the sum of two Crystal Ball functions [99]. A Crystal Ball function consists of a Gaussian core that transits into a power-law tail to one direction and is defined by

$$f(m;\alpha,n,\mu,\sigma) \propto \begin{cases} \exp(-\frac{(m-\mu)^2}{2\sigma^2}), & \text{for } \frac{m-\mu}{\sigma} > -\alpha \\ A\cdot(B-\frac{m-\mu}{\sigma})^{-n}, & \text{for } \frac{m-\mu}{\sigma} \leqslant -\alpha \end{cases} \tag{5.13}$$

with the parameters A and B

$$\begin{aligned} A &= \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) , \\ B &= \frac{n}{|\alpha|} - |\alpha| \ . \end{aligned} \tag{5.14}$$

The sign of the parameter α defines the direction of the tail, *i.e.* whether the tail is towards lower ($a > 0$) or higher ($a < 0$) masses. Further, the degree n defines the slope of the tail. The two Crystal Ball functions share a common mean, μ, but have different widths, σ_i, and tails to opposite directions. The parameters and the fraction between the Crystal Ball functions are obtained by a fit to the invariant $D^{*\pm}D^{\mp}$ mass distribution of simulated $B^0 \to D^{*\pm}D^{\mp}$ decays. All parameters are allowed to float except for n. As α and n are correlated, n is set to a fixed value for a more robust fit model. The resulting values of the fit parameters are reported in Table A.5 for all three data-taking periods. Fig. 5.10 shows the invariant $D^{*\pm}D^{\mp}$ mass distribution overlaid with the projection of the fitted PDF for the period 2017–2018. The distributions and fit projections of the other two data-taking periods are not shown at this point, but look similar.

The $B_s^0 \to D^{*\pm}D^{\mp}$ component is described by the same model with the same parameters. The only exception is the mean, which is shifted towards the mean of the B^0 component, μ_{B^0}, by the difference of the known B_s^0 and B^0 masses [34]. This is reasonable, as the widths of the shapes for both decay channels are compatible with each other on simulated data.

In fits to recorded data, the only floating parameters are the mean μ_{B^0} and a single scale factor, s, with which the widths are multiplied to correct them for inaccuracies in simulation. All other parameters are fixed to the values obtained with simulation.

Combinatorial background: The combinatorial background is described by an exponential function. The slope λ is floating in the fit to recorded data.

Partially reconstructed background: Backgrounds from partially reconstructed $B^0 \to D^{*+}D^{*-}$ and $B_s^0 \to D^{*+}D^{*-}$ decays are considered, where one of the $D^{*\pm}$ decays with a probability of about 30% into $D^{\pm}\pi^0$ and with 1.6% into $D^{\pm}\gamma$. If the π^0 or γ are not reconstructed, the final-state particles can be

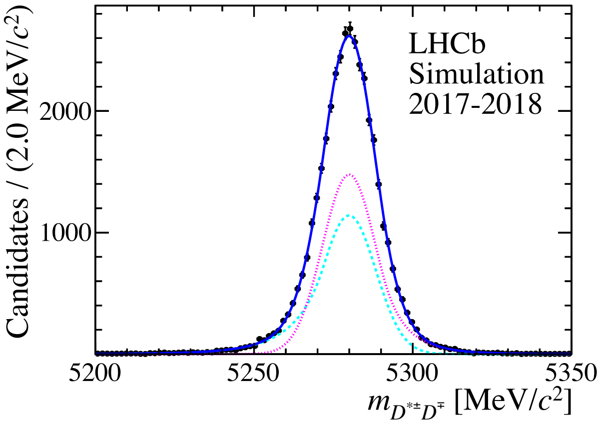

Figure 5.10: The $D^{*\pm}D^{\mp}$ mass distribution of simulated $B^0 \to D^{*\pm}D^{\mp}$ decays for the period 2017–2018 [100]. The full fit projection, shown as the blue solid line, and the two Crystal Ball functions, described by the cyan dashed and the magenta dotted lines, are overlaid.

falsely assigned to the signal $B^0 \to D^{*\pm}D^{\mp}$ or $B_s^0 \to D^{*\pm}D^{\mp}$ decays and cannot be easily suppressed. The shape of the $B^0 \to D^{*+}D^{*-}$ and $B_s^0 \to D^{*+}D^{*-}$ contributions depends on the polarisation of the $D^{*\pm}$ mesons and shows a double peak (longitudinal polarisation) or a single broad peak (transversal polarisation), on top of a broader shape. The double peak structure occurs in longitudinally polarised decays if the $D^{*\pm}$ meson decays into $D^{\pm}$ and π^0. Due to the small mass difference between the $D^{*\pm}$ and the combination of the $D^{\pm}$ and π^0, the π^0 only has a small momentum in its mother's rest frame. Thus, its flight direction is either approximately the same or the opposite to the fully reconstructed $D^{*\mp}$ in the same rest frame. This leads to a parabolic shape in the reconstructed B mass, which turns into the characteristic double-peak structure when accounting for resolution effects [101]. Assuming factorization, 55% of the $B^0 \to D^{*+}D^{*-}$ decays are with both $D^{*\pm}$ mesons longitudinal polarized [102].

To develop a model for the components, simulated $B^0 \to D^{*\pm}D^{*\mp}$ decays and $B_s^0 \to D^{*\pm}D^{*\mp}$ decays for the cases of pure longitudinal and pure transverse polarization are used. Only simulated samples corresponding to the 2012 and 2016 data taking conditions were available and are utilised for Run 1 and Run 2, respectively. The distributions between the samples of 2012 and 2016 are compatible. In case of the transversal polarisation, the single peak is modelled by two Gaussian functions, $\mathcal{G}_i$, with different means but a common width. The broad shape beneath the single peak is described by a function consisting of a uniform distribution between two values fixed by the kinematic of the decay, $\mathcal{U}_{\text{broad}}$, that is convolved with a Gaussian function, $\mathcal{R}$, which has a floating width and accounts for smearings at the kinematic boundaries. The resulting transversal PDF is defined by

$$\mathcal{P}^{\text{transversal}} = g_1 \mathcal{U}_{\text{broad}} \otimes \mathcal{R} + (1-g_1)\left[g_3 \mathcal{G}_3 + (1-g_3)\mathcal{G}_4\right] , \tag{5.15}$$

with the fractions g_i. In case of pure longitudinal polarisation, the two peaks are also modelled by two Gaussians that have different means but share a common width. The broader shape can be parameterised by the sum of two Gaussians with a common mean but different widths. The resulting longitudinal PDF is given by

$$\mathcal{P}^{\text{longitudinal}} = g_1 \mathcal{G}_1 + (1-g_1)\left\{g_2 \mathcal{G}_2 + (1-g_2)\left[g_3 \mathcal{G}_3 + (1-g_3)\mathcal{G}_4\right]\right\} . \tag{5.16}$$

A combination of the PDFs corresponding to pure longitudinal and pure transverse polarisation is used for the $B^0 \to D^{*\pm} D^{*\mp}$ as well as for the $B_s^0 \to D^{*\pm} D^{*\mp}$ component. In the fit to data, the relative fraction between the two helicity contributions are floating while the shape parameters are fixed to the values obtained with simulation. The resulting values of the parameters of the fits to simulation are listed in Table A.6 for 2012 and 2016. Fig. 5.11 shows the invariant $D^{*\pm} D^{\mp}$ mass distribution for the 2016 sample overlaid with the projection of the fitted PDF.

5.4.2 Mass Fit

The invariant $D^{*\pm} D^{\mp}$ mass distribution from recorded $B^0 \to D^{*\pm} D^{\mp}$ data and the fit projection of the unbinned extended maximum likelihood fit are shown in Fig. 5.12 for each data-taking period. The resulting values of the free fit parameters are reported in Table A.7. The free parameters in the fit are the mean μ of the $B^0 \to D^{*\pm} D^{\mp}$ component, the scale factor s for the widths of the signal, the slope λ of the exponential function, the relative fractions g between the two models corresponding to longitudinal and transverse polarised $D^{*\pm}$ mesons in $B^0 \to D^{*+} D^{*-}$ decays as well as in $B_s^0 \to D^{*+} D^{*-}$ decays, and the yields N of all components of the total PDF. All other parameters are fixed to their values obtained with simulation. The resulting yields of $B^0 \to D^{*\pm} D^{\mp}$ and $B_s^0 \to D^{*\pm} D^{\mp}$ decays are 466 ± 22 and 12 ± 4 in 2011–2012, 780 ± 29 and 34 ± 7 in 2015–2016, and 1263 ± 36 and 49 ± 8 in 2017–2018, respectively, where the quoted uncertainties are statistical only. Before including the blinded signal region, the fit model has been validated by pseudo-experiments with various assumptions of the number of B_s^0 candidates.

Based on this fit, sWeights are extracted with the *sPlot* method (see Sec. 4.3) to unfold the B^0 or the B_s^0 signal from the background. To verify that the B^0 and B_s^0 mesons decay into real $D^{*\pm}$ and $D^{\mp}$ mesons, the weighted distributions of the $D^{*\pm}$ mass, the $D^{\mp}$ mass and the mass difference of the $D^{*\pm}$ and the D^0 mesons, using either B_s^0 sWeights or B^0 sWeights, are examined and shown in Fig. 5.13. Clear peaks at the respective known masses [34] are visible.

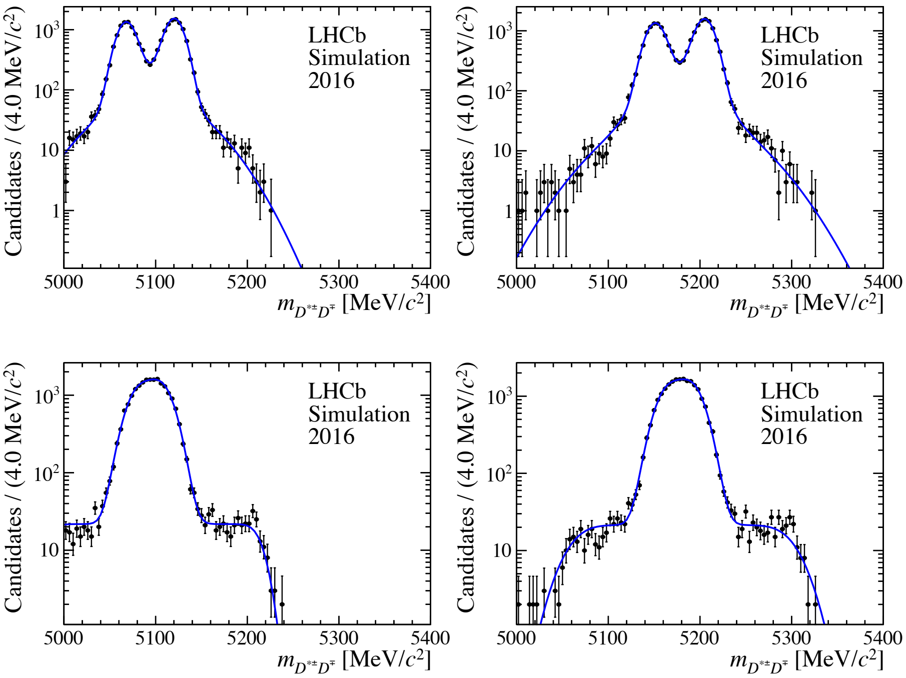

Figure 5.11: The mass distributions of simulated (left) $B^0 \to D^{*+}D^{*-}$ and (right) $B_s^0 \to D^{*+}D^{*-}$ decays reconstructed as signal overlaid with the projections of the fit for 2016 [100]. The upper row refers to pure longitudinal polarisations and the lower row corresponds to pure transverse polarisations of the two $D^{*\pm}$ mesons.

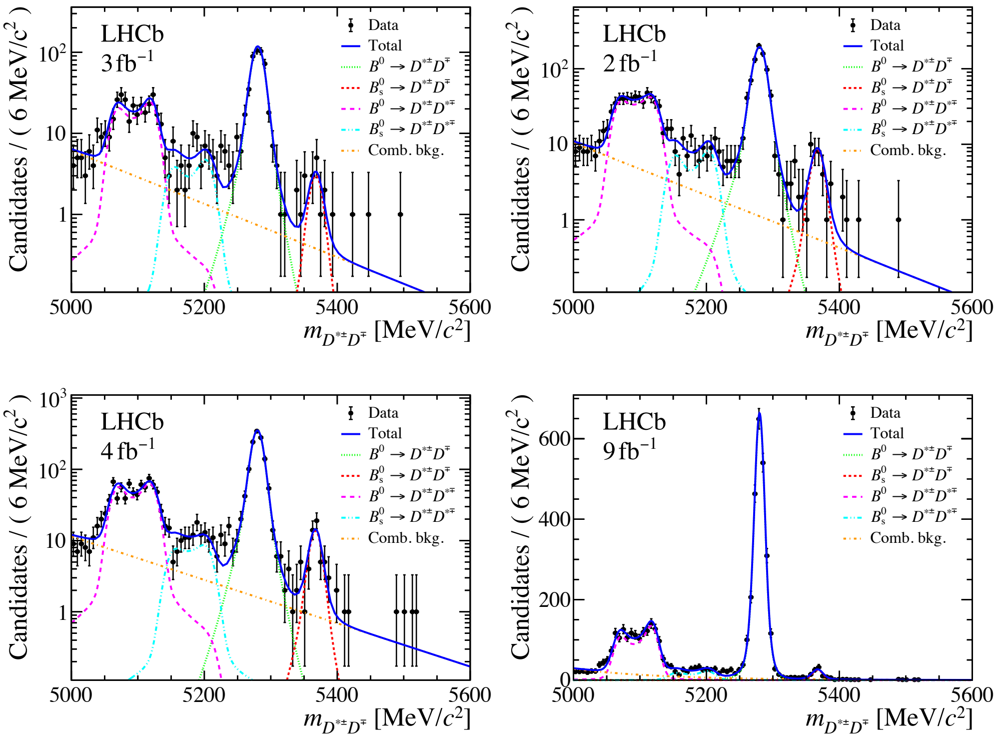

Figure 5.12: The $D^{*\pm}D^{\mp}$ mass distributions of recorded data for (top left) 2011–2012, (top right) 2015–2016, (bottom left) 2017–2018 data in logarithmic scale, and (bottom right) the combined data sample in linear scale [100]. The total fit projection, shown as the blue solid line, is overlaid. The signal contributions for the B^0 and B_s^0 decays are described by the green dotted and the red dashed lines, respectively. The orange dash-dotted line corresponds to the combinatorial background contribution. The magenta long-dashed and the cyan long-dashed-two-dotted lines correspond to the $B^0 \to D^{*\pm}D^{*\mp}$ and $B_s^0 \to D^{*\pm}D^{*\mp}$ background components.

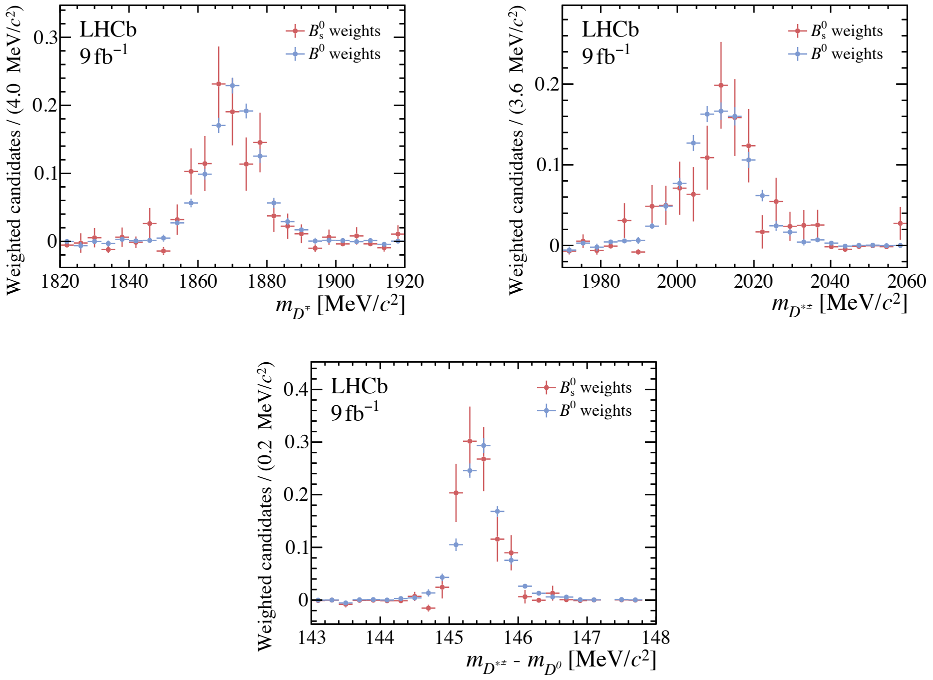

Figure 5.13: The normalised distributions of the weighted mass of the (top left) $D^\mp$ candidates and (top right) $D^{*\pm}$ candidates, and (bottom) the weighted mass difference of the $D^{*\pm}$ and D^0 candidates with the combined data sample using the *sPlot* technique (see Sec. 4.3) for B_s^0 candidates in red and B^0 candidates in blue [100].

5.5 Significance of the Observation

The significance of a measurement represents the probability that an excess of events is not a statistical fluctuation and is commonly calculated to get information about whether the measurement is an evidence or observation. A significance larger than 3σ is commonly called an evidence for the measurement, while a significance greater than 5σ is called an observation. The significance, given in standard deviations σ, is calculated with Wilks' theorem [103]

$$\text{significance} = \sqrt{-2\ln\left(\mathcal{L}_0/\mathcal{L}_{\text{free}}\right)}, \tag{5.17}$$

where $\mathcal{L}_{\text{free}}$ refers to the likelihood from the nominal fit to data with a floating B_s^0 yield and $\mathcal{L}_0$ corresponds to the likelihood of a fit where it is assumed that no B_s^0 component exists. The likelihoods of the three maximum likelihood fits in Sec. 5.4.2 are added together. In the fits, a systematic uncertainty on the fit model is taken into account by scaling the yields with the uncertainties reported in Sec. 5.6. The resulting negative log-likelihoods of the two fits are $-\ln\left(\mathcal{L}_{\text{free}}\right) = -541.818$ and $-\ln\left(\mathcal{L}_0\right) = -446.447$. Utilising Eq. (5.17) results in a significance of the measurement of the $B_s^0 \to D^{*\pm}D^{\mp}$ decay of 13.8σ. Thus, the $B_s^0 \to D^{*\pm}D^{\mp}$ decay is observed with a high significance.

A validation of the likelihood distribution is performed using so-called pseudo-experiments, which are discussed in the following chapter. For such a high significance it is not possible to validate the likelihood distribution in a reasonable time. Instead 10 million pseudo-datasets are generated of which none is expected to be compatible with the observed likelihood difference. This assumption is confirmed by analysing the pseudo-datasets.

5.6 Systematic Uncertainties

The extended maximum likelihood fit to the invariant mass of the reconstructed $B^0 \to D^{*\pm}D^{\mp}$ decays provides statistical uncertainties. But various other effects, which are not accounted for in the fit, can influence the result as well. These systematic uncertainties are discussed in this Chapter.

In the investigation of systematic uncertainties, so-called pseudo-experiments can be useful, where data samples are generated with the same number of candidates as observed in the fit to recorded data. The generated distributions are sampled according to a PDF so that the computing resource is largely reduced in contrast to a full physics simulation. Here, the candidates are generated using an alternative PDF to model a component of the mass distribution. These generated pseudo datasets are then fitted with the nominal PDF. This allows to verify a fit model or to evaluate different fit configurations. In a toy study, this process is repeated multiple times, while saving the result each time. For each step, the residual and the pull are calculated. The residual is calculated as the difference between the generation value and the fitted value. With the help of the residual the actual size of an existing bias is identified by the mean of the distribution of the residual for the parameter of interest. The pull is defined as the difference between the generation value and the fitted value normalised to the fit uncertainty. If no bias is present and the error estimation of the fit is correct, the pull follows a standard normal distribution. To study systematic effects, differences between the PDFs used in the generation and in the fit are introduced and possible biases can be quantified with the pull distribution for the parameter of interest.

5.6.1 Ratio of Hadronisation Fractions

One potential source of a systematic effect is due to the precision of external input parameters. The measurement of the ratio of branching fractions of the $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ decays (see Eq. (5.2)) is dependent on input from the measurement of the ratio of the hadronisation fractions of the B_s^0 and B^0 mesons, f_s/f_d. The values are reported in Eq. (5.3) for 2011–2012 and in Eq. (5.4) for 2015–2016 and 2017–2018. The uncertainties on these input values cause the dominant source of systematic uncertainty. In the combination of the different data-taking periods the values are treated as partially correlated as the two measurements share sources of systematic uncertainties.

5.6.2 Efficiency Ratio

The calculation of $\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})/\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})$ (see Eq. (5.2)) utilises the ratio of the efficiencies of the $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ decays, which are given in Table 5.2. Two sources of systematic uncertainties on this efficiency ratio are considered. The first one takes into account that the efficiencies are calculated with simulation samples that have a finite size caused by the limited computing power. As a result, the statistical precision on the ratio of the efficiencies leads to a systematic uncertainty. The second systematic effect originates from the fact that requirements on ProbNN variables are used in the selection. Though a correction of the simulated ProbNN variables has been performed, the distributions still do not match perfectly between simulation and data. Therefore, an effect on the measurement is estimated by varying the kernel density estimation in the transformation of the simulated ProbNN variables and performing the selection with the resulting corrected variables. The discrepancy of the resulting efficiency ratio to the nominal one is taken as an estimation of the systematic uncertainty.

5.6.3 Signal Model

A systematic uncertainty is considered that combines systematic effects due to the choice of a specific signal model that is used and the assumption that the B^0 and B_s^0 components have the identical shape in the mass fit. This is done using pseudo-experiments with 100,000 pseudo datasets. An alternative PDF that describes the signal component reasonably is the Hypatia function [104]. Therefore, candidates are generated with a mass distribution described by a Hypatia function in the toy study. The poisson fluctuation of signal and background yields is included in the generation. Different values of the parameters for the B^0 and B_s^0 components are used and determined by a fit to simulated $B^0 \to D^{*\pm}D^{\mp}$ and $B_s^0 \to D^{*\pm}D^{\mp}$ decays, respectively. The results of the fits to simulation are shown in Fig. 5.14 for the period 2017–2018. The nominal model is then fitted to the generated distributions. For each pseudo dataset, the result of the ratio of the branching fractions is calculated. Afterwards, the mean of all pseudo datasets and its residual are calculated for the three data-taking periods separately. The mean of the residual and its uncertainty are summed up in quadrature and the square root is calculated. The resulting value is assigned as the systematic uncertainty.

