

# Strange oscillating beauty-mesons

Measurement of the oscillation frequency  $\Delta m_s$  at LHCb

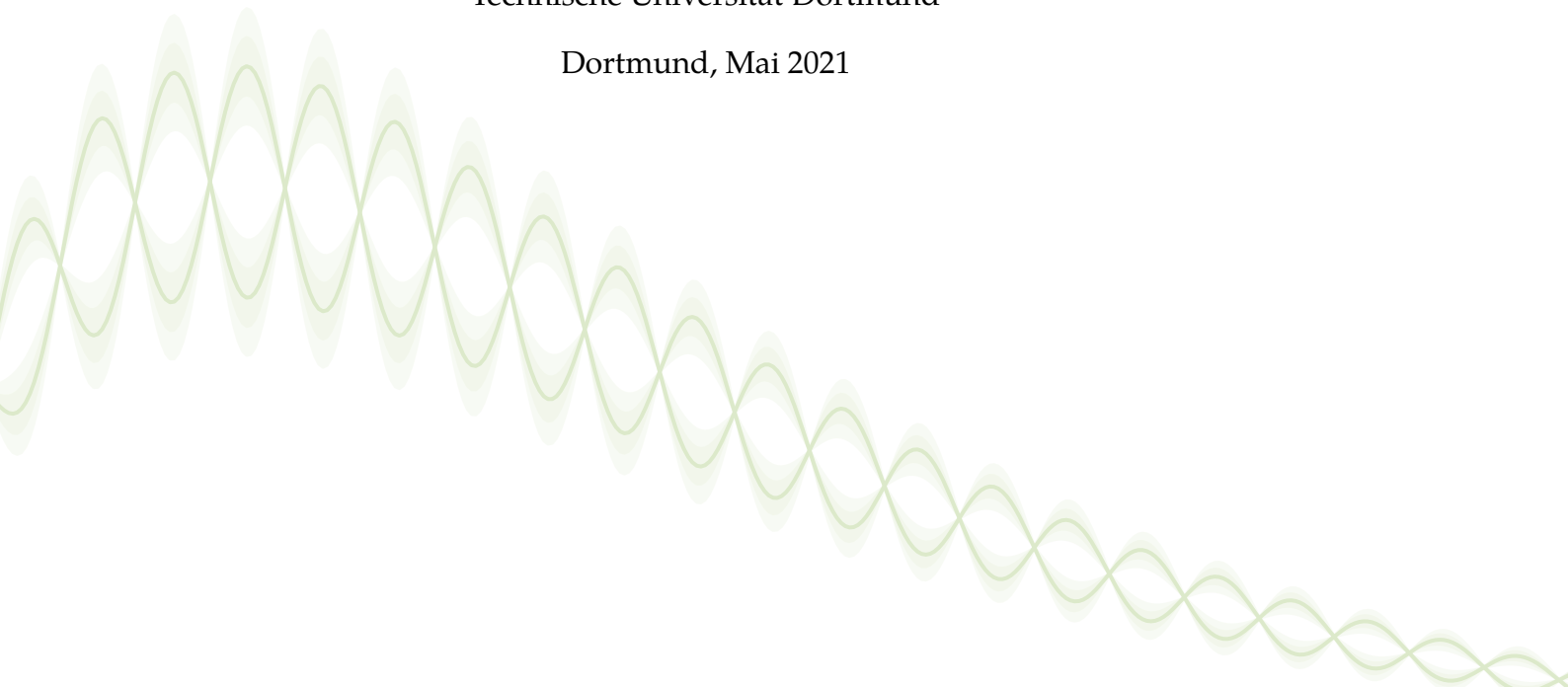
Arbeit zur Erlangung des akademischen Grades  
Dr. rer. nat.

vorgelegt von

Kevin Heinicke

Fakultät Physik  
Technische Universität Dortmund

Dortmund, Mai 2021



Der Fakultät Physik der Technischen Universität Dortmund zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften vorgelegte Dissertation.

Erstgutachter: Prof. Dr. Bernhard Spaan

Zweitgutachter: Prof. Dr. Dr. Wolfgang Rhode

Vorsitz der Prüfungskommission: Prof. Dr. Heinz Hövel

Vertreter der wissenschaftlichen Mitarbeiter: Dr. Michael Paulus

Die Arbeit wurde eingereicht am: 31.05.2021

Datum der mündlichen Prüfung: 16.09.2021

## Abstract

In this thesis, a measurement of the  $B_s^0\text{-}\bar{B}_s^0$  oscillation frequency,  $\Delta m_s$ , is presented. The oscillation frequency provides constraints on the parameters of the CKM matrix and is a crucial input for decay-time dependent measurements of  $CP$  violation with  $B_s^0$  mesons.

The study is performed with a data sample of 378 700 decays of  $B_s^0 \rightarrow D_s^- \pi^+$ , which is extracted from a sample of proton-proton collisions at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of  $6 \text{ fb}^{-1}$ . The data sample was recorded between 2015 and 2018 using the LHCb experiment. An unbinned, weighted maximum-likelihood fit is used to determine the frequency as

$$\Delta m_s = (17.7683 \pm 0.0051 \pm 0.0032) \text{ ps}^{-1}.$$

The result is one of the most precise measurements of the LHCb experiment to date. It is in agreement with current Standard Model predictions.

## Kurzfassung

In dieser Arbeit wird eine Messung der  $B_s^0\text{-}\bar{B}_s^0$  Oszillationsfrequenz  $\Delta m_s$  vorgestellt. Diese Frequenz ermöglicht die Bestimmung einiger Parameter der CKM-Matrix und dient als wichtiger externer Parameter für Messungen von  $CP$ -Verletzung mit  $B_s^0$ -Mesonen.

Die Analyse wird mit einem Datensatz von 378 700 Zerfällen von  $B_s^0 \rightarrow D_s^- \pi^+$  durchgeführt, was einer integrierte Luminosität von  $6 \text{ fb}^{-1}$  von Proton-Proton Kollisionen entspricht. Der Datensatz wurde zwischen 2015 und 2018 bei einer Schwerpunktsenergie von 13 TeV mit dem LHCb-Experiment aufgenommen. Die Oszillationsfrequenz wird mit Hilfe eines ungebinnten und gewichteten Maximum-Likelihood-Fits zu

$$\Delta m_s = (17.7683 \pm 0.0051 \pm 0.0032) \text{ ps}^{-1}$$

bestimmt. Das Ergebnis stellt eine der bisher präzisesten Messungen der LHCb-Kollaboration dar. Es ist im Einklang mit aktuellen Voraussagen des Standard Modells.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 What we believe: The Standard Model of Particle Physics</b>	<b>3</b>
2.1 Matter and Interactions . . . . .	3
2.2 Symmetries and Symmetry Violation . . . . .	5
2.3 The Quark Mixing Matrix . . . . .	6
2.4 Limits of the SM . . . . .	8
<b>3 Particle Mixing</b>	<b>9</b>
3.1 Time Evolution of Neutral Mesons . . . . .	10
3.2 Neutral Meson Decay Rates and $CP$ Violation . . . . .	11
3.2.1 Direct $CP$ Violation . . . . .	13
3.2.2 $CP$ Violation in the Mixing . . . . .	14
3.2.3 $CP$ Violation in the Interference of Decay and Decay After Mixing	15
3.3 Theoretical Assessment of $\Delta m_s$ . . . . .	15
3.3.1 New Physics Constraints from $\Delta m_s$ . . . . .	17
3.4 The Decay $B_s^0 \rightarrow D_s^- \pi^+$ . . . . .	18
<b>4 The Laboratory</b>	<b>21</b>
4.1 The Large Hadron Collider at CERN . . . . .	21
4.2 The LHCb Experiment . . . . .	23
4.2.1 Tracking . . . . .	25
4.2.2 Particle Identification . . . . .	27
4.2.3 Trigger and Data Processing . . . . .	29
4.2.4 Monte Carlo Simulation . . . . .	31
<b>5 Data Analysis Methods</b>	<b>33</b>
5.1 Multivariate Analysis . . . . .	33
5.1.1 Decision Trees and Boosting . . . . .	34
5.1.2 Neural Networks and Back Propagation . . . . .	35
5.1.3 Shapley Values and Interpretability of Multivariate Analysis Algorithms . . . . .	38
5.2 Maximum Likelihood Fitting . . . . .	39
5.2.1 Unfolding Data Distributions in Extended Likelihood Fits . . . . .	40
5.3 Analysis Automation and Reproducibility . . . . .	42

<b>6</b>	<b>Processing of <math>B_s^0 \rightarrow D_s^- \pi^+</math> Data</b>	<b>45</b>
6.1	Data Sample Properties . . . . .	45
6.1.1	Simulated Data Samples . . . . .	46
6.2	Signal Candidate Selection . . . . .	46
6.2.1	Distributed Data Preparation and Local Data Selection . . . . .	46
6.2.2	Multivariate Background Suppression . . . . .	47
6.2.3	Suppression of Misidentified Decays . . . . .	54
6.2.4	Suppression of Charmless Background Events . . . . .	56
6.3	Extraction of the Effective Signal Component . . . . .	60
<b>7</b>	<b>Flavour Tagging Studies</b>	<b>65</b>
7.1	Beauty Flavour Tagging at LHCb . . . . .	65
7.1.1	Mistag Prediction, Calibration and Combination . . . . .	68
7.1.2	LHCb Flavour Tagging for LHC Run 2 . . . . .	69
7.2	Early Studies of an Inclusive Flavour Tagging Algorithm . . . . .	71
7.3	Opposite Side Tagging Precalibration for $B_s^0 \rightarrow D_s^- \pi^+$ Data . . . . .	74
7.3.1	Multivariate Reweighting of $B^\pm \rightarrow J/\psi K^\pm$ Data . . . . .	75
7.3.2	Calibration and Combination of Opposite Side Tagging Algorithms . . . . .	75
<b>8</b>	<b>Measurement of the Oscillation Frequency <math>\Delta m_s</math></b>	<b>81</b>
8.1	Description of the Detector Acceptance with Cubic Splines . . . . .	81
8.2	Correction of the Decay-time and Decay-time Uncertainty . . . . .	82
8.2.1	Decay-time Uncertainty Calibration . . . . .	83
8.2.2	Decay-time Bias Correction . . . . .	83
8.3	Decay-time Fit to Flavour-tagged Data . . . . .	85
8.4	Model Validation . . . . .	89
8.4.1	Pseudo Experiments . . . . .	89
8.4.2	sPlot Validation: Correlation among Observables . . . . .	90
8.4.3	Data Sample Splits . . . . .	92
8.5	Perspective of Indirect $CP$ Violation in $B_s^0 \rightarrow D_s^- \pi^+$ . . . . .	94
<b>9</b>	<b>Systematic Uncertainties</b>	<b>95</b>
9.1	Reconstruction Effects . . . . .	95
9.2	Analysis Methods . . . . .	96
<b>10</b>	<b>Conclusions &amp; Outlook</b>	<b>99</b>
	<b>Appendix</b>	<b>101</b>
	<b>Bibliography</b>	<b>105</b>

# 1 Introduction

A precise description of nature's processes has always been the result of an interplay between theoretical concepts and experimental research. During the past century, and especially in recent decades, this has enabled the development of technologies that now shape many aspects of everyday life. Within the variety of research areas that – often indirectly – contribute to these developments, particle physics can be seen as an extreme: its concepts are usually far away from everyday experience, yet particle physics aims to explain the very fundamental processes that govern all known matter.

Today, the Standard Model of particle physics (SM) is the widely accepted theory to describe strong, weak and electromagnetic interactions of fundamental particles. Its current form evolved in the 1970s, but includes many concepts that reach back to the early 20<sup>th</sup> century. A central concept of the theory is the symmetry of all processes under simultaneous charge-, parity-, and time-inversion, referred to as *CPT* invariance. Small violations of the *P* and *CP* symmetries were found in 1956 and 1964 [1, 2], respectively, which could nicely be explained within the SM via the Cabibbo-Kobayashi-Maskawa (CKM) mechanism [3, 4]. Based on this, a third quark generation was predicted, which was impressively experimentally confirmed with the discovery of the heavy beauty quark in 1977 [5] and top quark in 1995 [6, 7]. The electroweak part of the theory was confirmed even before these discoveries, by the unexpected observation of neutral weak currents in 1973 [8] whose corresponding heavy bosons were discovered a decade later [9–12]. Ultimately, the discoveries of the tau neutrino in 2000 and the Higgs boson in 2012 confirmed the remaining predictions of the model, leading towards today's great acceptance of the SM [13–15].

Despite this success, the theory leaves several fundamental questions unanswered. So far, no model has been found to describe gravitation in a way that is coherent with the SM. Moreover, cosmological observations reveal that only 5% of the universe can be described within the SM [16]. In fact the theory makes no statements about the remaining components, namely Dark Matter and Dark Energy, which make up around 68% and 27%, respectively. Finally, *CP* violation as described in the SM could explain an asymmetry between matter and antimatter, but the SM effect is too small to explain the observed dominance of matter in our universe.

Due to this discrepancy between the great success of the theory and the significant open questions, a theory beyond the Standard Model is needed. Therefore, the Large Hadron Collider (LHC) at CERN pursues the search for processes that are not described by the SM, possibly hinting towards some kind of New Physics, which could lead to the development of such a new theory.

But so far, no new, fundamental particles have been found after the discovery of the Higgs boson. The goal of the LHC experiments has shifted towards precision measurement of SM parameters to identify weak spots of the existing theory, which could as well

indicate physics beyond the SM. Herein, the LHCb experiment, which is one of the four larger LHC experiments, is especially well equipped for this kind of indirect searches. Its focus on  $b$ - and  $c$ -hadrons with a very high vertex resolution and good particle identification capabilities allows to probe the flavour sector of the SM with unprecedented precision. Recently, an anomaly seen by the LHCb collaboration in rare decays of  $B$  mesons into electrons and muons provides first evidence from a single flavour physics measurement for a weak spot of the theory [17].

Besides the study of rare decays, the main topics of the LHCb physics programme include the study of  $CP$  violation. In particular, time-dependent  $CP$  measurements in decays such as  $B_s^0 \rightarrow D_s^\mp K^\pm$  can be performed. These measurements profit from the good vertex resolution of the LHCb detector. Moreover, they require information of the initial particle flavours, i.e. whether a meson was created as a particle or antiparticle. This task of flavour tagging is especially delicate due to the complex structure of the LHC collisions, while it profits from the good particle identification capabilities of LHCb.

In this work one of the most precise measurements that can be performed with the LHCb experiment, the measurement of the oscillation frequency  $\Delta m_s$  is presented. This frequency not only provides a direct constraint on the CKM parameters of the SM, but is also a crucial input for decay-time dependent  $CP$  violation measurements with  $B_s^0$  mesons.

The work presented in this thesis is performed with data collected with the LHCb detector during LHC's second data taking period between 2015 and 2018 (Run 2). The data sample corresponds to an integrated luminosity of  $6 \text{ fb}^{-1}$ , recorded at a centre-of-mass energy of  $\sqrt{s} = 13 \text{ TeV}$ . The  $B_s^0$ - $\bar{B}_s^0$  oscillation frequency  $\Delta m_s$  is measured using flavour-specific decays of  $B_s^0 \rightarrow D_s^- \pi^+$ , for which flavour tagging information is provided. At an early stage of this work, a significant amount of time has been spent to update and integrate a new version of the flavour tagging software into the LHCb software stack.

The analysis is a joint effort of LHCb colleagues from the Technische Universität Dortmund, Germany, the Polish Academy of Sciences, the National Institute for Subatomic Physics, Netherlands, the National Institute for Nuclear Physics, Padova – Italy, and the University of Manchester, United Kingdom. Besides the authors own contribution, this analysis has therefore profited from the work of other group members. Their corresponding contribution is mentioned in the respective sections and a paper has been submitted for publication [18]. Some additional studies – especially focusing on the various flavour tagging algorithms – have been supervised by the author and contributed indirectly or directly to this work. They are documented in the master's theses of Robin Eichhorn [19] and bachelor's thesis of Julian Jung [20].

In the course of this document, the main topic of this thesis is introduced first theoretically in Chapter 2. The specific formalism, relevant for the measurement of  $\Delta m_s$ , is presented in Chapter 3. The LHCb detector as a source for large data samples of  $B$  meson decays is described in more detail in Chapter 4, and several data analysis methods, that are used to study these data samples, are introduced in Chapter 5. The data processing and explicitly the flavour tagging studies for this measurement are described in Chapters 6 and 7, respectively. Finally, the fit to extract  $\Delta m_s$  is presented in Chapter 8 and the studied systematic effects are shown in Chapter 9. The final results of this work and a future outlook for further measurements are summarised in Chapter 10.



## 2 What we believe: The Standard Model of Particle Physics

In the course of the 20th century, a deep understanding of nature's fundamental processes emerged. This understanding almost always originates from theoretical descriptions of nature's processes that are formulated within some mathematical framework. Such descriptions evolve to scientific theories if they can repeatedly predict the outcome of some experimental study of the corresponding processes. Scientific theories are valid if they fit well with experimental data, and even more so if they achieve to predict yet unknown phenomena. As introduced previously, the SM is a good example of a very successful scientific theory.

While it is tempting to think of some valid theory as a "true" description of nature, it is very dangerous to forget that a scientific theory can only be a verifiable or falsifiable description behind some experimental data.

The Standard Model of particle physics is currently the only widely accepted fundamental theory of the field. Some core concepts that are relevant for this work are introduced in the following sections, starting by describing the fundamental building blocks of the theory in Section 2.1 and introducing the basic concepts of the theory on a more technical level in Sections 2.2 and 2.3. Some outstanding limits of the theory are afterwards introduced in Section 2.4.

### 2.1 Matter and Interactions

In the SM, all matter corresponds to some combination of twelve fundamental fermions: six quarks and six leptons. For each of these particles, an antiparticle exists, which only differs by its inverted charge properties. The fermions interact via bosons that describe the fundamental electromagnetic, weak, and strong forces. The corresponding coupling between fermions and the force-carrying bosons is determined by different quantum numbers such as the electric charge or spin numbers. Although most of these particles are massive, gravitation is not described within the SM. In fact, if compared to the weak(est) force of the SM, the gravitational force is smaller by a factor of  $10^{24}$  to  $10^{29}$  and therefore neglected [21].

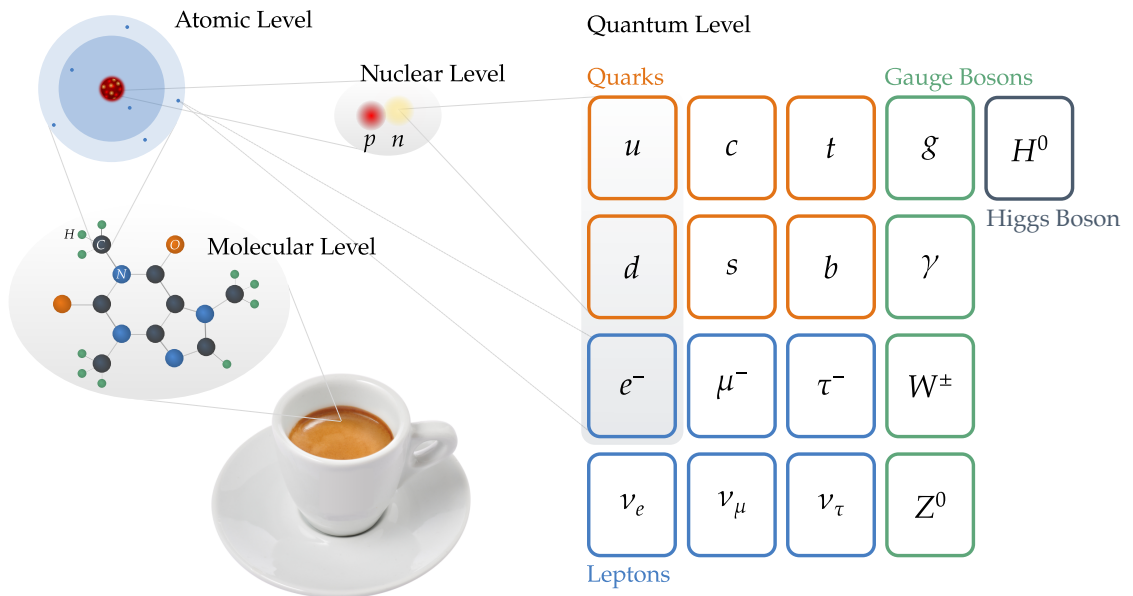
The quarks come along in three families which each contain one up-type quark with elementary charge of  $+\frac{2}{3}$  and one down-type quark with charge  $-\frac{1}{3}$ . The quarks also carry one of three colour charges, to which one of eight gluons can couple in strong interactions. The coupling between two quarks and the gluons in between is in fact so strong that no free quarks can be observed. They are also referred to as being "confined".

## 2 What we believe: The Standard Model of Particle Physics

Instead of emitting a single quark, a highly excited hadron will decay into additional hadrons by creating quark-anti-quark ( $q\bar{q}$ ) pairs from its energy budget. This process is called hadronization.

Similarly to quarks, the leptons are also grouped into three families within the SM, each including a charged lepton and a neutral neutrino partner. All charged particles interact via exchange of photons, which corresponds to the electromagnetic force. Moreover, quarks and leptons carry a weak isospin that is connected to the electric charge via the weak hypercharge and describes weak interactions. These are mediated by massive charged  $W^\pm$  and neutral  $Z^0$  bosons.

The particle constituents of the SM are shown in the right part of Fig. 2.1. The dominant part of known matter in the universe (excluding dark matter) is herein described by only the first generation of quarks and leptons. All fermions from the second or third family decay into the lightest families after some time. Taking the effect of containment into account, this leaves the proton, made from two up and one down quarks, the only stable hadron, with an upper limit for the half life time of  $\tau_p > 3.6 \cdot 10^{29}$  y [22].



**Figure 2.1:** Particles of the Standard Model of particle physics. For each quark and lepton, an antiparticle with opposite charge exists, and each quark can also have one of three colour charge values. For reference, several scales of an everyday object [23] are shown with corresponding visualisations: The molecular level, the atomic scale, the nuclear scale and the quantum scale, which is described by the SM.

If everyday experience is considered, most of the initially mentioned modern technology can well be described by referring to protons, neutrons, and electrons as the smallest structures. Moreover, a large amount of neutrinos exist in nature, which do nevertheless not interfere with common matter due to their small interaction rates. Particles from the heavier lepton families occasionally occur naturally during interactions of high energetic cosmic particles with the earth's atmosphere.

To visualise the level of abstraction which is described by the SM, Figure 2.1 contains depictions of various levels of detail of a common, cm-scale object. Each level can be described with a different set of theories, each of which corresponds to a field of research on its own. While the SM is describing the most fundamental processes that we are currently aware of, it is already unable to model most physical processes at nuclear levels or above.

## 2.2 Symmetries and Symmetry Violation

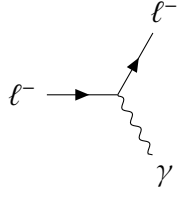
The Standard Model of particle physics is a Quantum Field Theory, in which each particle is associated with an excited state of an underlying quantum field. Symmetries are essential to the theory, as it is required to be invariant under any space-time dependent phase. Additionally, the discrete symmetries of time reversal,  $T$ , charge conjugation,  $C$ , and parity  $P$ , play a crucial role in determining the SM properties. As for any non-abelian gauge theory, the gauge symmetry of the SM implies symmetry under  $CPT$  transformation, while each individual transformation is not necessarily conserved. Phenomenologically this implies that masses of particles and antiparticles are exactly identical, which is experimentally confirmed with very high precision in a multitude of experiments, e.g. [24–27]. On the other hand, the weak interaction maximally violates parity  $P$  and charge  $C$  symmetries, such that only left-handed particles or right-handed anti-particles interact with the charged weak bosons. Moreover a small violation of the  $CP$  symmetry has been observed within the weak interaction in strange, charmed, and beauty mesons [28–31], which is explained in more detail in the next Chapter.