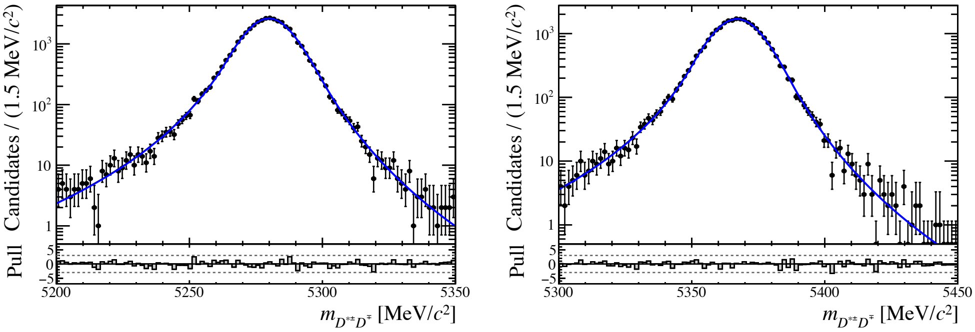

Figure 5.14: The $D^{*\pm}D^{\mp}$ mass distribution of simulated (left) $B^0 \to D^{*\pm}D^{\mp}$ and (right) $B_s^0 \to D^{*\pm}D^{\mp}$ decays for the period 2017–2018. The full fit projection, shown as the blue solid line, is overlaid. The used fit model consist of a Hypatia function.

5.6.4 Background Model

A systematic uncertainty due to the specific model of the combinatorial background is evaluated using pseudo-experiments with 100,000 pseudo datasets. In the generation of the candidates the parameters of the signal and partially reconstructed background models are set to the values found in the nominal fit. A different slope of the exponential function is used that describes the distribution of the combinatorial. It is extracted by a fit to recorded data with an enhanced contribution of combinatorial background which is achieved by applying a looser BDT requirement with a value of 0.25. The result of the fit for the period 2017–2018 is shown in Fig. 5.15. The generated candidates are then fitted with the nominal model. The systematic uncertainty is evaluated in the same way as the systematic uncertainty for the signal model.

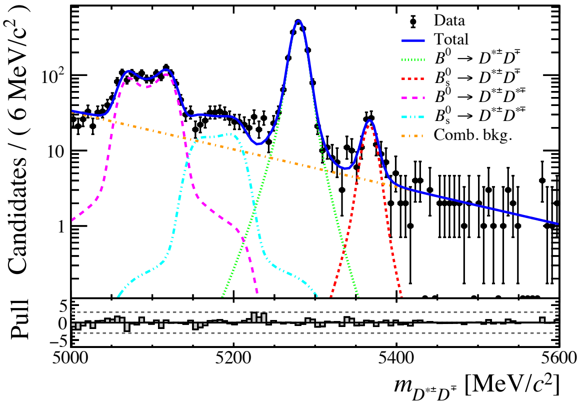

Figure 5.15: The $D^{*\pm}D^{\mp}$ mass distributions of recorded data for the period 2017–2018. A looser BDT requirement with a value of 0.25 has been applied to the data sample to enhance the contribution of combinatorial background. The resulting slope of the (orange dash-dotted) exponential function is utilised in the evaluation of the systematic uncertainty for the background model. The total fit projection is overlaid as the blue solid line. The signal contributions for the B^0 and B_s^0 decays are described by the green dotted and the red dashed lines, respectively. The magenta long-dashed and the cyan long-dashed-two-dotted lines correspond to the $B^0 \to D^{*\pm}D^{*\mp}$ and $B_s^0 \to D^{*\pm}D^{*\mp}$ background components.

5.6.5 Total Systematic Uncertainty

In Table 5.3 all relative systematic uncertainties on the ratio of branching fractions $\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})/\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})$ are listed. The individual uncertainties are summed up in quadrature to obtain the total systematic uncertainty per data-taking period. To determine the combined result of the three data-taking periods, a weighted average is calculated, where possible correlations of sources of systematic uncertainties between the different periods are taken into account. While the systematic uncertainty due to the size of the simulation sample is assumed to be uncorrelated, the systematic uncertainty due to the PID variables and the signal and background model are treated as fully correlated between the data-taking periods. The systematic uncertainty due to the f_s/f_d input is treated as partially correlated between Run 1 and Run 2 as the utilised measurements of f_s/f_d share sources of systematic uncertainties.

Table 5.3: Systematic uncertainties on $\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})/\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})$ given relative to the measured value.

Source	Systematic uncertainties [%]			
	2011–2012	2015–2016	2017–2018	Combined
f_s/f_d [93, 94]	5.8	4.9	4.9	4.6
Simulated data size	0.8	1.2	0.8	0.6
PID	0.7	0.7	0.8	0.7
Signal model	0.1	0.1	0.5	0.3
Background model	1.7	1.3	0.8	1.1
Total without f_s/f_d	2.0	1.9	1.5	1.5
Total	6.1	5.3	5.1	4.8

5.7 Calculation of the Branching Ratio

The ratio of the branching fractions $\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})/\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})$ is calculated with Eq. (5.2). The ratio of efficiencies of the $B^0 \to D^{*\pm}D^{\mp}$ and $B_s^0 \to D^{*\pm}D^{\mp}$ decays are given in Table 5.2, the number of B_s^0 and B^0 candidates are reported in Table A.7 and the external input value for the ratio of the hadronisation fractions can be found in Eqs. (5.3) and (5.4).

The individual ratios of branching fractions are calculated to

$$\frac{\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})}{\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})}_{2011\text{–}2012} = 0.093 \pm 0.032 \pm 0.002 \pm 0.005\,,$$
$$\frac{\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})}{\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})}_{2015\text{–}2016} = 0.168 \pm 0.034 \pm 0.003 \pm 0.008\,,$$
$$\frac{\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})}{\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})}_{2017\text{–}2018} = 0.149 \pm 0.024 \pm 0.002 \pm 0.007\,,$$

where the first uncertainty is statistical, the second systematic and the third is due to the uncertainty of the hadronisation fraction ratio f_s/f_d. The weighted average of the three data-taking periods is calculated using the quadratic sums of the uncertainties as weights and including the correlation of the systematic uncertainties. The resulting value is

$$\frac{\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})}{\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})} = 0.137 \pm 0.017 \pm 0.002 \pm 0.006\,.$$

Using the measured value of the $B^0 \to D^{*\pm}D^{\mp}$ branching fraction from Ref. [57], the $B_s^0 \to D^{*\pm}D^{\mp}$ branching fraction is found to be

$$\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp}) = (8.41 \pm 1.02 \pm 0.12 \pm 0.39 \pm 0.79) \times 10^{-5}\,,$$

where the fourth uncertainty is due to the $B^0 \to D^{*\pm}D^{\mp}$ branching fraction.

5.8 Results

The $B_s^0 \to D^{*\pm}D^{\mp}$ decay is observed for the first time with a high significance using data corresponding to $9\,\mathrm{fb}^{-1}$ collected by the LHCb experiment. The branching fraction is measured relative to the $B^0 \to D^{*\pm}D^{\mp}$ decay. The resulting ratio of branching fractions, for all data-taking periods combined, is

$$\frac{\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp})}{\mathcal{B}(B^0 \to D^{*\pm}D^{\mp})} = 0.137 \pm 0.017 \pm 0.002 \pm 0.006\,,$$

where the first uncertainty is statistical, the second systematic and the third is due to the uncertainty on the ratio of the hadronisation fractions of B_s^0 and B^0. The branching fraction of $B_s^0 \to D^{*\pm}D^{\mp}$ is found to be

$$\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp}) = (8.41 \pm 1.02 \pm 0.12 \pm 0.39 \pm 0.79) \times 10^{-5}\,,$$

where the fourth uncertainty is due to the $B^0 \to D^{*\pm}D^{\mp}$ branching fraction. While the result disagrees with a theory prediction using a perturbative QCD approach [24], it is in agreement with a prediction that assumes prominent contributions from rescattering from *e.g.* $D_s^{*\pm}D_s^{\mp}$ states [23]. This result can be used to assess contributions from W-exchange, penguin-annihilation and rescattering diagrams in other $B \to D\overline{D}$ decays, with the main goal to constrain these contributions in the measurement of ϕ_s in $B_s^0 \to D_s^+D_s^-$ decays [21, 22].

6 Measurement of *CP* Violation in $B_s^0 \to D_s^+ D_s^-$ Decays

This chapter presents the measurement of *CP* violation in the $B_s^0 \to D_s^+ D_s^-$ decay to measure the weak mixing phase ϕ_s. In parallel to this measurement, an analysis of the $B^0 \to D^+ D^-$ decay is carried out exploiting the similarities of the two decays. The result of the measurement of *CP* violation using $B^0 \to D^+ D^-$ decays is not presented here but in Ref. [105]. The $B^0 \to D_s^+ D^-$ and $B_s^0 \to D_s^- \pi^+$ decays are used as control channels for the flavour-tagging calibration and parametrisation of the decay-time resolution and decay-time-dependent efficiency. The measurement is performed in close collaboration with Louis Gerken and Philipp Ibis within the LHCb physics group from Dortmund. The author contributed to almost all parts of the analysis. Parts where the involvement of the author was less pronounced are the development of the BDT based selection steps and the final fit to the decay-time distribution to extract the *CP* observables. Smaller emphasis is put on the description of these parts. Significant contributions by the author comprise the development of the cut-based selection, the scan of the BDT requirement, the mass fits, the flavour-tagging calibration, and the parametrisations of the decay-time resolution and decay-time-dependent efficiency. The flavour-tagging calibration of the SS tagger also makes use of a bachelor thesis [106] supervised by the author. All parts are described to provide a comprehensive presentation of the complete measurement.

The analysis is in an advanced state but not complete yet. Hence, the central values of the measured *CP* parameters are kept blind to avoid an unconscious bias during the measurement. The sensitivity of the measurement is still known.

In Sec. 6.1, the selection of signal candidates is outlined and the fit to the invariant mass distribution is described, which is used to extract signal sWeights. The flavour-tagging calibration is discussed in Sec. 6.2. Afterwards, studies for the decay-time parametrisation of the B_s^0 meson are presented in Sec. 6.3 and the fit to measure the *CP* parameters is described in Sec. 6.4. The expected systematic uncertainties and the expected results are summarised in Sec. 6.5 and Sec. 6.6, respectively.

6.1 Data Preparation and Signal Extraction

The $B_s^0 \to D_s^+ D_s^-$, $B^0 \to D_s^+ D^-$ and $B_s^0 \to D_s^- \pi^+$ decays are reconstructed in a fully hadronic final state, which leads to various background contributions. To enable a precise measurement of the *CP* parameters, the background contributions have to be suppressed. For this, a selection consisting of multiple steps is developed, which is divided into the preselection (Sec. 6.1.2), vetoes to suppress physical backgrounds (Sec. 6.1.3), a multivariate selection to reduce combinatorial backgrounds (Sec. 6.1.4) and the selection of multiple candidates in single events (Sec. 6.1.5). To disentangle the signal contribution from the remaining background in the data sample, the *sPlot* method (see Sec. 4.3) is utilized by performing an extended maximum likelihood fit to the invariant $D_s^+ D_s^-$ mass (Sec. 6.1.6).

6.1.1 Data samples

The measurement is performed using the Run 2 dataset collected by the LHCb experiment from 2015 to 2018 at a centre-of-mass energy of 13 TeV. The whole dataset corresponds to an integrated luminosity of $6\,\mathrm{fb}^{-1}$. The data samples of each year of data taking are combined. The $B_s^0 \to D_s^+ D_s^-$ candidates are reconstructed in decays of the D_s^+ meson into three charged hadrons with the largest branching fractions, namely the final states $D_s^+ \to K^- K^+ \pi^+$, $D_s^+ \to \pi^- \pi^+ \pi^+$ and $D_s^+ \to \pi^- K^+ \pi^+$. If not stated otherwise, *CP*-conjugate decays are implied throughout this chapter. The three most probable D_s^+ final-state combinations are considered in this analysis and are reported in Table 6.1 along with their branching fractions. Besides recorded $B_s^0 \to D_s^+ D_s^-$ decays, additional samples are used at various parts of the analysis. $B^0 \to D_s^+ D^-$ decays are utilised as a control channel for the flavour-tagging calibration as well as to model detector effects like the decay-time acceptance. The same D_s^+ final states are considered as in the $B_s^0 \to D_s^+ D_s^-$ channel. The D^- meson is reconstructed in the hadronic final states $K^+ K^- \pi^-$ and $K^+ \pi^- \pi^-$. The flavour-tagging calibration also makes

Table 6.1: Considered final state combinations for the $B_s^0 \to D_s^+ D_s^-$, $B^0 \to D_s^+ D^-$ and $B_s^0 \to D_s^- \pi^+$ decay channels and the product of the branching fractions $\mathcal{B}$ of the D mesons taken from [34].

Decay channel	Final state		$\mathcal{B}_{D^+} \cdot \mathcal{B}_{D^-}$
$B_s^0 \to D_s^+ D_s^-$	$K^- K^+ \pi^+$	$K^+ K^- \pi^-$	$29 \cdot 10^{-4}$
	$K^- K^+ \pi^+$	$\pi^+ \pi^- \pi^-$	$12 \cdot 10^{-4}$
	$K^- K^+ \pi^+$	$\pi^+ K^- \pi^-$	$7.0 \cdot 10^{-4}$
$B^0 \to D_s^+ D^-$	$K^- K^+ \pi^+$	$K^+ \pi^- \pi^-$	$51 \cdot 10^{-4}$
	$\pi^- \pi^+ \pi^+$	$K^+ \pi^- \pi^-$	$10 \cdot 10^{-4}$
	$\pi^- K^+ \pi^+$	$K^+ \pi^- \pi^-$	$6.1 \cdot 10^{-4}$
	$K^- K^+ \pi^+$	$K^+ K^- \pi^-$	$5.2 \cdot 10^{-4}$
$B_s^0 \to D_s^- \pi^+$	$K^+ K^- \pi^-$		$5.4 \cdot 10^{-2}$
	$\pi^+ \pi^- \pi^-$		$1.1 \cdot 10^{-2}$
	$\pi^+ K^- \pi^-$		$6.5 \cdot 10^{-3}$

use of $B_s^0 \to D_s^- \pi^+$ decays including the same D_s^- final states as listed before. In addition, prompt $B_s^0 \to D_s^- \pi^+$ samples are utilised for the determination of the decay-time resolution, *i.e.* D_s^- candidates originating in the PV are combined with the so-called bachelor pions originating in the PV imitating a $B_s^0 \to D_s^- \pi^+$ decay. For this, only the the $D_s^- \to K^+ K^- \pi^-$ decay is reconstructed as other D_s^- meson final states are not available for the prompt data sample. The products of the branching ratios of the control channels are reported in Table 6.1. Lastly, simulated samples of all decays are used to parametrise models and train multivariate classifiers. The same selection steps are applied to all data samples if not stated otherwise.

Before applying the selection to the data samples, the final state and intermediate particles of the decays are sorted to simplify some selection steps. The following denotation is used for the $D_s^\pm \to h_0^\mp h_1^\pm h_2^\pm$ decay where it is necessary:

- $h_0^\mp$ is the kaon or pion with the opposite charge of the $D_s^\pm$ meson.
- $h_1^\pm$ is the hadron with the higher p_T if the two final state hadrons with the same charge are the same particles. Otherwise it is always the kaon.
- $h_2^\pm$ is the hadron with the lower p_T if the two final state hadrons with the same charge are the same particles. Otherwise it is always the pion.

The $B_s^0 \to D_1 D_2$ decay uses the following denotation:

- D_1 is the meson with the higher p_T if the D_s^+ and D_s^- mesons decay into the same final state. Otherwise it gets assigned to the $D_s^\pm$ meson that has more kaons in the final state.
- D_2 is the meson with the lower p_T if the D_s^+ and D_s^- mesons decay into the same final state. Otherwise it gets assigned to the $D_s^\pm$ meson that has more pions in the final state.

The same sorting schema can be applied to the control channels. Additionally, for the $B^0 \to D_s^+ D^-$ channel it applies that D_1 is always the D_s^+ meson and D_2 the D^- meson.

6.1.2 Preselection

In this analysis, the output of the `StrippingB02DDBeauty2CharmLine` Stripping line is used. The requirements are summarised in Table B.1. The Stripping line uses the same requirements as the one discussed in the $B_s^0 \to D^{*\pm} D^\mp$ analysis. A detailed description of the requirements can be found in Sec. 5.2.1. An additional requirement is applied on the difference of the z coordinate of the $D^\pm$ and B^0 mesons, $z(D^\pm) - z(B^0)$, to ensure a $D^\pm$ decay downstream of the B^0 decay. The efficiency of this centralised preselection is about 1.7%. For the $B_s^0 \to D_s^- \pi^+$ decay channel, the output of the Stripping line `B02DpiD2HHHBeauty2CharmLine` is used, that incorporates very similar requirements. Here, the efficiency is about 4.4%.

The requirements of the additional preselection are reported in Table 6.2. The same requirements are applied to the control channels taking into account that one of the charged D mesons is a $D^\pm$ meson in the $B^0 \to D_s^+ D^-$ decay. The invariant mass of the $D_s^\pm$ meson has to be in the range of ±45 MeV/c^2 around its nominal mass. This is a loose but efficient requirement to remove background without D_s^+ meson, *e.g.* combinatorial background or background with misidentified D_s^+ mesons, with a signal efficiency of about 98% for each final state combination. Further, the χ^2 of the $D_s^\pm$ flight distance with respect to the SV must be greater than 5. This requirement ensures that the $D_s^\pm$ decay vertex is displaced from its origin vertex. When applying this requirement on both $D_s^\pm$ mesons, about 62% of the $B_s^0 \to D_s^+ D_s^-$ candidates are kept for each final state combination. The signal efficiency is rather small, but studies of physical background contributions in Sec. 6.1.3 show that this global requirement is useful to successfully remove single-charm background with non-resonant $D_s^\pm$ mesons. In total, the offline preselection suppresses about 90% of the combinatorial background, calculated with the upper-mass sideband with $m_{B_s^0} > 5600\,\text{MeV}/c^2$. The preselection of the $B_s^0 \to D_s^- \pi^+$ sample uses the same requirements and adds requirements on the bachelor pion: the ProbNN$_\pi$ variable must be larger than 0.3 and candidates, which have hits in the muon chambers, are discarded.

Table 6.2: Requirements of the additional preselection. Candidates that do not satisfy the requirements are rejected. The PDG values are taken from Ref. [34].

Requirement	Unit	Definition
$\lvert m_{h^\pm h^\mp h^\mp} - m_{D_s^\pm,\mathrm{PDG}} \rvert < 45$	MeV/c^2	Invariant $D_s^\pm$ mass
$\chi^2_{D_s^\pm,\mathrm{FD}} > 5$		χ^2 of $D_s^\pm$ flight distance

6.1.3 Vetoes

Physical backgrounds from hadronic decays of $D^\pm$ mesons, Λ_c^+ hadrons and two-body decays of D^0 and ϕ mesons are considered, where one or multiple final state hadrons are misidentified leading to a false reconstruction of the decay as a signal $D_s^\pm$ decay. The mass hypotheses of a pion, kaon or proton are assigned to one or multiple final state hadrons of the $D_s^\pm$ decay and the mass of the hhh system is recalculated for all possible combinations. Additionally, single-charm backgrounds are investigated after all other physical backgrounds have been suppressed, *i.e.* decays where only one resonant charged D meson and three charged hadrons that originate from the SV are present. The invariant mass of the $D_{1,2}hhh$ system is recalculated taking all possible misidentifications of the hadrons into account. Due to the cut on the flight distance of the $D_s^\pm$ mesons in the preselection, all contributions from single-charm backgrounds are suppressed.

Cut-based Rejection of misidentified decays

In general, physical background appears as a resonant structure in the recalculated mass distribution. To suppress the background contributions, vetoes are applied in which requirements are made on the recalculated mass and the ratio

$$\mathrm{ProbNN}_{p_1,p_2} = \frac{\mathrm{ProbNN}_{p_1}}{\mathrm{ProbNN}_{p_1} + \mathrm{ProbNN}_{p_2}}\,, \tag{6.1}$$

in which ProbNN_{p_1} and ProbNN_{p_2} are the ProbNN variables of the original particle and the new particle hypothesis, respectively. With the use of the ProbNN ratio, it is possible to distinguish between two particle hypotheses. Though, in some cases, a requirement on only the recalculated mass distribution is needed to suppress the background with a high signal efficiency. For the simulated samples, the ProbNN variables are corrected to have a better agreement between simulation and recorded data, as explained in Sec. 5.2.2. In Table B.2 all cut-based vetoes for the $B_s^0 \to D_s^+ D_s^-$ decay channel are summarised. In the determination of the requirements, it was distinguished between D_1 and D_2, the D flavour as well as the different final states. The requirements in Table B.2 are also applied on the control channels. In addition, vetoes for the $D^- \to K^+\pi^-\pi^-$ final state of the $B^0 \to D_s^+ D^-$ decay are applied, which are reported in Table B.3. No simulated background decays are available to calculate the background rejection, therefore the requirements are determined on recorded data.

In the following paragraphs, the vetoes for each final state of the $D_s^\pm$ decays, that are considered in the analysis, are discussed following the order in Table B.2.

$D_{(s)}^+ \to K^- K^+ \pi^+$: Physical background due to a pion-kaon misidentification or vice versa arise from $D_s^+ \to K^-\pi^+K^+$ decays with a double misidentification of the last two final state hadrons and from $D^+ \to K^-\pi^+\pi^+$ decays with single misidentification. The latter one is not removed with a cut-based selection but with a multivariate selection explained below due to low signal efficiencies. Further, a background contribution from $\phi \to K^-K^+$ decays is visible in the mass of the K^-K^+ system, where the K^+ meson corresponds to the h_2 hadrons and gets misidentified as a pion. This background can emerge from $D_s^+ \to \phi\pi^+$ decays, where also the π^+ meson is misidentified as a kaon. Background contributions from decays of neutral D mesons are also considered, where instead of the charged $D^\pm$ mesons, a D^0 meson and a charged hadron are present. A contribution from $D^0 \to K^-K^+$ decays without any misidentification is visible in the invariant mass of the two oppositely charged kaons. In addition, contributions from Λ_c^+ decays stemming from $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays are considered. A proton-kaon misidentification leads to the background contribution of $\Lambda_c^+ \to K^- p \pi^+$

decays. Contributions of $\Lambda_c^+ \to K^-\pi^+p$ decays occur after a double misidentification of the last two final-state particles. No single-charm background is identified in this final state. In Fig. 6.1, the distributions of the recalculated masses are shown for the $D_1 \to K^-K^+\pi^+$ final state before and after applying the respective veto.

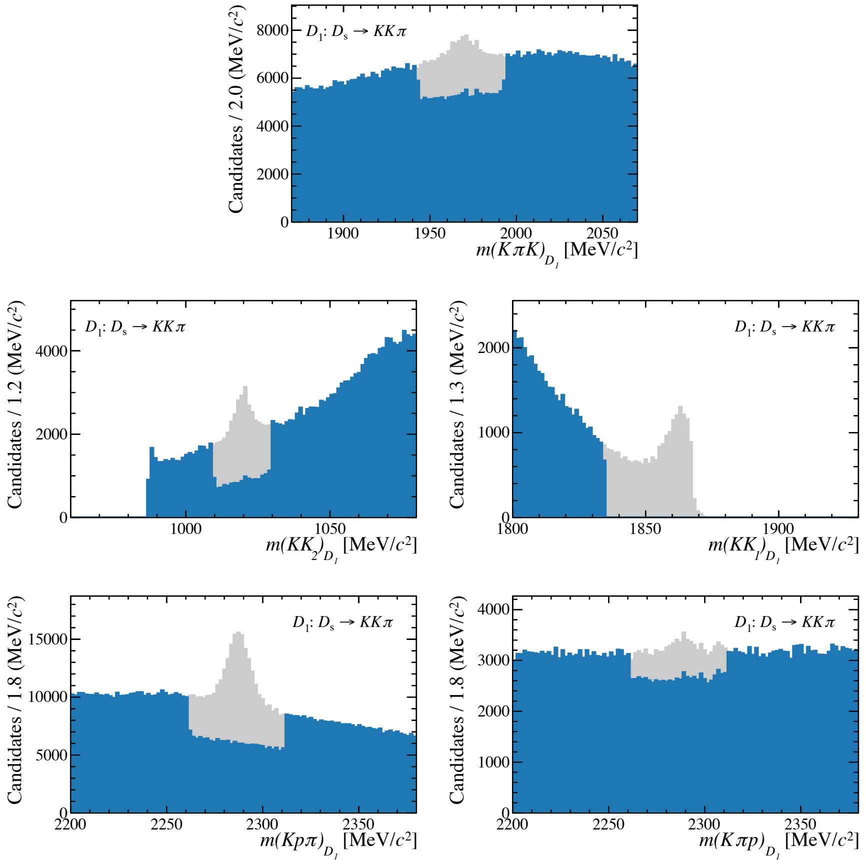

Figure 6.1: Invariant mass distributions of the $D_1 \to KK\pi$ final state recalculated as (top) $K\pi K$, (middle left) KK_2, (middle right) KK_1, (bottom left) $Kp\pi$ and (bottom right) $K\pi p$. Background contributions from D_s^+ decays, ϕ decays, D^0 decays, and in the last two plots from Λ_c^+ decays are visible. Candidates in the blue area are kept while candidates in the grey area get rejected by the veto requirement. The requirements of the vetoes are reported in Table B.2.

$D_s^+ \to \pi^-K^+\pi^+$: A pion-kaon misidentification of one or multiple final-state hadrons leads to contributions from $D^+ \to \pi^-\pi^+\pi^+$, $D^+ \to K^-\pi^+\pi^+$, $D_s^+ \to \pi^-\pi^+K^+$ and $D_s^+ \to K^-\pi^+K^+$ decays. Background contributions emerge from $\phi \to K^-K^+$ decays, where the K^- meson is misidentified as a π^- meson and the K^+ meson corresponds to either the h_1 or the h_2 hadron with an additional misidentification in the latter case. Several background contributions emerge from decays of neutral D mesons. Among them are $D^0 \to \pi^- K^+$ decays with no misidentification, $D^0 \to K^- K^+$ decays with a kaon-pion misidentification of the K^- meson and $D^0 \to K^- \pi^+$ decays with misidentifications of both hadrons. Further, a proton-kaon misidentification results into a background contribution of $\Lambda_c^+ \to \pi^- p \pi^+$ decays. The recalculated-mass distributions of the $D_2 \to \pi^- K^+ \pi^+$ final state are shown in Fig. 6.2 before and after applying the vetoes.

$D_s^+ \to \pi^-\pi^+\pi^+$: Decays of neutral D mesons result into several background contributions in this final state as well. First, $D^0 \to K^- \pi^+$ decays with the π^+ meson being either the h_1 or h_2 hadron and the K^- being misidentified as a pion. Second, $D^0 \to \pi^- \pi^+$ decays with the π^+ meson being either the h_1 or h_2 hadron, and no misidentification of any hadron. Lastly, $D^0 \to \pi^- K^+$ decays with a kaon-pion misidentification of the h_1 hadron. Furthermore, a background contribution from $\Lambda_c^+ \to \pi^- p \pi^+$ decays can occur after a proton-pion misidentification. The recalculated-mass distributions for the $D_2 \to \pi^- \pi^+ \pi^+$ final state are shown in Fig. 6.3 before and after applying the vetoes.

$D^+ \to K^-\pi^+\pi^+$: Additional vetoes are applied for the $D^+ \to K^- \pi^+ \pi^+$ final state in $B^0 \to D_s^+ D^-$ decays. The vetoes are summarised in Table B.3 and the effects of the vetoes are shown in Fig. B.1.

Efficiencies: For the $B_s^0 \to D_s^+ D_s^-$ decay channel, applying all cut-based vetoes results into signal efficiencies of about 95% for the $K^- K^+ \pi^+ \; K^+ K^- \pi^-$ final state, about 97% for the $K^- K^+ \pi^+ \;\; \pi^+ \pi^- \pi^-$ final state and about 86% for the $K^- K^+ \pi^+ \; \pi^+ K^- \pi^-$ final state.

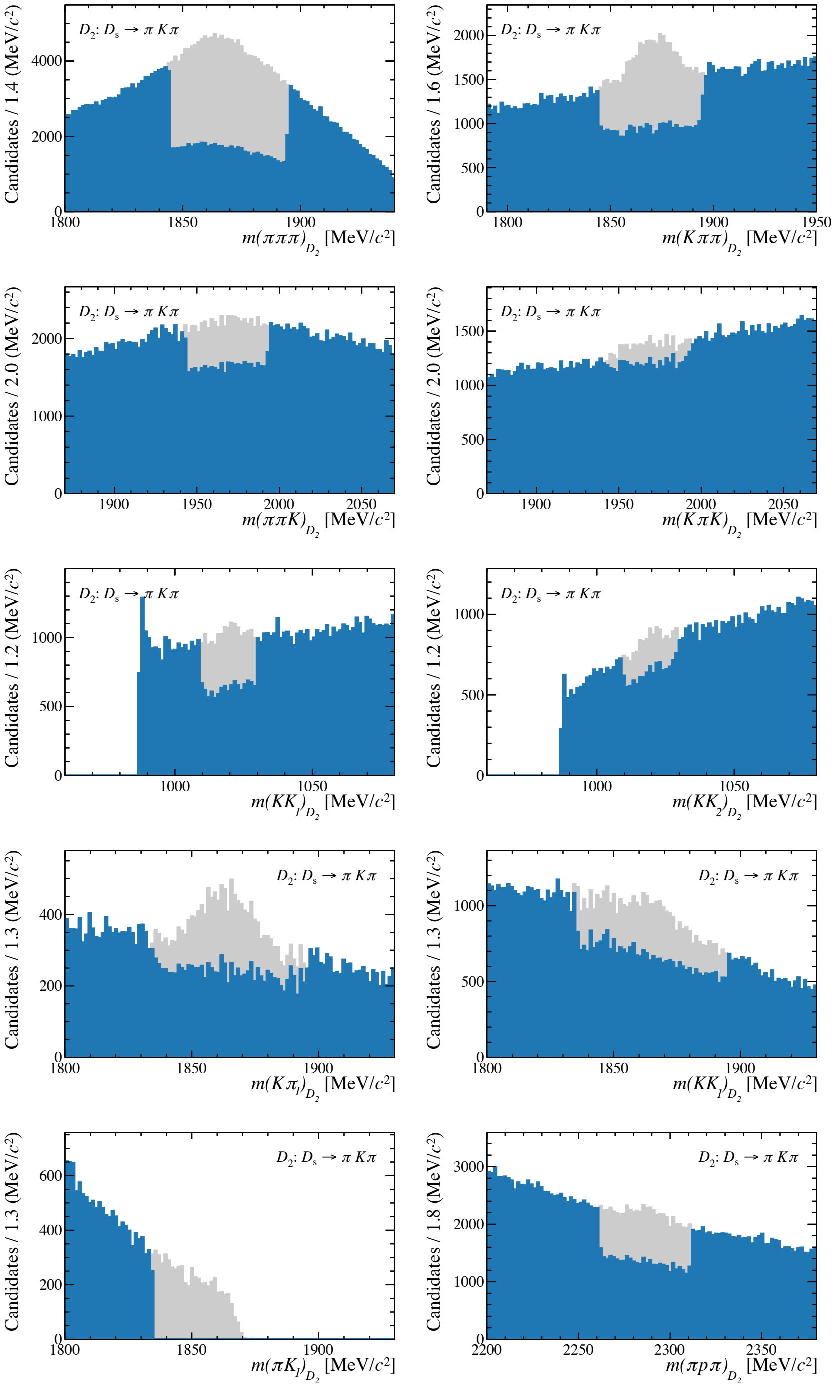

Figure 6.2: Recalculated invariant mass distributions for the $D_2 \to \pi K \pi$ final state. Background contributions from (top) D^+ decays, (second row) D_s^+ decays, (third row) ϕ decays, (fourth row and bottom left) D^0 decays and (bottom right) Λ_c^+ decays are visible. Candidates in the blue area are kept while candidates in the grey area get rejected by the veto requirement. The vetoes are reported in Table B.2.

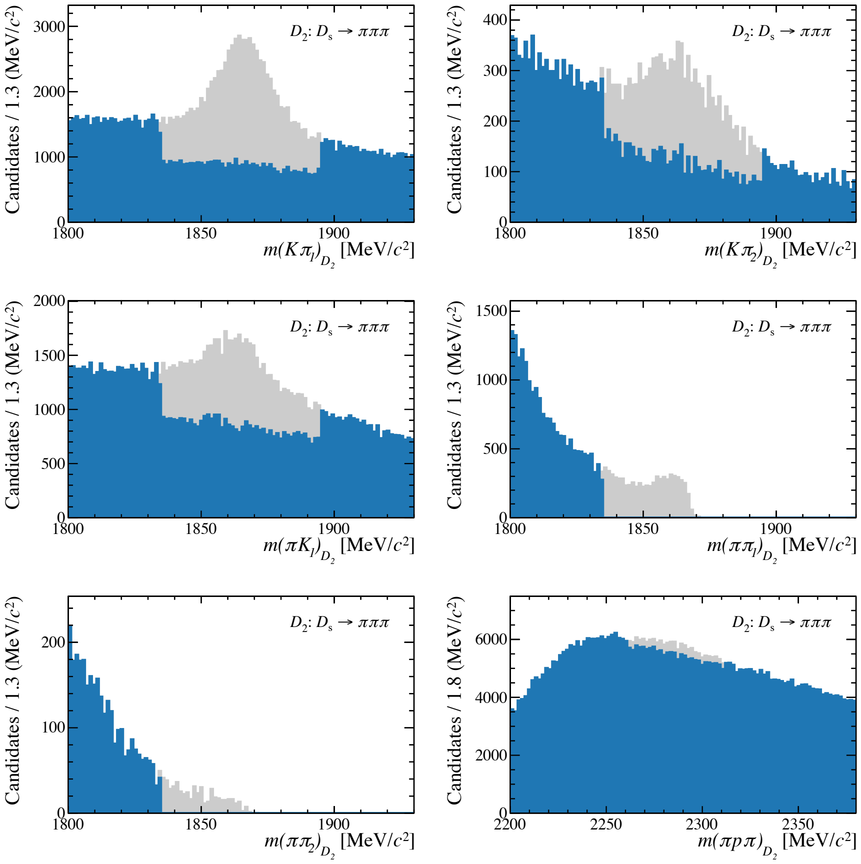

Figure 6.3: Recalculated invariant mass distributions for the $D_2 \to \pi\pi\pi$ final state. Background contributions of (top, middle and bottom left) D^0 decays and (bottom right) Λ_c^+ decays are visible. Candidates in the blue area are kept while candidates in the grey area get rejected by the veto requirement. The vetoes are reported in Table B.2.

Multivariate Rejection of misidentified decays

In the $D_s^+ \to K^- K^+ \pi^+$ final state, background of $D^- \to K^+ \pi^- \pi^-$ decays stemming from $B^0 \to D_s^+ D^-$ decays contributes after a pion-kaon misidentification. Inversely, in the final state $D^- \to K^+ \pi^- \pi^-$, background of $D_s^- \to K^+ K^- \pi^-$ decays stemming from $B_s^0 \to D_s^+ D_s^-$ decays can contribute after a kaon-pion misidentification. A veto using cut-based requirements on the mass distribution and PID information, similar to the previously discussed vetoes, is too inefficient. Therefore, a multivariate selection utilising a BDT is chosen to suppress these backgrounds.

A BDT is trained to separate $D_s^+ \to K^- K^+ \pi^+$ and $D^- \to K^+ \pi^- \pi^-$ decays. Simulated samples of $B_s^0 \to D_s^+ D_s^-$, $B^0 \to D_s^+ D^-$ and $B^0 \to D^+ D^-$ decays are merged to one sample and used in the training. For the BDT training, the `xgboost` package [88] is used, which utilises the gradient boosting algorithm. A k-folding with $k = 5$ and early stopping with $n = 10$ are used to avoid overtraining. As the validation metric the area under the ROC curve is used. The input features comprise several variables of the misresconstructed D meson: the flight distance to take into account that the D_s^+ meson has a smaller lifetime than the D^+ meson, ProbNN_K and ProbNN_π variables that consider the PID information of the two final-state particles that can be misidentified and various recalculated invariant two and three body masses of its final-state particles. The two body masses of the final-state particles are used to check for resonances, *e.g.* a $\overline{K}^*$ resonance decaying into $K^- \pi^+$ or a ϕ resonance decaying into $K^- K^+$. The ϕ resonance can only occur in the $D_s^+ \to K^- K^+ \pi^+$ decay and the amounts of the $\overline{K}^*$ resonances differ between both D decays [34]. All of the variables are listed in Table B.4. The agreement between the distributions of the BDT features in simulation and background-subtracted data is checked before using the features in the training.

In all folds of the BDT, the mass of the $K\pi\pi$ system is the most important feature, followed by the ProbNN_K of the two hadrons that can be misidentified and then the remaining three-body masses. The other features are less important with the flight distance variable and the two-body masses being the features with the smallest importance.

In Fig. 6.4, the BDT-classifier distribution for one fold is shown with the training and test sample superimposed. No overtraining is visible. Additionally, the ROC curve is shown, whose AUC value indicates a good BDT performance. The requirements are chosen, such that the background contributions are sufficiently suppressed. This can be seen in Fig. 6.5, which shows the invariant $D_s^+ D_s^-$ mass distribution before and after the application of the requirement on the veto BDT. In both cases, the previously discussed cut-based vetoes have been applied.

The signal efficiencies are about 87% for the $K^- K^+ \pi^+$ $K^+ K^- \pi^-$ final state, and 93% for the $K^- K^+ \pi^+$ $\pi^+ \pi^- \pi^-$ as well as $K^- K^+ \pi^+$ $\pi^+ K^- \pi^-$ final states of the $B_s^0 \to D_s^+ D_s^-$ decay. More than 99% of the background is suppressed.

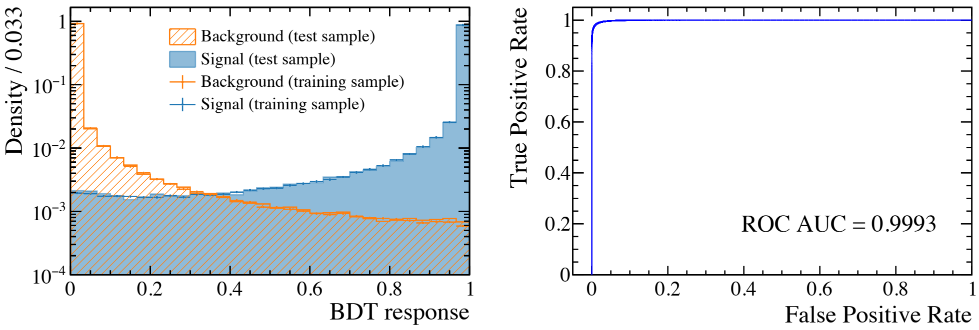

Figure 6.4: On the left, the distribution of the BDT response for (blue) $D_s^+ \to K^-K^+\pi^+$ and (orange) $D^+ \to K^-\pi^+\pi^+$ candidates from the (data points) training and (shaded areas) test sample of one fold is shown. On the right the corresponding ROC curve is shown. The value of the ROC AUC is close to one and indicates a good separation.

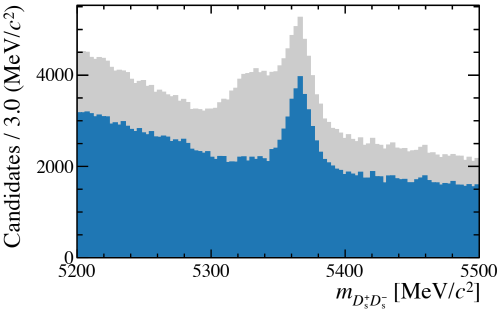

Figure 6.5: Invariant $D_s^+ D_s^-$ mass distribution (gray) before and (blue) after the application of the veto BDT for $D^+ \to K^-\pi^+\pi^+$ background decays in the $D_s^+ \to K^-K^+\pi^+$ final state. In both distributions, all other cuts against misidentified background contributions have been applied. In the grey distribution, a clear peaking structure on the left side of the B_s^0 peak is visible and is sufficiently suppressed after applying the veto BDT.

6.1.4 Multivariate Selection

After removing physical background, combinatorial background is suppressed with the use of a BDT.

BDT Training

The `xgboost` package [88] is used, which utilises the gradient boosting algorithm. In the training, the background sample is given by the upper-mass sideband of data with $m_{D_s^+ D_s^-} > 5600\,\mathrm{MeV}/c^2$ and simulated $B_s^0 \to D_s^+ D_s^-$, $B^0 \to D_s^+ D^-$ and $B^0 \to D^+ D^-$ decays are used as the signal sample. The set of input variables comprises various features of the B_s^0 and $D_s^\pm$ mesons and the final state hadrons: the χ^2_{IP} of the B_s^0 meson, and the χ^2 of the DTF, χ^2_{DTF}. Variables of the $D_s^\pm$ mesons that are included are the difference of the reconstructed $D_s^\pm$ mass and the nominal $D_s^\pm$ mass, the transverse momenta, p_{T}, the χ^2_{IP} and the $D_s^\pm$ flight distance, FD. Furthermore, the p_{T} of all final state hadrons and the cosine of the angle between the flight direction of the $D_s^\pm$ meson and one final-state particle, $\theta_{D_s^\pm, K/\pi}$, are included. All input variables are listed in Table B.5. Before using the features in the training, it is checked that their distributions are reasonably well described in the simulation sample using background-subtracted data. The importance of the variables during the BDT training is shown in Fig. 6.6 for one fold. The χ^2_{IP} of the B_s^0 meson is the most important feature, followed by the mass differences, the χ^2_{IP} of the $D_s^\pm$ mesons and χ^2_{DTF}. The other features are less important with the $\theta_{D_s^\pm, K/\pi}$ variables being the features with the smallest importance.

To prevent overtraining, the early stopping method with $n = 10$ and a k-folding with $k = 5$ is performed. From the five classifier outputs one classifier is randomly chosen for each event that is not used in the training. The area under the ROC curve is used as the validation metric. On the left in Fig. 6.7, the BDT classifier distribution is shown for one fold with the training and test sample superimposed. No overtraining is visible. On the right in Fig. 6.7, the corresponding ROC curve is shown. An AUC value of over 0.99 is achieved for all folds indicating a high BDT performance.

Optimisation of the BDT requirement

The BDT response is added to the data sample and the requirement on the BDT is optimised with a suitable figure of merit. The goal of the optimisation is to find a cut point that leads to the best sensitivity on the *CP* parameters with the given data sample. The FOM that is utilised here is motivated by [107, 108]. It represents the inverse variance of the *CP* parameters of interest and is maximised with a scan of different requirements on the BDT response. Hence, the FOM minimises the variance of the parameter ϕ_s^{eff}. As only the variance enters the FOM, the central values of the *CP* parameter remain unbiased. The FOM takes several effects into account that lead to a more precise measurement of the *CP*

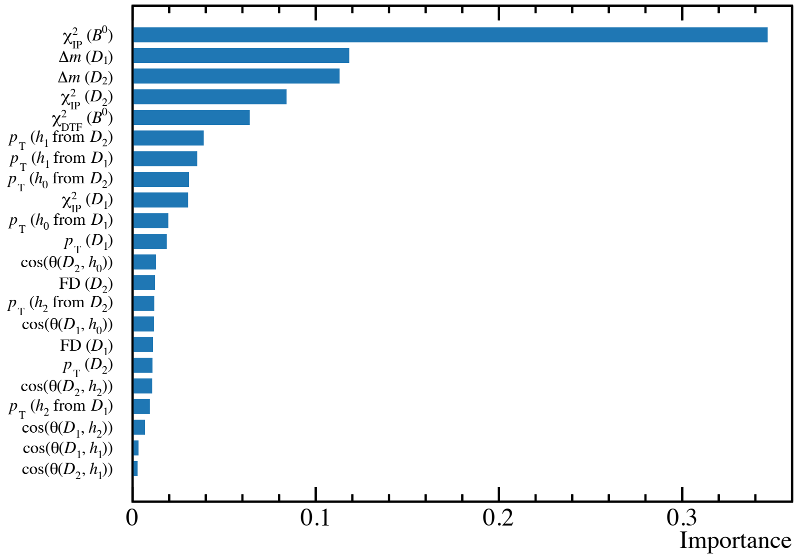

Figure 6.6: Feature importance for one fold. The features are defined in the text.

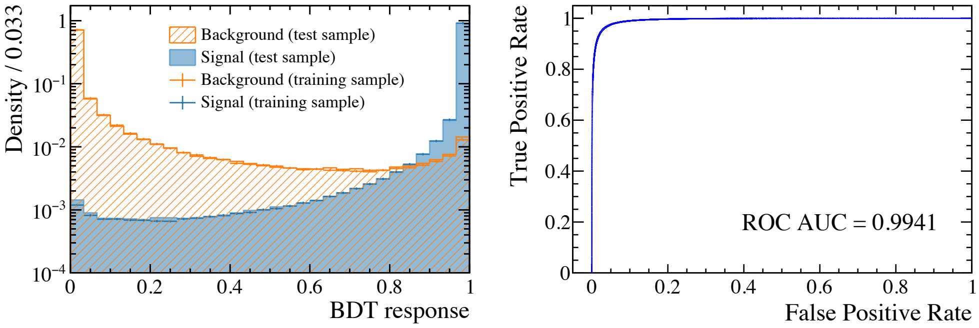

Figure 6.7: Classifier distribution for (blue) signal and (orange) background candidates from the (histogram) training and (data points) test sample for one fold on the left and the corresponding ROC curve on the right. The AUC value is close to one and indicates a good classification.

parameters: a high effective signal size, a good performance of the flavour tagging and small estimates of the decay time error. Furthermore, the sensitivity of the *CP* parameters depends on the reconstructed decay time of the B_s^0 candidates.