The explicit form of the SM Lagrangian is given in a condensed way in [32] and a general introduction into the theory can be found in the literature, e.g. [33, 34]. Some concepts can nevertheless be illustrated considering one of the SM predecessors, quantum electrodynamics (QED), which solely describes electromagnetic interactions. With a fermion field  $\psi$  of mass  $m$  and electric charge  $e$ , the QED Lagrangian reads

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - e\bar{\psi}\gamma^\mu A_\mu\psi, \quad (2.1)$$

where the first two terms describe the behaviour of a massive fermion and the electromagnetic field, respectively, where  $F_{\mu\nu}$  is the electromagnetic field tensor. The last term connects the fermion field  $\psi$  with the gauge field  $A_\mu$ , which are respectively interpreted as an interaction between the charged lepton and a photon. The strength of this interaction is proportional to the electric charge  $e$  of the lepton. The electromagnetic, as well as weak and strong interactions are implemented in an analogous way in the SM and the  $U(1)$  symmetry group of QED is incorporated in the  $SU(3) \times SU(2)_L \times U(1)_Y$  symmetry of the SM. Despite the similarity between QED and the SM, the  $U(1)_Y$  symmetry of the SM is described in terms of a weak hypercharge  $Y$  instead of the electric charge. The maximally violated parity is furthermore indicated by the index  $L$  of the  $SU(2)_L$  symmetry, meaning that the corresponding gauge bosons only couple to left-handed particles. Another major difference are the massive weak gauge bosons, for which corresponding mass terms only

appear in the Lagrangian via the Higgs mechanism. The discovery of the Higgs boson that was predicted by this formulation is a major confirmation of the theory.



**Figure 2.2:** Feynman graph for an electron-photon interaction.

A very illustrative, yet powerful way to describe SM interactions is provided with Feynman diagrams [35]. The Lagrangian calculus of the SM can be translated into a set of drawing rules for these diagrams. As an example, the interaction between a charged lepton and a photon, such as in Eq. (2.1) translates into the graph shown in Fig. 2.2: An incoming lepton  $\ell^-$  radiates a photon  $\gamma$ , and the momentum is conserved at the vertex. The rate of this process is proportional to the square of the coupling constant  $e$ .

### 2.3 The Quark Mixing Matrix

The weak interaction, as described in the SM, has several interesting features, one of which allows for mixing of quark flavours via the  $W^\pm$  boson. The mass terms of quarks that propagate freely with respect to the electroweak interactions, are acquired via Yukawa interactions with the Higgs field. By calculating the (measurable) quark mass eigenstates of the Lagrangian, a matrix of the form

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.2)$$

is obtained that can mix the mass eigenstates  $(d, s, b)$  and the flavour eigenstates  $(d', s', b')$ . The matrix  $V_{\text{CKM}}$  is named after Cabibbo, Kobayashi, and Maskawa who introduced this application of Yukawa interactions into the SM [3]. A correspond coupling term, exemplary for positively charged  $W^+$  bosons, is proportional to

$$\frac{1}{\sqrt{2}} \frac{-e}{\sin \theta_w} \bar{u}_{Lj} \gamma^\mu W_\mu^+ V_{\text{CKM}} d_{Lk}, \quad (2.3)$$

with the left-handed quark spinors  $u_{Lj}$  and  $d_{Lk}$  for up-type and down-type quarks respectively, the gamma matrices  $\gamma^\mu$ , and the gauge fields  $W_\mu^+$ . The only free parameters in this term are contained in the CKM matrix, the electromagnetic coupling constant  $e$  and the Weinberg angle  $\theta_w$ , which relates electromagnetic and weak couplings.

The CKM matrix for  $N = 3$  quark generations is a complex, unitary  $3 \times 3$  matrix. Since the decay rates that correspond to Eq. (2.3) only depend on a common global phase of the six quark fields, the CKM matrix can be fully described with 4 real parameters. These are often identified with three Euler angles  $\theta_{ij}$ , the complex phase  $\delta_{13}$ , and using the abbreviations  $c_{ij} = \cos(\theta_{ij})$  and  $s_{ij} = \sin(\theta_{ij})$ , such that

$$V_{\text{CKM}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{i\delta_{13}} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta_{13}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (2.4)$$

An expansion in  $\lambda = s_{12}$  allows to define the Wolfenstein parametrisation, where all parameters are of similar order [36]:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \begin{pmatrix} -\frac{1}{8}\lambda^4 & 0 & 0 \\ \frac{1}{2}A^2\lambda^5(1 - 2(\rho + i\eta)) & -\frac{1}{8}\lambda^4(1 + 4A^2) & 0 \\ \frac{1}{2}A\lambda^5(\rho + i\eta) & \frac{1}{2}A\lambda^4(1 - 2(\rho + i\eta)) & -\frac{1}{2}A^2\lambda^4 \end{pmatrix} + \mathcal{O}(\lambda^6). \quad (2.5)$$

Moreover, the unitary conditions of the CKM matrix can be represented as triangles in the complex plane. The  $B_d^0$  and  $B_s^0$  system allows to measure parameters of the conditions

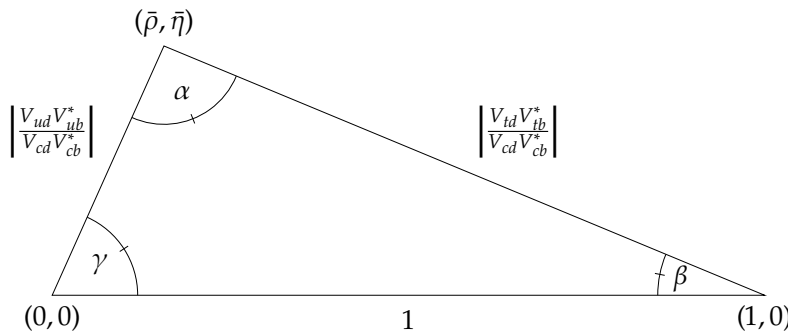
$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad \text{and} \quad V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0, \quad (2.6)$$

respectively. Figure 2.3 shows the left relation of Eq. (2.6), normalized to  $V_{cd}V_{cb}^*$ . All angles and sides of the triangle can be measured experimentally in different physical processes, and therefore provide a stringent test of the SM. As of today, fits to combined experimental results yield

$$A \approx 0.823, \quad \lambda \approx 0.225, \quad \bar{\rho} \approx 0.157, \quad \text{and} \quad \bar{\eta} \approx 0.35, \quad (2.7)$$

where the least known parameter is  $\bar{\rho}$ , with a relative uncertainty of approximately 6% [37]. The complex phase of the CKM matrix is the only dominant source of  $CP$  violation which is explained in the next chapter. The amount of  $CP$  violation therefore corresponds to the area of the CKM triangles which is commonly denoted by the Jarlskog number  $J/2$  [38], where  $J \approx \lambda^6 A^2 \eta \sim 3 \cdot 10^{-5}$ , in the aforementioned parametrisation.

As explained with more detail in Section 3.2, the here presented measurement of  $\Delta m_s$  gives a strong constraint on the length of one sides of the CKM triangles. It is further an important ingredient to other measurements in the  $B_s^0$  system with which also the complex phase of the CKM matrix and the amount of  $CP$  violation can be determined.



**Figure 2.3:** A CKM triangle, obtained from the unitarity requirement of the CKM matrix, Eq. (2.2). To test the SM prediction, the apex of the triangle can be over-constrained by various measurements of the lengths and angles of the triangle.

## 2.4 Limits of the SM

While the SM successfully describes processes at the smallest scales, it fails to do so at cosmological scales. Here, gravitation plays a central role, which is not considered in the SM. The sun for example burns its fuel via processes like weak decays of  $p \rightarrow \ell^+ \nu$  which are fundamentally described by the SM, but are driven by extreme gravitational pressure. An universal description would need to take both sides of the scale into account.

The search for New Physics must therefore pave the way for a unified theory, that includes general relativity and describes some – if not all – remaining open questions. In the following, two popular unsolved problems are briefly described.

### Dark Matter and Dark Energy

The expansion of the universe is parametrised with the Hubble parameter  $H$  which is currently usually modelled within the  $\Lambda$ CDM theory [39]. Herein, the time-dependence is described via a dimensionless scale factor  $a$  and several cosmological density parameters  $\Omega_x$ :

$$H(a) = H_0 \sqrt{(\Omega_c + \Omega_b)a^{-3} + \Omega_{\text{rad}}a^{-4} + \Omega_\Lambda}, \quad (2.8)$$

where  $\Omega_b$  and  $\Omega_c$  are the baryon density and cold dark matter density, respectively,  $\Omega_{\text{rad}}$  is the radiation density and  $\Omega_\Lambda$  is called dark energy density. From the measurement of these densities via the cosmic microwave background, the relative fraction of baryonic matter, which is the sole component described within the SM, is determined to be 5%, while dark matter and dark energy account for 27% and 68%, respectively [16].

This suggests the existence of some type of extremely rarely interacting, but abundant particles that could be described in an extension or successor of the SM.

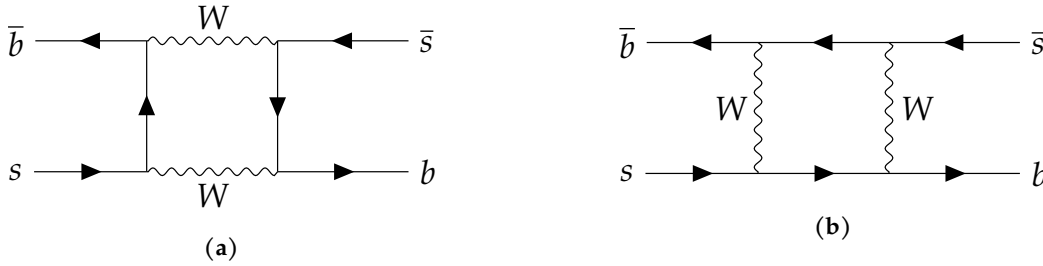
### Baryon Asymmetry

According to the Big Bang theory, and again based on cosmological observations of primordial  ${}^4\text{He}$  and deuterium, as well as the cosmic microwave background, the ratio of baryons to photons in the universe is  $\eta_{\text{BB}} \approx 6 \cdot 10^{-10}$  [16, 40]. Therefore, the early stage of the universe must have been dominated by a small excess of baryons over antibaryons.

Several conditions for this to happen can be formulated, three of which were first proposed in 1967 by Sakharov [41]: First of all the baryon number could be violated, which is strongly excluded experimentally, for example by the lower limit of the proton life time; Moreover, a departure from thermal equilibrium during the inflation of the universe could explain the baryon asymmetry. Finally, the interaction rate of baryons and antibaryons could differ, which corresponds to  $C$  or  $CP$  violation. As stated before, this effect has been observed experimentally, but is not large enough to explain the observed baryon asymmetry.

### 3 Particle Mixing

The quark mixing phenomenon introduced in the previous chapter is able to describe mixing of neutral meson states such as  $|B_s^0\rangle = |\bar{b}s\rangle$  with their charge-conjugate states  $|\bar{B}_s^0\rangle = |b\bar{s}\rangle$ . The process can be described via box diagrams as shown in Fig. 3.1. The SM



**Figure 3.1:** Box diagrams describing the leading processes that contribute to neutral meson mixing. The vertices are governed by CKM matrix elements, making these processes sensitive to  $CP$  violation.

further allows to predict the time dependent decay rates of these mesons into lighter particles. When normalized, this corresponds to a probability density function (PDF)  $\Gamma(P^0(t) \rightarrow f)$  for a meson  $P^0$  to decay into the final state  $f$  after a time  $t$ . This can be experimentally accessed by collecting large samples of  $N$  individual meson decays and comparing their decay-time distribution to the PDF,

$$N(P^0 \rightarrow f)(t) = N \cdot \Gamma(P^0(t) \rightarrow f). \tag{3.1}$$

In this thesis, decays of  $B_s^0$  mesons into  $D_s^- \pi^+$  final states are studied. A measurement of the initial flavour states –  $|B_s^0\rangle$  or  $|\bar{B}_s^0\rangle$  – and the decay time  $t$  of the  $B_s^0$  mesons allows to determine the right side of Eq. (3.1). In the Dirac notation, this is described as

$$\Gamma(P^0(t) \rightarrow f) = |\langle f | \mathcal{H} | P^0(t) \rangle|^2, \tag{3.2}$$

for which an explicit form is deduced in the next sections.

First, in Section 3.1 the propagation of a  $B_s^0$  meson is described based on principles of Quantum Mechanics. This corresponds to the time-dependent term  $|P^0(t)\rangle$  of above equation. In Section 3.2, different types of  $CP$  violation are introduced and the relevant connections to SM parameters are made. This will allow to describe the dynamics of Eq. (3.2), which is governed by the Hamiltonian  $\mathcal{H}$ . Finally the functions that are used in Chapter 8 to measure and visualize  $\Delta m_s$  are introduced in Section 3.4 and the theoretical assessment of the parameter are briefly discussed.

Since the methodology of Sections 3.1 and 3.2 is also valid for other neutral mesons, the  $P^0$  state refers to both,  $B_s^0$  and  $B_d^0$  mesons, as well as neutral strange or charm mesons,  $K^0$  or  $D^0$ , respectively. This chapter is mostly based on [42], while some conventions of [43] are adopted.

### 3.1 Time Evolution of Neutral Mesons

The time evolution of a quantum mechanical state  $|\psi\rangle$  is governed by the Schrödinger equation

$$i \frac{d}{dt} |\psi\rangle = \mathbf{H} |\psi\rangle . \quad (3.3)$$

Here, the Hamilton operator  $\mathbf{H}$  is a hermitian matrix, that yields the probability to observe one of the possible states as the square of its eigenvalues and the corresponding quantum states as its eigenvectors.

Within the Schrödinger picture, the time propagation of neutral mesons can therefore be described as a two-state system, with flavour states  $|P^0\rangle$  and  $|\bar{P}^0\rangle$ . Using the Wigner-Weisskopf approximation [44], the mesons are also allowed to decay as it is observed in nature for any excited state. The transition matrix for this is not necessarily hermitian, but can be any complex  $2 \times 2$  matrix. It is usually split into a hermitian part  $\mathbf{M}$  and an anti-hermitian part  $i\mathbf{\Gamma}$ , such that the time propagation and decay of a neutral meson can be described by solving

$$i \frac{d}{dt} \begin{pmatrix} |P^0(t)\rangle \\ |\bar{P}^0(t)\rangle \end{pmatrix} = (\mathbf{M} - i\mathbf{\Gamma}/2) \begin{pmatrix} |P^0(t)\rangle \\ |\bar{P}^0(t)\rangle \end{pmatrix} . \quad (3.4)$$

Any non-zero off-diagonal elements of  $\mathbf{M} - i\mathbf{\Gamma}/2$  describe the flavour mixing. Their dominant amplitudes can be calculated by evaluating the Feynman graphs in Fig. 3.1, which is briefly described in Section 3.3. As stated previously,  $CPT$  invariance of Eq. (3.4) is required. With the additional requirement that  $\mathbf{M}$  and  $\mathbf{\Gamma}$  are hermitian, the matrices need to fulfil

$$\begin{aligned} & M_{11} = M_{22} \equiv m \quad \text{and} \quad \Gamma_{11} = \Gamma_{22} \equiv \Gamma , \\ \text{as well as} \quad & M_{12} = M_{21}^* \quad \text{and} \quad \Gamma_{12} = \Gamma_{21}^* . \end{aligned} \quad (3.5)$$

These requirements allow to calculate the eigenvalues of Eq. (3.4), such that the corresponding eigenstates can be expressed as superpositions of the flavour states,

$$\begin{aligned} |P_H\rangle &= p |P^0\rangle + q |\bar{P}^0\rangle \\ \text{and} \quad |P_L\rangle &= p |P^0\rangle - q |\bar{P}^0\rangle , \end{aligned} \quad (3.6)$$

with some complex coefficients  $p$  and  $q$  that obey  $|p|^2 + |q|^2 = 1$ . By convention, the states  $P_L$  and  $P_H$  are referred to as light and heavy mass eigenstate, respectively.



The eigenvalues are

$$\mu_{H/L} = m - \frac{i}{2}\Gamma \pm \sqrt{|M_{12}|^2 - |\Gamma_{12}|^2/4 - i \operatorname{Re}(M_{12}\Gamma_{12}^*)} \quad (3.7)$$

$$= m \pm \frac{\Delta m}{2} - \frac{i}{2} \left( \Gamma \pm \frac{\Delta\Gamma}{2} \right) \quad (3.8)$$

$$\equiv m_{H/L} - \frac{i}{2}\Gamma_{H/L}, \quad (3.9)$$

where the heavy and light index H and L correspond to the + and – signs, respectively. The second identity, Eq. (3.8), is true as long as

$$\Delta m^2 - \Delta\Gamma^2/4 = 4|M_{12}|^2 - |\Gamma_{12}|^2 \quad \text{and} \quad \Delta m\Delta\Gamma = 4 \operatorname{Re}(M_{12}\Gamma_{12}^*). \quad (3.10)$$

With the definitions in Eq. (3.9), the time evolution of the mass eigenstates is given by

$$|P_{H/L}(t)\rangle = e^{-i(m_{H/L} - \frac{i}{2}\Gamma_{H/L})t} |P_{H/L}(t=0)\rangle. \quad (3.11)$$

The amplitude square of this already predicts the decay-time distribution of a neutral meson system. Even if no information about the initial meson flavour is available, an experiment can observe flavour mixing from Eq. (3.11).

The equation can now be expressed in the flavour states, via the eigenvectors of Eq. (3.6). These vectors define the matrix

$$\mathbf{Q} = \begin{pmatrix} p & p \\ q & -q \end{pmatrix} \quad \text{and its inverse} \quad \mathbf{Q}^{-1} = \frac{1}{2pq} \begin{pmatrix} q & p \\ q & -p \end{pmatrix}, \quad (3.12)$$

such that

$$\mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \mathbf{Q} \begin{pmatrix} e^{-i(m_H - i\Gamma_H/2)t} & 0 \\ 0 & e^{-i(m_L - i\Gamma_L/2)t} \end{pmatrix} \mathbf{Q}^{-1} = \begin{pmatrix} g_+(t) & \frac{q}{p}g_-(t) \\ \frac{p}{q}g_-(t) & g_+(t) \end{pmatrix} \quad (3.13)$$

solves Eq. (3.4) with  $g_{\pm}(t) = (e^{-i\mu_H t} \pm e^{-i\mu_L t})/2$ . The time evolution of each flavour state now includes a possibly non-zero amplitude of the opposite flavours and evolves like

$$\begin{aligned} |P^0(t)\rangle &= g_+(t) |P^0\rangle + \frac{q}{p}g_-(t) |\bar{P}^0\rangle \\ \text{and} \quad |\bar{P}^0(t)\rangle &= \frac{p}{q}g_-(t) |P^0\rangle + g_+(t) |\bar{P}^0\rangle. \end{aligned} \quad (3.14)$$

## 3.2 Neutral Meson Decay Rates and CP Violation

As stated at the beginning of this chapter, neutral mesons are not observed directly. Instead, only differential decay rates, such as Eq. (3.2), are experimentally accessible. Usually, exclusive decays are studied, where the initial meson,  $P^0$  or  $\bar{P}^0$ , is reconstructed

### 3 Particle Mixing

in a final state  $f$ , or its anti-particle state  $\bar{f}$ . Therefore, four experimentally accessible amplitudes can be defined, which parametrise the dynamics of these decays,

$$\begin{aligned} A_f &= \langle f | \mathcal{H} | P^0(t) \rangle, & A_{\bar{f}} &= \langle \bar{f} | \mathcal{H} | P^0(t) \rangle, \\ \bar{A}_f &= \langle f | \mathcal{H} | \bar{P}^0(t) \rangle, & \bar{A}_{\bar{f}} &= \langle \bar{f} | \mathcal{H} | \bar{P}^0(t) \rangle. \end{aligned} \quad (3.15)$$

The Hamiltonian  $\mathcal{H}$  describes the weak interaction within the SM. While strong and electromagnetic forces are symmetric under  $C, P$ , and  $T$  transformation, particle and anti-particle states must be transformed into one another via the combined  $CP$  transformation under the weak interaction:

$$\begin{aligned} CP |P^0\rangle &= e^{+i\xi_{p0}} |\bar{P}^0\rangle, & CP |f\rangle &= e^{+i\xi_f} |\bar{f}\rangle, \\ CP |\bar{P}^0\rangle &= e^{-i\xi_{p0}} |P^0\rangle, & CP |\bar{f}\rangle &= e^{-i\xi_f} |f\rangle, \end{aligned} \quad (3.16)$$

such that  $(CP)^2 = 1$ . The phases  $\xi_{p0}$  and  $\xi_f$  depend on the quark content of the states. They are only predicted by the quark model but cannot be measured due to the flavour symmetry of the strong force. The decay amplitudes Eq. (3.15) transform under  $CP$  like

$$A_f \rightarrow \langle f | CP^+ \mathcal{H} CP | P^0(t) \rangle = e^{i(\xi_f - \xi_{p0})} A_f = \bar{A}_{\bar{f}}, \quad (3.17)$$

such that the absolute amplitudes are invariant,  $|A_f| = |\bar{A}_{\bar{f}}|$ , only if  $CP$  is conserved.