The FOM is defined as

$$\mathrm{FOM} = \frac{(\sum_i s_{wi})^2}{\sum_i s_{wi}^2} \mathrm{D}\,. \tag{6.2}$$

Here, s_{wi} indicates the sWeights that are determined by performing an extended maximum likelihood fit to the invariant $D_s^+ D_s^-$ mass distribution (see Sec. 6.1.6). The term D is a dilution term and takes the decay-time uncertainty estimates $\sigma(t_i)$, where t_i denotes the decay time, and the mistag probabilities ω_i into account. It is averaged over all candidates and is defined by

$$\mathrm{D} = \frac{1}{\sum_i s_{wi}} \sum_i (1 - 2\omega_i)^2 \mathrm{e}^{-(\Delta m_s \sigma(t_i))^2} \cdot \mathrm{X}_i \cdot s_{wi}\,, \tag{6.3}$$

where Δm_s denotes the mass difference in the $B_s^0 - \bar{B}_s^0$ system and is an external constant input. The term X_i is defined by

$$\mathrm{X}_i = \left[\frac{2 d_i |\lambda| s}{1 + |\lambda|^2 + d_i (1 - 2\omega_i) \mathrm{e}^{-(\Delta m_s \sigma(t_i))^2/2} (-2|\lambda| s \sin\phi_s^{\mathrm{eff}} - (1 - |\lambda|^2) c)} \right]^2 , \tag{6.4}$$

where d_i indicates the tagging decision for the production flavour of the B_s^0 mesons, and $s = \sin(\Delta m_s t_i)$ and $c = \cos(\Delta m_s t_i)$ are introduced as abbreviations. It includes the parameter $\sin\phi_s^{\mathrm{eff}}$, which is also an external and constant input.

The FOM can be divided into four individual components, which are normalised separately: The effective signal size described by the sWeight ratio

$$\mathrm{FOM}_{\mathrm{Signal}} \equiv \frac{(\sum_i s_{wi})^2}{\sum_i s_{wi}^2}\,, \tag{6.5}$$

the effective flavour-tagging dilution term given by

$$\mathrm{FOM}_{\mathrm{Tagging}} \equiv \frac{1}{\sum_i s_{wi}} \sum_i (1 - 2\omega_i)^2 \cdot s_{wi}, \tag{6.6}$$

the decay-time-resolution dilution term described by

$$\mathrm{FOM}_{\mathrm{Resolution}} \equiv \frac{1}{\sum_i s_{wi}} \sum_i \mathrm{e}^{-(\Delta m_s \sigma(t_i))^2} \cdot s_{wi}\,, \tag{6.7}$$

and a term associated to $\sin\phi_s^{\mathrm{eff}}$

$$\mathrm{FOM}_{CP} \equiv \frac{1}{\sum_i s_{wi}} \sum_i X_i \cdot s_{wi}\,. \tag{6.8}$$

The FOM is obtained from a BDT requirement scan for the three final state combinations separately. For each requirement, a maximum likelihood fit to the invariant $D_s^+ D_s^-$ mass distribution is performed (for details see Sec. 6.1.6) and the FOM is calculated. Here, the calibrated mistags ω_i (see Sec. 6.2) and the calibrated per-event decay-time uncertainty (see Sec. 6.3.1) enter the FOM. Further, the value for Δm_s is taken from [34], $|\lambda|$ is set to 1, and ϕ_s is set to -0.051 [53]. Fig. 6.8 shows the FOM and its contributions for the different final state combinations. For the final-state combination with the largest branching fraction, $KK\pi\ KK\pi$, a plateau around a requirement of 0.95 is present. This requirement is near the maximum of the other two final-state combinations with smaller branching fractions and is therefore chosen as the BDT requirement for all three final states. Signal efficiencies of about 93-95% are achieved, whereas about 97-98% of the combinatorial background is removed. The same BDT requirements are applied to the $B^0 \to D_s^+ D^-$ sample.

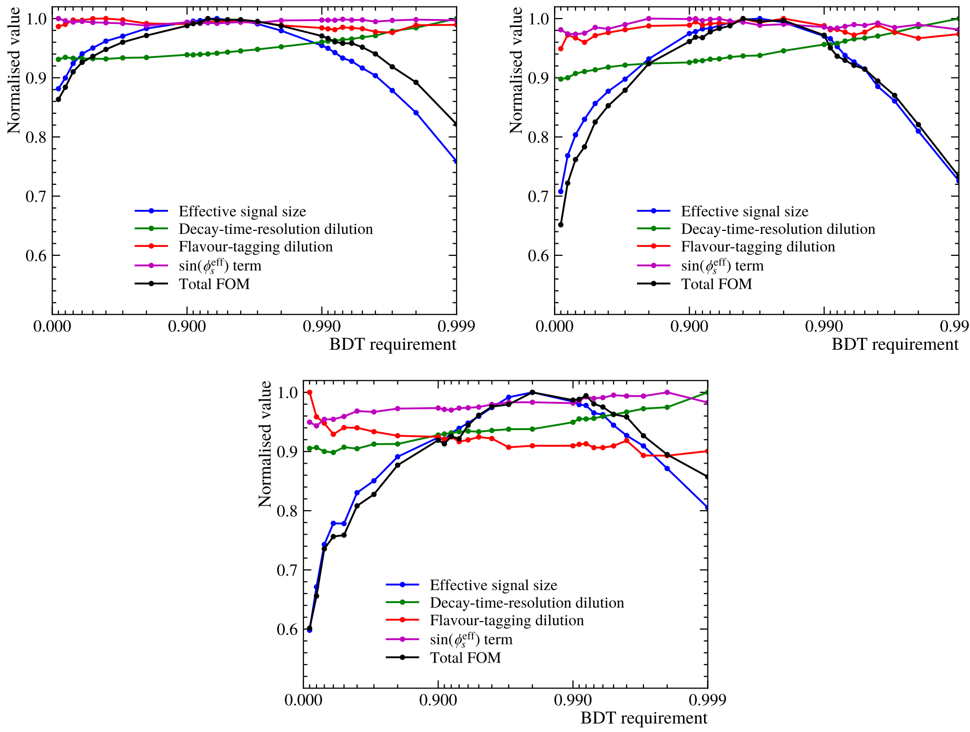

Figure 6.8: Figure of merit scan for the final state combinations (top left) $KK\pi\ KK\pi$, (top right) $KK\pi\ \pi\pi\pi$ and (bottom) $KK\pi\ \pi K\pi$. At each BDT requirement the (black) total FOM is calculated according to Eq. (6.2). In addition, the individual components are shown that reflect different effects on the measurement: (blue) the sWeight ratio, (green) the decay-time resolution dilution, (red) the tagging dilution and (violet) the sin(ϕ_s) term. All contributions are normalised to their highest value.

A similar BDT is trained for the $B_s^0 \to D_s^- \pi^+$ sample. The optimal BDT requirement for this sample is found with a BDT requirement scan, where the B_s^0 yield divided by its uncertainty, determined with mass fits, is used as the FOM.

6.1.5 Multiple Candidates

It is possible that more than one candidate remains in an event. After all selection steps, the fraction of events with multiple candidates is about 1.3–1.75% depending on the final-state combination. All of these candidates are equally likely to be a signal candidate, therefore one of the candidates is randomly selected and the others are discarded.

6.1.6 Mass Fit and Extraction of Signal Weights

In order to fit the decay time of signal candidates, the signal distribution has to be extracted. This is done by performing an extended maximum likelihood fit to the invariant $D_s^+ D_s^-$ mass for each final state combination separately, and calculating sWeights. The fit on recorded data is performed in the range of 5300 to 5600 MeV/c^2 to eliminate partially reconstructed backgrounds at low masses. Different PDFs are needed to describe the signal component and the component for the remaining combinatorial background.

To describe the signal component, a double-sided Hypatia function [104] is used, which is composed of a generalised hyperbolic core and a power-law tail to each side. The Hypatia function is defined as

$$\mathcal{I}(x;\mu,\sigma,\lambda,\zeta,\beta,a_1,n_1,a_2,n_2) \propto \begin{cases} \frac{G(\mu-a_1\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{(1-x/(n_1R_1-a_1\sigma))^{n_1}}, & -a_1 > \frac{x-\mu}{\sigma}\,, \\ \frac{G(\mu+a_2\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{(1-x/(-n_2R_2-a_2\sigma))^{n_2}}, & a_2 < \frac{x-\mu}{\sigma}\,, \\ G(m,\mu,\sigma,\lambda,\zeta,\beta), & \text{otherwise}\,. \end{cases} \tag{6.9}$$

with the generalised hyperbolic core

$$G(x,\mu,\sigma,\lambda,\zeta) = \\ \left((x-\mu)^2 + A_\lambda^2(\zeta)\sigma^2\right)^{\frac{1}{2}\lambda-\frac{1}{4}} \mathrm{e}^{\beta(x-\mu)} K_{\lambda-\frac{1}{2}}\left(\zeta\sqrt{1+\left(\frac{m-\mu}{A_\lambda(\zeta)\sigma}\right)^2}\right) \tag{6.10}$$

and the ratio of G and its derivative G'

$$R_i = \frac{G(\mu-a_i\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{G'(\mu-a_i\sigma,\mu,\sigma,\lambda,\zeta,\beta)}\,. \tag{6.11}$$

Cylindrical harmonics are denoted by the parameter $K_\lambda(\zeta)$, which is used to parametrise the term

$$A_\lambda^2(\zeta) = \frac{\zeta K_\lambda(\zeta)}{K_{\lambda+1}(\zeta)}\,. \tag{6.12}$$

The core is described by the mean μ, the width σ, the asymmetry parameter β and the parameters λ and ζ, which model the shape of the peak. The parameters a_i and n_i describe power-law tails to both sides of the distribution. In this analysis, the paramters ζ and λ are fixed to zero to achieve a more stable fit. For the fit to the invariant $D_s^+D_s^-$ mass, the parameters a_i, n_i, and λ are fixed to the values determined by a fit to simulated $B_s^0 \to D_s^+D_s^-$ decays, while all other parameters are floating. The result of the fit to simulated $B_s^0 \to D_s^+D_s^-$ decays is shown in Fig. 6.9.

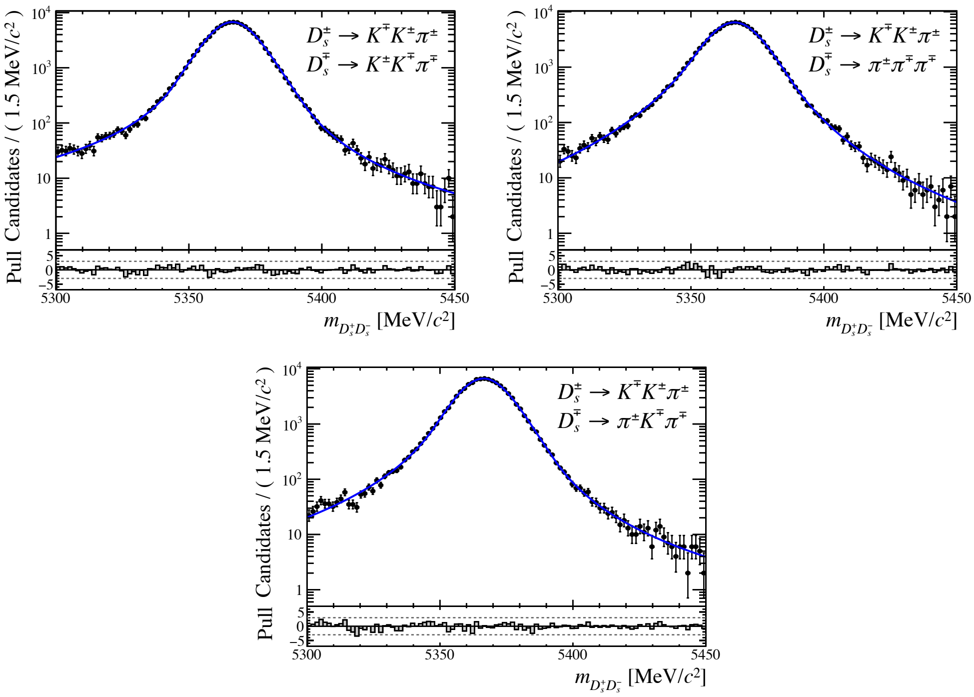

Figure 6.9: Invariant $D_s^+D_s^-$ mass distributions of simulated $B_s^0 \to D_s^+D_s^-$ decays for the final states (top left) $KK\pi$ $KK\pi$, (top right) $KK\pi$ $\pi\pi\pi$ and (bottom) $KK\pi$ $\pi K\pi$.

The component for the combinatorial background is described by an exponential function with the exponent λ_{exp}.
The invariant $D_s^+D_s^-$ mass distribution of recorded data and the fit projection are shown in Fig. 6.10. The resulting values of the floating parameters are reported in Table B.6. The total number of B_s^0 candidates is 13277 ± 135.

Using the result of the fit, sWeights are calculated using the method described in Sec. 4.3.

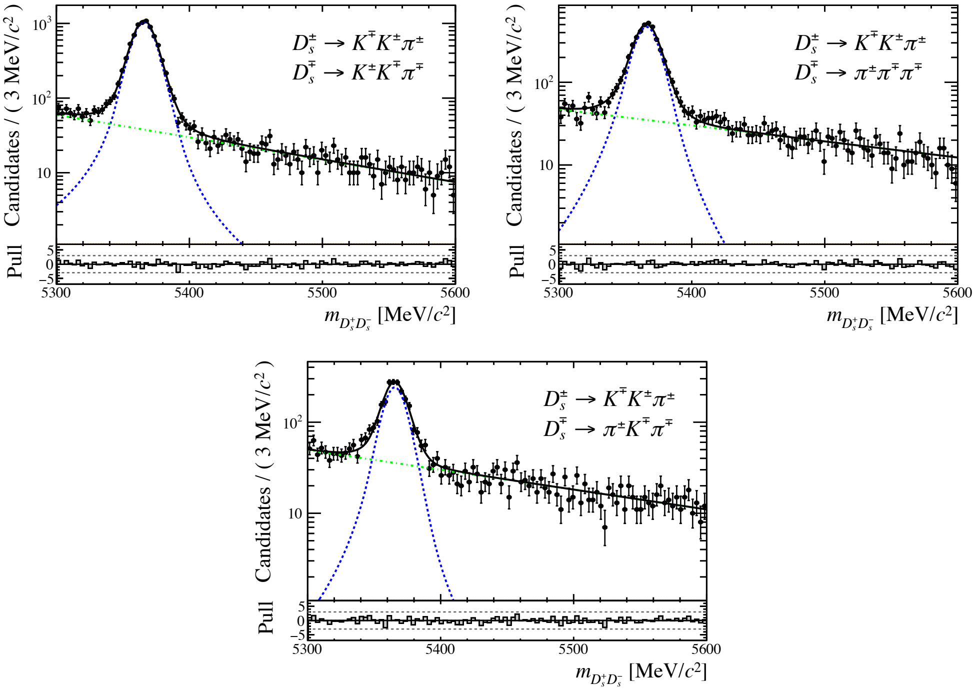

Figure 6.10: Invariant $D_s^+ D_s^-$ mass distributions of recorded data for the final states (top left) $KK\pi\ KK\pi$, (top right) $KK\pi\ \pi\pi\pi$ and (bottom) $KK\pi\ \pi K\pi$. The total fit projection is overlaid in black. The blue line describes the signal contribution and the green line corresponds to the combinatorial background contribution.

In the invariant $D_s^+ D^-$ mass distribution, a B^0 and a B_s^0 contribution are present. Hence, two signal PDFs composed of a double-sided Hypatia function are used to describe these two components. All parameters are shared between both PDFs except for the yield and the mean, where the difference of the mean is fixed to the known values [34]. The results of the fits are shown in Fig. 6.11. The total number of B^0 and B_s^0 candidates are 133175 ± 392 and 1292 ± 59, respectively.

In the fit to the invariant $D_s^- \pi^+$ distribution, the parametrisation consists of a double-sided Hypatia function to model the signal component and the sum of a constant function and an exponential function to describe the combinatorial background component. The parameter ζ is fixed to zero and the parameters n_i are fixed to values obtained with simulation. All other parameters are free in the fit. The result of the fit is shown in Fig. 6.12. The total number of $B_s^0 \to D_s^- \pi^+$ candidates is 305525 ± 4803.

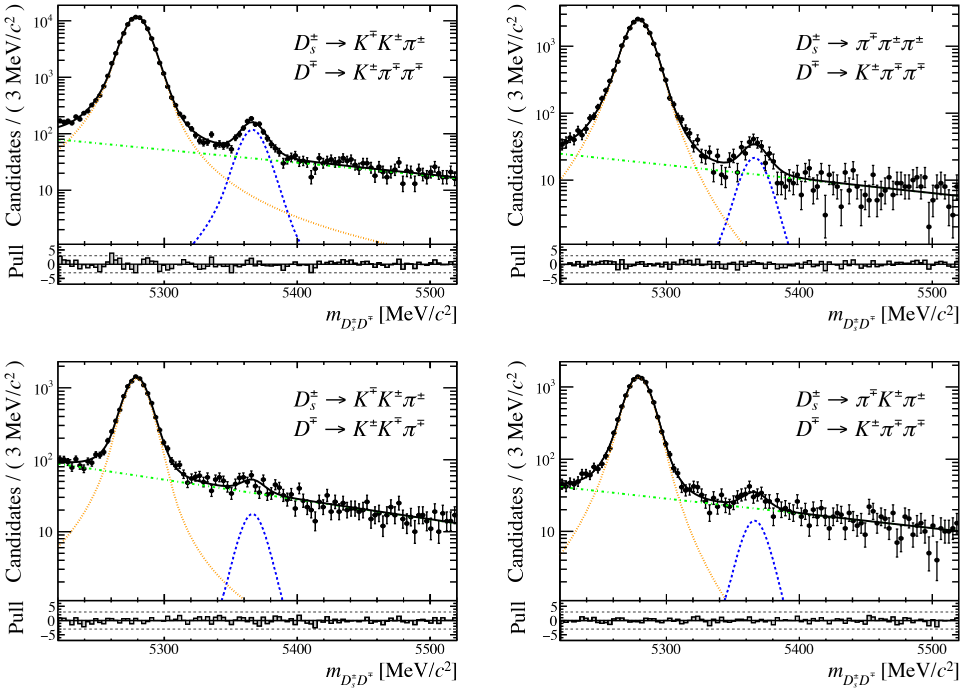

Figure 6.11: Invariant $D_s^+ D^-$ mass distributions of recorded data for the final states (top left) $KK\pi$ $K\pi\pi$, (top right) $\pi\pi\pi$ $K\pi\pi$, (bottom left) $KK\pi$ $KK\pi$ and (bottom right) $\pi K\pi$ $K\pi\pi$. The total fit projection is overlaid in black. The orange and blue lines describe the $B^0 \to D_s^+ D^-$ and $B_s^0 \to D_s^+ D^-$ contributions, respectively. The green line corresponds to the combinatorial background contribution.

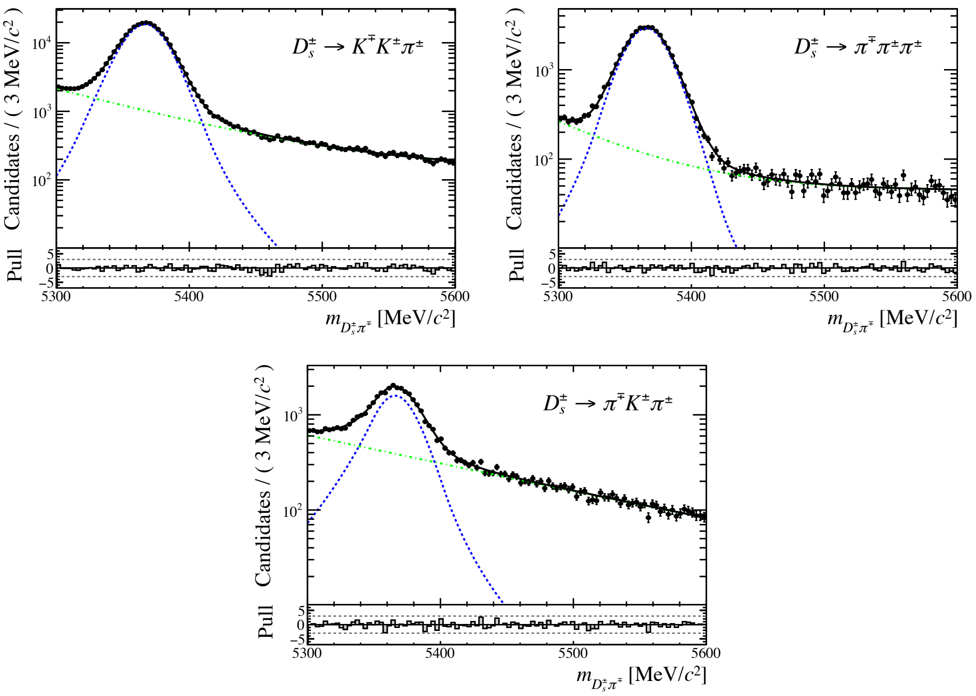

Figure 6.12: Invariant $D_s^+\pi^-$ mass distributions of recorded data for the D_s^+ final states (top left) $KK\pi$, (top right) $\pi\pi\pi$ and (bottom) $\pi K\pi$. The total fit projection is overlaid in black. The blue line describes signal contribution and the green line corresponds to the combinatorial background contribution.

6.2 Flavour-tagging Calibration

The flavour-tagging calibration of the single taggers is performed using the Espresso Performance Monitor (EPM), a tool developed by the flavour tagging group at LHCb [109]. For the calibration, either self-tagging charged B-meson decays or flavour-specific decays of neutral B-mesons are needed. In case of neutral B mesons, which are able to oscillate, a fit to the decay time is performed during the calibration taking into account the known oscillation frequency and the decay time. A suitable calibration channel is one that has similar kinematic properties to the signal channel to reduce systematic uncertainties from the flavour tagging. To also reduce the statistical uncertainty on the calibration parameters, the calibration channel should have a large number of candidates. In addition, the selection of signal and calibration channel should be similar to avoid large systematic uncertainties.