An important parameter to describe effects of  $CP$  violation is the relative phase between the previously introduced ratio  $q/p$  and the decay amplitudes from Eq. (3.15). It is often included terms of

$$\lambda_f = \frac{1}{\lambda_{\bar{f}}} = \frac{q}{p} \frac{\bar{A}_f}{A_f} \quad \text{and} \quad \bar{\lambda}_{\bar{f}} = \frac{1}{\lambda_{\bar{f}}} = \frac{q}{p} \frac{A_{\bar{f}}}{\bar{A}_{\bar{f}}}. \quad (3.18)$$

Using these parameters, the solutions in Eq. (3.14) for the time evolution can be used to obtain an explicit form of the differential decay rates,

$$\begin{aligned} \Gamma(P^0(t) \rightarrow f) &= |A_f|^2 \left[ |g_+(t)|^2 + |\lambda_f|^2 |g_-(t)|^2 + 2 \operatorname{Re}(\lambda_f g_+^*(t) g_-(t)) \right], \\ \Gamma(\bar{P}^0(t) \rightarrow f) &= |A_f|^2 |p/q|^2 \left[ |g_-(t)|^2 + |\lambda_f|^2 |g_+(t)|^2 + 2 \operatorname{Re}(\lambda_f g_+(t) g_-^*(t)) \right], \\ \Gamma(P^0(t) \rightarrow \bar{f}) &= |\bar{A}_{\bar{f}}|^2 |q/p|^2 \left[ |g_-(t)|^2 + |\bar{\lambda}_{\bar{f}}|^2 |g_+(t)|^2 + 2 \operatorname{Re}(\bar{\lambda}_{\bar{f}} g_+(t) g_-^*(t)) \right], \\ \Gamma(\bar{P}^0(t) \rightarrow \bar{f}) &= |\bar{A}_{\bar{f}}|^2 \left[ |g_+(t)|^2 + |\bar{\lambda}_{\bar{f}}|^2 |g_-(t)|^2 + 2 \operatorname{Re}(\bar{\lambda}_{\bar{f}} g_+^*(t) g_-(t)) \right]. \end{aligned} \quad (3.19)$$

The decay-time independent parameters of these decay rates model all effects that originate from the complex phase of the CKM matrix. Within the SM, this phase is the dominating origin of  $CP$  violation, which is why Eq. (3.19) is often simplified further by expanding all time-dependent terms, and defining the coefficients of the trigonometric and hyperbolic terms as the  $CP$  observables [43]

$$\begin{aligned} A_f^{\Delta\Gamma} &= -\frac{2 \operatorname{Re} \lambda_f}{1 + |\lambda_f|^2}, & C_f &= \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2}, & S_f &= \frac{2 \operatorname{Im} \lambda_f}{1 + |\lambda_f|^2}, \\ \text{and} \quad A_{\bar{f}}^{\Delta\Gamma} &= -\frac{2 \operatorname{Re} \lambda_{\bar{f}}}{1 + |\lambda_{\bar{f}}|^2}, & C_{\bar{f}} &= -\frac{1 - |\lambda_{\bar{f}}|^2}{1 + |\lambda_{\bar{f}}|^2}, & S_{\bar{f}} &= -\frac{2 \operatorname{Im} \lambda_{\bar{f}}}{1 + |\lambda_{\bar{f}}|^2}. \end{aligned} \quad (3.20)$$

These coefficients are normalized, such that

$$(A_f^{\Delta\Gamma})^2 + (C_f)^2 + (S_f)^2 = (A_{\bar{f}}^{\Delta\Gamma})^2 + (C_{\bar{f}})^2 + (S_{\bar{f}})^2 = 1.$$

Within this convention, the decay rates read like

$$\Gamma(P^0(t) \rightarrow f) = \frac{1}{2}e^{-\Gamma t} |A_f|^2 (1 + |\lambda_f|^2) \cdot \left[ \cosh\left(\frac{\Delta\Gamma t}{2}\right) + A_f^{\Delta\Gamma} \sinh\left(\frac{\Delta\Gamma t}{2}\right) + C_f \cos(\Delta m t) - S_f \sin(\Delta m t) \right], \quad (3.21)$$

$$\Gamma(\bar{P}^0(t) \rightarrow f) = \frac{1}{2}e^{-\Gamma t} |A_f|^2 (1 + |\lambda_f|^2) \left|\frac{p}{q}\right|^2 \cdot \left[ \cosh\left(\frac{\Delta\Gamma t}{2}\right) + A_f^{\Delta\Gamma} \sinh\left(\frac{\Delta\Gamma t}{2}\right) - C_f \cos(\Delta m t) + S_f \sin(\Delta m t) \right], \quad (3.22)$$

$$\Gamma(P^0(t) \rightarrow \bar{f}) = \frac{1}{2}e^{-\Gamma t} |\bar{A}_{\bar{f}}|^2 (1 + |\bar{\lambda}_{\bar{f}}|^2) \left|\frac{q}{p}\right|^2 \cdot \left[ \cosh\left(\frac{\Delta\Gamma t}{2}\right) + A_{\bar{f}}^{\Delta\Gamma} \sinh\left(\frac{\Delta\Gamma t}{2}\right) + C_{\bar{f}} \cos(\Delta m t) - S_{\bar{f}} \sin(\Delta m t) \right], \quad (3.23)$$

$$\Gamma(\bar{P}^0(t) \rightarrow \bar{f}) = \frac{1}{2}e^{-\Gamma t} |\bar{A}_{\bar{f}}|^2 (1 + |\bar{\lambda}_{\bar{f}}|^2) \cdot \left[ \cosh\left(\frac{\Delta\Gamma t}{2}\right) + A_{\bar{f}}^{\Delta\Gamma} \sinh\left(\frac{\Delta\Gamma t}{2}\right) - C_{\bar{f}} \cos(\Delta m t) + S_{\bar{f}} \sin(\Delta m t) \right]. \quad (3.24)$$

These decay rates provide a generalized parametrisation of  $CP$  violation in any decay from  $P^0$  or  $\bar{P}^0$  to  $f$  or  $\bar{f}$ , and are therefore sometimes referred to as “master equations of  $CP$  violation”.

Moreover, the decay rates include the mixing parameters  $\Delta\Gamma$  and  $\Delta m$ , latter of which can be measured to obtain additional constraints on the CKM matrix. This is explained for the oscillation frequency of  $B_s^0$  mesons,  $\Delta m_s$ , in Section 3.3. As described in Section 3.4, a measurement of the parameter is possible with flavour-specific decays of  $B_s^0$  mesons, such as the decays of  $B_s^0 \rightarrow D_s^- \pi^+$  that are studied in this work. In the prospect of largely increased statistics these measurement could moreover provide constraints on indirect  $CP$  violation. In addition to that, the parameter  $\Delta m_s$  is an important input for time-dependent measurements of  $CP$  violation in other decays of  $B_s^0$  mesons, which is why the different classes of  $CP$  violation are briefly introduced in the following.

### 3.2.1 Direct CP Violation

Several parameters of the just introduced decay rate PDFs can describe  $CP$  violation. The most general type can be identified if the instantaneous decay amplitudes  $|A_f|^2$  and  $|\bar{A}_{\bar{f}}|^2$

### 3 Particle Mixing

differ. This is possible for any decay amplitude to which multiple processes with different weak and strong phases contribute. This leads to a time-independent asymmetry of the respective decay rates,

$$A_{CP} = \frac{\Gamma(\bar{P} \rightarrow \bar{f}) - \Gamma(P \rightarrow f)}{\Gamma(\bar{P} \rightarrow \bar{f}) + \Gamma(P \rightarrow f)} = \frac{|\bar{A}_{\bar{f}}/A_f|^2 - 1}{|\bar{A}_{\bar{f}}/A_f|^2 + 1}. \quad (3.25)$$

Since the mixing formalism is not taken into account in this case, the particle  $P$  can also refer to hadrons in general, and is not restricted to neutral mesons. As a consequence, direct  $CP$  violation is the only type of  $CP$  violation that can occur in decays of any type of meson or baryon.

Direct  $CP$  violation has first been observed in the kaon system, and is now also well established in the  $B$  meson system. Recent results from the LHCb collaboration [45] report

$$A_{CP}(B \rightarrow K^+ \pi^-) = -0.0824 \pm 0.0033(\text{stat.}) \pm 0.0033(\text{syst.})$$

and  $A_{CP}(B_s^0 \rightarrow K^- \pi^+) = 0.236 \pm 0.013(\text{stat.}) \pm 0.011(\text{syst.})$ .

Moreover, LHCb reported first observation of  $CP$  violation in decays of  $D^0 \rightarrow K^- K^+$  and  $D^0 \rightarrow \pi^- \pi^+$ , where the reported value of  $\Delta A_{CP} = (-15.4 \pm 2.9) \cdot 10^{-4}$  is dominated by direct  $CP$  violation [46].

#### 3.2.2 $CP$ Violation in the Mixing

If the mixing formalism of neutral mesons is considered, the rates  $P^0 \rightarrow \bar{P}^0$  and  $\bar{P}^0 \rightarrow P^0$  could differ, which corresponds to  $|q/p| \neq 1$  as can be seen in Eq. (3.14). The origin of these parameters can be deduced from the time evolution formalism in Section 3.1, if the solution Eq. (3.14) is transformed back into the mass basis, where  $\mathbf{Q}^{-1}(\mathbf{M} - i\mathbf{\Gamma}/2)\mathbf{Q}$  is diagonal. The parameters are then related to the transition matrix via

$$\left(\frac{q}{p}\right)^2 = \frac{2M_{12}^* - i\Gamma_{12}^*}{2M_{12} - i\Gamma_{12}} = \frac{M_{12}^*}{M_{12}} \frac{1 + i|\frac{\Gamma_{12}}{2M_{12}}|e^{+i\phi}}{1 + i|\frac{\Gamma_{12}}{2M_{12}}|e^{-i\phi}}, \quad (3.26)$$

where any phase  $\phi = \arg(-M_{12}/\Gamma_{12}) \neq 0$  leads to  $CP$  violation in the mixing which is often also referred to as indirect  $CP$  violation.

The experimental implication is a difference between the decay-time dependent decay rates of a neutral meson into an exclusive final state. These can be measured in semi-leptonic meson decays as

$$a_{sl} = \frac{\Gamma(P^0(t) \rightarrow \ell^+ X) - \Gamma(\bar{P}^0(t) \rightarrow \ell^- \bar{X})}{\Gamma(P^0(t) \rightarrow \ell^+ X) + \Gamma(\bar{P}^0(t) \rightarrow \ell^- \bar{X})} \approx 1 - |q/p|^2. \quad (3.27)$$

The asymmetry is expected to be small in the SM, containing suppression factors of the order  $10^{-5}$  [47]. The current world average of  $a_{sl}$  is compatible with the SM, such that  $|q/p| \approx 1$  holds experimentally in both, the  $B_d^0$  and  $B_s^0$  system [22]. The most recent measurements of indirect  $CP$  violation were performed by the LHCb experiment [48, 49].

### 3.2.3 CP Violation in the Interference of Decay and Decay After Mixing

Some neutral mesons decay into a final state, which is common to both initial flavour states. Therefore, both aforementioned processes can interfere, such that two decay paths must be considered: The mesons can decay directly via transitions  $P^0 \rightarrow f$  or they can mix before they decay,  $P^0 \rightarrow \bar{P}^0 \rightarrow f$ . This allows to define another type of CP violation, if

$$\arg \lambda_f + \arg \lambda_{\bar{f}} \neq 0. \quad (3.28)$$

If CP violation occurs neither in the mixing,  $|q/p| = 1$ , nor as direct CP violation in the decay amplitudes,  $|A_f| = |\bar{A}_{\bar{f}}|$ , a relative phase between these processes is still possible. This leads to a non-vanishing, time-dependent CP asymmetry.

If the final state is a CP eigenstate  $f_{CP}$ , such that  $CP|f_{CP}\rangle = \pm|f_{CP}\rangle$ , the requirement in Eq. (3.28) simplifies to

$$\text{Im} \lambda_{f_{CP}} \neq 0. \quad (3.29)$$

In this case, asymmetries of the form

$$A_{f_{CP}}(t) = \frac{\Gamma(\bar{P}^0(t) \rightarrow f_{CP}) - \Gamma(P^0(t) \rightarrow f_{CP})}{\Gamma(\bar{P}^0(t) \rightarrow f_{CP}) + \Gamma(P^0(t) \rightarrow f_{CP})} = \frac{S_{f_{CP}} \sin(\Delta mt) - C_{f_{CP}} \cos(\Delta mt)}{\cosh(\Delta\Gamma t/2) + A_{f_{CP}}^{\Delta\Gamma} \sinh(\Delta\Gamma t/2)} \quad (3.30)$$

can be measured experimentally. This equation is further simplified if  $\Delta\Gamma \approx 0$ , which is a good approximation in the  $B_d^0$  system, or if neither direct nor indirect CP violation occurs, in which case  $C_{f_{CP}} = 0$ . Here, the CKM angle  $\beta$  defined in the left side of Eq. (2.6), is directly related to the CP coefficients via  $S \approx \sin(2\beta)$ .

This approximation is valid in  $b \rightarrow c\bar{c}s$  transitions. Experimental data from the  $b$ -factories and hadron colliders yield  $S = 0.691 \pm 0.017$  and  $C \approx 0$ , which is in good agreement with the theory expectation [50]. The latest measurement, coming from LHCb, is not yet included in these averages. It determines  $C = -0.017 \pm 0.029$  and  $S = 0.760 \pm 0.034$  by combining several  $B_d^0 \rightarrow [c\bar{c}]K_S^0$  data samples [51], which is on its own in agreement with the SM expectation.

## 3.3 Theoretical Assessment of $\Delta m_s$

So far, neutral meson mixing has been described in a Schrödinger picture, while the oscillating terms are only explained conceptually by QFT. In fact, each element of the transition matrix in Eq. (3.4) can be interpreted in terms of the SM Lagrangian, which is briefly outlined in this section.

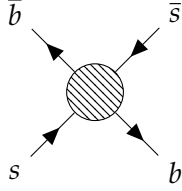
The entries of  $M$  refer to off-shell transitions in which energy is on average conserved within the hadron. In contrast, the  $\Gamma$  matrix contains all on-shell transitions into lighter states. In the SM, the diagonal entries of the transition matrix originates from the fermion mass terms and the QCD binding energy of the hadron. The off-diagonal elements however are governed by the weak interaction, to which several largely different energy scales  $m_t \approx m_W \gg m_b \gg \Lambda_{\text{QCD}}$  contribute, where  $\Lambda_{\text{QCD}} \sim 0.4 \text{ GeV}$  is the scale at which QCD binding forces act. The latter is especially hard to estimate because perturbation

### 3 Particle Mixing

theory cannot be applied at low energies. The different processes are therefore calculated separately, and combined in an effective field theory, where the process in Fig. 3.4 is governed by the effective Hamiltonian

$$\mathcal{H}_{\text{eff}} = \mathcal{H}^{\text{QCD}} + \mathcal{H}^{\text{QED}} + \mathcal{H}^{|\Delta B|=2}. \quad (3.31)$$

The weak interaction – responsible for the mixing – is herein described with an effective four-fermion interaction in  $\mathcal{H}^{|\Delta B|=2}$ , which is depicted in Fig. 3.2.



**Figure 3.2:** An effective four-fermion operator.

Due to the equally small up and charm quark masses  $m_u \sim m_c \sim 0$ , several terms cancel, once the unitarity requirements Eq. (2.6) of the CKM matrix are implied. This effect – referred to as GIM suppression – has first been used in a four quark model to explain mixing effects in the neutral kaon system [52]. Due to the GIM mechanism, the process is dominated by terms related to the top quark and the matrix element  $M_{12}$  turns out as

$$M_{12} = \frac{\langle B_s^0 | \mathcal{H}^{|\Delta B|=2} | \bar{B}_s^0 \rangle}{2m_{B_s^0}} = \frac{G_F^2}{12\pi^2} \eta_B m_{B_s^0} \hat{B}_{B_s^0} f_s^2 m_W^2 S\left(\frac{m_t^2}{m_W^2}\right) (V_{tb} V_{ts}^*)^2.$$

Here, several effects are factorised, firstly into the Fermi coupling constant  $G_F$ , as well as the meson and boson mass terms  $m_{B_s^0}$  and  $m_W$ , respectively. The function  $S(x_i)$  is one of the Inami-Lim functions which are obtained from the evaluation of the box diagrams in Fig. 3.4 [53, 54]. When neglecting the up and charm masses against the heavy top mass, it reduces to

$$S(x) = \frac{1}{4} \left[ 1 + \frac{9}{1-x} - \frac{6}{(1-x)^2} - \frac{6x^2}{(1-x)^3} \log x \right]. \quad (3.32)$$

This term is corrected for additional internal gluon interactions with an approximately constant QCD correction factor  $\eta_B$  [55]. The bag factor  $\hat{B}_{B_s^0}$  and the  $B_s^0$  meson decay constant  $f_s$  can be estimated using non-perturbative lattice QCD calculations, in which the QCD dynamics is computed numerically at different discrete points in space-time. Finally,  $V_{tb} V_{ts}^*$  are the CKM elements from Section 2.3 which allow for  $CP$  violation.

Revisiting the definitions of  $CP$  violation in the mixing, the ratio  $|q/p|$  can be expressed in terms of a small,  $CP$  violating amplitude  $a$ , which approximately corresponds to the  $CP$  violation parameter  $a_{\text{sl}}$ :

$$|q/p|^2 = 1 - a. \quad (3.33)$$

To identify  $a$ , the relation Eq. (3.26) can be expanded in terms of  $|\Gamma_{12}/M_{12}|$ , such that the relations Eq. (3.10) can be solved for  $\Delta m$  and  $\Delta\Gamma$ . This finally allows to connect the time evolution parameters of the neutral meson with SM predictions for the mixing process via

$$\Delta m_s \approx 2|M_{12}|. \quad (3.34)$$

Recent calculations yield an average SM prediction of  $\Delta m_s = 18.4_{-1.2}^{+0.7} \text{ ps}^{-1}$  which is in good agreement with the experimental average, as well as the here presented measurement [56]. Here, the theoretical uncertainties are dominated by the non-perturbative

QCD factors, which are currently estimated as  $f_s^2 \hat{B}_{B_s^0} = (0.0452 \pm 0.0014) \text{ GeV}^2$ , but the uncertainty is expected to improve to  $\sim 3\%$  until the mid of this decade.

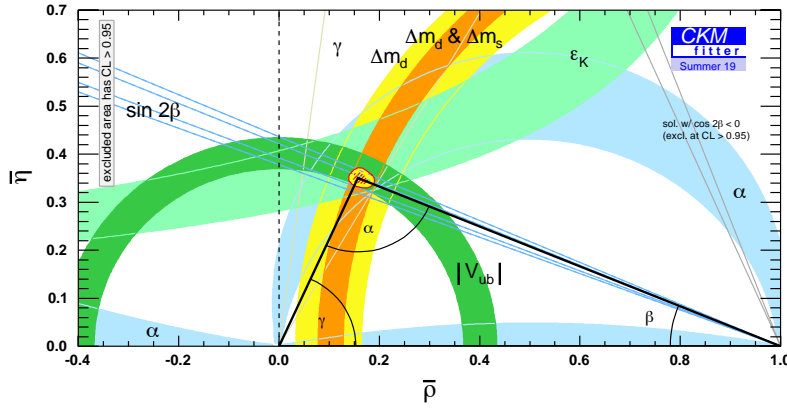
A more precise prediction is possible for the ratio of  $\Delta m_d/\Delta m_s$ , in which many uncertainties cancel:

$$\frac{\Delta m_d}{\Delta m_s} = \left| \frac{V_{td}^2}{V_{ts}^2} \right| \frac{m_{B_d^0} f_d \hat{B}_{B_d^0}}{m_{B_s^0} f_s \hat{B}_{B_s^0}} \equiv \left| \frac{V_{td}^2}{V_{ts}^2} \right| \frac{m_{B_d^0}}{m_{B_s^0}} \xi^{-2}. \quad (3.35)$$

If recent values of the CKMFitter group are considered for  $V_{td}/V_{ts}$  [37, 57], a comparison of the experimental and theoretical ratios can be made, which is again in agreement with the SM:

$$\left( \frac{\Delta m_d}{\Delta m_s} \right)_{\text{exp.}} = 0.0285 \pm 0.0001 \quad \text{and} \quad \left( \frac{\Delta m_d}{\Delta m_s} \right)_{\text{th.}} = 0.0298^{+0.0005}_{-0.0009},$$

where the experimental uncertainty is dominated by the uncertainty of the experimental average of  $\Delta m_d = (0.5065 \pm 0.0019) \text{ ps}^{-1}$ . A similar agreement with the SM is seen if the experimental ratio Eq. (3.35) is used to extract the ratio  $|V_{td}/V_{ts}|$  of CKM elements. The latest global fit results of the CKMFitter group is shown in Fig. 3.3, where the sensitivity obtained from Eq. (3.35), is highlighted as an orange areas.



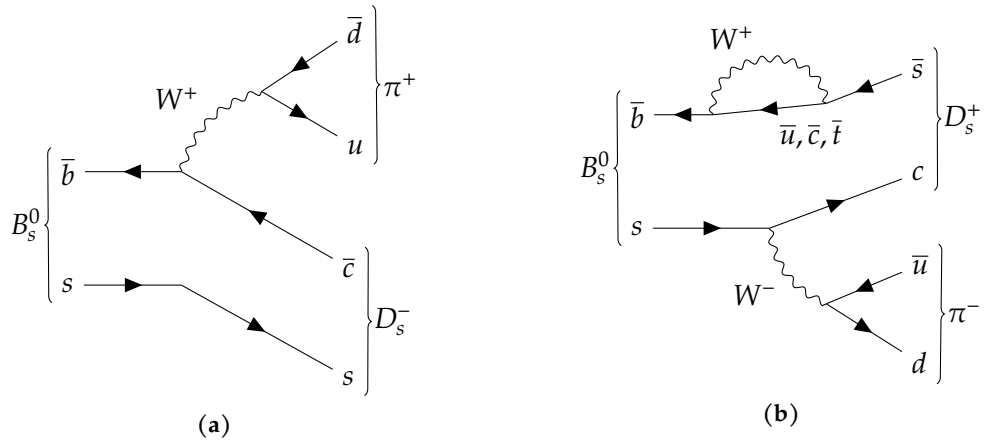
**Figure 3.3:** CKM triangle, obtained through a global fit of the CKM parameters, using experimental results as of summer 2019 [57]. Good sensitivity on the length of the side  $|V_{td}V_{tb}^*|$  is obtained from the combined experimental values of  $\Delta m_s$  and  $\Delta m_d$ .

### 3.3.1 New Physics Constraints from $\Delta m_s$

Recently, some anomalies in  $b \rightarrow s \ell \ell$  transitions have been strengthened by the LHCb measurement of  $R_K^{\text{LHCb}} = 0.846^{+0.044}_{-0.041}$  [17], which deviates  $3.1 \sigma$  from the SM expectation,  $R_K^{\text{SM}} = 1$ . Several theoretical models are able to explain these deviations by introducing additional vector bosons that couple to both quarks and fermions. As investigated in [56], the measurement of  $\Delta m_s$  can put different constraints on these models, depending on the further development of measurements such as  $R_K$ , and the evolution of the theoretical predictions.

### 3.4 The Decay $B_s^0 \rightarrow D_s^- \pi^+$

To measure the neutral meson oscillation frequency, the trigonometric terms of Eqs. (3.21) to (3.24) must be resolved. These equations can be individually compared to data if both, the initial and final  $CP$  states are known. As later described in Chapter 7, the initial flavour can be measured with various flavour tagging techniques. Similarly, some final states  $f_{\pm}$  that include charged particles are dominated by a single decay amplitude. In this case, the decay  $P^0 \rightarrow f_{\pm}$  is flavour-specific, which allows to assign each reconstructed decay to one of the decay rate equations unambiguously.



**Figure 3.4:** Leading-order Feynman diagrams for decays of  $B_s^0 \rightarrow D_s^- \pi^+$  (a) and  $B_s^0 \rightarrow D_s^+ \pi^-$  (b). The latter is heavily suppressed due to the additional weak loop in the diagram.

This is the case for decays of  $B_s^0 \rightarrow D_s^- \pi^+$ , for which the leading Feynman graph is shown in Fig. 3.4a. The decay  $B_s^0 \rightarrow D_s^+ \pi^-$  with the oppositely charged final state particles is shown in Fig. 3.4b and largely suppressed due to the additional massive vector bosons, introducing a suppressive factor of  $G_F^2 \propto \alpha_w^2/M_W^4 \sim 10^{-10} \text{ GeV}^4$ . This translates to negligible amplitudes  $|A_{\bar{f}}| = |\bar{A}_f| \approx 0$  while the remaining amplitudes are referred to as  $|A_f| \equiv |A_{D_s^- \pi^+}|$  and  $|\bar{A}_{\bar{f}}| \equiv |A_{D_s^+ \pi^-}|$ . A complex phase in Eq. (3.18) due to  $CP$  violation in the interference between decay and decay after mixing is undetectable for the same reason. Moreover, no indirect  $CP$  violation has been observed in the  $B_s^0$  system, such that  $|q/p| \approx 1$ . Even if small deviations from one are allowed, the  $CP$  coefficients for decays of  $B_s^0 \rightarrow D_s^- \pi^+$  can be much simplified to  $A_f^{\Delta I} = \bar{A}_{\bar{f}}^{\Delta I} = S_f = S_{\bar{f}} = 0$  and  $C_f = -C_{\bar{f}} = 1$ .