The flavour-specific control channel $B^0 \to D_s^+ D^-$ is utilised for the calibration of the OS tagger as it is kinematically very similar to the signal channel. For the SS however, a decay of a B_s^0 meson is necessary. Therefore, flavour-specific $B_s^0 \to D_s^- \pi^+$ decays are used for the calibration of the SS kaon tagger, even though kinematic differences can be present. To account for these kinematic differences, the sWeighted distributions of the pseudorapidity η and transverse momentum p_T of the B_s^0 candidates, the number of PVs and the number of tracks in the event are reweighted to match the distributions of the signal sample using a BDT based method, *i.e.* Gradient Boosting Reweighting [110]. These inputs are chosen due to the known dependence of the mistag on these variables. The distributions of the original and reweighted $B_s^0 \to D_s^- \pi^+$ candidates and $B_s^0 \to D_s^+ D_s^-$ candidates are shown in Fig. 6.13. In addition, the predicted mistag distribution of the SS kaon tagger is shown.

The calibration plots of the OS combination and the SS Kaon tagger determined on the calibration channels are shown in Fig. 6.14. The resulting calibration parameters are listed in Table 6.3 and enter the fit to the decay-time distribution via Gaussian constraints. In Table 6.4 the calibration performances of the OS combination and SS Kaon tagger as well as the combination of both determined on the signal sample are summarised. The effective tagging efficiency is 5.55%.

Table 6.3: Calibration parameters of the OS combination and the SS kaon tagger determined on the calibration channels.

Tagger	$\langle\eta\rangle$	p_0	p_1	Δp_0	Δp_1
OS combination	0.343	0.018	0.878	0.003	0.032
SS kaon	0.413	0.028	0.851	0.004	0.052

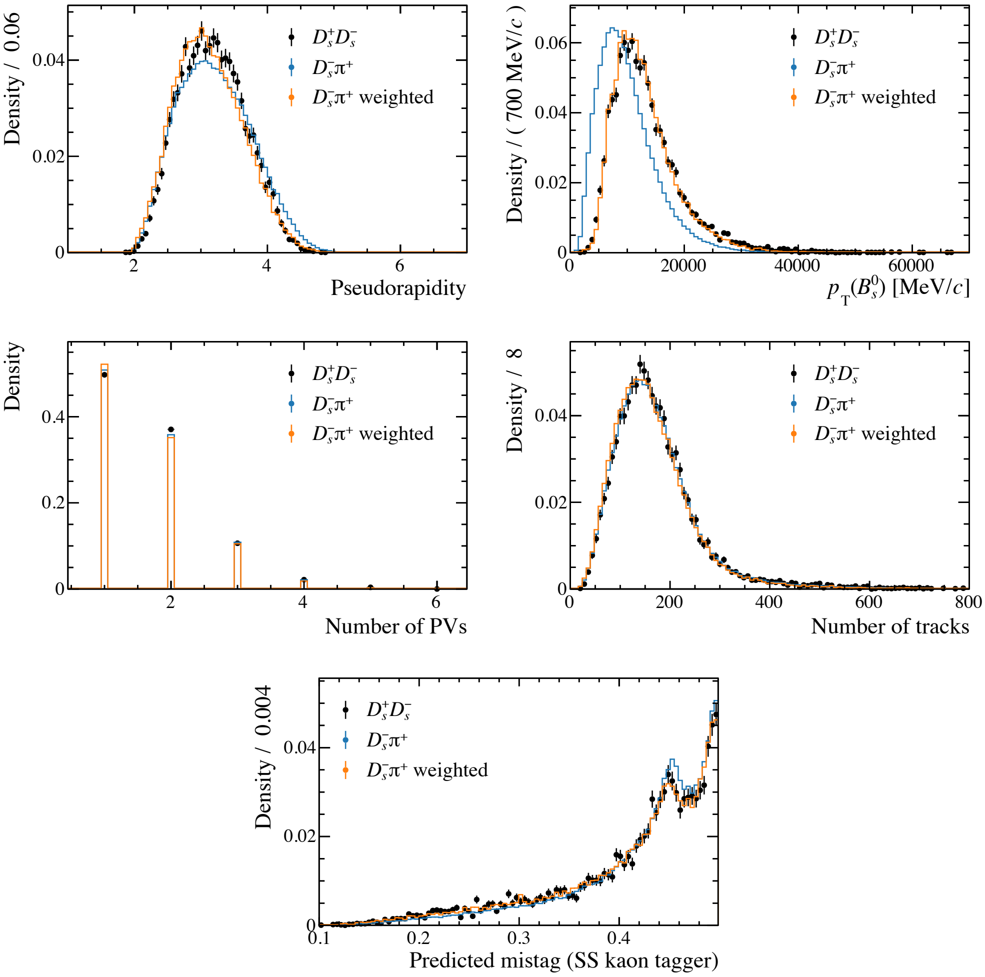

Figure 6.13: Normalised distributions for the (top left) pseudorapidity, (top right) transverse momentum p_{T} of the B_s^0 candidate, (middle left) number of PVs, (middle right) number of tracks, and (bottom) predicted mistag of the SS kaon tagger for (black) $B_s^0 \to D_s^+ D_s^-$ data, (blue) original $B_s^0 \to D_s^- \pi^+$ data and (orange) reweighted $B_s^0 \to D_s^- \pi^+$ data. Only the first four variables are inputs for the reweighting.

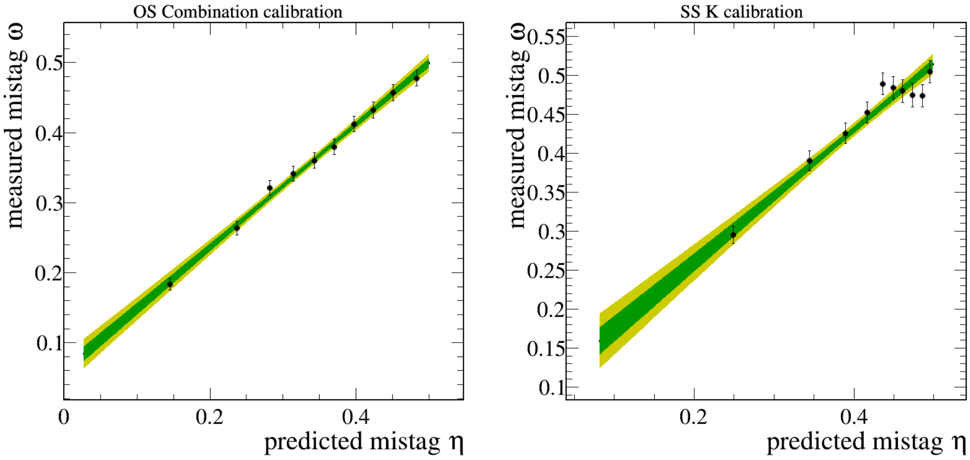

Figure 6.14: Results of the flavour-tagging calibration on the (left) $B^0 \to D_s^+ D^-$ sample for the OS combination and on the (right) $B_s^0 \to D_s^- \pi^+$ sample for the SS kaon tagger. The measured mistag ω is presented as a function of the predicted mistag η. The light and dark green shaded areas represent the one and two sigma confidence intervals.

Table 6.4: The average mistag probability estimate, the tagging efficiency and tagging power of the OS combination and the SS kaon tagger and the combination of both determined on the signal sample.

Tagger	$\langle\eta\rangle$	$\varepsilon_{\text{tag}}[\%]$	$\varepsilon_{\text{eff}}[\%]$
OS	0.342	34.71 ± 0.33	3.56 ± 0.06
SS	0.412	58.64 ± 0.34	2.27 ± 0.03
OS+SS		72.05 ± 0.31	5.55 ± 0.07

6.3 Decay-time Parameterisation

To describe the measured decay time on data, the theoretical PDF has to be modified to also consider various detector effects. The PDF to describe the distribution of the reconstructed decay-time, t, with the tags $\vec{d} = (d^{OS}, d^{SS})$ and the per-event mistag probability estimates $\vec{\eta} = (\eta^{OS}, \eta^{SS})$ is defined by

$$P(t, \vec{d}|\vec{\eta}) = \varepsilon(t) \int \mathcal{P}(t', \vec{d}|\vec{\eta}) R(t - t') dt' \,. \tag{6.13}$$

Here, $\varepsilon(t)$ is a decay-time-dependent efficiency function caused by the reconstruction and selection. The term $\mathcal{P}(t', \vec{d}|\vec{\eta})$ describes the distribution of the true decay-time, t'. The term $R(t - t')$ describes the decay-time resolution. The finite precision of the vertex measurements of the LHCb detector causes a deviation of the reconstructed decay time from the true decay time. Hence, the resolution term is convolved with the $\mathcal{P}(t', \vec{d}|\vec{\eta})$ term.

The PDF describing the true decay time, $\mathcal{P}(t', \vec{d}|\vec{\eta})$, is obtained using the theoretical decay rates from Eq. (2.23) simplified by the assumption of $|q/p|^2 \approx 1$, *i.e.* no indirect *CP* violation is present, which is compatible with the current experimental precision [53]. Further, the production asymmetry of the B_s^0 and $\overline{B}_s^0$ mesons has to be taken into account. It is defined as

$$A_{\text{prod}} = \frac{\sigma(\overline{B}_s^0) - \sigma(B_s^0)}{\sigma(\overline{B}_s^0) + \sigma(B_s^0)}\,, \tag{6.14}$$

where σ is the production cross-section for B_s^0 and $\overline{B}_s^0$ mesons. The decay rates from Eq. (2.23) change to

$$\begin{aligned}
\mathcal{P}(t', B_s^0) \propto e^{-\Gamma t}(1 - A_{\text{prod}})\cdot& \\
\mathcal{P}\left[\cosh(\frac{\Delta\Gamma}{2}t) + A^{\Delta\Gamma}\sinh(\frac{\Delta\Gamma}{2}t) + C\cos(\Delta m t) - S\sin(\Delta m t)\right]&\,, \\
\mathcal{P}(t', \overline{B}_s^0) \propto e^{-\Gamma t}(1 + A_{\text{prod}})\cdot& \\
\left[\cosh(\frac{\Delta\Gamma}{2}t) + A^{\Delta\Gamma}\sinh(\frac{\Delta\Gamma}{2}t) - C\cos(\Delta m t) + S\sin(\Delta m t)\right]&\,.
\end{aligned} \tag{6.15}$$

Next, the flavour tagging has to be included in the PDF. The decisions of the OS combination and the SS kaon tagger are combined in the PDF. This way, instead of fixing the calibration parameters in the fit, they can be constrained and their uncertainties are included in the statistical uncertainty of the *CP* observables. For a given true production flavour d, the probability to obtain the tagging decisions d^i of tagger i is by definition

$$P^i(d^i, b) = \begin{cases} \varepsilon^i(1 - \omega_b^i) & \text{if } d^i = b\,, \\ \varepsilon^i \omega_b^i & \text{if } d^i = -b\,, \\ (1 - \varepsilon^i) & \text{if } d^i = 0\,, \end{cases} \tag{6.16}$$

with b indicating a B_s^0 or $\overline{B}_s^0$ meson, the tagging efficiency ε and the calibrated mistag ω_b, which is a function of the predicted mistag η. The PDF describing the observed tag and decay time is given by

$$\mathcal{P}(t', \vec{d}|\vec{\eta}) = P(d|B_s^0)\mathcal{P}(t', B_s^0) + P(d|\overline{B}_s^0)\mathcal{P}(t', \overline{B}_s^0)\,. \tag{6.17}$$

Further, the parameter $\Delta^\pm$ is introduced

$$\Delta^\pm = P(B_s^0) \pm P(\overline{B}_s^0)\,, \tag{6.18}$$

where

$$P(b) = P^{\text{SS}}(b)P^{\text{OS}}(b)\,, \tag{6.19}$$

can be calculated using Eq. (6.16). The PDF in Eq. (6.15) changes to

$$\begin{aligned} P(t', \vec{d}|\vec{\eta}) = \mathrm{e}^{-\Gamma t}\Big[&(\Delta^{+} - A_{\text{prod}}\Delta^{-})\Big(\cosh\big(\frac{\Delta\Gamma}{2}t\big) + A^{\Delta\Gamma}\sinh\big(\frac{\Delta\Gamma}{2}t\big)\Big) \\ &+ (\Delta^{-} - A_{\text{prod}}\Delta^{+})\Big(C\cos(\Delta m t) - S\sin(\Delta m t)\Big)\Big]\,. \end{aligned} \tag{6.20}$$

The description of the decay-time resolution and decay-time-dependent efficiency is discussed in Sec. 6.3.1 and Sec. 6.3.2, respectively.

6.3.1 Decay-time Resolution

In general, the decay-time resolution can be described by a Gaussian function with mean μ and width σ. The width is dependent on the predicted per-event decay-time uncertainty δ_t. Due to the fact that δ_t is not well described, it has to be calibrated. A linear calibration function given by

$$\sigma(\delta_t) = p_0 + p_1 \cdot \delta_t\,, \tag{6.21}$$

with the calibration parameters p_0 and p_1, is found to be suitable. The calibration parameters are determined using prompt data samples, where the final state is produced in the PV, but not necessarily in the decay of a B_s^0 meson. No prompt $D_{(s)}^{+}D_{(s)}^{-}$ data samples exist, therefore prompt $D_s^{-}\pi^{+}$ candidates are utilised, where a prompt D_s^{-} candidate is combined with a random π^{+} originating from the PV. The difference of the resolution between $D_s^{-}\pi^{+}$ and $D_s^{+}D_s^{-}$ is corrected by the scale factor s, which is determined with simulated $B_s^0 \to D_s^{+}D_s^{-}$ and $B_s^0 \to D_s^{-}\pi^{+}$ samples. The linear function changes to

$$\sigma(\delta_t) = s \cdot (p_0 + p_1 \cdot \delta_t)\,. \tag{6.22}$$

Only the $D_s^{-} \to K^{+}K^{-}\pi^{-}$ final state is present in the prompt $D_s^{-}\pi^{+}$ sample. To suppress background, only the relative clean resonant $D_s^{-} \to \phi(\to K^{+}K^{-})\pi^{-}$ mode is considered by excluding candidates where the invariant $K^{+}K^{-}$ mass does not lie in a mass window of $\pm 10\,\mathrm{MeV}/c^2$ around the known ϕ mass [34]. In addition, only events with one PV are considered to prevent resolution effects from additional PVs.

The resolution is determined by performing fits to the decay-time distribution of the prompt sample in 50 bins of the predicted decay-time uncertainty. Each bin is chosen such that the height of the distribution is the same in all bins. This allows the same parametrisation of the resolution to be used in all bins and gives the most robust fit results. In every bin the same procedure is conducted. A fit to the D_s^{-} mass is performed to extract sWeights using a Hypatia function to describe the signal component and a linear function to model the background. In Fig. 6.15, the result of the fit for the two bins with the lowest and highest values of δ_t are shown. Afterwards, a fit to the sWeighted decay time is performed. The positive tail of the decay-time distribution is neglected to discard background

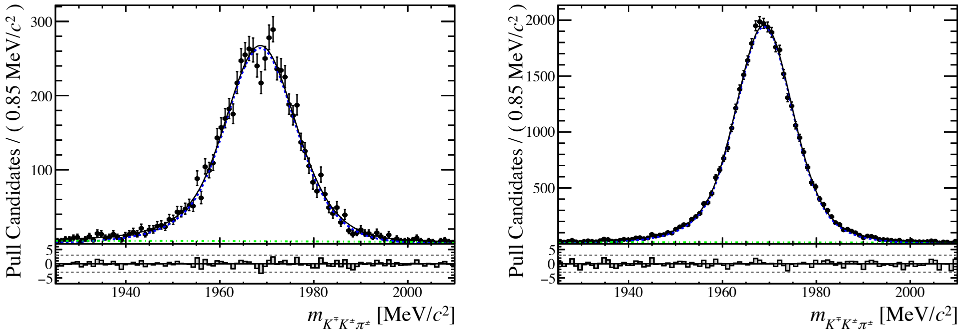

Figure 6.15: The $K^{\mp}K^{\pm}\pi^{\pm}$ mass distribution of the prompt $D_s^+\pi^-$ sample with the (black) full fit projection overlaid for two bins with the (left) lowest respectively (right) highest decay-time uncertainty. The signal component described by an Hypatia function and the combinatorial background component described by a linear function are shown in blue and green, respectively.

from long-lived decays. The fit range is recalculated in every bin. The lower boundary is always $-0.5\,\mathrm{ps}$. The upper boundary is equal to the mean of the decay-time uncertainty in the current bin to include the $1\,\sigma$ region of the resolution. The resolution model comprises three Gaussian functions with a shared mean and individual widths. The result of the fit to the decay time is shown in Fig. 6.16 for the two previously mentioned bins. The effective resolution can be inferred from the widths of the decay-time distribution. The individual widths are combined using the decay-time dilution D

$$D = \sum_i f_i e^{-\sigma_i^2 {\Delta m_s}^2/2}\,, \tag{6.23}$$

where f and σ are the individual fractions and widths of the Gaussian functions, and Δm_s is the B_s^0 oscillation frequency. Afterwards, the effective width σ_{eff} representing the measured resolution is calculated with

$$\sigma_{\text{eff}} = \sqrt{(-2/{\Delta m_s}^2)\ln D}\,. \tag{6.24}$$

Finally, the results of the 50 bins are used to calibrate the measured resolution with Eq. (6.21). The result of a χ^2 minimisation fit is shown in Fig. 6.17. The resulting calibration parameters are:

$$\begin{aligned} p_0 &= (0.0088 \pm 0.0008)\,\mathrm{ps}\,, \\ p_1 &= 1.061 \pm 0.024\,. \end{aligned} \tag{6.25}$$

To determine the scale factor from Eq. (6.22), simulated samples are utilised. Here, the true decay time is known and the resolution can be determined by a fit to the difference of the reconstructed and true decay time $t - t_{\text{true}}$. To

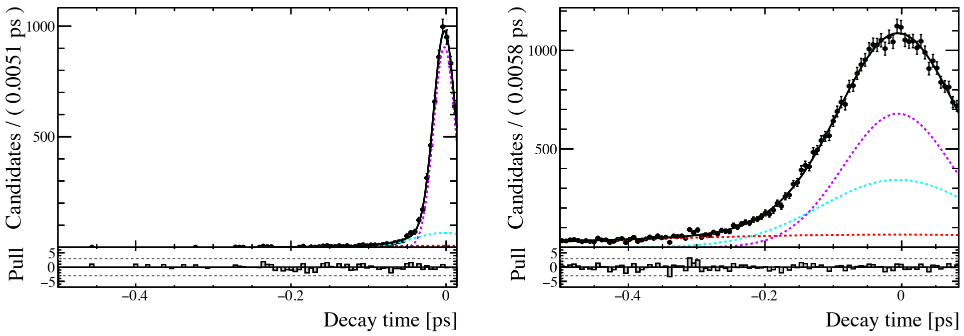

Figure 6.16: The sWeighted decay-time distribution of the prompt $D_s^+\pi^-$ sample for two bins with the (left) lowest respectively (right) highest decay time error. The full fit projection of the resolution is shown in solid black, while the three Gaussian components are shown in red, blue and violet.

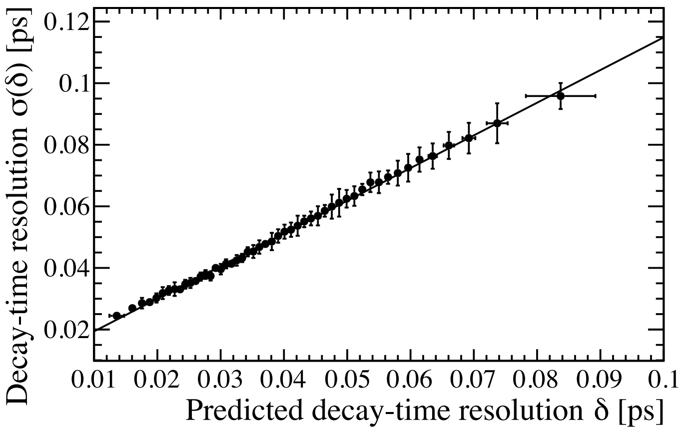

Figure 6.17: Measured decay-time resolution σ determined from fits to the decay time as a function of the predicted decay-time uncertainty δ. The linear calibration function is overlaid.

find the resolution of $B_s^0 \to D_s^- \pi^+$ decays, a fit using $B_s^0 \to D_s^- \pi^+$ simulation is performed. The resolution model comprises the sum of three Gaussian functions. Afterwards, a fit is performed using $B_s^0 \to D_s^+ D_s^-$ simulation with the model from the $B_s^0 \to D_s^- \pi^+$ fit, where the widths are multiplied with a single scale factor. Here, only the mean and the scale factor are floating in the fit. The scale factor accounts for resolution differences between $B_s^0 \to D_s^- \pi^+$ and $B_s^0 \to D_s^+ D_s^-$ and its resulting value is

$$s = 1.222 \pm 0.011\,. \tag{6.26}$$

The fits to $B_s^0 \to D_s^- \pi^+$ and $B_s^0 \to D_s^+ D_s^-$ simulation are shown in Fig. 6.18.

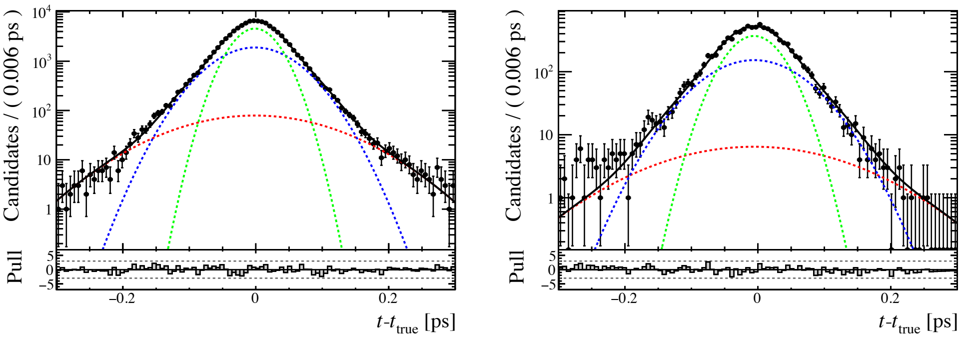

Figure 6.18: Distribution of the difference between the reconstructed and true decay time of simulated (left) $B_s^0 \to D_s^- \pi^+$ and (right) $B_s^0 \to D_s^+ D_s^-$ decays. The full fit projection is shown as the solid black line. The resolution is modelled by three Gaussian functions, which are shown in dashed blue, red and green. For the $B_s^0 \to D_s^+ D_s^-$ fit, the model from the $B_s^0 \to D_s^- \pi^+$ fit is used and the widths are multiplied with a single scale factor, which accounts for resolution differences between $B_s^0 \to D_s^- \pi^+$ and $B_s^0 \to D_s^+ D_s^-$ decays.