Conclusively, the decay rates for these flavour-specific decays can be described via

$$\Gamma(B_s^0(t) \rightarrow D_s^- \pi^+) = \frac{1}{2} e^{-\Gamma_s t} |A_{D_s^- \pi^+}|^2 \cdot \left[ \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \cos(\Delta m_s t) \right], \quad (3.36)$$

$$\Gamma(\bar{B}_s^0(t) \rightarrow D_s^- \pi^+) = \frac{1}{2} e^{-\Gamma_s t} |A_{D_s^- \pi^+}|^2 \left| \frac{q}{p} \right|^2 \cdot \left[ \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) - \cos(\Delta m_s t) \right], \quad (3.37)$$

$$\Gamma(B_s^0(t) \rightarrow D_s^+ \pi^-) = \frac{1}{2} e^{-\Gamma_s t} |A_{D_s^+ \pi^-}|^2 \left| \frac{p}{q} \right|^2 \cdot \left[ \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) - \cos(\Delta m_s t) \right], \quad (3.38)$$

$$\Gamma(\bar{B}_s^0(t) \rightarrow D_s^+ \pi^-) = \frac{1}{2} e^{-\Gamma_s t} |A_{D_s^+ \pi^-}|^2 \cdot \left[ \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \cos(\Delta m_s t) \right]. \quad (3.39)$$

A mixing asymmetry  $A_{\text{mix}}$  can be defined in which the decay amplitudes and exponential terms cancel, leading to a clean observable

$$A_{\text{mix}}(t) = \frac{\Gamma_u(t) - \Gamma_m(t)}{\Gamma_u(t) + \Gamma_m(t)} = \frac{(2+a) \cos(\Delta m_s t) + a \cosh\left(\frac{\Delta\Gamma_s t}{2}\right)}{(2-a) \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + a \cos(\Delta m_s t)}, \quad (3.40)$$

where  $\Gamma_u$  and  $\Gamma_m$  are the sums of the unmixed and mixed decay rates, respectively, and the indirect  $CP$  asymmetry  $a$  from Eq. (3.33) is used. This term is usually expanded in  $a$ ,

$$A_{\text{mix}}(t) = \frac{\cos(\Delta m_s t)}{\cosh\left(\frac{\Delta\Gamma_s t}{2}\right)} + \frac{a}{2} \left[ 1 - \frac{\cos^2(\Delta m_s t)}{\cosh^2\left(\frac{\Delta\Gamma_s t}{2}\right)} \right] + \mathcal{O}(a^2), \quad (3.41)$$

where to date all but the first term have been neglected in measurements of  $\Delta m_s$ . As described in Section 8.5, additional terms might need to be considered in the upcoming decade.

Given the current experimental result for the decay constant  $\Gamma_s = 0.66 \text{ ps}^{-1}$ , the  $B_s^0$  meson is expected to oscillate approximately 3 times per half life time. This allows to measure several full oscillation periods, which in turn improves the statistical power to determine  $\Delta m_s$  from the cosine-terms of Eqs. (3.36) to (3.39). When compared to typical LHC measurements, the frequency measurement of  $\Delta m_s$  provides a high precision by construction.

To give some perspective, even more precise measurements of oscillation effects are possible with different experimental set ups. Recently this has been demonstrated with a new measurement of the anomalous magnetic moment of the muon,  $g - 2$ , where a relative uncertainty below  $10^{-6}$  was achieved [58]. The measurement furthermore adds more evidence for New Physics in the lepton sector, as the value deviates from the current, best SM prediction.



## 4 The Laboratory

Neutral  $B$  mesons are not commonly found in nature and to study all of the so far introduced effects, a large amount of these particles must be measured. Large samples of unstable particles can be created at a hadron collider, where charged hadrons – usually protons – are accelerated to very high velocities, before they are forced to collide.

Within the SM, the collision is dominantly described as an interaction of two or more gluons, which originate from the hadrons and carry some fraction of their kinetic energy. This interaction is also referred to as a “gluon fusion”, in which the interaction energy is condensed again into other gluons or pairs of quarks and anti-quarks. Alternatively, quarks and other partons of the colliding hadrons can interact, which will again transform the kinetic energy into other particles. These interactions initiate showers of particle decays, from heavier to lighter particles until the initial energy is fully transformed into stable protons, neutrons, or electrons which are ultimately absorbed by the environment. A fraction of these hadron showers also contain  $B$  mesons which need to be identified, after which they can be studied.

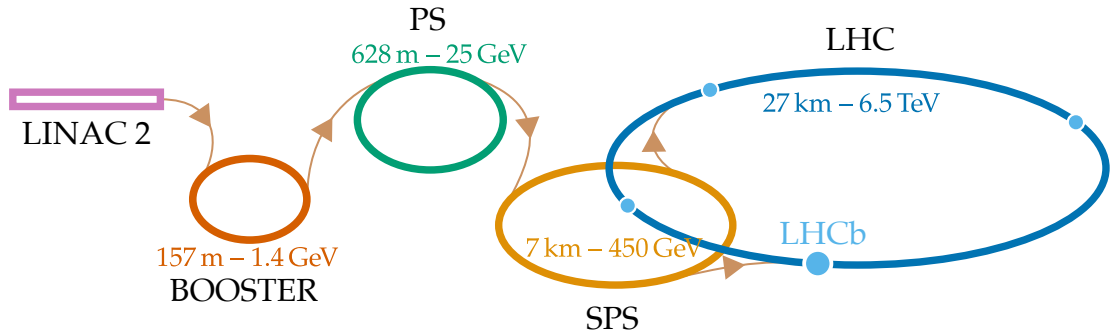
In contrast to this,  $B$  mesons can be created in a  $e^+e^-$  collider, such as the new Belle II experiment [59]. Here, the centre-of-mass energy is chosen to initiate  $e^+e^- \rightarrow b\bar{b}$  processes at the  $\Upsilon(4S)$  resonance, in which the  $b$  quarks will subsequently hadronise into  $B^\pm$  or  $B_q^0$  mesons. The Belle II experiment, as well as its predecessor Belle and the similar  $BABAR$  experiment are therefore also called  $B$ -factories. The advantage of a very clean experimental environment comes at the price of a reduced variety of particles that can be studied. As an example,  $B$  factories cannot generate strange  $B_s^0$  mesons with their default energy configuration.

As of today, the LHCb experiment is therefore the best experiment that is able to measure  $B_s^0$  meson oscillation. The Large Hadron Collider, where the LHCb experiment is run, is introduced in Section 4.1. The experiment is described in more detail in Section 4.2. If not noted otherwise, the following sections are based on [60, 61].

### 4.1 The Large Hadron Collider at CERN

The Large Hadron Collider (LHC) at the European Organisation for Nuclear Research (CERN) is the currently worldwide largest existing particle physics facility, with a circumference of the main accelerator of 27 km. It is designed to collide beams of protons or heavy ions at several different interaction points at a centre-of-mass energy of up to  $\sqrt{s} = 14$  TeV. These energy levels are reached with a chain of pre-accelerators as depicted in Fig. 4.1.

Protons are obtained by ionizing a stream of hydrogen atoms in an electric field after which they are accelerated using radio frequency (RF) cavities in a linear accelerator



**Figure 4.1:** The accelerator chain for the Large Hadron Collider (LHC) at CERN. Hydrogen atoms are ionized and the protons are accelerated in a linear accelerator (LINAC 2). The beam of protons is then transferred into a chain of circular accelerators via several transfer lines. Within each accelerator – Proton Synchrotron Booster, Proton Synchrotron (PS), Super PS (SPS) – the beam energy is increased until the nominal LHC energy is reached. Major experiments reside at four LHC interaction points (IPs), starting at LHCb, followed in clockwise direction by ATLAS, ALICE and CMS. The nominal beam energy and ring circumference is given for each accelerator.

(LINAC2). The beam of protons enters the Proton Synchrotron Booster with an energy of 50 MeV and is accelerated to 1.4 GeV before entering the Proton Synchrotron (PS), which again increases the energy up to 25 GeV. At this stage, the proton beam travels at a velocity close to the speed of light,  $v/c \approx 99.93\%$ , and enters the Super Proton Synchrotron (SPS), which accelerates the proton beam further and injects it into the LHC at an energy of 450 GeV. A second beam is injected into a separate LHC beam pipe in opposite direction and both beams are accelerated further by eight RF cavities to reach the nominal beam energies of 6.5 TeV. The acceleration via RF cavities splits the beam into several bunches of  $\sim 10^{11}$  protons each. Some gaps are maintained between some of these bunches, leading to “bunch trains”, which allow for controlled beam dumps by ramping up kicker magnets within several  $\mu\text{s}$  between two trains.

The beams are kept on their trajectory with 1232 superconducting dipole magnets operating at a current of 11 kA to create magnetic fields of up to 8.3 T, and additional support magnets. The beams cross at several interaction points, resulting in a bunch crossing rate of up to 40 MHz, with an average number of 50  $pp$  collisions per bunch crossing. This corresponds to a nominal instantaneous luminosity of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ .

In addition to the proton beams, a different linear accelerator (LINAC3) is able to provide heavy ion beams for the LHC. Moreover, the beam pipe can be filled with different gases at the LHCb interaction points, resulting in an experimental set up similar to a fixed-target.

First proton beams circulated in the LHC in 2008, after which a major fault of the magnet splines made a revisiting of all superconducting magnets necessary. First  $pp$  collisions were then achieved in 2010, marking the start of the first data-taking period, LHC Run 1, spanning until 2012. During this time, the beam energies were first limited to 3.5 TeV and 4 TeV in until 2011 and in 2012, respectively. Between first and second data

taking period, the LHC was prepared for a higher beam energy of 13 TeV and a two-fold higher instantaneous luminosity of up to  $1.9 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  [62]. Correspondingly, some experiments were upgraded and all existing hardware has undergone consolidation. The second data taking period, Run 2, ranged from 2015 until 2018.

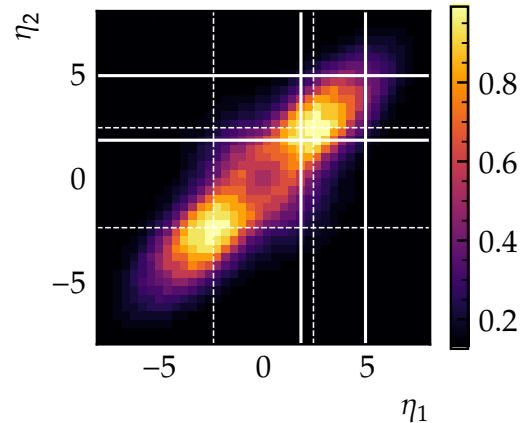
Currently, the LHC is preparing Run 3, which is planned to start in 2021. A stepping up of the beam energy to the design target of 14 TeV is discussed. Between Run 2 and Run 3, major upgrades are being installed in the LHCb experiment. The experiment is described in the following sections, and a thorough review of the detector and the planned design for Run 3 are given in [60, 63–66].

Besides the LHCb experiment, several other experiments are operated at the LHC. The largest experiments, ATLAS and CMS, run general purpose detectors covering the full  $4\pi$  angular region around their respective interaction points. They are well equipped for direct discoveries of heavy particles, such as the Higgs boson. Similarly, the ALICE detector covers a  $4\pi$  region around the interaction point, but it is focused on studies of heavy ion collisions. Additionally, several smaller experiments study the hadronic environment in the extreme forward region, complementing the general purpose detectors. These include LHCf, TOTEM, and FASER, which will start taking data during Run 3 [67–69].

## 4.2 The LHCb Experiment

The LHCb Experiment is specifically designed to study beauty and charm hadrons. Due to the ratio of the average collision energy and the rest mass of these particles, they obtain a large boost, as shown in the simulated distribution of pseudo rapidity of  $b\bar{b}$  quark pairs in Fig. 4.2. The LHCb detector is therefore designed as a single-arm forward spectrometer, focusing on the outer region of this distribution, in which roughly one quarter of all produced  $b\bar{b}$  pairs can be detected. Furthermore, the high LHC luminosity is levelled to a constant value of  $4 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$  at the interaction point, by adjusting the beam cross sections. Thereby, the amount of inelastic  $pp$  scattering in the detector is kept low, providing a cleaner environment and improving the reconstruction quality. As a result, the LHCb detector allows to study beauty and charm hadrons with a high vertex resolution and outstanding particle identification.

The LHCb detector is placed  $\sim 100$  m below surface in the former DELPHI cavern at the LHC interaction point 8. It covers an angular acceptance from 10 to 300 mrad horizontally, and up to 250 mrad vertically. As highlighted in Fig. 4.2, the



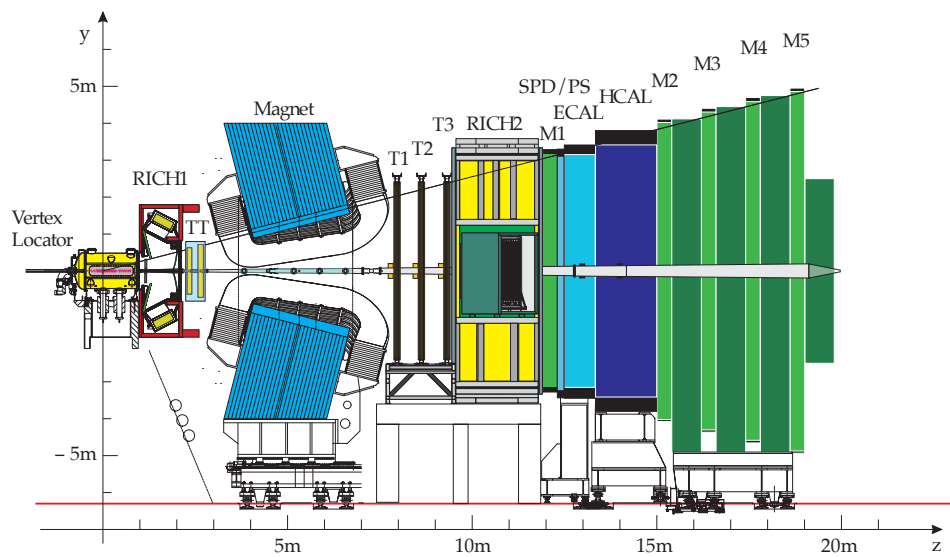
**Figure 4.2:** Simulated distribution of  $b\bar{b}$  quark pairs from  $pp$  collisions at 14 TeV centre-of-mass energy, based on [70]. The LHCb detector acceptance is marked with solid lines. A general purpose detector acceptance is marked with dashed lines.

## 4 The Laboratory

corresponding range of  $b$ -quark pseudo rapidity spans from 1.6 to 4.9. The LHCb detector and its sub-components are depicted in Fig. 4.3, where the interaction point is positioned at the centre of the coordinate system and the  $z$ -axis is pointing along the beam pipe (downstream) into the other detector components. Within the detector, the LHC vacuum is enclosed in the vertex locator housing, a beryllium pipe between the VELO and the calorimeters, and a stainless steel pipe downstream of the calorimeters.

The detector components can be generally grouped into a tracking system and a particle identification system, described in Section 4.2. The detector information is read out and fully reconstructed in real time in a three-stage trigger system, which defines the data stream that is stored to disk. The trigger and the successive data processing is described in Section 4.2.3. Moreover, the full data acquisition is implemented in a Monte Carlo simulation which is briefly explained in Section 4.2.4.

The large energy of the proton beams – around  $\sim 300$  MJ per beam – could potentially have devastating effects on the detector material if it escapes from the nominal trajectory. Therefore, a Beam Condition Monitor (BCM) is constantly measuring the charged particle flux close to the beam pipe and initiates a beam dump before the detector components are endangered [71].



**Figure 4.3:** Side view of the LHCb detector at LHC. The detector is a single-arm forwards spectrometer, with a vertex locator surrounding the interaction point, followed by (in  $z$ -direction, parallel to the beam) one of two RICH detectors, a tracking station (TT), a bending magnet, a second tracking station (T1-T3), a second RICH detector, a calorimeter system (SPD/PS, ECAL, HCAL), and a muon system (M1-M5) [72].

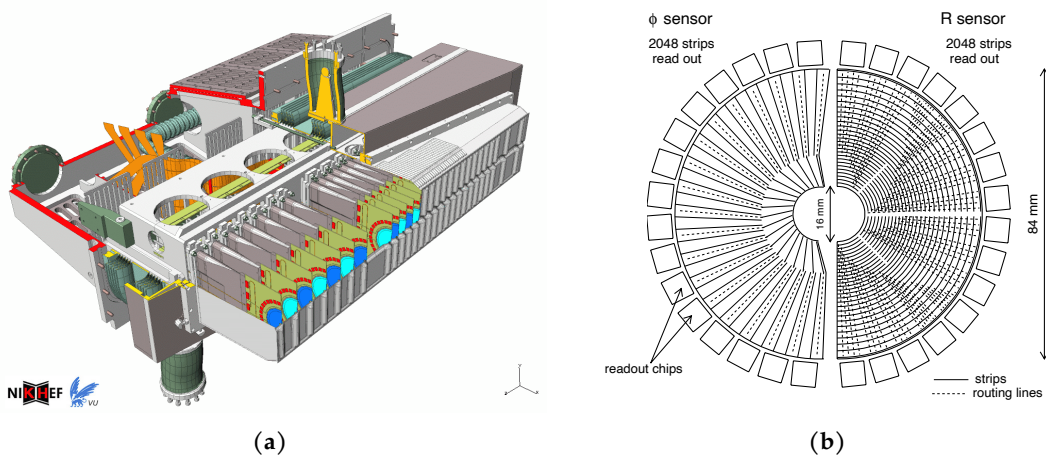
### 4.2.1 Tracking

Charged particles which are created in the  $pp$  collisions will interact electromagnetically with the detector material. By measuring the positions of these interactions, the particle tracks can be reconstructed, which is the main purpose of the LHCb tracking system. It consists of a silicon Vertex Locator (VELO) with a very high spacial resolution, and several tracking stations (TT and T1–T3), as well as a bending magnet.

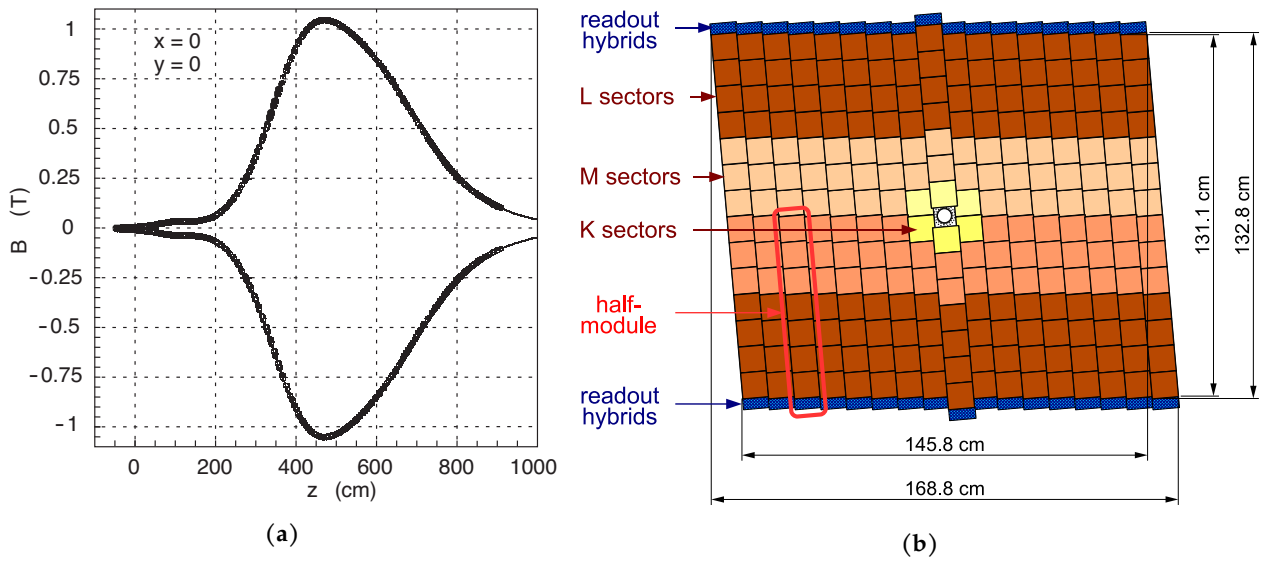
The reconstructed particle tracks furthermore allow to identify decay vertices where multiple tracks are in close proximity to each other. Due to momentum conservation, the vector-sum of the corresponding momenta furthermore define the direction and momentum of a parent track. This way, the initial  $pp$  interaction point can be defined as the primary vertex (PV), and further particle decays manifest as secondary vertices (SVs).

#### The Vertex Locator

The Vertex Locator is based on silicon strip sensors, which are distributed over 42 half disks with 84 mm diameter. On each side of the disks, silicon strips are arranged alternately in a concentric pattern ( $r$ -sensors) and a radial pattern ( $\phi$ -sensors). The disks are mounted on a support structure on both sides of the beam, as shown for one half in Fig. 4.4a. Here, the VELO disks are marked with dark and light blue colours. A 0.3 mm thick  $\text{AlMg}_3$  foil separates the beam vacuum from the VELO's own vacuum and shields the VELO and LHC beams against mutual radio-frequency influences. The VELO halves can be moved several cm in the  $x$ -direction, such that the sensor disks can approach up to 8 mm to the beam during stable beam conditions. The halves are “parked” several cm away from the beam otherwise.



**Figure 4.4:** A rendering of one half of the vertex locator with partly transparent housing is shown in (a). The different types of silicon strip sensors are marked in blue. A schematic view of the half discs is shown in (b) [73, 74].



**Figure 4.5:** Magnetic field strength along the beam pipe, for both magnet polarities (a), and one layer of the TT station (b) [60].

### The Magnet

To infer momentum information from the reconstructed tracks, the Lorentz force on charged particles is exploited. A water-cooled dipole magnet is installed between the tracking stations, creating an integrated magnetic field of 4 T m in between the tracking stations. It is operated at room temperature with a nominal current of 5.85 kA. The curvature of the tracks within the magnetic field allows to measure the track momenta with a relative momentum resolution between 0.4% at  $p = 5$  GeV and 0.6% at  $p = 100$  GeV. The magnet polarity is occasionally switched during a LHC Run to prevent detection asymmetries due to the magnetic field.

### The Tracking Stations

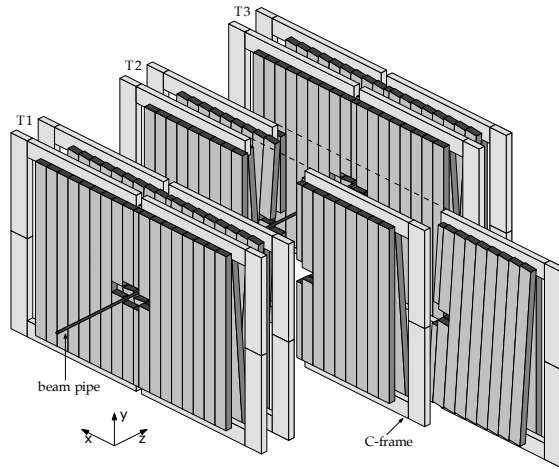
Tracking stations are located upstream (TT) and downstream (T1–T3) of the Magnet. The TT, for which one layer is shown in Fig. 4.5b, and the inner part of the downstream tracking stations (IT) use silicon strip detectors with a strip pitch of  $200 \mu\text{m}$  that allow for a spatial resolution for individual hits of  $50 \mu\text{m}$ . The TT and IT cover an active area of  $8.4 \text{ m}^2$  and  $4 \text{ m}^2$ , respectively.

The outer part of the downstream tracking stations (OT) is designed as an array of straw-tube modules as shown in Fig. 4.6. The drift tubes have an inner diameter of 4.9 mm and are filled with a mixture of 70% Argon and 30%  $\text{CO}_2$ , which results in drift times below 50 ns and single cell resolution of up to  $200 \mu\text{m}$ .

Each of the total three OT and IT stations consists of four layers of modules, where the inner layers ( $u$  and  $v$ ) are tilted by  $\pm 5^\circ$  around the beam axis, allowing to reconstruct



$y$  coordinates of particles tracks. The TT contains two stations, arranged in the same manner.



**Figure 4.6:** The outer tracker (OT) system with support structures. The dark grey OT modules contain drift tubes to measure the  $x$ -coordinates of particle tracks. The inner two modules of each station are tilted to gain additional  $y$  resolution.

## 4.2.2 Particle Identification

To identify the different exclusive final states that were introduced in Chapter 3, an excellent particle identification information is needed. As the lightest charged particles, electrons leave a distinct signature in the detector, emitting a large amount of bremsstrahlung and usually being fully absorbed in the calorimeter system. Even cleaner signatures are produced by muons, as they behave as minimally ionizing particles, have a long life time of  $2 \mu\text{s}$ , and usually a high momentum. Therefore, muons often pass through the whole detector, making up the dominant fraction of particles that can reach the muon chambers downstream of the calorimeter system. In contrast, hadrons are heavier than electrons and radiate significantly less bremsstrahlung. But due to the strong interaction, they are usually also completely stopped within the calorimeter system. This signature is similar for all hadrons, and to further distinguish these, the LHCb detector includes two Ring-Imaging-Cherenkov (RICH) detectors.