Finally, using the calibration parameters from Eq. (6.25) and the scale factor from Eq. (6.26), in the decay-time fit to $B_s^0 \to D_s^+ D_s^-$ data, the resolution of each candidate is calibrated with

$$\sigma_{D_s^+ D_s^-} = s \cdot (p_0 + p_1 \cdot \delta_t) = 1.222 \cdot (0.0088\,\text{ps} + 1.061 \cdot \delta_t)\,. \tag{6.27}$$

With this, the effective decay-time resolution is measured to be $\sigma_{\text{eff}} \approx 67\,\text{fs}$.

6.3.2 Decay-time-dependent Efficiency

The reconstruction and selection requirements lead to a decay-time-dependent efficiency, which appears as a deviation from a pure exponential decay-time distribution. The selection mostly rejects backgrounds stemming from the PV that have small reconstructed decay times. Thus, signal candidates with smaller decay times are more often rejected.

The decay-time-dependent efficiency is modelled by the sum of seven cubic B-splines b_i, which are parametrised by five knots and their coefficients c_i

$$\varepsilon(t) = \sum_i c_i b_i(t)\,. \tag{6.28}$$

The knots define the boundaries of the range where a B-spline is defined. The knot positions of 0.3, 0.5, 2.7, 6.3 and 10.3 ps are found to provide a good description. The first and last positions correspond to the lower and upper range used in the decay-time fit. The coefficients of the first two respectively the last two B-splines are set to the same values for a more robust fit.

The decay-time-dependent efficiency is not well described in simulation, thus the determination is purely data-driven using the $B^0 \to D_s^+ D^-$ channel. All final states are merged. This can be done as the decay time distribution does not differ between the final states, as shown with simulated $B_s^0 \to D_s^+ D_s^-$ decays in Fig. 6.19.

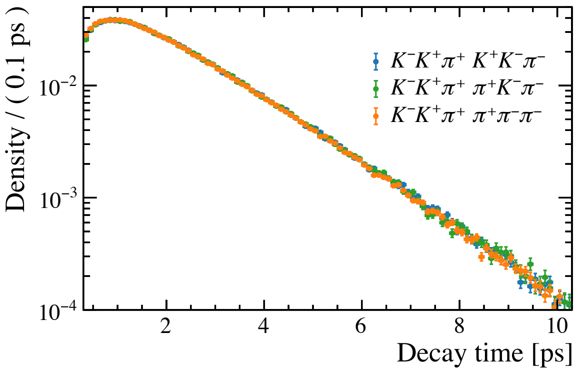

Figure 6.19: Normalised decay-time distribution of simulated $B_s^0 \to D_s^+ D_s^-$ decays separately for each final state. The distributions of the three final-state categories (blue) $K^-K^+\pi^+$ $K^+K^-\pi^-$, (green) $K^-K^+\pi^+$ $\pi^+K^-\pi^-$, and (orange) $K^-K^+\pi^+$ $\pi^+\pi^-\pi^-$ are in agreement with each other. Thus, the final states are merged in the modelling of the decay-time-dependent efficiency.

The parametrisation of the decay-time-dependent efficiency is found using Eq. (6.13), where the term containing the true decay time t' is replaced by an exponential function $\exp(t'/\tau_{B^0})$ that decreases with the fixed B^0 lifetime. A maximum likelihood fit to the decay time distribution is performed on recorded $B^0 \to D_s^+ D^-$ data. The largest coefficient is fixed to unity to fix the overall normalisation. The resulting coefficients are reported in Table 6.5. Fig. 6.20 shows the sWeighted decay-time distribution with the result of the fit overlaid. These values are fixed in the decay-time fit to determine the *CP* parameters, because of their high correlation to the *CP* parameter $A_{\Delta\Gamma}$. In addition, the result of a maximum likelihood fit to the decay time distribution on simulated

Table 6.5: Coefficients of the individual B-splines describing the decay-time-dependent efficiency resulting from a fit to the decay-time distribution using $B^0 \to D_s^+ D^-$ data. The values reflect the characteristic distribution of the decay-time-dependent efficiency: at shorter decay times the efficiency goes to zero, it increases to a plateau at intermediate decay times, and decreases at longer decay times.

Coefficient	Value
c_1	0.330 ± 0.006
c_2	$\equiv c_1$
c_3	0.924 ± 0.021
c_4	1.0
c_5	0.973 ± 0.042
c_6	0.729 ± 0.033
c_7	$\equiv c_6$

$B^0 \to D_s^+ D^-$ decays is shown together with the individual spline components where the effect of the exponential decay is eliminated.

To verify the portability from $B^0 \to D_s^+ D^-$ to $B_s^0 \to D_s^+ D_s^-$ data, two checks are performed. First, the decay-time-dependent efficiency model obtained with $B^0 \to D_s^+ D^-$ data is applied to the decay-time distribution of $B_s^0 \to D_s^+ D_s^-$ data. As the $D_s^+ D_s^-$ final state is *CP*-even, only the *CP*-even light B_s^0 mass eigenstate is allowed to perform this decay when neglecting *CP* violation in mixing. Thus, the lifetime comprised in the exponential function for the decay-time-dependent efficiency is fixed to the effective lifetime of the light mass eigenstate of the B_s^0 system. The result is shown in Fig. 6.21 on the left. Second, a maximum likelihood fit to the decay-time distribution is performed using $B_s^0 \to D_s^+ D_s^-$ data, and the fit projection and decay-time distribution are shown in Fig. 6.21 on the right. Both checks confirm that the model from the $B^0 \to D_s^+ D^-$ fit can be used to describe the decay-time-dependent efficiency function of $B_s^0 \to D_s^+ D_s^-$ decays.

6.4 Determination of *CP* Parameters

A maximum likelihood fit to the sWeighted decay-time distribution is performed to extract ϕ_s^{eff} and $|\lambda|$. The *CP* observables $S_{D_s^+ D_s^-}$, $C_{D_s^+ D_s^-}$ and $A^{\Delta\Gamma}_{D_s^+ D_s^-}$ are not directly extracted from the *CP* fit, instead ϕ_s^{eff} and $|\lambda|$ enter the *CP* observables

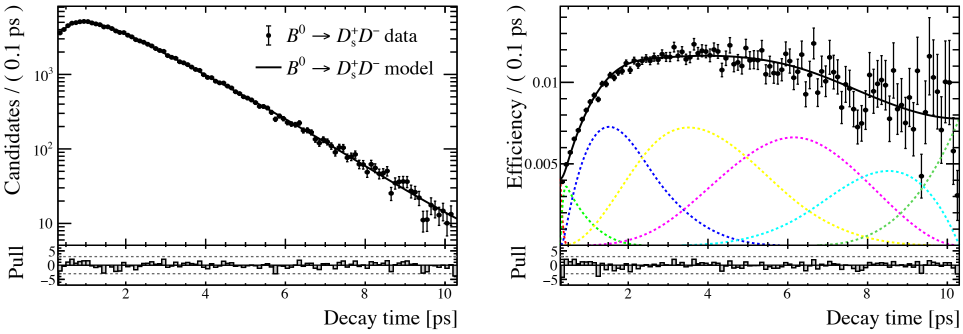

Figure 6.20: (Left) decay-time distribution in sWeighted $B^0 \to D_s^+ D^-$ data and the fit projection, and (right) the decay-time distribution of $B^0 \to D_s^+ D^-$ simulation with the projection of the decay-time-dependent efficiency, where the effect of the exponential decay is eliminated. The different B-spline components are shown in different colours.

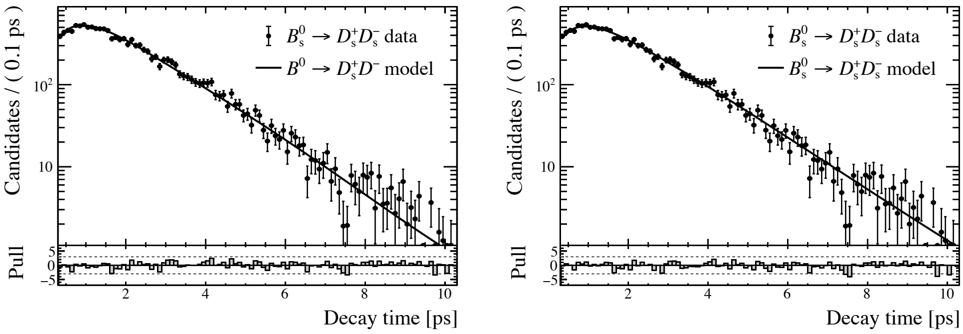

Figure 6.21: (Left) decay-time distribution using sWeighted $B_s^0 \to D_s^+ D_s^-$ data and the fit projection obtained with $B^0 \to D_s^+ D^-$ data, and (right) the decay-time distribution using sWeighted $B_s^0 \to D_s^+ D_s^-$ data and the fit projection obtained from a fit with $B_s^0 \to D_s^+ D_s^-$ data. Both models look almost identical, proving that the results from the $B^0 \to D_s^+ D^-$ data can be used to fix the decay-time-dependent efficiency function.

from Eq. (2.41) via

$$\begin{aligned} A^{\Delta\Gamma}_{D_s^+ D_s^-} &= -\frac{2\cos(\phi_s^{\text{eff}})}{1+|\lambda|^2}\,, \\ C_{D_s^+ D_s^-} &= \frac{1-|\lambda|^2}{1+|\lambda|^2}\,, \\ S_{D_s^+ D_s^-} &= -\frac{2\sin(\phi_s^{\text{eff}})}{1+|\lambda|^2}\,. \end{aligned} \tag{6.29}$$

Gaussian constraints are applied to external parameters: the average lifetime $\tau_{B_s^0}$ [53], the mixing frequency Δm_s [53], and the decay-width difference $\Delta\Gamma_s$ [53]. In addition, Gaussian constraints on the flavour-tagging-calibration parameters (see Table 6.3) are applied. The coefficients on the decay-time-dependent efficiency (see Table 6.5) are fixed. The parameters of the decay-time resolution calibration (see Eq. (6.25)) as well as the scale factor (see Eq. (6.26)) are also fixed in the fit.

In Fig. 6.22 the decay-time distribution and the projection of the fit PDF are shown. The following uncertainties are obtained from the fit:

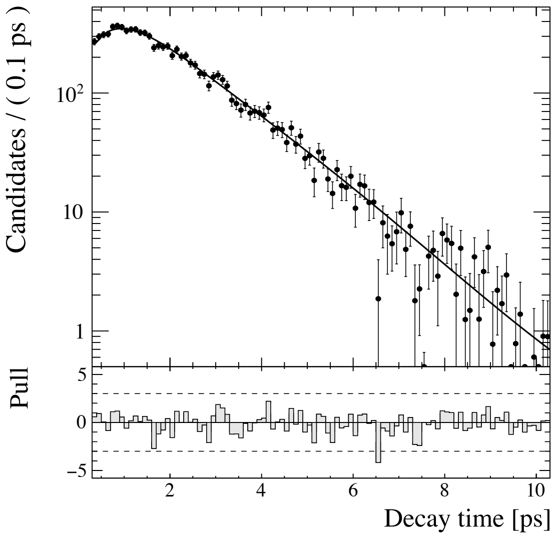

Figure 6.22: The sWeighted decay-time distribution of $B_s^0 \to D_s^+ D_s^-$ decays in logarithmic scale. The full fit projection is overlaid in black.

$$\begin{aligned} \sigma_\lambda &= 0.08\,, \\ \sigma_{\phi_s^{\text{eff}}} &= 0.13\,. \end{aligned} \tag{6.30}$$

The statistical correlation is important when combining these results with other measurements. The correlation between the λ and ϕ_s^{eff} is $\rho(\lambda, \phi_s^{\text{eff}}) = -0.37$. Additionally, significant correlations to nuisance parameters have the values $\rho(\lambda, \Delta m_s) = -0.05$, $\rho(\lambda, \tau_{B_s^0}) = -0.09$ and $\rho(\phi_s^{\text{eff}}, \tau_{B_s^0}) = 0.24$. All other correlations are in the sub-percent level.

6.5 Expected Systematic Uncertainties

The systematic uncertainties of the measurement of the *CP* parameters are currently being evaluated. This section gives an estimation of the sizes of systematic uncertainties that can be expected in this analysis using the previous

Run 1 analysis [16]. In the previous analysis, the systematic uncertainty was about 11% of the statistical uncertainty and came from three major sources.

The largest source of systematic uncertainty in the previous analysis is associated with the decay-time-resolution calibration. It yields a systematic uncertainty of about 9% and 10% of the statistical uncertainty for ϕ_s and $|\lambda|$, respectively. The decay-time resolution was similar to the presented measurement calibrated in bins of the per-event predicted decay-time uncertainty, but the measured resolution was determined on signal simulation. The predicted decay-time uncertainty is not well described in simulated samples. The assigned systematic uncertainty takes the data-simulation disagreements of the predicted decay-time uncertainty into account. It was determined by performing pseudo-experiments where alternative calibration parameters obtained with signal data are used in the generation of pseudo datasets, which are then fitted assuming the nominal resolution model. In the presented analysis, prompt $D_s^-\pi^+$ data samples are used instead, in which the predicted decay-time uncertainty is well described. Simulated samples are used to scale the resolution from $B_s^0\to D_s^-\pi^+$ to $B_s^0\to D_s^+D_s^-$ decays. The scale factor potentially differs between simulation and data. A systematic uncertainty will be evaluated using alternative methods to scale the $D_s^-\pi^+$ resolution to the signal resolution. Due to the high impact of the resolution, this source is still expected to lead to the largest systematic uncertainty.

The second largest source of systematic uncertainties in the previous analysis is the parametrisation of the mass PDF. The values of the systematic uncertainties are about 4% and 1% of the statistical uncertainty for ϕ_s and $|\lambda|$, respectively. A different model was used to parametrise the mass distribution than in the presented analysis. Misidentified backgrounds from $B^0\to D_s^+D^-$ and $\Lambda_b^0\to\Lambda_c^+D_s^-$ and single charm background from $B_s^0\to D_s^+K^+K^-\pi^-$ decays have not been suppressed leading to additional components in the PDF. In the evaluation of the systematic uncertainty, the mass range has been enlarged to lower masses and an additional background component from partially-reconstructed decays is included in the PDF. In the presented analysis misidentified backgrounds from *e.g.* $B^0\to D_s^+D^-$ and $\Lambda_b^0\to\Lambda_c^+D_s^-$ decays are suppressed to a negligible level with the use of vetoes. In addition, possible single-charm backgrounds are suppressed with requirements on the flight distance of the $D_s^\pm$ mesons. Due to the fact that these backgrounds do not need to be taken into account in the mass model, it is expected that this source of systematic uncertainty is notably smaller.

The third largest source of systematic uncertainty in the previous analysis is due to the parametrisation of the decay-time-dependent efficiency and yields about 2.7% of the statistical uncertainty. It has been evaluated in pseudo-experiments by generating data sets with the decay-time-dependent efficiency model obtained from $B_s^0\to D_s^+D_s^-$ simulation and fitting the generated data sets with the nominal model derived from $B^0\to D_s^+D^-$ data. In the presented analysis, the decay-time-dependent efficiency model is also derived from $B^0\to D_s^+D^-$

data. Therefore a similar systematic uncertainty can be expected.

Other systematic uncertainties are negligible compared to the described ones. With these estimations, it is expected that the result in the presented analysis will still be statistically limited.

6.6 Expected Results

The measurement of *CP* violation in $B_s^0 \to D_s^+ D_s^-$ decays is not complete yet, but in an advanced state. The selection of the signal candidates and the extraction of the blind *CP* observables has been performed. The systematic uncertainties are currently being evaluated. After finalising all studies of systematic uncertainties, the analysis will be thoroughly reviewed by the physics working group and the whole LHCb collaboration. Only once after the review is finished, the result will be published. Although the measurement of the *CP* observables is still blind, the statistical uncertainties are evaluated to be

$$\begin{aligned} \sigma_{\lambda} &= 0.08\,, \\ \sigma_{\phi_s^{\text{eff}}} &= 0.13\,. \end{aligned} \tag{6.31}$$

Sources of systematic uncertainties can be derived from a previous analysis and indicate that the measurement will be statistically limited. The dominant source of systematic uncertainty is expected to be due to the decay-time resolution, which has a high impact on *CP*-violation measurements in B_s^0 decays. However, it is expected that the amount of this uncertainty will be lower than in the previous analysis. In the updated analysis, the determination of the decay-time resolution has been optimised by utilising prompt data instead of signal simulation, which should describe the resolution more accurately. Furthermore, the selection is optimised to suppress various background components, reducing the second largest systematic uncertainty due to the mass parametrisation in the previous analysis. The results are expected to be the most precise measurement of ϕ_s in the $B_s^0 \to D_s^+ D_s^-$ decay channel.

7 Discussion and Outlook

The LHCb experiment provides powerful possibilities for indirect searches of New Physics by performing precision measurements of SM observables. The family of $B \to D\overline{D}$ decays gives access to *CP*-violating observables, like the mixing phase ϕ_s. With increasing experimental precision, higher-order SM contributions will become prominent, but can be constrained by relating measurements of different $B \to D\overline{D}$ decays [21, 22]. Two measurements of this class of decays are presented in this thesis.

The first analysis is a search for the $B_s^0 \to D^{*\pm}D^{\mp}$ decay using data corresponding to an integrated luminosity of $9\,\text{fb}^{-1}$ collected at centre-of-mass energies of 7, 8 and 13 TeV. The $B_s^0 \to D^{*\pm}D^{\mp}$ decay is observed for the first time and its branching fraction is measured to be

$$\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp}) = (8.41 \pm 1.02 \pm 0.12 \pm 0.39 \pm 0.79) \times 10^{-5}\,,$$

where the first uncertainty is statistical, the second systematic, the third is due to the uncertainty of the hadronisation fraction ratio f_s/f_d and the fourth uncertainty is due to the $B^0 \to D^{*\pm}D^{\mp}$ branching fraction. The measurement has been published in Ref. [25]. The result disagrees with a theory prediction from a perturbative QCD approach, $\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp}) = (3.6 \pm 0.6) \times 10^{-3}$ [24], but confirms a prediction that assumes prominent contributions from rescattering, $\mathcal{B}(B_s^0 \to D^{*\pm}D^{\mp}) = (6.1 \pm 3.6) \times 10^{-5}$ [23]. This measurement indicates the importance of rescattering processes for making accurate predictions for other $B \to D\overline{D}$ decays.

The second analysis is a decay-time-dependent measurement of *CP* violation in $B_s^0 \to D_s^+ D_s^-$ decays to measure ϕ_s. It utilises data corresponding to $6\,\text{fb}^{-1}$ collected at 13 TeV. The measurement is currently ongoing. The statistical sensitivities on the *CP* parameters are evaluated to be

$$\begin{aligned} \sigma_\lambda &= 0.08\,, \\ \sigma_{\phi_s^{\text{eff}}} &= 0.13\,. \end{aligned} \tag{7.1}$$

The measurement is an update of a previous measurement of the mixing phase ϕ_s in the same decay channel [16] and will be the most precise measurement in this decay mode. Compared to the previous measurement, the selection is optimised by reducing misidentification background to a negligible level through vetoes and a multivariate classifier. In addition, the decay-time resolution is determined using prompt data instead of signal simulation, which yields a more accurate description and reduces the respective systematic uncertainty. Systematic uncertainties are assumed to originate from similar sources as in the previous analysis. The total systematic uncertainty in the previous analysis is $\sigma_{\text{syst}} = 0.02$ for both *CP* parameters. Due to the optimisations of the selection and decay-time-resolution determination, the systematic uncertainty in the presented analysis is expected to be smaller. As a result, the new measurement will still be statistically limited.

Measurements of *CP* violation in $B_s^0 \to D_s^+ D_s^-$ decays serve as an important cross-check for similar measurements of ϕ_s in B decays to charmonium final states. While the latter decays profit from higher efficiencies, measurements in $B_s^0 \to D_s^+ D_s^-$ decays give important complementary results with different contributions of penguin amplitudes [17]. Furthermore, using additional $B \to D\overline{D}$ decays allows to control the pollution from higher-order contributions and gives access to the phase shift of ϕ_s. This can be done using additional *CP*-violation measurements. In parallel to the $B_s^0 \to D_s^+ D_s^-$ analysis, the author contributed to a *CP*-violation measurement in $B^0 \to D^+ D^-$ decays. The result of this measurement can be transferred to the $B_s^0 \to D_s^+ D_s^-$ analysis by exploiting the U-spin symmetry between the two decays [21, 22]. Additionally, the results of branching-ratio measurements, *e.g.* the presented measurement of $B_s^0 \to D^{*\pm} D^{\mp}$ decays, can be related to other $B \to D\overline{D}$ decays with the goal of quantifying subleading contributions in measurements of *CP* violation in $B \to D\overline{D}$ decays [21, 22]. Additional measurements to complement the understanding of the $B \to D\overline{D}$ family include the branching-ratio measurement of $B_s^0 \to D^{*+} D^{*-}$ decays [111] and the measurement of *CP* violation in $B^0 \to D^{*+} D^{*-}$ decays. The author contributed to these measurements giving advisory input. The selection in the $B_s^0 \to D_s^+ D_s^-$ and the $B^0 \to D^+ D^-$ analysis can be exploited to update previous measurements of branching ratios of the $B_s^0 \to D^+ D^-$ and $B_s^0 \to D_s^+ D^-$ decays [112, 113] and to search for the yet unobserved $B^0 \to D_s^+ D_s^-$ decay [114].

Furthermore, $B \to D\overline{D}$ decays will especially profit from the upgrade of the LHCb detector. The hardware-based L0 trigger will be removed so that the trigger will be fully software based [115]. $B \to D\overline{D}$ decays are inefficient at the L0 trigger and benefit considerably from the removal compared to muon-triggered modes. In addition, a new flavour tagging algorithm is currently developed, which utilises a deep neural network instead of individual taggers [116]. The tagging power is expected to increase significantly leading to an improvement of the sensitivity of the measurement. During Run 3, LHCb will collect data corresponding to an integrated luminosity of $50\,\text{fb}^{-1}$. The expected sensitivity of ϕ_s using a combination of all decay modes available at LHCb [117], will be at the same level as the current SM prediction. At this point, deviations from the SM predictions will become apparent or constraints on possible New Physics contributions can be imposed.

A Additional Material for the $B_s^0 \to D^{*\pm} D^{\mp}$ Analysis

Table A.1: `StrippingB02DstDBeauty2CharmLine`. If the requirements differ, the first refers to Run 1 and the second in parentheses refers to Run 2.