### The RICH Detectors

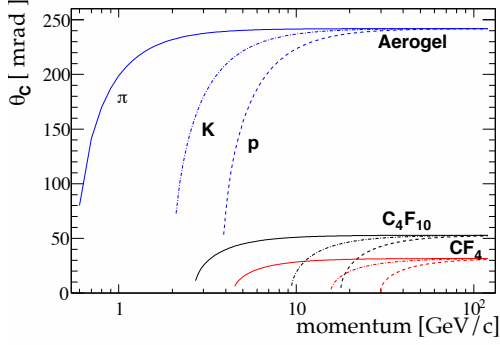
An especially good identification of pions, kaons and protons can be achieved by exploiting the Cherenkov effect [75]: Charged particles that pass through a dielectric medium will locally polarize the material. If their velocity  $v$  is larger than the speed of light in that medium,  $v > c/n$ , where  $n$  is the refractive index of the medium, they will induce a polarization wave front which corresponds to light, emitted at the Cherenkov angle  $\theta_C$ ,

## 4 The Laboratory

with

$$\cos \theta_C = \frac{1}{n\beta}, \quad \text{where} \quad \beta = \frac{v}{c} \stackrel{h=c=1}{\simeq} \frac{p}{E}. \quad (4.1)$$

At a given momentum  $p$ , which can be measured with the tracking system, this angle only depends on the charged particle mass, resulting in a characteristic curve for each type of particle.



**Figure 4.7:** The expected Cherenkov angle  $\theta_C$  at different particle momenta [76]. Different particle hypotheses and RICH detector materials are shown.

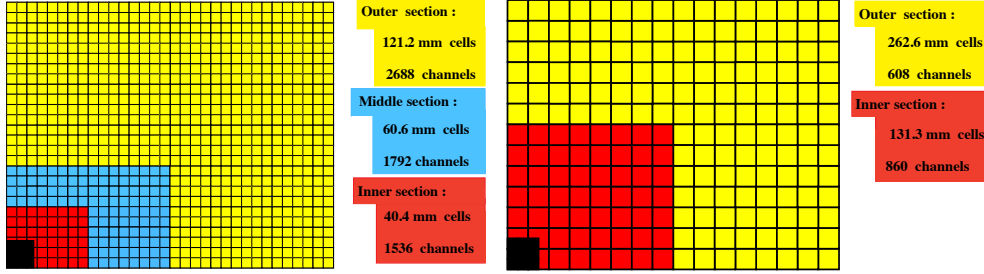
Within the LHCb RICH detectors, the emitted light cones of these particles are collected via a set of parabolic mirrors and projected as light rings onto an array of photo multipliers. The Cherenkov angle can then be measured by reconstructing the radius of these rings, resulting in the characteristic distributions shown in Fig. 4.7. The RICH1 detector, located upstream of the magnet, contains Aerogel and  $C_4F_{10}$ , providing good sensitivity in momentum regions from 1 GeV to  $\sim 60$  GeV. The RICH2 detector, located downstream of the magnet, uses  $CF_4$  and is therefore optimized for particles with larger momenta between 15 GeV and 100 GeV. While both detectors are close to the magnet, the magnetic field at the photo multipliers inside the RICH detectors is below  $\sim 2$  mT, preventing electromagnetic disturbances (see Fig. 4.5a).

### The Calorimeters

Besides distinguishing electrons and hadrons, as mentioned above, the calorimeter system also allows to measure the energy of these particles with additional spatial information. Furthermore, neutral particles, such as photons and  $\pi^0$  can be detected. The calorimeters are positioned between the first and second muon chamber, M1 and M2, downstream of the tracking station T3. First, a scintillating pad detector (SPD), followed by a thin led layer and an almost identical pre-shower detector (PS) help to distinguish neutral and charged tracks. After that, the electromagnetic calorimeter (ECAL) absorbs the energy of most photons and electrons, allowing to measure their total energy. Most hadrons on the other hand only deposit a small fraction of their energy within these systems but will be fully absorbed by the hadronic calorimeter, which is installed directly after the ECAL.

All calorimeters use scintillating pads and tiles based on polystyrene. They are connected via optical fibres with their respective photo multiplier readout electronics. The SPD/PS pads have a high granularity, being read out in 12 032 detection channels. The led layer between SPD and PS corresponds to 2.5 radiation lengths ( $X_0$ ), where  $1 \cdot X_0$  describes the average distance that a high-energy electron propagates within a material, until its energy is reduced by a factor  $1/e$  due to bremsstrahlung. The ECAL is built in a shashlik design, where each detector element combines 66 scintillating pads, which are joined by led plates and in total correspond to  $25 \cdot X_0$ . The HCAL is made from longer

scintillating tiles and iron plates, which are aligned parallel to the beam pipe. On average, hadrons scatter inelastically 5.7 times within the HCAL. The arrangement of ECAL and HCAL elements is depicted for one detector quadrant in Fig. 4.8.



**Figure 4.8:** Layout of one quadrant of the ECAL (left) and HCAL (right) [60]. The detectors measure the energy of electrons, photons, and hadrons, while providing some spacial resolution.

## The Muon System

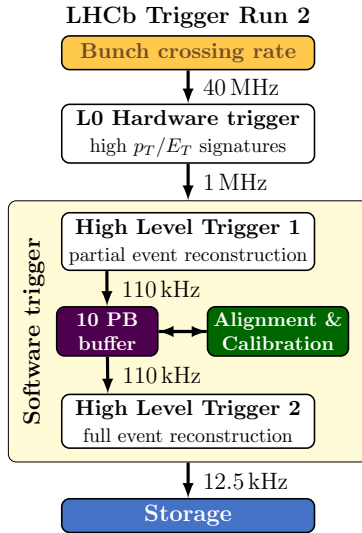
A total of 5 muon stations is used to identify muons and reconstruct their tracks. The first muon station M1 is positioned upstream of the calorimeters, while the stations M2 – M5 are placed downstream of the calorimeter system. All stations use multi-wire proportional chambers, except for the inner region of the first station, which utilizes a more radiation-hard gas electron multiplier. Between each downstream muon station M2 – M5, iron of 80 cm thickness is installed to increase the total absorber thickness. Including the calorimeters, this corresponds to 20 radiation lengths.

### 4.2.3 Trigger and Data Processing

The majority of events within the 40 MHz rate of  $pp$  collisions do not include any  $b$  or  $c$  quarks and therefore hardly contain any interesting information. The cross-section of  $pp \rightarrow B^\pm X$  interactions has been measured to  $\sigma(pp \rightarrow B^\pm X) = (86.6 \pm 6.4) \mu\text{b}$  [77]. This can be combined with the average instantaneous luminosity of  $4 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ , as well as the ratio of the  $b$  quark hadronization fractions into mesons or any baryon,  $f_q$  and  $f_B$  respectively, to estimate the rate of events containing a  $b$  quark. Using the hadronization ratios given in [43], and ignoring acceptance effects and correlations, the average rate of  $b$  hadrons can be estimated to be  $F_b \sim 46 \text{ kHz}$ . Similarly, a lower bound for the average rate of  $c$  hadrons can be estimated with the prompt charm production cross-sections reported in e.g. [78], which yields a rate of charm hadrons  $F_c > 0.8 \text{ MHz}$ .

To identify these interesting events and store them to disk for later analysis, a multi-staged trigger is implemented at the LHCb detector, as shown in Fig. 4.9. The maximum average bandwidth that can be stored to disk is essentially defined by the experiment budget and poses a stringent requirement for the trigger. During Run 2, the output rate of the LHCb trigger system was 12.5 kHz, corresponding to a data rate of  $\sim 600 \text{ MB/s}$

[79] To meet these external constraints, the bunch crossing rate is first reduced to 1 MHz using a hardware trigger (L0) that is based on field programmable gate arrays, which are directly connected to the detector components. They allow to select events based on high energy signatures from the calorimeters, high transverse momentum signatures from the tracking systems or muon signatures from the muon system.



**Figure 4.9:** LHCb Run 2 trigger schema [80]. The 40 MHz bunch crossing rate is reduced to a 12.5 kHz event rate to meet storage bandwidth limitations.

The L0-triggered events are afterwards processed in a two-staged high level software trigger, HLT1 and HLT2. The software trigger is running on an event filter farm (EFF) with  $\sim 1700$  compute nodes, each containing between 24 to 32 logical processors, which allow to combine and evaluate information from all detector components in software. The HLT1 reconstructs charged particle tracks from the VELO and tracking stations, selecting  $\sim 110$  kHz of events with at least one track of good quality. A full event reconstruction is afterwards performed in HLT2, allowing to implement more specific selection requirements for certain types of decays. Different sets of HLT2 selection requirements – referred to as “trigger lines” – can be applied to the complete particle decay trees, including fully reconstructed vertices and particle tracks. The HLT1 output is buffered on 5.2 PB of EFF storage, corresponding to two weeks of LHCb data taking, which allows to perform online alignment and calibration of the detector system. Moreover, the computing resources can be dynamically allocated between HLT1 and HLT2 to increasing the overall efficiency of the farm.

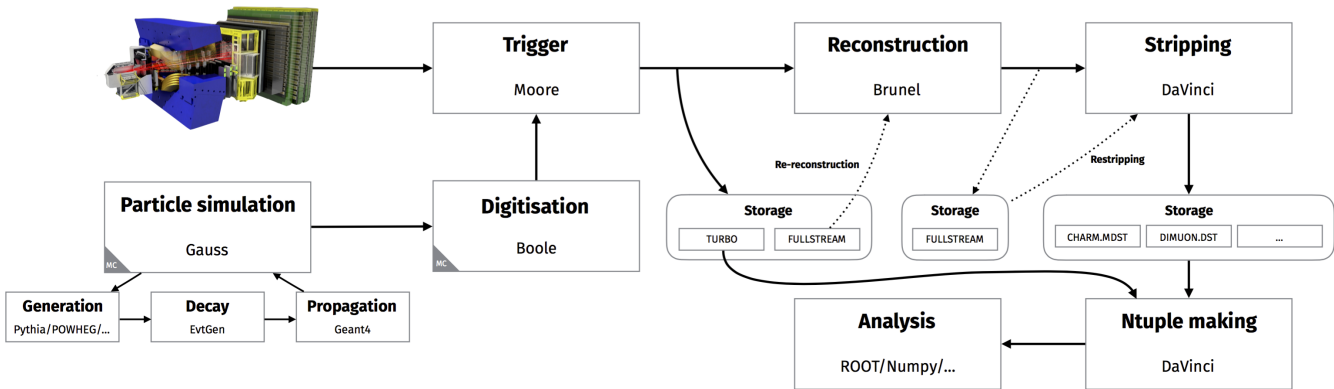
The recorded LHCb Run 2 data sample corresponds to an integrated luminosity of  $6 \text{ fb}^{-1}$ . Several digital copies of the data sample are stored at CERN and multiple external server farms (Tier-1 nodes) of the worldwide LHC computing grid (WLCG)

[81]. The raw data samples are stored on tape and amount to  $\sim 38$  PB of storage capacity. The high level data sets, including simulated events and the user data, are distributed across the WLCG and correspond to  $\sim 33$  PB [82].

To make the data samples accessible for physics analyses, the detector output is translated into a high level output format via a chain of different custom software packages, as shown in Fig. 4.10. As mentioned before, the event information is already fully reconstructed during the HLT2 trigger evaluation, but the event size of the high level data format exceeds the bandwidth capacity of the storage farms. Therefore only a fraction of the reconstructed tracks are stored in a directly accessible TURBO stream, while the majority of events are stored in a raw format and re-reconstructed offline with the BRUNEL framework. Due to computing limitations, the resulting fully reconstructed event samples are “stripped” into several smaller sub-samples (streams), which are split by rough event categories, like the DIMUON stream which contains events with two muons in the final state. These samples are afterwards queried for events that fulfil analysis-specific preselection requirements, referred to as “stripping lines” which are centrally managed and only executed infrequently. The stripping streams and lines are implemented in the DAVINCI framework by the analysis teams. The same framework is then used by analysts

for on-demand reprocessing and downloads of the data of one or more stripping lines in the ROOT data format [83]. This final step usually generates sample sizes of up to several TB, which can be handled on local computers or small computing clusters to perform more specialized data analyses. All LHCb-specific software packages – including the HLT2 software, which is bundled in the MOORE package – are built upon the GAUDI framework [84].

A variety of software packages is used for the detailed offline analysis step, historically making use of the ROOT framework, but in recent years increasingly utilizing the python data science landscape, which can be seen in Fig. 4.11. The software stack that has been used for the here presented analysis is described in more detail in Section 5.3.



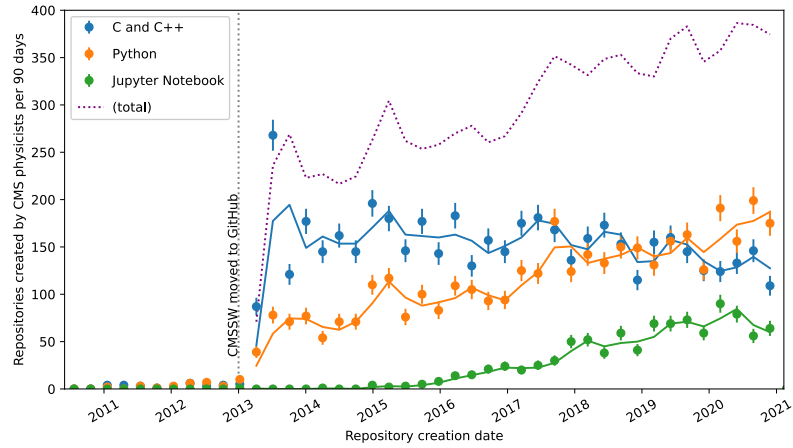
**Figure 4.10:** The data processing pipeline for the LHCb experiment [85]. Several custom software packages are used to transform the detector response into a reduced data sample with high-level features. Different simulation frameworks are furthermore capable to simulate a detector output with current best knowledge of the underlying physics processes.

#### 4.2.4 Monte Carlo Simulation

Today’s good understanding of particle physics processes allows to numerically simulate a large portion of the experiment. These simulations are not only essential to the development of the physics data analyses, but indispensable for the planning, design and operation of a particle physics experiment.

Starting at the  $pp$  collisions, a spectrum of QCD interactions is generated using Monte Carlo techniques, reproducing real LHC data to best knowledge. Such particle generators are implemented in several software packages, out of which PYTHIA is commonly used for LHCb simulations [87]. The hadronization and decay of the particles are modelled with the EVTGEN framework, capable of good description of  $b$  hadron decays, including effects such as  $CP$  violation [88]. Interactions with the detector material are afterwards simulated using GEANT4 [89]. Within the LHCb software stack, all of the aforementioned packages are bundled inside the GAUSS package, as shown in the lower part of Fig. 4.10, which is also based on the GAUDI framework.

## 4 The Laboratory



**Figure 4.11:** Popularity of python and C-languages among CMS analysts [86]. A trend towards python-based data analysis is clearly visible and similarly expected across the field.

Ultimately, the `BOOLE` framework uses the Monte Carlo simulation to mimic a detector response that can be piped into the nominal data reconstruction software. The resulting simulated samples are afterwards structurally indistinguishable from real data samples, except for optionally accessible simulated information. Moreover, the resource-intensive simulations are limited to exclusive physical processes, making it infeasible for the simulation of combinatorial background.

## 5 Data Analysis Methods

Despite the data preprocessing that has been described in the previous Chapter, LHCb data samples usually contain a variety of different decay processes, including large fractions of non-signal decays. To allow a thorough analysis of the physics processes, and therefore resolve the left sides of Eqs. (3.36) to (3.39), these non-signal decays must be further suppressed from the data samples.

As a starting point, signal events are usually selected based on specific requirements for the exclusive decay channels. This selection step is often supported with different machine learning algorithms that allow to classify data points based on multiple data features simultaneously, exploiting possible correlations among those features. Two kinds of such methods are introduced in Section 5.1.1 and Section 5.1.2, first of which is used in the analysis presented here, as described in Section 6.2.2. The second set of algorithms is relevant for different flavour tagging algorithms which are described in Chapter 7.

The extraction of the physics parameters is usually performed using a Maximum Likelihood method, which is explained in Section 5.2. It can be further used to statistically extract a pure sample of signal events from mixtures of signal and non-signal events, as described in Section 5.2.1.

In practice, the exact combination and parametrisation of these tools must be studied extensively and the concrete analysis strategy only emerges in the course of the analysis. This process can be simplified if modern tools for software automation are adapted. A set of tools that have proven helpful to the author and lessons that have been learned are described in Section 5.3. Following these principles not only simplifies the data analysis development, but also helps to achieve good reproducibility.

### 5.1 Multivariate Analysis

Many particle physics analyses profit from the precise understanding of the underlying processes, such that powerful data features can be worked out. As a result, it is possible to obtain very clean data samples by applying individual selection requirements to these parameters. As explained in Section 6.2.3, this is also exploited for the measurement of  $\Delta m_s$ .

In addition to that, multivariate machine learning methods allow to identify correlations in higher dimensions of the feature space. These multivariate methods proved to be a powerful extension to the tool set, and evolved to a crucial ingredient to most particle physics analysis. While the applications of these algorithms is extremely versatile, the following sections will focus on classification problems of labelled data samples.

Here, two classes are defined (often labelled as signal and background), which the algorithms learn to distinguish, based on a sample of training data with a set of training features. This is achieved by adjusting the internal parameters of the algorithms such that it estimates the labels of the training data. The quality of the estimations is based on an objective function, which is often equal to a loss function  $L$ , which describes the difference between the true and the predicted classes. Common choices of loss functions include the mean squared error, or the logistic loss,

$$L_{\text{mse}}(\hat{y}, y) = \sum_i (y_i - \hat{y}_i)^2, \quad (5.1)$$

$$\text{or } L_{\text{log}}(\hat{y}, y) = \sum_i [y_i \ln(1 + e^{-\hat{y}_i}) + (1 - y_i) \ln(1 + e^{\hat{y}_i})], \quad (5.2)$$

where the predicted label of the  $i$ -th data point is denoted as  $y_i$ , the true label is  $\hat{y}_i$  and the sum runs over the data sub-sample for which the function is evaluated.

Machine learning algorithms are generally prone to the effect of over-training: Instead of identifying distinctive high level structures in the data, the algorithms can potentially memorize the individual training data points. To prevent over-training, the data sample is usually split into one training sample and one or more validation samples for which the trained algorithm predicts the class labels. The loss function must be monitored simultaneously on training and test sample and the training must be stopped once the loss on both samples diverges. Most machine learning algorithms can be configured with a set of hyper parameters, which influence the performance and amount of overtraining that is achieved on a given data sample.

### 5.1.1 Decision Trees and Boosting

Several powerful machine learning libraries are based on decision tree ensembles. A decision tree categorizes data by applying successive selection criteria, as depicted in Fig. 5.1: Different features are used to split the data into sub-trees, while minimizing the loss function, until a desired tree depth is reached and the data categories can be read off the leaves. In the case of binary classification, a weight can be assigned to each leaf, such that positive and negative numbers correspond to one or the other class label, respectively. The good predictive power of decision tree algorithms originates from the combination of many trees and the way in which weights are assigned to the leaves. The summation of multiple decision trees turns out to yield especially powerful predictions, if each additional tree further reduces the loss of the existing trees. This type of algorithm is referred to as Boosted Decision Trees (BDTs), where the prediction  $\hat{y}$  is the simple sum of  $t$  decision trees  $f_k(x)$ :

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i). \quad (5.3)$$

The predictive power of BDTs can further be improved by using different regularization methods, such as training on random sub-samples of the data, or penalizing complex decision trees. The latter approach is implemented in the XGBoost framework that is also used later in this work [90].



Here, the regularization term  $\Omega$  is added to the objective function, taking into account the number  $T$  of leaves per tree, the weight  $\omega_j$  that is assigned to each leaf, and the regularization factors  $\gamma$  and  $\lambda$ :

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2.$$

In the case of binary class labels, the sum of the decision tree outputs is transformed to a probability estimate via the logistic function

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (5.4)$$

The example shown in Fig. 5.1 corresponds to a two-tree BDT, trained with XGBoost using a small fraction of the data sample of  $B_s^0 \rightarrow D_s^- \pi^+$  decays, which is described in more detail in Section 6.2.2. Only two features are considered for training: The transverse momentum  $p_T$  of the leftover pion candidate and the quality of the primary vertex impact parameter fit,  $\chi^2(\text{IP})$ . The BDT is trained to distinguish reconstructed invariant  $m(D_s^\mp \pi^\pm)$  masses around the true  $B_s^0$  mass against  $B_s^0$  candidates with a mass greater than  $5450 \text{ MeV}/c^2$ . The confusion matrix obtained for this two-tree ensemble is shown in Table 5.1a. The columns of the matrix contain the true number of data points per class, the rows contain the predicted number of data points. This simple example yields a signal efficiency of  $\sim 65\%$  at a background rejection of  $\sim 70\%$ .

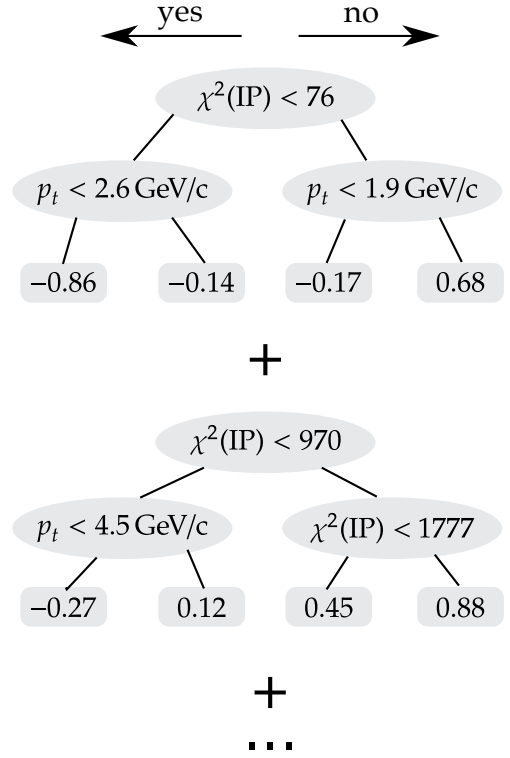
The distributions of the input features is shown in Fig. 5.2, together with the mass distribution of the classified samples. The latter furthermore shows the mass distributions obtained with a neural network, another classification algorithm described in the next section.

The performance of MVA classifiers is often compared via the area under the receiver-operator characteristic (ROC) curve, for which the true positive rate is calculated for different values of the false positive rate. It highlights the fact that any specific choice of MVA working point is a trade-off between the properties.