Candidate	Variable	Requirement	Unit
	#longtracks	< 500	
	`Hlt2Topo` or `Hlt2IncPhi`		
$\pi^{\pm}/K^{\pm}$	$\chi^2_{\text{track}}/\text{ndf}$	$< 3 (< 4)$	
	p_{T}	> 100	MeV/c
	p	> 1000	MeV/c
	$\min(\chi^2_{\text{IP,PV}})$	> 4	
	p_{ghost}	< 0.4	
$\pi^{\pm}$	$\text{DLL}_{K\pi}$	< 20	
$K^{\pm}$	$\text{DLL}_{K\pi}$	> -10	
$D^{*\mp}/D^0/D^{\pm}$	$\Sigma_{\text{daughters}} p_{\mathrm{T}}$	> 1800	MeV/c
	$\lvert m_{D^{*\pm}} - m_{D^{*\pm},\text{PDG}} \rvert$	$< 50 (< 600)$	MeV/c^2
	$\lvert m_{D^0} - m_{D^0,\text{PDG}} \rvert$	< 200	MeV/c^2
	$m_{D^{\pm}}$	$\in [1769.62, 2068.49]$	MeV/c^2
	$m_{D^*} - m_{D^0}$	$-(< 200)$	MeV/c^2
	DOCA	< 0.5	mm
	$\chi^2_{\text{vtx}}/\text{ndf}$	< 10	
	$\chi^2_{\text{distance to any PV}}$	> 36	
	$\text{DIRA}_{\text{best PV}}$	> 0	
one daughter hadron	p_{T}	> 500	MeV/c
	p	> 5000	MeV/c
	$\chi^2_{\text{track}}/\text{ndf}$	$< 2.5 (< 4)$	
B^0	$\Sigma_{\text{daughters}} p_{\mathrm{T}}$	> 5000	MeV/c
	m	$\in [4750, 6000]$	MeV/c^2
	$\chi^2_{\text{vtx}}/\text{ndf}$	< 10	
	t	> 0.2	ps
	$\chi^2_{\text{IP,PV}}$	< 25	
	$\text{DIRA}_{\text{best PV}}$	> 0.999	
one hadron in decay chain	p_{T}	> 1700	MeV/c
	p	> 10000	MeV/c
	$\min(\chi^2_{\text{IP to any PV}})$	> 16	
	$\min(\text{IP}_{\text{any PV}})$	> 0.1	mm

Table A.2: Requirements for the D_s^+, ϕ veto and the Λ_c^+ veto and their signal efficiency for the $B_s^0 \to D^{*\pm}D^{\mp}$ and $B^0 \to D^{*\pm}D^{\mp}$ decays. The vetoes are optimised to have a background rejection of 95%. Particles satisfying either requirements are kept. The bold marked hadron gets misidentified as either of the pions of the $D^+ \to K^-\pi^+\pi^+$ decay, which are sorted by their transverse momenta. The variable Δm is the difference of the mass hypothesis and the nominal mass of the D_s^+ or ϕ meson, or Λ_c^+ baryon [34]. The variable $\mathrm{ProbNN}_{h\pi}$ is defined in Eq. (5.5).

Misidentified decay	Requirements		Signal efficiency [%]	
	$\lvert\Delta m\rvert$	$\mathrm{ProbNN}_{h\pi}$	B_s^0	B^0
2011-2012				
$D_s^+ \to K^-\boldsymbol{K}^{\boldsymbol{+}}\pi^+$	$> 22.0\,\mathrm{MeV}/c^2$	< 0.22	91.41 ± 0.10	91.51 ± 0.08
$D_s^+ \to K^-\pi^+\boldsymbol{K}^{\boldsymbol{+}}$	$> 19.0\,\mathrm{MeV}/c^2$	< 0.02		
$\phi\pi^+ \to K^-\boldsymbol{K}^{\boldsymbol{+}}\pi^+$	$> 12.0\,\mathrm{MeV}/c^2$	< 0.03		
$\phi\pi^+ \to K^-\pi^+\boldsymbol{K}^{\boldsymbol{+}}$	$> 16.0\,\mathrm{MeV}/c^2$	< 0.35		
$\Lambda_c^+ \to K^-\boldsymbol{p}\pi^+$	$> 25.0\,\mathrm{MeV}/c^2$	< 0.17	98.87 ± 0.04	98.87 ± 0.03
$\Lambda_c^+ \to K^-\pi^+\boldsymbol{p}$	$> 14.0\,\mathrm{MeV}/c^2$	< 0.06		
2015-2016				
$D_s^+ \to K^-\boldsymbol{K}^{\boldsymbol{+}}\pi^+$	$> 20.0\,\mathrm{MeV}/c^2$	< 0.08	91.83 ± 0.18	91.71 ± 0.12
$D_s^+ \to K^-\pi^+\boldsymbol{K}^{\boldsymbol{+}}$	$> 20.0\,\mathrm{MeV}/c^2$	< 0.01		
$\phi\pi^+ \to K^-\boldsymbol{K}^{\boldsymbol{+}}\pi^+$	$> 16.0\,\mathrm{MeV}/c^2$	< 0.03		
$\phi\pi^+ \to K^-\pi^+\boldsymbol{K}^{\boldsymbol{+}}$	$> 17.0\,\mathrm{MeV}/c^2$	< 0.35		
$\Lambda_c^+ \to K^-\boldsymbol{p}\pi^+$	$> 22.0\,\mathrm{MeV}/c^2$	< 0.13	99.30 ± 0.06	99.30 ± 0.04
$\Lambda_c^+ \to K^-\pi^+\boldsymbol{p}$	$> 13.0\,\mathrm{MeV}/c^2$	< 0.03		
2017-2018				
$D_s^+ \to K^-\boldsymbol{K}^{\boldsymbol{+}}\pi^+$	$> 24.0\,\mathrm{MeV}/c^2$	< 0.10	91.47 ± 0.11	91.66 ± 0.09
$D_s^+ \to K^-\pi^+\boldsymbol{K}^{\boldsymbol{+}}$	$> 24.0\,\mathrm{MeV}/c^2$	< 0.01		
$\phi\pi^+ \to K^-\boldsymbol{K}^{\boldsymbol{+}}\pi^+$	$> 22.0\,\mathrm{MeV}/c^2$	< 0.02		
$\phi\pi^+ \to K^-\pi^+\boldsymbol{K}^{\boldsymbol{+}}$	$> 16.0\,\mathrm{MeV}/c^2$	< 0.30		
$\Lambda_c^+ \to K^-\boldsymbol{p}\pi^+$	$> 26.0\,\mathrm{MeV}/c^2$	< 0.17	99.34 ± 0.03	99.33 ± 0.03
$\Lambda_c^+ \to K^-\pi^+\boldsymbol{p}$	$> 15.0\,\mathrm{MeV}/c^2$	< 0.07		

Table A.3: Requirements for the vetoes against single-charm background and their signal efficiency. Particles satisfying the requirements are kept.

Period	$\chi^2_{D^\pm,\mathrm{FD}}$	Signal efficiency [%]	
		$B_s^0 \to D^{*\pm}D^\mp$	$B^0 \to D^{*\pm}D^\mp$
2011-2012	> 9	87.49 ± 0.12	87.40 ± 0.10
2015-2016	> 6	88.44 ± 0.21	88.28 ± 0.14
2017-2018	> 6	88.43 ± 0.13	88.10 ± 0.11

Table A.4: List of variables used in the training of the BDT. The variables are discussed in Sec. 5.2.3. Their importance in the training is reflected by the order in the list.

variable
$\log(\chi^2_{\mathrm{IP}})$ (B^0)
$m_{D^*} - m_{D^0}$
$\mathrm{ProbNN}_{K\pi}$ ($K^\pm$ from $D^\mp$)
$\log(\chi^2_{\mathrm{FD}})$ $(D^\pm)$
$\log(\chi^2_{\mathrm{DTF\ Fit}})$ (B^0)
$\mathrm{ProbNN}_{K\pi}$ (${\pi^\pm}_1$ from $D^\mp$)
$\log(\chi^2_{\mathrm{FD}})$ (D^0)
$\mathrm{ProbNN}_{K\pi}$ ($K^\pm$ from D^0)
$\mathrm{ProbNN}_{K\pi}$ (slow pion from $D^{*\pm}$)
$t/\sigma_t(D^\mp)$
p_{T} ($K^\pm$ from $D^\mp$)
$\log(\chi^2_{\mathrm{IP}})$ $(D^\pm)$
p_{T} (B^0)

Table A.5: Results of the fit to the invariant $D^{*\pm}D^{\mp}$ mass distribution with simulated $B^0 \to D^{*\pm}D^{\mp}$ samples for 2011–2012, 2015–2016 and 2017–2018. The parameters are discussed in Sec. 5.4.1. The indices of the parameters denote the two Crystal Ball functions. The parameter f is the relative fraction of the two Crystal Ball functions.

Parameter	Value 2011–2012	 2015–2016	 2017–2018
μ [MeV/c^2]	5279.70 ± 0.07	5280.00 ± 0.10	5279.90 ± 0.08
σ_1 [MeV/c^2]	8.64 ± 0.31	8.31 ± 0.48	8.43 ± 0.24
σ_2 [MeV/c^2]	8.42 ± 0.13	8.64 ± 0.17	8.36 ± 0.17
α_1	0.90 ± 0.06	0.77 ± 0.07	0.97 ± 0.05
α_2	-1.42 ± 0.04	-1.36 ± 0.05	-1.29 ± 0.05
f	0.34 ± 0.04	0.32 ± 0.04	0.45 ± 0.05
n	10	10	10

Table A.6: Fit results of the mass fit to simulated $B^0 \to D^{*\pm}D^{*\mp}$ and $B_s^0 \to D^{*\pm}D^{*\mp}$ samples generated with pure longitudinal and transverse polarisations. The parameters are discussed in Sec. 5.4.1. The indices of the parameters denote the Gaussian of the model in Eq. (5.15) and Eq. (5.16) they correspond to.

	2012		2016	
Parameter	Longitudinal	Transverse	Longitudinal	Transverse
	$B^0 \to D^{*\pm}D^{*\mp}$			
μ_1 [MeV/c^2]	5094.9 ± 2.6	–	5095.0 ± 1.2	–
μ_2 [MeV/c^2]	$\equiv \mu_1$	–	$\equiv \mu_1$	–
μ_3 [MeV/c^2]	5067.4 ± 0.3	5083.1 ± 0.6	5067.7 ± 0.2	5081.8 ± 0.3
μ_4 [MeV/c^2]	5120.7 ± 0.3	5106.3 ± 0.7	5120.7 ± 0.2	5105.7 ± 0.3
σ_1 [MeV/c^2]	48.26 ± 3.5	10.1 ± 1.9	45.4 ± 1.6	9.9 ± 1.1
σ_2 [MeV/c^2]	19.0 ± 2.5	–	19.49 ± 1.6	–
σ_3 [MeV/c^2]	9.65 ± 0.18	13.0 ± 0.35	9.68 ± 0.11	13.14 ± 0.17
σ_4 [MeV/c^2]	$\equiv \sigma_3$	$\equiv \sigma_3$	$\equiv \sigma_3$	$\equiv \sigma_3$
g_1 [%]	08.7 ± 1.3	0.061 ± 0.005	10.7 ± 0.9	5.7 ± 0.2
g_2 [%]	10.0 ± 4.3	–	10.1 ± 4.0	–
g_3 [%]	44.0 ± 1.1	0.508 ± 0.022	46.6 ± 0.5	46.3 ± 1.1
	$B_s^0 \to D^{*\pm}D^{*\mp}$			
μ_1 [MeV/c^2]	5177.6 ± 1.8	–	5178.4 ± 1.0	–
μ_2 [MeV/c^2]	$\equiv \mu_1$	–	$\equiv \mu_{\mathrm{MC}}1$	–
μ_3 [MeV/c^2]	5150.9 ± 0.4	5165.5 ± 0.5	5151.0 ± 0.2	5165.7 ± 0.3
μ_4 [MeV/c^2]	5205.1 ± 0.3	5191.0 ± 0.5	5205.3 ± 0.2	5190.6 ± 0.3
σ_1 [MeV/c^2]	44.1 ± 2.4	15.8 ± 2.3	53.1 ± 2.4	18.8 ± 1.3
σ_2 [MeV/c^2]	25.9 ± 2.9	–	26.0 ± 1.6	–
σ_3 [MeV/c^2]	10.06 ± 0.23	12.9 ± 0.28	9.99 ± 0.14	13.44 ± 0.17
σ_4 [MeV/c^2]	$\equiv \sigma_3$	$\equiv \sigma_3$	$\equiv \sigma_3$	$\equiv \sigma_3$
g_1 [%]	11.3 ± 2.1	0.054 ± 0.004	07.5 ± 1.3	6.0 ± 0.2
g_2 [%]	10.0 ± 12.4	–	15.5 ± 2.2	–
g_3 [%]	42.6 ± 0.9	0.480 ± 0.017	45.7 ± 0.5	47.4 ± 1.0

Table A.7: Results of the floating parameters from the fit to the invariant $D^{*\pm}D^{\mp}$ mass distribution from recorded data. The parameters are discussed in Sec. 5.4.1. The B_s^0 mean is shifted by the fixed value of 87.26 MeV/c^2 [34] to the B^0 mean.

Parameter	2011-2012	2015–2016	2017–2018
μ [MeV/c^2]	5280.4 ± 0.5	5279.8 ± 0.4	5279.9 ± 0.3
s	1.05 ± 0.05	1.04 ± 0.04	0.95 ± 0.03
λ [c^2/ MeV]	-0.008 ± 0.001	-0.008 ± 0.001	-0.007 ± 0.001
$g_{B^0 \to D^{*\pm}D^{*\mp}}$	0.59 ± 0.07	0.47 ± 0.06	0.60 ± 0.04
$g_{B_s^0 \to D^{*\pm}D^{*\mp}}$	0.57 ± 0.19	0.66 ± 0.14	0.45 ± 0.15
N_{B^0}	466 ± 22	780 ± 29	1263 ± 36
$N_{B_s^0}$	12 ± 4	34 ± 7	49 ± 8
$N_{B^0 \to D^{*\pm}D^{*\mp}}$	257 ± 20	475 ± 26	682 ± 31
$N_{B_s^0 \to D^{*\pm}D^{*\mp}}$	51 ± 10	88 ± 14	101 ± 15
N_{Comb}	134 ± 21	216 ± 28	267 ± 31

B Additional Material for the $B_s^0 \to D_s^+ D_s^-$ Analysis

Table B.1: `StrippingB02DDBeauty2CharmLine`. If the requirements differ, the first refers to 2015–2016 and the second in parentheses refers to 2017–2018.

Candidate		Variable	Requirement	Unit
		#longtracks	< 500	
		`Hlt2Topo` or `Hlt2IncPhi`		
$\pi^\pm/K^\pm$		$\chi^2_{\text{track}}/\text{ndf}$	< 4	
		p_{T}	> 100	MeV/c
		p	> 1000	MeV/c
		$\min(\chi^2_{\text{IP,PV}})$	> 4	
		p_{ghost}	< 0.4	
$\pi^\pm$		$\text{DLL}_{K\pi}$	< 20	
$K^\pm$		$\text{DLL}_{K\pi}$	> -10	
$D^\pm$		$\Sigma_{\text{daughters}} p_{\text{T}}$	> 1800	MeV/c
		$m_{D^\pm}$	$\in [1769.62, 2068.49]$	MeV/c^2
		DOCA	< 0.5	mm
		$\chi^2_{\text{vtx}}/\text{ndf}$	< 10	
		$\chi^2_{\text{distance to any PV}}$	> 36	
		$\text{DIRA}_{\text{best PV}}$	> 0	
	one daughter hadron	p_{T}	> 500	MeV/c
		p	> 5000	MeV/c
		$\chi^2_{\text{track}}/\text{ndf}$	< 4	
B^0		$\Sigma_{\text{daughters}} p_{\text{T}}$	> 5000	MeV/c
		m	$> 4900 (> 4950)$	MeV/c^2
		m	< 6000	MeV/c^2
		$\chi^2_{\text{vtx}}/\text{ndf}$	< 10	
		t	> 0.2	ps
		$\chi^2_{\text{IP,PV}}$	< 25	
		$\text{DIRA}_{\text{best PV}}$	> 0.999	
		$z(\text{D}_{1,2}) - z(B^0)$	> -1.5	mm
	one hadron in decay chain	p_{T}	> 1700	MeV/c
		p	> 10000	MeV/c
		$\min(\chi^2_{\text{IP to any PV}})$	> 16	
		$\min(\text{IP}_{\text{any PV}})$	> 0.1	mm

Table B.2: Requirements for the cut-based vetoes. Particles satisfying either the mass or the ProbNN requirement are kept. The vetoes are applied on both $D_{1,2}$ mesons. If the requirements differ between D_1 and D_2, the cut for D_1 is given and the cut for D_2 is noted in parentheses. The variable Δm is the difference of the mass hypothesis and the nominal mass of the mother particle of the background decay taken from [34]. The variable $\text{ProbNN}_{p_1 p_2}$ is defined in Eq. (6.1).

Final state $h_0 h_1 h_2$	Background	Mass window $[\text{MeV}/c^2]$	PID requirement
$K^\mp K^\pm \pi^\pm$	$D_s^\pm \to K^\mp \pi^\pm K^\pm$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{\pi,K}(h_1) < 0.1$
			$\text{ProbNN}_{\pi,K}(h_2) > 0.9$
	$\phi \to K^\mp K^\pm{}_2$	$\lvert\Delta m\rvert > 10$	$\text{ProbNN}_{\pi,K}(h_2) > 0.98$
	$D^0 \to K^\mp K^\pm{}_1$	$\lvert\Delta m\rvert > 30$	-
	$\Lambda_c^+ \to K^\mp \pi^\pm p$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{\pi,K}(h_1) < 0.1$
			$\text{ProbNN}_{\pi,p}(h_2) > 0.9$
	$\Lambda_c^+ \to K^\mp p \pi^\pm$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{K,p}(h_1) > 0.7$
$\pi^\mp K^\pm \pi^\pm$	$D^\pm \to K^\mp \pi^\pm K^\pm$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{\pi,K}(h_0) > 0.9$
			$\text{ProbNN}_{\pi,K}(h_1) < 0.1$
			$\text{ProbNN}_{\pi,K}(h_2) > 0.9$
	$D^\pm \to K^\mp \pi^\pm \pi^\pm$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{\pi,K}(h_0) > 0.9(0.95)$
			$\text{ProbNN}_{\pi,K}(h_1) < 0.1$
	$D^\pm \to \pi^\mp \pi^\pm \pi^\pm$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{\pi,K}(h_1) < 0.3$
	$D^\pm \to \pi^\mp \pi^\pm K^\pm$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{\pi,K}(h_1) < 0.2$
			$\text{ProbNN}_{\pi,K}(h_2) > 0.8(0.9)$
	$\phi \to K^\mp K^\pm{}_1$	$\lvert\Delta m\rvert > 10$	$\text{ProbNN}_{\pi,K}(h_0) > 0.8(0.9)$
	$\phi \to K^\mp K^\pm{}_2$	$\lvert\Delta m\rvert > 10$	$\text{ProbNN}_{\pi,K}(h_0) > 0.95$
			$\text{ProbNN}_{\pi,K}(h_2) > 0.95$
	$D^0 \to K^\mp K^\pm{}_1$	$\lvert\Delta m\rvert > 30$	$\text{ProbNN}_{\pi,K}(h_0) > 0.8$
	$D^0 \to K^\mp \pi^\pm{}_1$	$\lvert\Delta m\rvert > 30$	$\text{ProbNN}_{\pi,K}(h_0) > 0.9$
			$\text{ProbNN}_{\pi,K}(h_1) < 0.1$
	$D^0 \to \pi^\mp K^\pm{}_1$	$\lvert\Delta m\rvert > 30$	-
	$\Lambda_c^+ \to \pi^\mp p \pi^\pm$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{K,p}(h_1) > 0.5(0.6)$
$\pi^\mp \pi^\pm \pi^\pm$	$D^0 \to K^\mp \pi^\pm{}_1$	$\lvert\Delta m\rvert > 30$	$\text{ProbNN}_{\pi,K}(h_0) > 0.95$
	$D^0 \to K^\mp \pi^\pm{}_2$	$\lvert\Delta m\rvert > 30$	$\text{ProbNN}_{\pi,K}(h_0) > 0.9$
	$D^0 \to \pi^\mp K^\pm{}_1$	$\lvert\Delta m\rvert > 30$	$\text{ProbNN}_{\pi,K}(h_1) > 0.9$
	$D^0 \to \pi^\mp \pi^\pm{}_1$	$\lvert\Delta m\rvert > 30$	-
	$D^0 \to \pi^\mp \pi^\pm{}_2$	$\lvert\Delta m\rvert > 30$	-
	$\Lambda_c^+ \to \pi^\mp p \pi^\pm$	$\lvert\Delta m\rvert > 25$	$\text{ProbNN}_{\pi,p}(h_1) > 0.3(0.2)$

Table B.3: Additional cut-based vetoes applied for the $B^0 \to D_s^+ D^-$ control channel. Particles satisfying either the mass or the ProbNN requirement are kept. The variable Δm is the difference of the mass hypothesis and the nominal mass of the mother particle of the background decay taken from [34]. The variable $\mathrm{ProbNN}_{p_1 p_2}$ is defined in Eq. (6.1).

Final state $h_0 h_1 h_2$	Background	Mass window [MeV/c^2]	PID requirement
$K^{\mp}\pi^{\pm}\pi^{\pm}$	$\phi \to K^{\mp}K^{\pm}{}_1$	$\lvert\Delta m\rvert > 10$	-
	$\phi \to K^{\mp}K^{\pm}{}_2$	$\lvert\Delta m\rvert > 10$	$\mathrm{ProbNN}_{\pi,K}(h_2) > 0.95$
	$D^0 \to K^{\mp}K^{\pm}{}_1$	$\lvert\Delta m\rvert > 30$	$\mathrm{ProbNN}_{\pi,K}(h_1) > 0.8$
	$\Lambda_c^+ \to K^{\mp}p\pi^{\pm}$	$\lvert\Delta m\rvert > 25$	$\mathrm{ProbNN}_{\pi,p}(h_1) > 0.8$

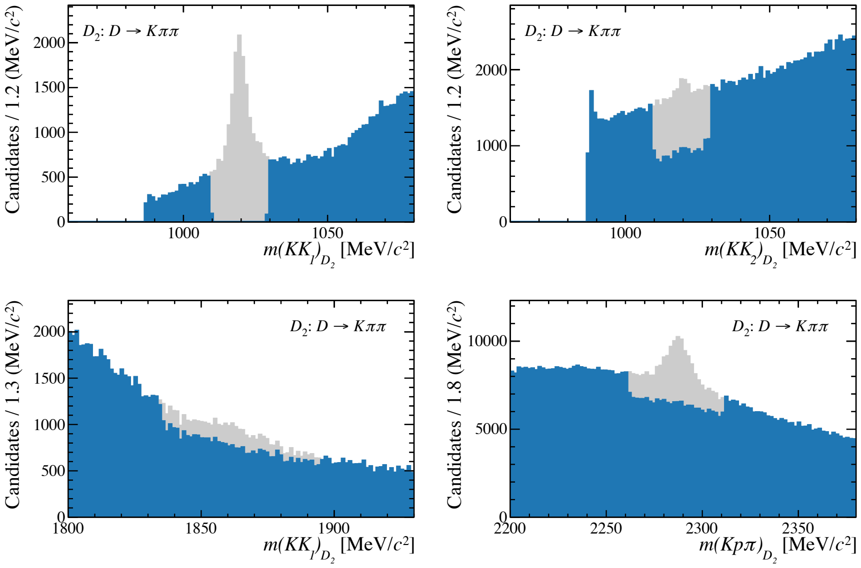

Figure B.1: Invariant mass distributions of the $D_2 \to K\pi\pi$ final state of the $B^0 \to D_s^+ D^-$ decay recalculated as (top left) KK_1, (top right) KK_2, (bottom left) KK_1 and (bottom right) $Kp\pi$. Background contributions of (top) ϕ decays, (bottom left) D^0 decays and (bottom right) Λ_c^+ decays are visible. Candidates in the blue area are kept while candidates in the grey area get rejected by the veto requirement. The vetoes are reported in Table B.3.

Table B.4: Input variables for the veto BDT. The variables are discussed in Sec. 6.1.3.

variable
FD $(D^{\pm}_{(s)})$
ProbNN$_K$ $(p_{(1,2)})$
ProbNN$_\pi$ $(p_{(1,2)})$
$m(K_0\pi_1\pi_2)$
$m(K_0K_1\pi_2)$
$m(K_0\pi_1K_2)$
$m(K_0K_{(1,2)})$
$m(K_0\pi_{(1,2)})$

Table B.5: Input variables for the BDT used to reject combinatorial background. The variables are discussed in Sec. 6.1.4.

Particle	Variable
B_s^0	χ^2_{DTF}
	χ^2_{IP}
$D_s^{\pm}$	$\lvert m_{D_s^{\pm}} - m_{D_s^{\pm},\mathrm{PDG}}\rvert$
	p_{T}
	χ^2_{IP}
	FD
$K^{\pm}/\pi^{\pm}$	p_{T}
	$\cos\left(\theta_{D_s^{\pm},K/\pi}\right)$

Table B.6: Results of the floating parameters from the extended maximum likelihood fit to the invariant $D_s^+ D_s^-$ mass distribution from recorded data for the final states $KK\pi\ KK\pi$, $KK\pi\ \pi\pi\pi$ and $KK\pi\ \pi K\pi$. The parameters are discussed in Sec. 6.1.6.

parameter	$KK\pi\ KK\pi$	$KK\pi\ \pi\pi\pi$	$KK\pi\ KK\pi$
μ [MeV/c^2]	5365.80 ± 0.12	5366.0 ± 0.20	5365.1 ± 0.29
σ [MeV/c^2]	10.30 ± 0.14	10.92 ± 0.24	10.35 ± 0.38
$N_{B_s^0}$	7696 ± 97	3781 ± 75	1800 ± 57
N_{bkg}	2507 ± 66	2595 ± 66	2560 ± 63
λ_{exp} [c^2/ MeV]	−0.0069 ± 0.0003	−0.0045 ± 0.0003	−0.0050 ± 0.0003

Bibliography

[1] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[2] A. Salam, *Weak and electromagnetic interactions*, Conf. Proc. **C680519** (1968) 367.

[3] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.

[4] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.

[5] F. Englert and R. Brout, *Broken symmetry and the masses of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.

[6] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125* GeV *with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, `arXiv:1207.7235`.

[7] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, `arXiv:1207.7214`.

[8] SNO collaboration, Q. R. Ahmad *et al.*, *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301, `arXiv:nucl-ex/0204008`.

[9] Super-Kamiokande collaboration, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, `arXiv:hep-ex/9807003`.

[10] Planck collaboration, N. Aghanim *et al.*, *Planck 2018 results. VI. Cosmological parameters*, Astron. Astrophys. **641** (2020) A6, Erratum ibid. **652** (2021) C4, `arXiv:1807.06209`.

[11] A. D. Sakharov, *Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32.

[12] ATLAS collaboration, G. Aad *et al.*, *Measurement of the CP-violating phase* ϕ_s *in* $B_s^0 \to J/\psi\,\phi$ *decays in ATLAS at 13* TeV, Eur. Phys. J. **C81** (2021), no. 4 342, `arXiv:2001.07115`.

[13] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of the CP-violating phase* ϕ_s *in the* $B_s^0 \to J/\psi\,\phi(1020) \to \mu^+\mu^- K^+K^-$ *channel in proton-proton collisions at* $\sqrt{s} = 13\,\text{TeV}$, Phys. Lett. B **816** (2021) 136188, `arXiv:2007.02434`.

[14] LHCb collaboration, R. Aaij *et al.*, *Updated measurement of time-dependent CP-violating observables in* $B_s^0 \to J/\psi\, K^+K^-$ *decays*, Eur. Phys. J. **C79** (2019) 706, Erratum ibid. **C80** (2020) 601, `arXiv:1906.08356`.

[15] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP-violating phase* ϕ_s *from* $B_s^0 \to J/\psi\,\pi^+\pi^-$ *decays in 13* TeV *pp collisions*, Phys. Lett. **B797** (2019) 134789, `arXiv:1903.05530`.

[16] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP-violating phase* ϕ_s *in* $\bar{B}_s^0 \to D_s^+ D_s^-$ *decays*, Phys. Rev. Lett. **113** (2014) 211801, `arXiv:1409.4619`.

[17] R. Fleischer, *Exploring CP violation and penguin effects through* $B_d^0 \to D^+D^-$ *and* $B_s^0 \to D_s^+D_s^-$, Eur. Phys. J. C **51** (2007) 849, `arXiv:0705.4421`.

[18] R. Fleischer, *Extracting* γ *from* $B_{s(d)} \to J/\psi\, K_S$ *and* $B_{d(s)} \to D^+_{d(s)} D^-_{d(s)}$, Eur. Phys. J. **C10** (1999) 299, `arXiv:hep-ph/9903455`.

[19] A. Datta and D. London, *Extracting* γ *from* $B_d^0(t) \to D^{(*)+}D^{(*)-}$ *and* $B_d^0 \to D^{(*)+}D^{(*)-}$ *decays*, Phys. Lett. **B584** (2004) 81, `arXiv:hep-ph/0310252`.

[20] J. Albert, A. Datta, and D. London, *A measurement of* γ *from the decays* $B_d^0(t) \to D^{(*)+}D^{(*)-}$ *and* $B_d^0 \to D^{(*)+}D^{(*)-}$, Phys. Lett. **B605** (2005) 335, `arXiv:hep-ph/0410015`.

[21] M. Jung and S. Schacht, *Standard model predictions and new physics sensitivity in* $B \to DD$ *decays*, Phys. Rev. **D91** (2015) 034027, `arXiv:1410.8396`.

[22] L. Bel *et al.*, *Anatomy of* $B \to D\bar{D}$ *decays*, JHEP **07** (2015) 108, `arXiv:1505.01361`.

[23] M. Gronau, D. London, and J. L. Rosner, *Rescattering contributions to rare b-meson decays*, Phys. Rev. **D87** (2013), no. 3 036008, `arXiv:1211.5785`.

[24] Y. Li, C.-D. Lu, and Z.-J. Xiao, *Rare decays* $B^0 \to D_s^{(*)+}D_s^{(*)-}$ *and* $B^0 \to D^{(*)+}D^{(*)-}$ *in perturbative QCD approach*, J. Phys. **G31** (2005) 273, `arXiv:hep-ph/0308243`.

[25] LHCb collaboration, R. Aaij *et al.*, *Observation of the* $B_s^0 \to D^{*\pm}D^{\mp}$ *decay*, JHEP **03** (2021) 099, `arXiv:2012.11341`.

[26] D. J. Griffiths, *Introduction to elementary particles*, Wiley, 2008.

[27] D. H. Perkins, *Introduction to High Energy Physics*, Cambridge University Press, 2000.

[28] K. G. Wilson, *Confinement of quarks*, Phys. Rev. D **10** (1974) 2445.

[29] Belle, S. K. Choi *et al.*, *Observation of a narrow charmonium-like state in exclusive* $B^\pm \to K^\pm\pi^+\pi^- J/\psi$ *decays*, Phys. Rev. Lett. **91** (2003) 262001, `arXiv:hep-ex/0309032`.

[30] LHCb collaboration, R. Aaij *et al.*, *Observation of exotic* $J/\psi\,\phi$ *structures from amplitude analysis of* $B^+ \to J/\psi\,\phi K^+$ *decays*, Phys. Rev. Lett. **118** (2017) 022003, `arXiv:1606.07895`.

[31] LHCb collaboration, R. Aaij *et al.*, *Amplitude analysis of* $B^+ \to J/\psi\,\phi K^+$ *decays*, Phys. Rev. **D95** (2017) 012002, `arXiv:1606.07898`.

[32] LHCb collaboration, R. Aaij *et al.*, *Observation of structure in the* J/ψ*-pair mass spectrum*, Science Bulletin **65** (2020), no. 23 1983, `arXiv:2006.16957`.

[33] LHCb collaboration, R. Aaij *et al.*, *Observation of* $J/\psi p$ *resonances consistent with pentaquark states in* $\Lambda_b^0 \to J/\psi p K^-$ *decays*, Phys. Rev. Lett. **115** (2015) 072001, `arXiv:1507.03414`.

[34] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01.

[35] G. C. Branco, L. Lavoura, and J. P. Silva, *CP violation*, Clarendon Press, 1999.

[36] I. I. Bigi and A. I. Sanda, *CP violation*, Cambridge University Press, 2000.

[37] L. L. Chau and W. Y. Keung, *Comments on the parametrization of the Kobayashi-Maskawa matrix. The parametrization of the Kobayashi-Maskawa matrix*, Phys. Rev. Lett. **53** (1984) 1802.

[38] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[39] C. Jarlskog, *Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Violation*, Phys. Rev. Lett. **55** (1985) 1039.

[40] CKMfitter group, J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. **C41** (2005) 1, `arXiv:hep-ph/0406184`, updated results and plots available at `http://ckmfitter.in2p3.fr/`.

[41] K. Anikeev *et al.*, *B physics at the Tevatron: Run II and beyond*, `arXiv:hep-ph/0201071`.

[42] G. Lüders, *On the equivalence of invariance under time reversal and under particle-antiparticle conjugation for relativistic field theories*, Dan. Mat. Fys. Medd. **28** (1954) 1.

[43] W. Pauli and L. F. Curtiss, *Niels bohr and the development of physics*, Am. J. Phys. **24** (1956) 292.

[44] O. W. Greenberg, *CPT violation implies violation of Lorentz invariance*, Phys. Rev. Lett. **89** (2002) 231602, `arXiv:hep-ph/0201258`.

[45] T. Inami and C. S. Lim, *Effects of superheavy quarks and leptons in low-energy weak processes* $K_l \to \mu\overline{\mu}$*,* $K^+ \to \pi^+\nu\overline{\nu}$ *and* $K^0 \leftrightarrow \overline{K}^0$, Prog. Theor. Phys. **65** (1981) 297, Erratum ibid. **65** (1981) 1772.

[46] C. S. Wu *et al.*, *Experimental test of parity conservation in beta decay*, Phys. Rev. **105** (1957) 1413.

[47] 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.

[48] BaBar collaboration, B. Aubert *et al.*, *Observation of CP violation in the* B^0 *meson system*, Phys. Rev. Lett. **87** (2001) 091801, `arXiv:hep-ex/0107013`.

[49] LHCb, R. Aaij *et al.*, *Observation of CP Violation in Charm Decays*, Phys. Rev. Lett. **122** (2019), no. 21 211803, `arXiv:1903.08726`.

[50] KTeV, A. Alavi-Harati *et al.*, *Observation of direct CP violation in* $K_{S,L} \to \pi\pi$ *decays*, Phys. Rev. Lett. **83** (1999) 22, `arXiv:hep-ex/9905060`.

[51] NA48, V. Fanti *et al.*, *A New measurement of direct CP violation in two pion decays of the neutral kaon*, Phys. Lett. B **465** (1999) 335, `arXiv:hep-ex/9909022`.

[52] LHCb, R. Aaij *et al.*, *First observation of* CP *violation in the decays of* B_s^0 *mesons*, Phys. Rev. Lett. **110** (2013), no. 22 221601, `arXiv:1304.6173`.

[53] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and* τ*-lepton properties as of 2018*, Eur. Phys. J. **C81**

(2021) 226, arXiv:1909.12524, updated results and plots available at https://hflav.web.cern.ch.

[54] BaBar collaboration, B. Aubert *et al.*, *Measurement of branching fractions and CP-violating charge asymmetries for B meson decays to* $D^{(*)}\bar{D}^{(*)}$*, and implications for the CKM angle* γ, Phys. Rev. **D73** (2006) 112004, arXiv:hep-ex/0604037.

[55] BaBar collaboration, B. Aubert *et al.*, *Measurements of time-dependent CP asymmetries in* $B^0 \to D^{(*)+}D^{(*)-}$ *decays*, Phys. Rev. **D79** (2009) 032002, arXiv:0808.1866.

[56] Belle collaboration, S. Fratina *et al.*, *Evidence for CP violation in* $B^0 \to D^+D^-$ *decays*, Phys. Rev. Lett. **98** (2007) 221802, arXiv:hep-ex/0702031.

[57] Belle collaboration, M. Röhrken *et al.*, *Measurements of branching fractions and time-dependent CP violating asymmetries in* $B^0 \to D^{(*)\pm}D^{\mp}$ *decays*, Phys. Rev. **D85** (2012) 091106, arXiv:1203.6647.

[58] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation in* $B^0 \to D^+D^-$ *decays*, submitted to PRL (2016) arXiv:1608.06620.

[59] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation in* $B^0 \to D^{*\pm}D^{\mp}$ *decays*, JHEP **03** (2020) 147, arXiv:1912.03723.

[60] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP-violating phase* ϕ_s *in* $\bar{B}^0_s \to D^+_s D^-_s$ *decays*, Phys. Rev. Lett. **113** (2014), no. 21 211801, arXiv:1409.4619.

[61] L. Evans and P. Bryant, *LHC machine*, JINST **3** (2008) S08001.

[62] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[63] ATLAS collaboration, G. Aad *et al.*, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[64] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[65] CMS collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[66] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, https://cds.cern.ch/record/2684277, 2019.

[67] LHCb collaboration, C. Elsässer, $\bar{b}b$ *production angle plots*, .

[68] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, `arXiv:1405.7808`.

[69] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2013) 2431, `arXiv:1211.6759`.

[70] LHCb collaboration, R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, `arXiv:1211.3055`.

[71] LHCb HLT Project, J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, J. Phys. Conf. Ser. **513** (2014) 012001, `arXiv:1310.8544`.

[72] LHCb, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019), no. 04 P04013, `arXiv:1812.10790`.

[73] *The* GAUDI *Project*, `http://proj-gaudi.web.cern.ch/proj-gaudi/`, August, 2016.

[74] R. Aaij *et al.*, *The* MOORE *project*, `http://lhcbdoc.web.cern.ch/lhcbdoc/moore/`.

[75] R. Aaij *et al.*, *The* BRUNEL *project*, `http://lhcbdoc.web.cern.ch/lhcbdoc/brunel/`.

[76] R. Aaij *et al.*, *The* DAVINCI *project*, `http://lhcbdoc.web.cern.ch/lhcbdoc/davinci/`.

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

[78] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[79] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`; T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, `arXiv:hep-ph/0603175`.

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

[81] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[82] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, `arXiv:hep-ph/0506026`.

[83] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270; Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[84] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[85] R. Aaij *et al.*, *The* BOOLE *project*, `http://lhcbdoc.web.cern.ch/lhcbdoc/boole/`.

[86] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, 1984.

[87] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[88] 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, p. 785–794, Association for Computing Machinery, 2016. doi: 10.1145/2939672.2939785.

[89] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*, Teubner, Stuttgart, 1998.

[90] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81.

[91] F. James and M. Roos, *Minuit - a system for function minimization and analysis of the parameter errors and correlations*, Computer Physics Communications **10** (1975), no. 6 343.

[92] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, `arXiv:physics/0402083`.

[93] LHCb collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP **04** (2013) 001, `arXiv:1301.5286`, f_s/f_d value updated in LHCb-CONF-2013-011.

[94] LHCb collaboration, R. Aaij *et al.*, *Measurement of b-hadron fractions in 13* TeV *pp collisions*, `arXiv:1902.06794`, submitted to Phys. Rev. Lett.

[95] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*, JINST **10** (2015), no. 02 P02011, `arXiv:1411.5528`.

[96] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, `arXiv:physics/0703039`.

[97] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, `arXiv:physics/0308063`.

[98] K. De Bruyn *et al.*, *Branching ratio measurements of B_s decays*, Phys. Rev. D **86** (2012) 014027.

[99] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[100] LHCb, R. Aaij *et al.*, *Observation of the $B_s^0 \to D^{*\pm}D^{\mp}$ decay*, JHEP **03** (2021) 099, `arXiv:2012.11341`.

[101] P. Gandini, *Measurement of the B_s^0 meson lifetime in the flavour specific hadronic decay $B_s^0 \to D_s^-\pi^+$*, LHCb-ANA-2014-015.

[102] J. L. Rosner, *Determination of pseudoscalar-charmed-meson decay constants from b-meson decays*, Phys. Rev. D **42** (1990) 3732.

[103] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9** (1938) 60.

[104] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. **A764** (2014) 150, `arXiv:1312.5000`.

[105] P. Ibis, *Measurements of CP violation in $B \to D\overline{D}$ decays at the LHCb experiment*, to be published, Phd thesis, Technische Universität Dortmund, 2022.

[106] A. Kordt, *Flavour Tagging Studien für eine Messung von CP-Verletzung im Zerfallskanal $B_s^0 \to D_s^+D_s^-$ mit Daten des LHCb Experiments*, B. Sc. thesis, Technische Universität Dortmund, 2020.

[107] Y. Xie, *Principles to optimize event selections for measurements of CP asymmetries*, tech. rep., CERN, Geneva, Jul, 2009.

[108] LHCb, R. Aaij *et al.*, *First study of the CP -violating phase and decay-width difference in $B_s^0 \to \psi(2S)\phi$ decays*, Phys. Lett. B **762** (2016) 253, `arXiv:1608.04855`.

[109] J. Wimberly *et al.*, *EspressoPerformanceMonitor*, `http://gitlab.cern.ch/lhcb-ft/EspressoPerformanceMonitor`.

[110] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, J. Phys. Conf. Ser. **762** (2016), no. 1 012036, `arXiv:1608.05806`.

[111] J. Langer, P. Ibis, and A. Mödden, *Search for the decay* $B_s^0 \to D_s^{*+} D_s^{*-}$, LHCb-PUB-2019-010.

[112] LHCb collaboration, R. Aaij *et al.*, *First observations of* $\overline{B}_s^0 \to D^+ D^-, D_s^+ D^-$ *and* $D^0 \overline{D}^0$ *decays*, Phys. Rev. **D87** (2013) 092007, `arXiv:1302.5854`.

[113] LHCb collaboration, R. Aaij *et al.*, *Study of beauty hadron decays into pairs of charm hadrons*, Phys. Rev. Lett. **112** (2014) 202001, `arXiv:1403.3606`.

[114] P. Alfsmann, *Suche nach dem Zerfall* $B^0 \to D_s^+ D_s^-$ *mit Daten des LHCb-Experiments*, B. Sc. thesis, Technische Universität Dortmund, 2021.

[115] R. Aaij *et al.*, *Allen: A high level trigger on GPUs for LHCb*, Comput. Softw. Big Sci. **4** (2020), no. 1 7, `arXiv:1912.09161`.

[116] T. Likhomanenko, D. Derkach, and A. Rogozhnikov, *Inclusive flavour tagging algorithm*, J. Phys. Conf. Ser. **762** (2016), no. 1 012045, `arXiv:1705.08707`.

[117] LHCb collaboration, *Physics case for an LHCb Upgrade II — Opportunities in flavour physics, and beyond, in the HL-LHC era*, `arXiv:1808.08865`.

Danksagung

An dieser Stelle mich ich mich bei einigen Menschen bedanken, ohne die es nicht möglich gewesen wäre diese Arbeit zu schreiben.

An erster Stelle möchte ich ein paar dankende Worte Herrn Prof. Dr. Bernhard Spaan widmen. Er hat mich seit der Bachelorarbeit begleitet und mir die Promotion an seinem Lehrstuhl ermöglicht. Dadurch konnte ich in einem spannenden und internationalen Umfeld arbeiten und mich weiterentwickeln. Während dieser wertvollen Zeit hat er mich stets unterstützt und beraten, aber dabei immer den großen Freiraum für die wissenschaftliche Arbeit ermöglicht. Dass er nicht mehr da ist, ist ein großer Verlust für den Lehrstuhl.

Besonderer Dank gilt Prof. Dr. Johannes Albrecht dafür, dass er sich bereiterklärt hat als Erstgutachter einzuspringen und die Betreuung meiner Promotion in den letzten Monaten übernommen hat.

Außerdem danke ich Herrn Prof. Dr. Kevin Kröninger dafür, dass er sich als Zweitgutachter zur Verfügung gestellt hat.

Ich möchte mich herzlich bei der Dortmunder $B \to D\overline{D}$ Gruppe für die sehr gute und erfolgreiche Zusammenarbeit und hilfreichen Diskussionen bedanken. Es war mir immer eine Freude mit euch zusammen zu arbeiten: Jan, Louis, Philipp und Sophie. Desweiteren möchte ich mich bei allen bedanken, die Teile der Arbeit Korrektur gelesen haben. Eure Kommentare waren eine große Hilfe beim Vervollständigen dieser Arbeit: Vukan, Martin, Quentin, Jan, Louis und besonders Philipp und Sophie.

Ein großer Dank geht an den gesamten Lehrstuhl E5 für das harmonische Lehrstuhlleben und die Zusammenarbeit, die zahlreichen Kaffeepausen, und die vielen Aktivitäten außerhalb des Unialltags, die auch teilweise während der Pandemie virtuell fortgeführt werden konnten. Ebenfalls danken möchte ich dem Admin-Team für das pflegen der lokalen E5 Infrastruktur, sowie Britta für das erledigen der bürokratischen Dinge. Explizit erwähnen möchte ich auch all die Bürokollegen: Kevin, Quentin, Louis und Jan E. Danke für die angenehme Arbeitsatmosphäre und zahlreichen Gespräche im Büro.

Zu guter Letzt gilt meiner Familie und meinen Freunden besonderer Dank, die mich während meines Studiums in jeglicher Art unterstützt und wenn nötig abgelenkt haben. Vom Herzen möchte ich Philipp danken. Danke für deinen Rückhalt und deine Motivation vor allem in stressigen Zeiten und dass du für mich da bist.