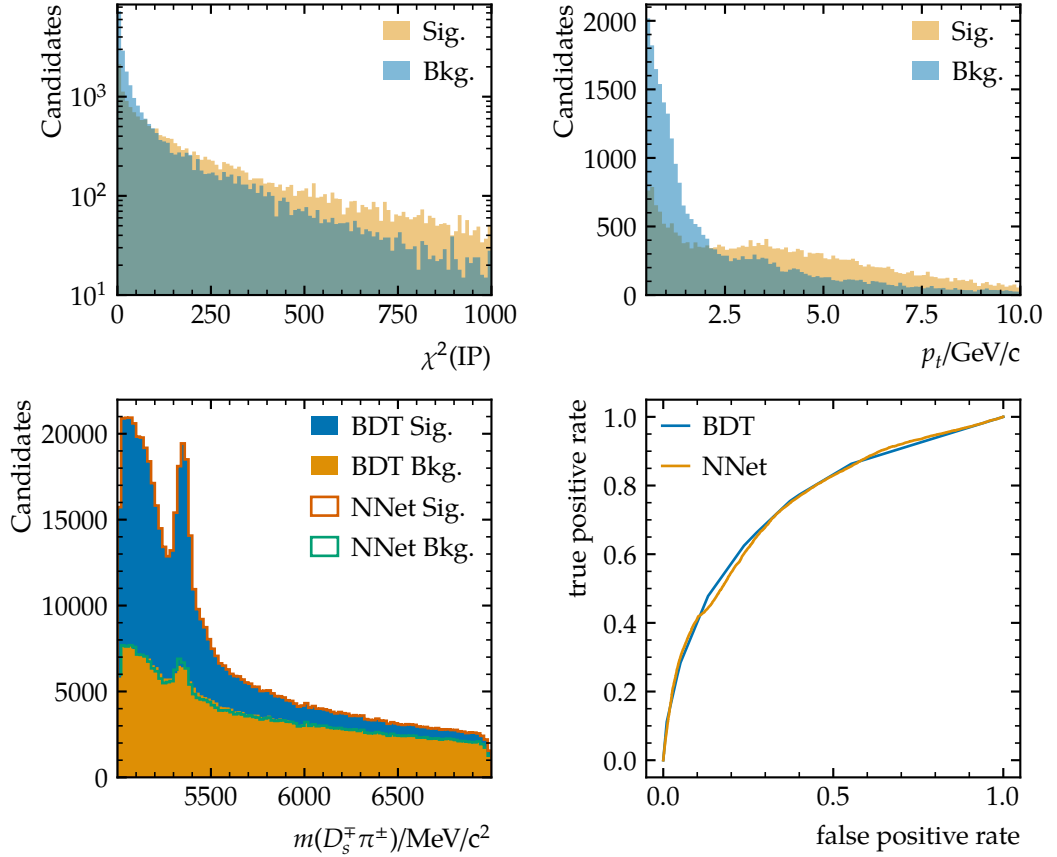
### 5.1.2 Neural Networks and Back Propagation

Artificial Neural Networks are similarly wide-spread in particle physics analyses, obtaining comparably good results as BDTs. Despite the similar interface, the internal structure of a neural networks is very different. The basic building block of a neural network is a linear combination of a vector  $x$  of training features, weighted with a weight matrix  $W$ ,

$$f(x) = Wx. \quad (5.5)$$



**Figure 5.1:** A minimal XGBoost BDT, using two internal tree. The BDT response is obtained by summation of the leaf weights.



**Figure 5.2:** Input feature distributions (top) and invariant  $m(D_s^\mp \pi^\pm)$  mass distribution for predicted events (bottom) by a minimal BDT and neural network. Both perform very similar, correctly classifying a large fraction of events in the distribution's tail as background.

**Table 5.1:** Confusion matrices of an minimal BDT (a) and a minimal neural network (b). Both algorithms are trained on a small sample of  $B_s^0 \rightarrow D_s^- \pi^+$  candidates, to suppress combinatorial background.

(a) BDT.			
Pred. \ True	Pos.	Neg.	
Pos.	9070	3840	
Neg.	4824	10 166	

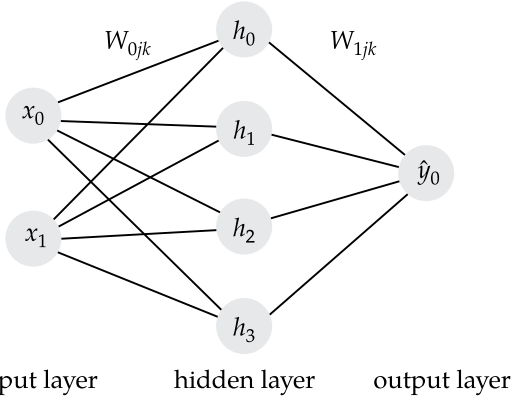
(b) Neural network.			
Pred. \ True	Pos.	Neg.	
Pos.	9260	4078	
Neg.	4634	9928	

Usually, a bias vector  $b$  is added to this term and a non-linear function  $\phi$  is applied, also referred to as activation function. A network is obtained by nesting these functions multiple times. A two-staged network can therefore be written as

$$\hat{y} = f(x) = \phi_1(\mathbf{W}_1 \phi_0(\mathbf{W}_0 x + b_0) + b_1). \quad (5.6)$$

This structure allows to model non-linear dependencies between the input vector  $x$  and output function  $f$ , depending on the weights  $\mathbf{W}_j$ , the bias  $b_j$ , and the specific form of the activation function  $\phi$ . A network of the form Eq. (5.6) is referred to as a fully-connected network, with a single input, output and hidden layer. Additional recursion steps add additional hidden layers and increase the complexity of the functions that can be modelled. A common representation of this network is shown in Fig. 5.3, where a similarity between synapses and neurons of a biological brain can be seen, giving rise to the name of these algorithms.

$$h = \phi_0((\mathbf{W}_0 x + b_0)_j) \quad \hat{y} = \phi_1((\mathbf{W}_1 h + b_1)_j)$$



(a) Structure of a minimal neural network. The mathematical form is shown at the top. Weights are represented as edges in the graph and the evaluation of the activation function and the bias are shown as nodes.

$$\mathbf{W}_0 = \begin{pmatrix} 0.001 & -0.057 \\ -0.035 & -0.028 \\ -0.04 & -0.026 \\ -0.001 & -0.001 \end{pmatrix}, \quad b_0 = \begin{pmatrix} 2.384 \\ -0.133 \\ -0.021 \\ 3.761 \end{pmatrix}$$

$$\mathbf{W}_1 = \begin{pmatrix} -1.105 \\ 0.009 \\ 0.016 \\ -1.809 \end{pmatrix}, \quad b_1 = (1.357),$$

$$\phi_1(x) = \phi_2(x) = \frac{1}{1 + e^{-x}}.$$

(b) Weights of the network shown in (a), which yield a similar performance as the BDT shown in Fig. 5.1.

**Figure 5.3:** The structure (a) and internal parameters (b) of a minimal neural network.

The predictive power is obtained by adjusting the internal parameters (weights and biases) of the network to minimize a loss function, such as Eq. (5.2). This can be achieved by building the gradients of the loss function  $L$ , with respect to all internal parameters  $W_{ijk}$  and  $b_{ijk}$ , which is referred to as gradient descent, and similarly relevant in Section 5.2. While the number of parameters can be large, the chain rule allows to calculate the gradient with respect to any parameter as the product of all contributing local gradients, which is also referred to as back propagation. The internal parameters can then be adjusted towards a minimal loss function,

$$W'_{ijk} = W_{ijk} - \alpha \frac{d}{dW_{ijk}} L(f(x_i), y_i), \quad (5.7)$$

with the learning rate  $\alpha$ . A model with the structure depicted in Fig. 5.3 has been trained on the same minimal data sample, as described in the previous Section. The number of 17 free parameters of the neural network is similar to the number of 14 free BDT parameters. Despite this, the neural network is more sensitive to the set of hyper parameters, and choice of activation functions, such that a similar performance to the BDT is only achieved with a specific configuration. The confusion matrix for this example is shown in Table 5.1b.

While the maths of fully connected neural networks is straight forward, the general idea has been modified in various ways in the past. A comprehensive review of the field is given in [91]. Neural networks have been increasingly popular over the last years, which is substantially driven by the relatively low cost of parallel computing capabilities.

### 5.1.3 Shapley Values and Interpretability of Multivariate Analysis Algorithms

The examples of a BDT and neural network in the previous sections are very simplified, such that the algorithms can be followed conceptually. Despite this, they provide reasonably good predictions. While this is an example for the large potential of machine learning techniques, it is hard to find any causality for the internal weights of these examples. This is even more the case in practice, where usually BDTs with several 100 trees or deep neural networks with more than  $10^6$  to  $10^9$  parameters are used.

One approach to explain the prediction process of such tools is based on a game theory method to assign fair rewards to competitors of a game [92]. This can be transferred to machine learning algorithms, where a feature corresponds to a competitor trying to “win” the model estimation. The corresponding value for a feature  $i$  is then obtained by retraining the algorithm using all possible subsets  $S$  of the set of features  $F$ , and comparing the models with and without the feature,

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} (f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)). \quad (5.8)$$

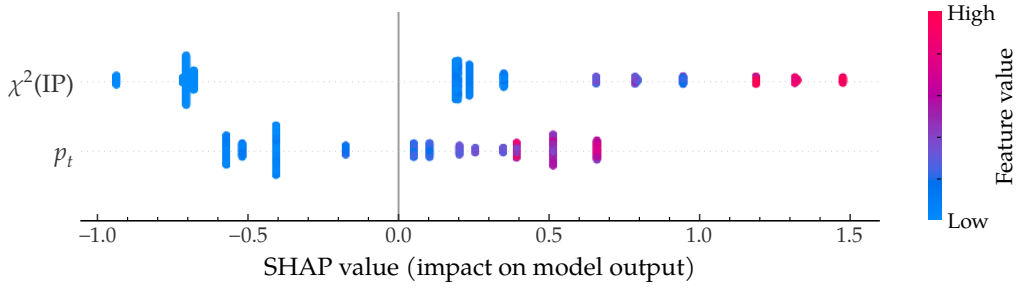
This was proposed by Lloyd Shapley in the early 1950s for which – among other contributions – he was awarded the nobel price of economics in 2012.

Based on this idea, SHAP values have been proposed as a model-agnostic measure of feature importance [93]. For a given data point of any classification algorithm, an explanation model can be found which resembles the original prediction as a linear combination of the binary feature vector  $z'$ ,

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \approx f(x). \quad (5.9)$$

The values  $\phi_i$  can be determined especially efficiently for tree ensemble algorithms. As an example, the SHAP values for each test data point of the BDT example of Section 5.1.1 is shown in Fig. 5.4.

This technique is utilized in a study of new flavour tagging algorithms, which is explained further in Section 7.2.



**Figure 5.4:** SHAP values swarm plot for 1000 classified events. The feature values are colour-coded. Negative SHAP values clearly correlate with small feature values. For positive SHAP values, a high level structure appears for the transverse momentum  $p_t$  (different colours at SHAP  $\sim 0.5$ ).

## 5.2 Maximum Likelihood Fitting

As shown in Section 3.4, the mathematical description of the decay-time distribution depends on several physical parameters, connecting the theory predictions with the measurable decay-time distribution of neutral mesons. Therefore, a measurement of the parameters of interest can be performed by finding the set of parameters that best describe the data distributions. Analogous to the previous sections, the level of agreement between theoretical model and data is described by a loss function that is minimal at perfect agreement.

Within the Maximum Likelihood method, any positive, normalized function can be interpreted as a probability  $\mathcal{P}(x_i, \lambda)$  to measure a vector of observables  $x_i$ , where the vector  $\lambda$  describes all other parameters of the function. The best set of parameters is then obtained by maximizing the product of the likelihood for all measurements,

$$\hat{\lambda} = \arg \max \mathcal{L}(\lambda), \quad \text{with} \quad \mathcal{L}(\lambda) = \prod_{i=1}^N \mathcal{P}(x_i, \lambda). \quad (5.10)$$

For large samples  $N$ , the product cannot be evaluated numerically. Moreover, numerical optimizers conventionally minimize a function, which is why in practice the negative logarithm of the likelihood is minimized to obtain the measurement

$$\hat{\lambda} = \arg \min(-\log \mathcal{L}(\lambda)), \quad \text{with} \quad \log \mathcal{L}(\lambda) = \sum_{i=1}^N \log \mathcal{P}(x_i, \lambda). \quad (5.11)$$

Various algorithms have been proposed to determine the best parameters  $\hat{\lambda}$ . A common choice for a minimization framework is MINUIT, which dynamically combines different algorithms [94], is integrated in the ROOT framework and also available with a python API [95].

### 5.2.1 Unfolding Data Distributions in Extended Likelihood Fits

As shown in Fig. 5.2, a data selection often requires a trade-off between high purity and high statistics. Therefore, usually some irreducible fraction of background events remain in a data sample, which can potentially bias the measured parameters.

The sPlot method [96] allows to statistically extract a pure signal component from the data sample. For a discriminating variable, such as well-known mass distributions, each component in the data sample can be modelled via a likelihood function. This can be exploited to extract a set of weights, which can be applied to any control variable that is uncorrelated with the discriminating variable. The weighted distributions correspond to statistically pure samples of the respective component.

The method relies on the result of an extended maximum likelihood fit, where the likelihood Eq. (5.10) is extended with a Poisson term,

$$L(\lambda) \rightarrow L_{\text{ext}}(\lambda, \nu) = \frac{\nu^N}{N!} e^{-\nu} \prod_{i=1}^N \mathcal{P}(x_i, \lambda), \quad (5.12)$$

such that the log-likelihood reads

$$\log L_{\text{ext}}(\lambda, \nu) = -\nu + \sum_{i=1}^N \log(\nu \mathcal{P}(x_i, \lambda)) + C, \quad (5.13)$$

with a constant factor  $C$  that can be ignored for the minimization. A fit model with  $N_s$  components can then be described with the sum of  $N_s$  log-likelihood terms. The sWeights for the  $n$ -th component are then defined as

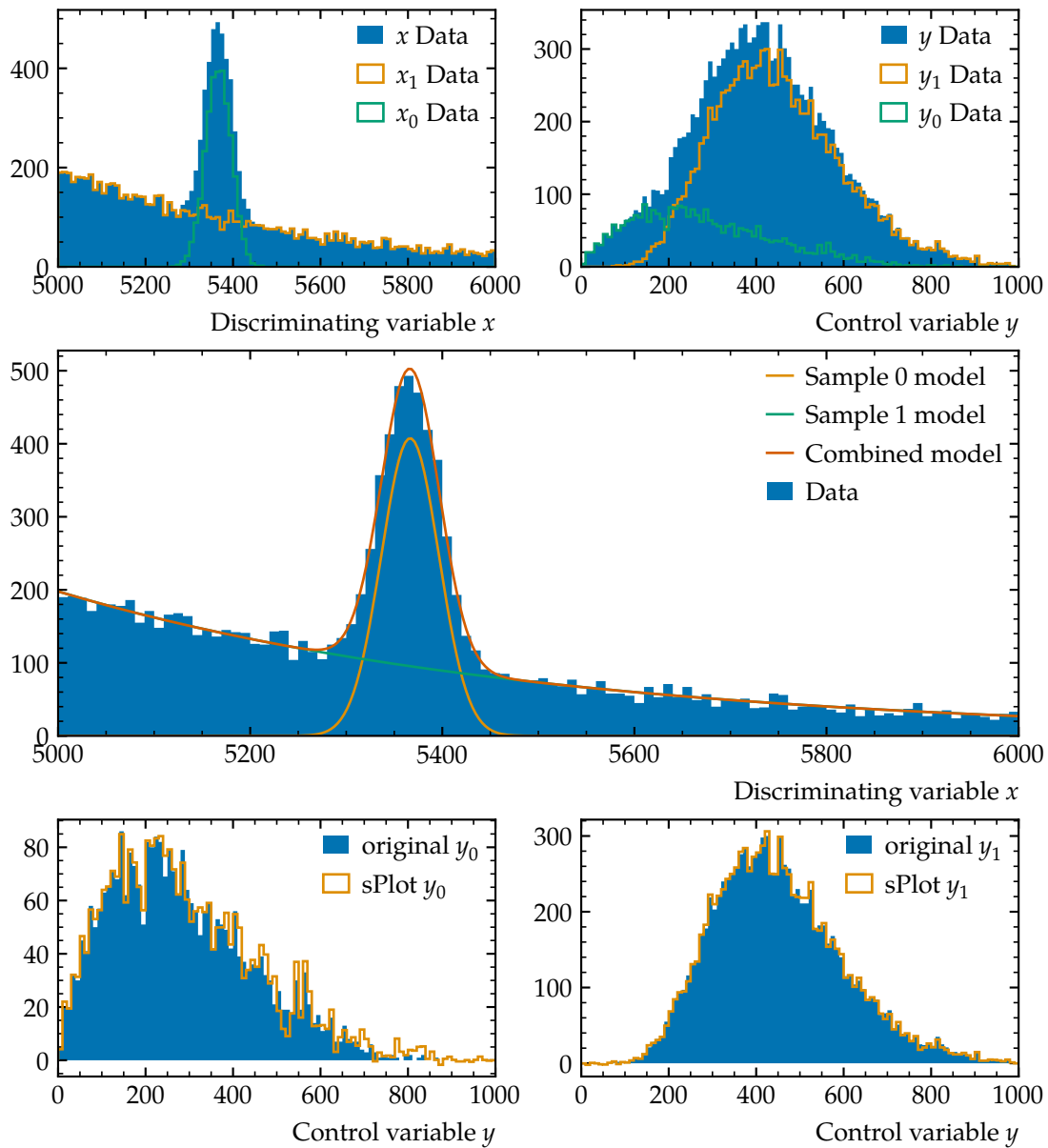
$${}_s w_n(x_i) = \frac{\sum_{j=1}^{N_s} V_{nj} \mathcal{P}_j(x_i)}{\sum_{k=1}^{N_s} N_k \mathcal{P}_k(x_i)}, \quad (5.14)$$

where the inverse of the matrix  $V$  is given by

$$V_{nj}^{-1} = \sum_{i=1}^N \frac{\mathcal{P}_n(x_i) \mathcal{P}_j(x_i)}{\left( \sum_{k=1}^{N_s} \nu_k \mathcal{P}_k(x_i) \right)^2}. \quad (5.15)$$

An example of the sPlot method is shown in Fig. 5.5: Two data samples are generated, each with two variables. The discriminating variable is drawn from an exponential and a Gauss distribution which can be clearly distinguished in the upper left histogram. The control variable is drawn from similar beta and gamma distributions, as seen in the upper right histogram. The result of a likelihood fit to the control variable is shown in the middle plot, where the fit model can well describe the data sample. The initial histograms, as well as the sWeighted histograms of the test variable for both data sample components are shown in the lower plots. The distributions are in good agreement.

An essential requirement for the sPlot method is the independence of control and test variables. While a small correlation between the two does not necessarily invalidate the method, the applicability must be rigorously checked, which is described for the  $\Delta m_s$  measurement in Sections 6.3 and 8.4.2.



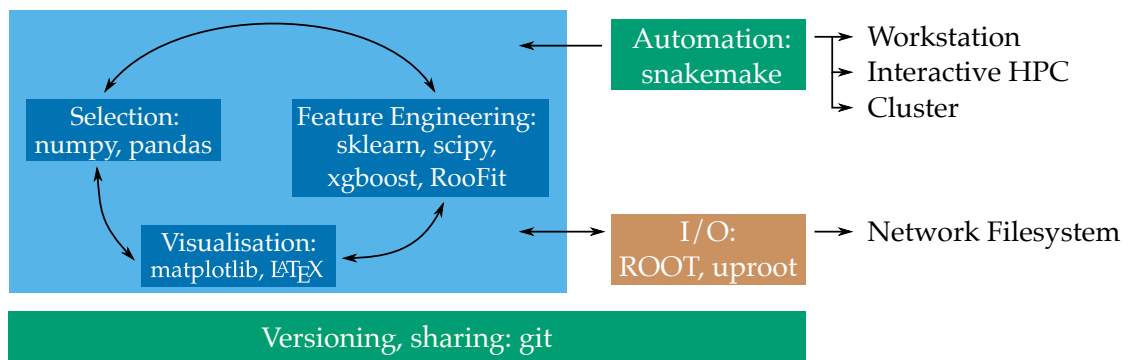
**Figure 5.5:** Example of the sPlot technique. The discriminating variable  $x$  (top left) and control variable  $y$  (top right) are randomly drawn from two different models. The first sample  $(x_0, y_0)$  originates from a Gaussian and a beta distribution; the second sample  $(x_1, y_1)$  is drawn from an exponential and a gamma function. A fit of the initial data model to the  $x$  distribution allows to extract the statistically pure components from the combined  $y$  data.

### 5.3 Analysis Automation and Reproducibility

Modern particle physics analyses profit from a variety of tools and methods to extract interesting parameters from big data samples. With increasing precision however, these tools have a potentially increasing effect on the systematic uncertainties and hence need to be controlled and tested rigorously. This leads to very dynamic development processes of physics analyses. The data pipelines grow in complexity as the number of effects that need to be taken into account increases.

This section lists a set of practices that have proven useful to the author. While many of these have also been formulated in a more rigorous and abstract way in [97], this remains a personal point of view.

A generalized sketch of a high energy physics data analysis is shown in Fig. 5.6, together with various software frameworks that are used in this thesis. The blue box of the diagram represents the abstract analysis steps that are iteratively applied during the analysis and generally apply to all data analyses. The brown box represents the input/output system that is conventionally defined via the central experiment software. The green boxes represent optional design principles, aiming to improve various aspects of the analysis.



**Figure 5.6:** Basic steps of a general particle physics data analysis (blue boxes), the common data I/O interface (brown box), and optional tools that improve reproducibility (green boxes). Different software packages that are used at the respective steps are listed.

Modern particle physics analyses should aim for a high level of analysis automation. One obvious reason for this is an improved reproducibility: While it is in principle possible to reconstruct a particle physics analysis from the detailed internal notes, it is in practice crucial to have access to the analysis software, software environment, and execution instructions. Especially the latter is hard to maintain, as it highly depends on the personal work flow and software environments that might evolve during the course of an analysis. Automation solves the first issue, by adding the execution instruction to the analysis pipeline itself. It furthermore adds an incentive to maintain compatibility of all parts of the analysis with a common software environment.

The python ecosystem provides an especially good tool belt for the development of



flexible, file-centric data analyses. Due to the good integration with other languages and libraries it is a good candidate as a central data analysis language. The `SNAKEMAKE` package [98] is extensively used in this analysis as a central automation tool, allowing to steer data flow within different software environments. It scales well from prototype analyses to a modularized analysis pipeline such as presented in this thesis. The dependency graph for the main part of the analysis is shown in Fig. 5.7, where the entry point includes the results of the next two Chapters, and the final rule collects several central figures of this thesis.



**Figure 5.7:** Dependency graph of the analysis described in the next Chapters. The entry point of the graph (top) corresponds to the analysis steps explained in Section 7.3, after which the selection requirements from Section 6.2 are applied. The central node “`run_md_fit`” corresponds to Section 6.3, followed by the decay-time fit, described in Chapter 8, and various visualisation tasks.



## 6 Processing of $B_s^0 \rightarrow D_s^- \pi^+$ Data

The LHCb dataset, recorded during LHC Run 2 corresponds to a integrated luminosity of  $\mathcal{L}_{\text{int}} = 6 \text{ fb}^{-1}$ . This includes data from the order of  $10^{14}$  bunch crossings in the detector from which  $10^9$  proton-proton collisions have been recorded, as described in Section 4.2. Considering the hadronization rate of  $B_s^0$  mesons at the centre-of-mass energy  $\sqrt{s} = 13 \text{ TeV}$ , as well as the decay fraction  $\Gamma(B_s^0 \rightarrow D_s^- \pi^+)/\Gamma(\text{tot})$  of the signal channel [99], and the detector efficiency  $\varepsilon_{\text{total}}$  [80], this data sample is expected to contain around  $10^7$  signal candidates.

The data sample is described further in the following Section 6.1. Afterwards, the data processing steps needed to extract a clean sample of  $B_s^0 \rightarrow D_s^- \pi^+$  candidates is described. This sample is finally used to extract sWeights, for which the fit procedure is described in Section 6.3, and which are ultimately used to measure the oscillation frequency  $\Delta m_s$  as described in Chapter 8.

### 6.1 Data Sample Properties

The notation of “an event” is usually referred to as a single  $B$  meson, decaying into a well-defined final state from which a signal can be reconstructed. In this case the corresponding data attributes can easily be stored in a tabular manner. In reality though, several effects make a hierarchical data structure necessary: With an average of  $\mu = 1.1$   $pp$  interactions per per bunch crossing, some events contain multiple PVs (the other LHC experiments have many more, with  $\mu \sim \mathcal{O}(100)$ ); the reconstructed tracks can be combined in various ways, yielding signal candidates of various quality; and similarly, some reconstruction algorithms might return multiple solutions for a given set of information. Therefore the data is stored in a custom data format that allows to represent the complex data dependencies for every event.

A set of selection requirements, which is described in the upcoming sections, is therefore applied after which only a small fraction of  $f_{\text{mult}} = 0.5\%$  of multiple candidates remain per event. Expanding these candidates leaves a tabular data sample, with a single row per candidate that can be easily passed down the analysis pipeline, as described in Section 6.3.

The data sample is generally split into the years of data taking, while the 2015 and 2016 periods are combined due to the small statistics of the first one. Moreover, the data sample is split into four different  $D_s^-$  final states, three of which are reconstructed from a  $K^- K^+ \pi^-$  signature, namely  $\phi\pi^-$ ,  $K^{*0}K^-$ , and non-resonant (n.r.)  $K^- K^+ \pi^-$ , as well as a  $\pi^- \pi^+ \pi^-$  signature. As explained in more detail in Section 6.2.3, the modes are distinguished via the invariant masses of some of the  $D_s^-$  children, as well as their PID information.

Splitting the data sample like this allows for different selection requirements and fit parametrisations, depending on the kinematic properties of the particle hypotheses. In the following, the final state particles are sometimes indexed with a number, such as  $B_s^0 \rightarrow \pi_1^+(D_s^- \rightarrow K_3^- K_4^+ \pi_5^-)$ , where in case of the  $\pi^- \pi^+ \pi^-$  signature, the higher momentum  $\pi^-$  is assigned the lower index. As the symbols for the  $B_s^0$  and  $D_s^-$  mesons are unambiguous, their respective indices are omitted.

Due to the large computational requirements (see Section 4.2.3), the initially recorded data sample cannot be treated with real-time, client-based data analysis algorithms. The initial data sample contains  $9 \cdot 10^8$  candidates with a set of 604 features per candidate, summing to a total disk size of 1 TB. The final data sample, produced after the fit described in Section 6.3, contains  $4 \cdot 10^5$  event candidates, with 20 features, occupying 5 GB on disk, and therefore being well manageable on modern personal computers.

The choice of reducing a data sample to fit into a single workstation's memory comes at the cost of increased data analysis iteration times and data dependency.

### 6.1.1 Simulated Data Samples

Samples of simulated events are used for several aspects of this analysis. They are obtained as described in Section 4.2.4. In total, 18 M simulated signal events are available for this analysis. In addition, simulated samples of  $B_d^0 \rightarrow D^- \pi^+$ ,  $\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+$ ,  $B_s^0 \rightarrow D_s^{*-} \pi^+$ , and  $B_s^0 \rightarrow D_s^\mp K^\pm$  are used for different studies of background candidates in this analysis (see Sections 6.2.3, 6.3, 8.2 and 8.4.2). All selection requirements that are described in the upcoming sections are also applied to simulated events, if not stated otherwise.

In a study, performed by Alessandro Bertolin, event weights are extracted to correct the PID information of the simulated samples, which is explained in Ref. [100, 101].

## 6.2 Signal Candidate Selection

The fraction of non-signal events is reduced by first requiring the signal candidates to fulfil a set of preselection requirements (see Section 4.2.3), after which a more specific set of offline selection requirements is applied. An additional multivariate selection with a BDT is used to suppress a large fraction of combinatorial background.

The full set of offline selection requirements is listed in Table 6.3, at the end of this section. More details about the individual selection requirements are listed in the following.

### 6.2.1 Distributed Data Preparation and Local Data Selection

As described in Section 4.2.3, the data preselection is distributed into the CERN world wide computing grid, and the selected data samples are downloaded and combined for further usage. The preselection requirements are summarized in Table 6.1.

During preselection, the particle decay trees are re-evaluated using a Kalman filter [102], also referred to as decay-tree fitter. This enables the application of different, additional constraints, one of which requires the  $D_s^\mp$  children to resemble the known

invariant  $D_s^\mp$  mass of  $1968 \text{ MeV}/c^2$ . The corresponding value of the  $B_s^0$  mass is later used for fits to the invariant  $m_{B_s^0}$  mass distribution (see Sections 6.2.2 and 6.3). Another constraint requires the  $D_s^\mp$  meson momentum to point back to the primary vertex. The correspond decay-time is used for fits to the  $B_s^0$  decay-time distribution (Chapter 8).

All final state particles are required to have a large momentum of  $p > 1 \text{ GeV}/c$ , and a transverse momentum of  $p_T > 100 \text{ MeV}/c$ . Their ghost probability must be low,  $P_{\text{NN}}^{\text{ghost}} < 0.4$ , the vertex separation to any primary vertex must fulfil  $\text{IP}\chi^2 > 4$ , and the track reconstruction quality should be good, with  $\chi^2/\text{ndf} < 4$ . In addition, the leftover track must have a large momentum of  $p > 5 \text{ GeV}/c$  and  $p_T > 500 \text{ MeV}/c$ . The distance of closest approach between any of the tracks considered to form a charm candidate must be below  $0.5 \text{ mm}$  and the sum of their transverse momenta must be greater than  $1.8 \text{ GeV}/c$ . Moreover the invariant mass must be within  $\pm 100 \text{ MeV}/c^2$  of the known  $D_{(s)}^\pm$  masses. The vertex fit for these candidates must then yield a reasonable quality,  $\chi^2/\text{ndf} < 10$  and good vertex separation to any PV,  $\chi^2 > 16$ . Additionally, at least one  $D_{(s)}^\pm$  child must fulfil the same requirements as the leftover pion track. During preselection, the  $B_{(s)}^0$  candidate is required to form a good secondary vertex (SV) with the leftover pion, with a vertex  $\chi^2/\text{ndf} < 10$ , and it must be displaced from the PV, such that the reconstructed decay time is larger than  $0.2 \text{ ps}$ . Its flight trajectory needs to point to its associated PV, with an  $\text{IP}\chi^2 < 25$ , and the momentum vector must be parallel to the trajectory, such that the cosine of the angle between both is larger than  $0.999$ . Moreover a BDT based on the  $B_{(s)}^0$  transverse momentum, its vertex separation  $\chi^2$ , and the combined  $B_{(s)}^0$  and  $D_{(s)}^\pm$  vertex  $\chi^2/\text{ndf}$ , is used to reduce non-signal candidates with a signal efficiency close to 1 and background rejection around 90% [103]. Ultimately either the event candidates must have triggered a HLT2 topological or inclusive  $\phi$  (IncPhi) trigger lines (TOS) or any other decay in the event must have triggered one of these lines (TIS). The mass- and time distributions are shown in Fig. 6.1. Despite the preselection, especially the mass distribution shows a remaining vast amount of combinatorial background that is contained in the sample, which originates from random track combinations. This type of background consists of mistakenly combined tracks that pass the aforementioned selection, but do not originate from signal decays. They are instead randomly (and falsely) combined to a  $B_s^0$  meson. On top of that, a broad structure at a lower mass than the  $B_s^0$  mass at  $m_{B_s^0} = 5367 \text{ MeV}/c^2$  are clearly visible. In contrast, the decay time distribution does not contain clearly distinguishable features.

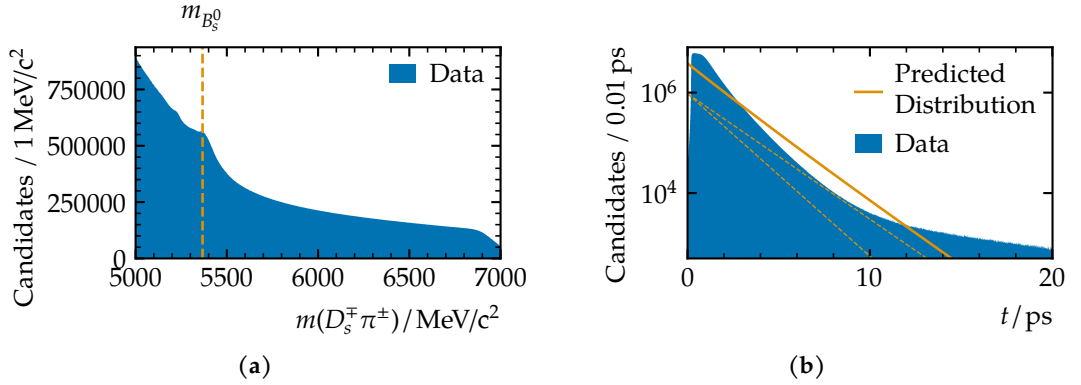
While it is computationally expensive to simulate combinatorial background data, machine learning algorithms such as BDTs prove to be extremely powerful to reduce this kind of data, which is described in more detail in the next section.

## 6.2.2 Multivariate Background Suppression

A data-driven BDT training is performed to efficiently suppress combinatorial background in the data sample. To prevent overtraining, the data sample is split in two, equally sized sub-samples using the normally distributed event number. The resulting sub-samples are referred to as even and odd sample, and a BDT is trained on each sample,

**Table 6.1:** Preselection requirements to select  $B_{(s)}^0 \rightarrow D_{(s)}^- \pi^+$  candidates.

Applied to	Feature	Requirement	Unit
Final state tracks lab{1, 3, 4, 5}	track $\chi^2/\text{ndf}$	< 4	
	$p_{\text{NN}}^{\text{ghost}}$	< 0.4	
	$p$	> 1	GeV/c
	$p_{\text{T}}$	> 100	MeV/c
	$\text{IP}\chi^2$	> 4	
leftover pion track lab1	$p$	> 5	GeV/c
	$p_{\text{T}}$	> 500	MeV/c
$D_{(s)}^\pm$ children lab{3, 4, 5}	$\sum p_{\text{T}}$	> 1.8	GeV/c
	$\text{DOCA}(i, j)$	< 0.5	mm
	$\text{abs}(m(h^- h^+ h^-) - m_{D_s^\mp}^{\text{PDG}})$	< 100	MeV/c <sup>2</sup>
	$\text{abs}(m(h^- h^+ h^-) - m_{D^-}^{\text{PDG}})$	< 100	MeV/c <sup>2</sup>
	track $\chi^2/\text{ndf}$	< 10	
	min. $\text{IP}\chi^2$	> 16	
at least one child	$p$	> 5	GeV/c
	$p_{\text{T}}$	> 500	MeV/c
$B_{(s)}^0$ candidate	vertex $\chi^2/\text{ndf}$	< 10	
	$t$	> 0.2	ps
	$\text{IP}\chi^2$ with PV	< 25	
	$\cos(p, \text{flight}), \text{DIRA}$	> 0.999	



**Figure 6.1:** Distribution of the invariant  $m(D_s^\mp \pi^\pm)$  mass (a) and the decay time (b), after preselection. A large amount of non-signal events is contained in the data sample. The expected position of a signal peak is marked in (a). The predicted slope of the decay-time distribution is shown in (b), where the dashed lines correspond to the heavy and light decay constants.

yielding one even and odd BDT. The respective other data sample is used as a validation and selection sample, such that the even BDT is evaluated to reject combinatorial background events from the odd sample, and vice versa.

The training features are chosen similarly to the LHCb Run 1 measurement of the CKM angle  $\gamma$  with  $B_s^0 \rightarrow D_s^\mp K^\pm$  decays [104]. In addition to features that are used in the preselection, they include the radial flight distance (RFD) of the  $B_s^0$  and  $D_s^\mp$  candidates, the vertex fit qualities,  $\chi^2/\text{ndf}$ , for both unconstrained and lifetime-constrained decay tree fits, and the angle  $\theta_{B_s^0}$  between a final state track and the parent  $B_s^0$  flight direction. All features are listed in Table 6.2.

Due to the narrow spikes in the DIRA distribution, the variable is transformed via

$$\text{DIRA}' = -\text{sign}(\text{DIRA}) \cdot \log(1 - |\text{DIRA}|), \quad (6.1)$$

preserving the feature's monotony within its definition range, while yielding more stable behaviour of the boosting algorithm.

**Table 6.2:** Features used for BDT training of even/odd BDT. The features selection originates from the previous LHCb  $\gamma$  measurement with  $B_s^0 \rightarrow D_s^\mp K^\pm$  [104].

Feature	
$B_s^0$ candidate:	leftover $\pi$ candidate:
minimum $\log(\text{IP}\chi^2)$	minimum $\log(\text{IP}\chi^2)$
DIRA w.r.t. PV, transformed	$p_T$
RFD	$\cos(\theta_{B_s^0})$
unconstrained vertex $\log(\chi^2/\text{ndf})$	Minimum within all $D_s^\mp$ children:
lifetime-constrained vertex $\log(\chi^2/\text{ndf})$	minimum $p_T$
$D_s^\mp$ candidate:	minimum $\log(\text{IP}\chi^2)$
DIRA w.r.t. SV, transformed	All final state particles:
minimum $\log(\text{IP}\chi^2)$	maximum $P_{\text{NN}}^{\text{ghost}}$
RFD	
unconstrained vertex $\log(\chi^2/\text{ndf})$	

The algorithm is implemented in the XGBoost package, version v0.82 [90], and default hyper-parameters are used except for the regularization parameter  $\gamma_{\text{XGB}} = 0.2$  and fixed learning rate  $\eta_{\text{XGB}} = 0.1$ . A fixed size of  $n_{\text{tree}} = 500$  trees with a maximum depth of  $d_{\text{max}} = 3$  and a minimum weight of  $w_{\text{min}} = 1000$  per node is used for both BDTs. While early stopping generally provides good regularization, the number of trees is fixed due to otherwise observed large differences between even and odd BDT's feature importance.

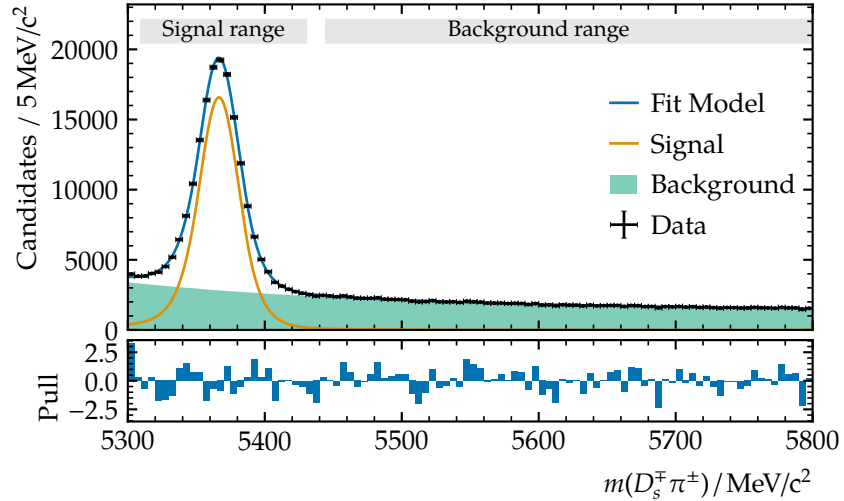
In addition to the data-driven BDT strategy, described in the following sub-section, a simulation-based training strategy has been tested. This shows similar performance, but comes with an additional need for data/simulation corrections. Therefore, and due to the good results obtained with the data-driven approach in previous analyses, latter is chosen as the nominal training strategy.

### Data-driven Training of a Boosted Decision Tree

The training labels for even and odd data sub-samples are assigned via different  $D_s^\mp \pi^\pm$  mass ranges, as shown in Fig. 6.2. Events in the upper mass side-band  $m(D_s^\mp \pi^\pm) \in [5445, 5800] \text{ MeV}/c^2$  are labelled as background and events around the known  $B_s^0$  mass,  $m(D_s^\mp \pi^\pm) \in [5310, 5430] \text{ MeV}/c^2$  are labelled as signal. As seen in Fig. 6.1, the signal range is clearly polluted with background data, which will negatively impact the distinctive power of the BDT.

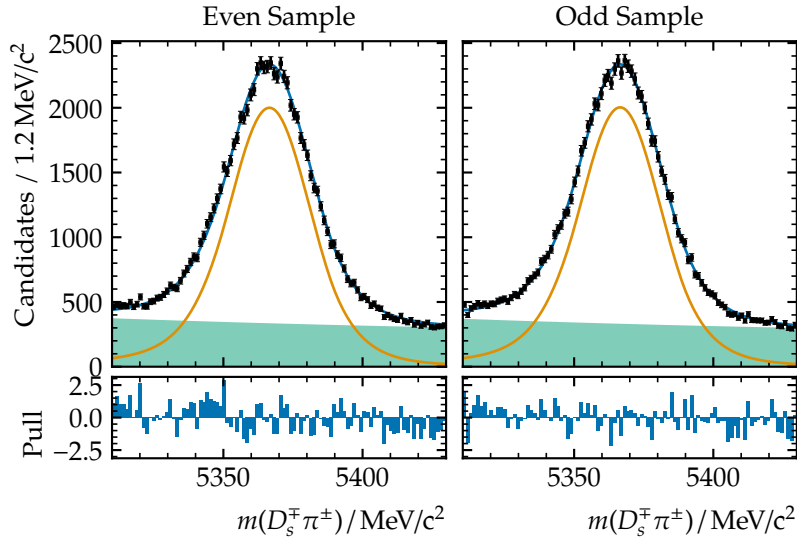
Therefore a simple set of pre-BDT requirements for PID and kinematic features is applied: Both kaons need to fulfil  $\text{PID}_K > 0$ ; a  $D^-$  veto is applied by requiring  $\text{PID}_K > 10$  for the same sign kaon  $K_3^-$ , or having the charm mass under  $D^-$  hypothesis outside of  $1839 \text{ MeV}/c^2$  to  $1899 \text{ MeV}/c^2$ ; and a similar  $\Lambda_c^+$  veto is applied, requiring  $\text{PID}_K - \text{PID}_p > 5$  for the same charge kaon, or having a charm mass under  $\Lambda_c^+$  mass hypothesis outside of the  $2255 \text{ MeV}/c^2$  to  $2315 \text{ MeV}/c^2$  range.

The remaining background pollution is suppressed in the BDT training by applying sWeights. To extract these weights, a simple PDF is fitted to the  $m(D_s^\mp \pi^\pm)$  mass distribution in the nominal fit range from  $5300 \text{ MeV}/c^2$  to  $5800 \text{ MeV}/c^2$ . The signal component is modelled with a double crystal Ball function, and its shape parameters are extracted from simulated events. The background component is modelled with an exponential function and a constant. The fit is also shown in Fig. 6.2. To obtain a correct sWeights normalization, the fit is repeated with fixed shape parameters in the signal mass range, separately for even and odd data sample. The corresponding refitted PDFs are shown in Fig. 6.3. Figure 6.4 shows the BDT accuracy on test and training data samples, for



**Figure 6.2:** Distribution of the  $m(D_s^\mp \pi^\pm)$  mass, after applying pre-BDT selection requirements. Different mass ranges are used to label BDT training data. As explained in Section 6.2.2, sWeights are applied to the signal labelled data.





**Figure 6.3:** Distribution of  $m(D_s^\mp \pi^\pm)$  data for even and odd data sub-samples, with refitted, mass PDF for a correct normalization.

increasing number of trained trees. The feature importance for even and odd BDT are shown in Fig. 6.5

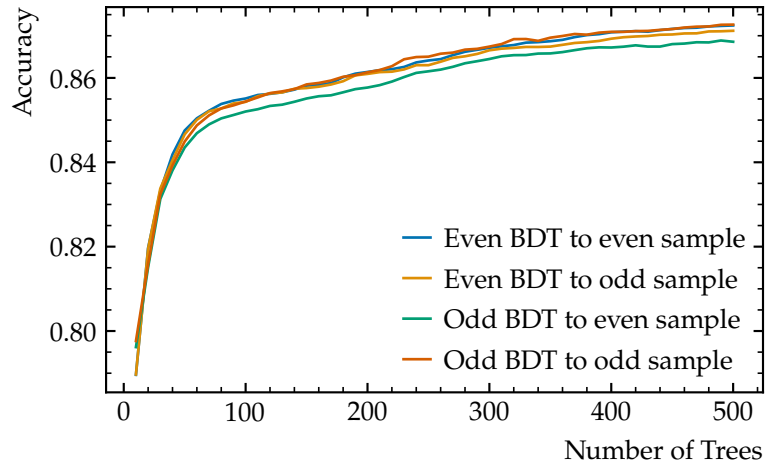
As discussed in Section 5.1.3, Shapley values describe a more generic way to interpret the relevance of individual features for multivariate classification. The discrete behaviour of decision trees and – compared to neural networks – small number of parameters allow to explicitly calculate these values for the here presented scenario. The average SHAP values for all features of 10 000 validation events are shown in Fig. 6.6. The distributions for the most important features, judging by XGBoost ranking and SHAP values, are shown in Fig. 6.7.

### Optimisation of Boosted Decision Tree Requirement

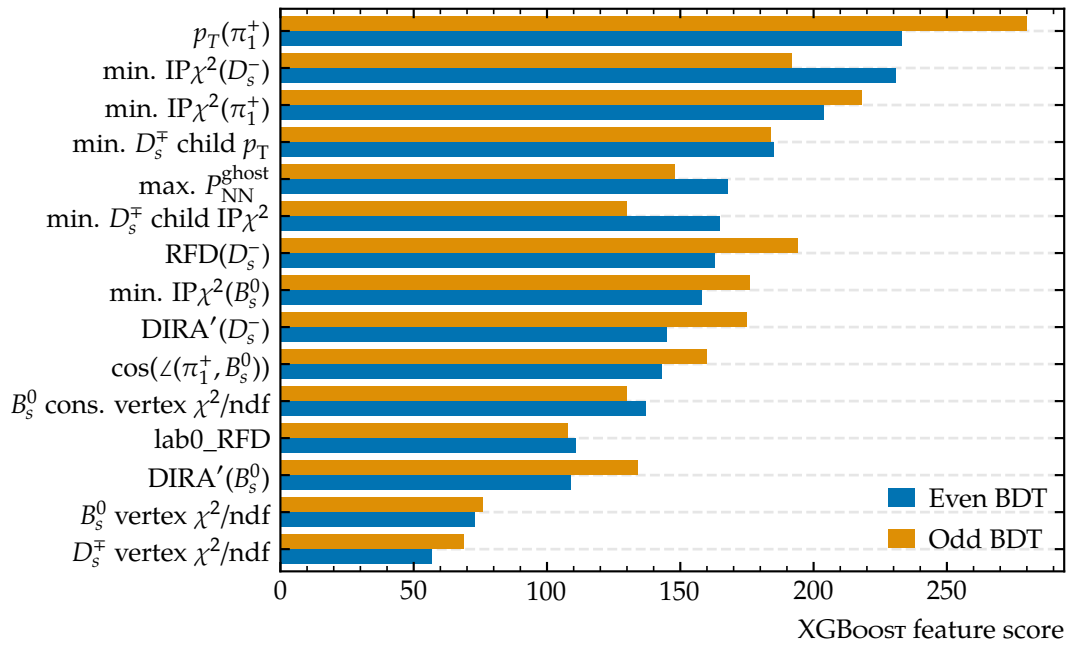
While the BDT training offers a good “guess” to classify events, the working point can be further tuned by adjusting the underlying BDT requirement. Since the decay mode  $B_s^0 \rightarrow D_s^- \pi^+$  is also well-suited control mode for studies of decays from  $B_s^0 \rightarrow D_s^\mp K^\pm$ , this BDT is also planned to be used for an upcoming measurement of the CKM angle  $\gamma$  in the aforementioned decay mode. Therefore, an ad-hoc figure of merit that has been used in the previous  $\gamma$  measurement is evaluated for different BDT requirements:

$$\text{FOM} = \varepsilon_{\text{sig.}} \cdot \frac{N_{\text{sig.}}}{\sqrt{N_{\text{sig.}} + N_{\text{bkg.}}}} \quad (6.2)$$

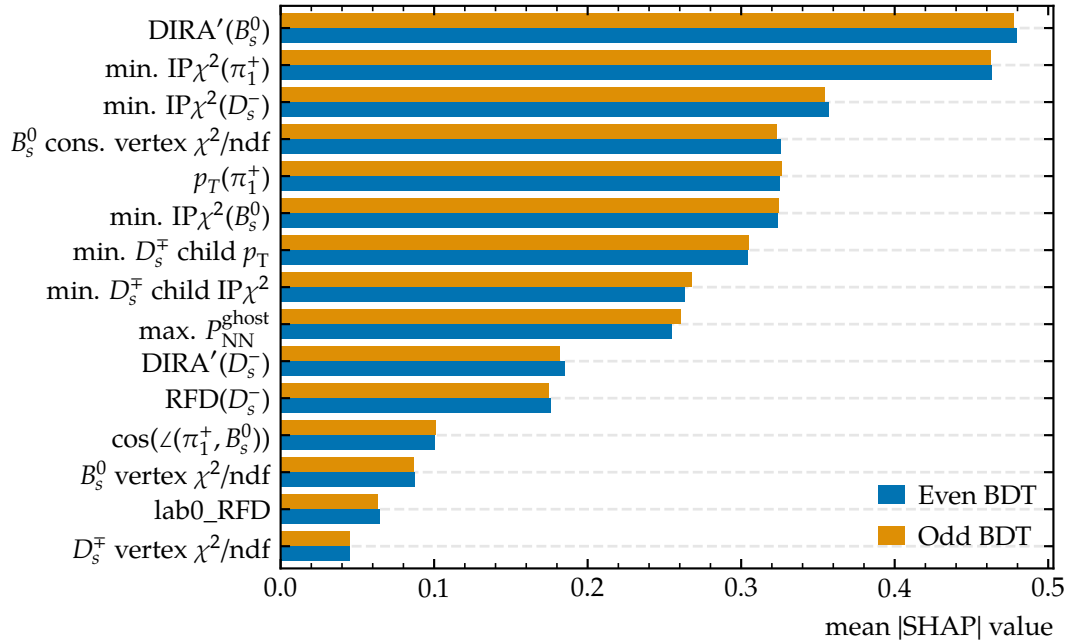
The figure of merit is hereby evaluated using a preliminary fit to the invariant  $m(D_s^\mp K^\pm)$  mass, where the all background yields are combined in  $N_{\text{bkg.}}$  and the signal yield is  $N_{\text{sig.}}$ . The efficiency is calculated from the signal yield with and without BDT selection



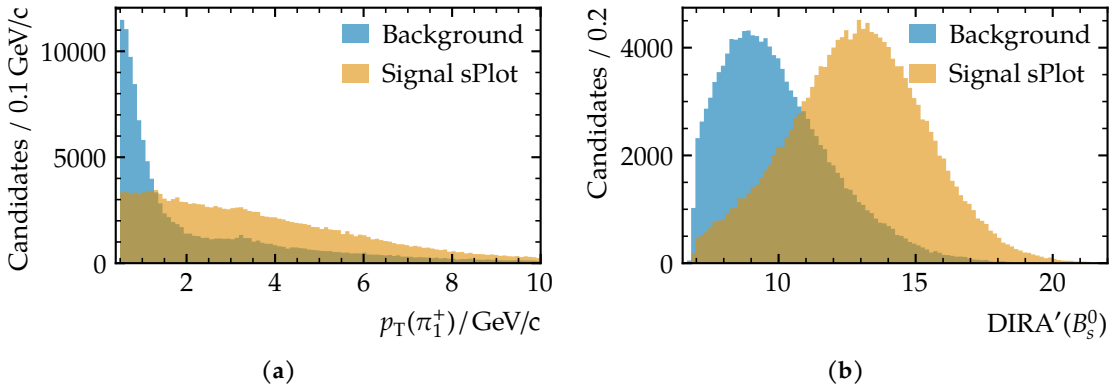
**Figure 6.4:** BDT accuracy during training, on cross-validated data samples. The performance is similar on both training samples, while the performance of the odd BDT applied to the even sample is lower than for the opposite case. Despite this, no over-training is observed for both set ups.



**Figure 6.5:** Feature importance of even (blue) and odd (orange) BDT, after training  $n_{tree} = 500$  trees. The default importance score is varies significantly between both splits.

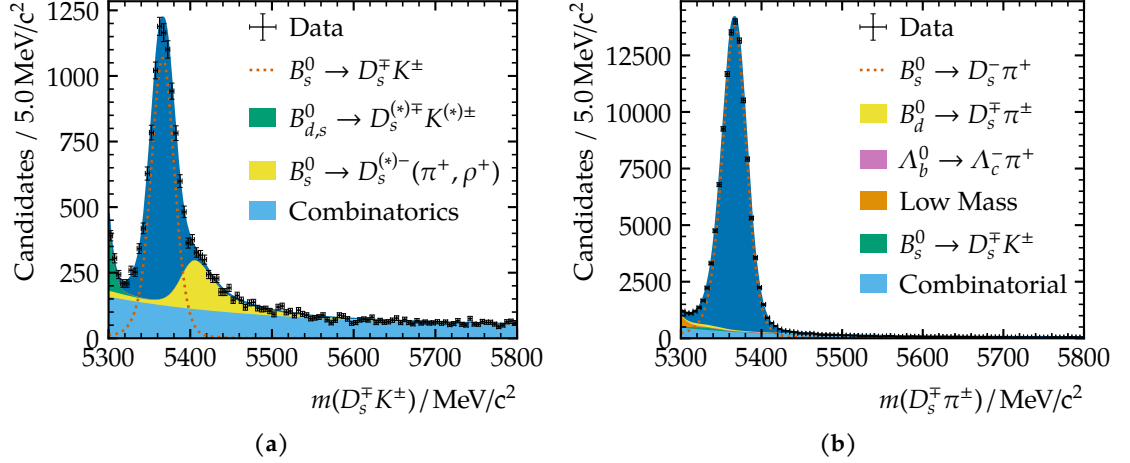


**Figure 6.6:** Shapley values of even (blue) and odd (orange) BDTs, evaluated on the respective test sample. The ranking is very similar for both BDTs.



**Figure 6.7:** Distributions for the variables with highest XGBoost feature ranking (a) and highest average SHAP value (b). The sWeights from the refitted model in Fig. 6.3 are applied for the signal sample.

applied. Figure 6.8a shows the fit to the  $m(D_s^\mp K^\pm)$  invariant mass at the BDT working point, selecting events with a BDT response  $f_{\text{BDT}} \geq 0.475$ . The same fit model as for the Run 1  $CP$  violation measurement with  $B_s^0 \rightarrow D_s^\mp K^\pm$  decays is used [104]. The data sample corresponds to the  $B_s^0 \rightarrow D_s^- \pi^+$  sample with inverted particle identification requirement and adjusted kaon hypothesis of the leftover hadron,  $\text{PID}_K(K_1^+) > 0$ .



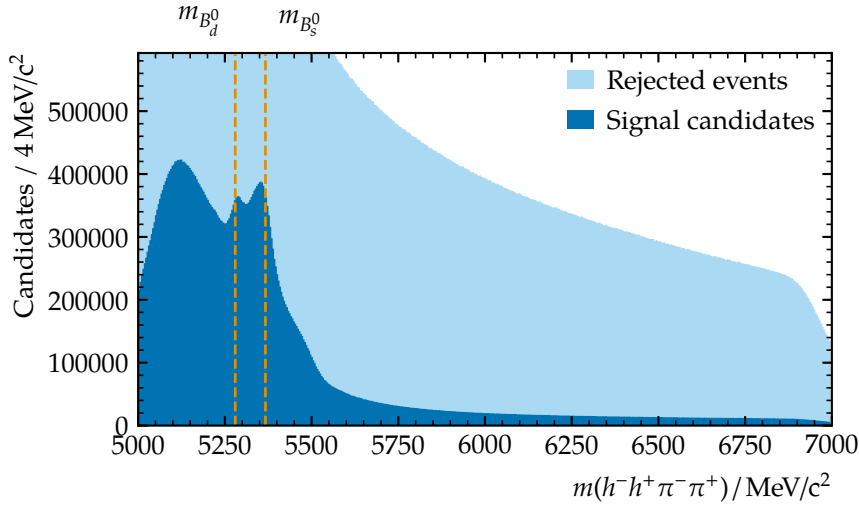
**Figure 6.8:** Preliminary fit to the invariant  $m(D_s^\mp K^\pm)$  mass (a) and the invariant  $m(D_s^\mp \pi^\pm)$  mass (b) of the combined 2015–2016  $D_s^\mp \rightarrow K^- K^+ \pi^-$  data samples, after applying the BDT working point selection.

In a later stage of the analysis, this working point is checked to give no different result for the measurement of  $\Delta m_s$  within the reported numerical precision.

### 6.2.3 Suppression of Misidentified Decays

After applying the just introduced BDT requirement, a significant fraction of non-signal events remains in the data sample, which can be seen in Fig. 6.9. These are originating mainly from different physical processes that are expected to produce a signal-like signature in the detector. Most prominently due to wrong particle identifications, and a small contribution due to some non-reconstructed final state particles. Several background candidates are considered in this analysis and further suppressed by specific selection requirements. Afterwards, as described in Section 6.3, the remaining fractions of these backgrounds are statistically subtracted with the sPlot method.

Decays of  $B_d^0 \rightarrow D^- \pi^+$  with  $D^- \rightarrow \pi^- K^+ \pi^-$  can be mistaken as signal candidates if one of the negatively charged pions, originating from the  $D^-$ , meson is misidentified as a kaon. This background populates the  $m(D_s^\mp \pi^\pm)$  mass distribution at the  $B_d^0$  meson mass of  $m_{B_d^0} = 5280 \text{ MeV}/c^2$ , clearly visible in Fig. 6.9. These decays are rejected by requiring either the final state kaon with same charge as the  $D_s^\mp$  meson to fulfil  $\text{PID}_K > 10$  or the  $D_s^\mp$  invariant mass under  $D^-$  hypothesis to be different from the known mass,  $m(K^- K^+ \pi^-)_{D^-} < 1839 \text{ MeV}/c^2$  and  $m(K^- K^+ \pi^-)_{D^-} > 1899 \text{ MeV}/c^2$ .



**Figure 6.9:** Invariant  $m(h^-h^+\pi^-\pi^+)$  mass distribution after preselection and BDT requirements. Peaking non-signal structure are visible, most dominantly at the  $B_d^0$  mass.

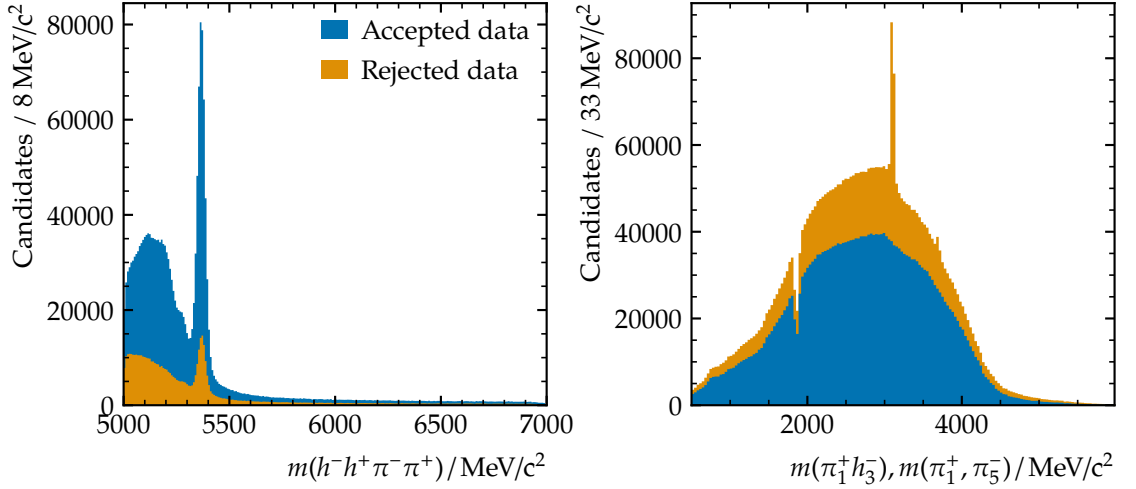
Similarly, decays of  $B_s^0 \rightarrow D_s^\mp K^\pm$  can be misidentified as signal candidates if the leftover kaon is identified as a pion. This background is harder to distinguish since it is expected to peak slightly below the  $m_{B_s^0}$  mass, due to the wrong pion mass hypotheses. Due to LHCb's good particle identification, it is sufficiently suppressed by requiring the leftover particle to fulfil  $\text{PID}_K < 0$ .

Finally, decays of  $\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+$ , with  $\Lambda_c^- \rightarrow \bar{p}K^+\pi^-$  can generate a signal signature in the detector. This is the case if either the proton is misidentified as a kaon or the proton and pion are simultaneously misidentified as pion and kaon, respectively. The first case will be referred to as single misid., the second as double misid. Background events due to single misid. are rejected similarly to  $B_d^0 \rightarrow D^- \pi^+$  background, by requiring the kaon with same charge as the  $D_s^\mp$  to fulfil  $\text{PID}_K - \text{PID}_p > 5$  or the  $D_s^-$  with  $\Lambda_c^-$  hypothesis to have a mass different from the know mass,  $m(K^-K^+\pi^-)_{\Lambda_c^-} < 2255 \text{ MeV}/c^2$  and  $m(K^-K^+\pi^-)_{\Lambda_c^-} > 2315 \text{ MeV}/c^2$ . The case of double misid. is suppressed by requiring the pion with same charge as the  $D_s^-$  to fulfil  $\text{PID}_p < 10$ .

Partially reconstructed decays of  $B_s^0 \rightarrow D_s^{*-} \pi^+$ , where a photon from the decay  $D_s^{*-} \rightarrow D_s^- \gamma$  is not reconstructed, is peaking at the low end of the  $m(D_s^\mp \pi^\pm)$  mass range. A large fraction of these decays is suppressed by the minimum considered  $B_s^0$  mass. Decays of  $B_d^0 \rightarrow D_s^\mp \pi^\pm$  lead to a very similar  $m(D_s^\mp \pi^\pm)$  shape, and they are similarly suppressed by the fit range. Both background contributions are handled in the MDfit, as described in Section 6.3.

In addition to the aforementioned requirements, an extensive study of the invariant mass distributions for all possible final state particle combinations has shown several signatures of charmonium states. These are produced by two muons, that are misidentified as final state pions and originate from decays of e.g.  $B_d^0 \rightarrow J/\psi X$ , where  $J/\psi X$

denotes some possible final state. As shown in Fig. 6.10, these are efficiently suppressed by requiring all final state particles to have no hits in the muon chambers. A small, remaining fractions of misidentified electrons from decays of intermediate  $J/\psi \rightarrow e^+ e^-$  states is suppressed by requiring the leftover pion and the pion with same charge as the  $D_s^-$  meson to fulfil  $\text{PID}_e < 5$ . In the  $D_s^- \rightarrow \pi^- \pi^+ \pi^-$  final state, two additional pions could be misidentified, which is why the additional high momentum pion with same charge as the  $D_s^-$  meson is also required to fulfil  $\text{PID}_e < 5$ .



**Figure 6.10:** Invariant  $m(h^- h^+ \pi^- \pi^+)$  mass distribution (left) and combined, invariant  $m(\pi_1^+ h_3^-)$  and  $m(\pi_1^+ \pi_5^-)$  mass distribution (right), of events that are accepted (blue) or rejected (orange) by the charmonium rejection requirements. A narrow signature of  $J/\psi$  states is removed. The additional requirements to suppress  $D^0$  is visible around the  $D^0$  mass of  $\sim 1864 \text{ MeV}/c^2$ .

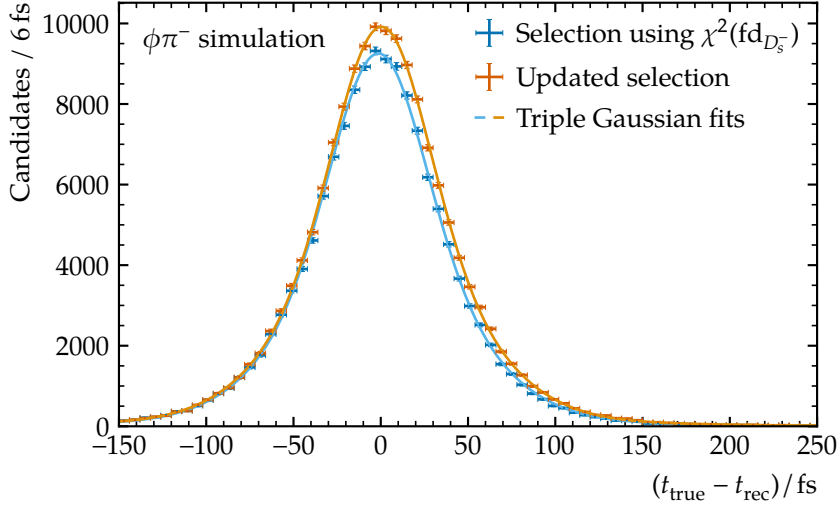
A similar signature originates from decays of  $B_{(s)}^0 \rightarrow D^0 X$ , where one of the decay products of  $D^0 \rightarrow K^- \pi^+$  or  $D^0 \rightarrow \pi^- \pi^+$  can be mistaken as the leftover pion of the  $B_s^0 \rightarrow D_s^- \pi^+$  candidates. These decays are suppressed by rejecting all events where the invariant mass of the leftover pion and any oppositely charged final state hadron is close to the  $D^0$  mass.

## 6.2.4 Suppression of Charmless Background Events

Previous analyses of the same decay channel  $B_s^0 \rightarrow D_s^- \pi^+$  efficiently suppressed charmless and charmonium backgrounds by requiring the  $D_s^-$  flight distance, with respect to the  $B_s^0$  decay vertex, to fulfil  $\text{fd}_{D_s^-} > 0$ , the  $D_s^-$  decay time to fulfil  $t_{D_s^-} > 0 \text{ ps}$ , and the  $D_s^-$  vertex separation quality with respect to the  $B_s^0$  decay vertex, to fulfil  $\chi^2(\text{fd}_{D_s^-}) > 2$  [104].

While these requirements are powerful to reject events containing fake charm mesons, a study of the decay time resolution (see Section 8.2.2) reveals that they introduce a decay time bias, which directly translates into a bias of  $\Delta m_s$ . The previously described study of charmonium backgrounds was initially triggered by this finding. The bias introduced by

the  $D_s^-$  flight distance and vertex quality requirements is shown in Fig. 6.11 for simulated signal events. A fit of a triple Gaussian function to the difference of the true and reconstructed decay-time yields a mean shift of  $(-1.42 \pm 0.11)$  fs. The selection requirements developed in this analysis prevent this bias, reducing the shift to  $(0.35 \pm 0.10)$  fs, which is also obtained if no additional requirements are applied. At the same time, a similar suppression of background events as for the previous selection requirements is maintained.

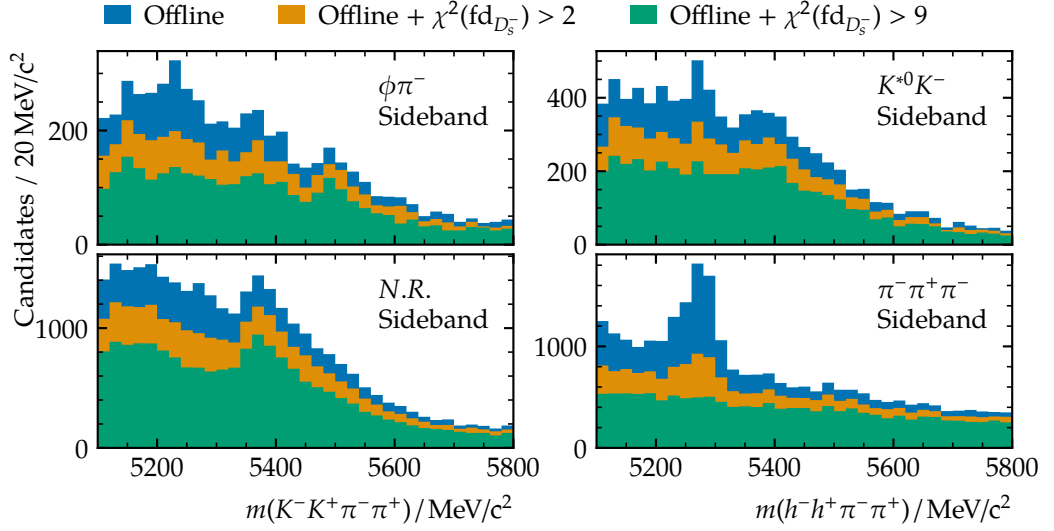


**Figure 6.11:** Distribution of the difference of true and reconstructed decay-times, obtained from a simulated sample of  $B_s^0 \rightarrow D_s^- \pi^+$  decays. As further explained in the text, a significant shift can be seen if  $\chi^2(\text{fd}_{D_s^-})$  requirements are used to suppress charmless background decays.

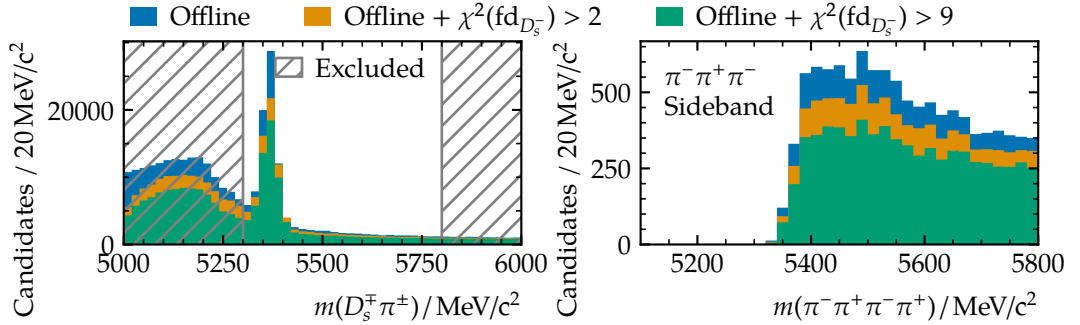
To verify that charmless background events are still sufficiently suppressed with the updated selection requirements, the  $B_s^0$  mass distribution is studied within the invariant  $D_s^-$  mass side-band, above  $m(D_s^-) \geq 2015 \text{ MeV}/c^2$ . In addition to the updated offline selection, a more stringent requirement to the  $D_s^-$  vertex separation quality is applied. Any remaining charmless background is expected to peak in the nominally selected distribution, while the same structure is significantly reduced with the additional  $D_s^-$  vertex requirement.

The corresponding data distributions are shown in Fig. 6.12. A signature of charmless background from  $B_d^0 \rightarrow X \rightarrow \pi^- \pi^+ \pi^- \pi^+$ , without intermediate  $D_s^\mp$ , is only visible in the  $D_s^- \rightarrow \pi^- \pi^+ \pi^-$  decay mode. Since the  $B_s^0$  mass fit range starts above the peaking structure, it is considered negligible for the further analysis. Figure 6.13 shows the upper  $D_s^\mp$  mass side-band, for data from the nominal  $B_s^0$  mass fit range. No structures of charmless background events can be identified in the other in the other  $D_s^- \rightarrow K^- K^+ \pi^-$  decay modes, after applying the more stringent  $\chi^2(\text{fd}_{D_s^-})$  requirement. The remaining resonant structures therefore assumed to originate from particle misidentification and

are statistically subtracted in the multidimensional mass fit.



**Figure 6.12:** Invariant  $m(h^- h^+ \pi^- \pi^+)$  mass distributions in the upper  $m(D_s^-)$  mass sideband of each  $D_s^-$  final state. Only the  $D_s^- \rightarrow \pi^- \pi^+ \pi^-$  decay mode contains a signature of charmless background decays.



**Figure 6.13:** Invariant  $m(\pi^- \pi^+ \pi^- \pi^+)$  mass distributions in the upper  $m(D_s^-)$  mass sideband of the  $D_s^- \rightarrow \pi^- \pi^+ \pi^-$  final state. Only data from the nominal  $m(D_s^\mp \pi^\pm)$  fit range is shown, which sufficiently suppresses charmless background contributions, seen in Fig. 6.12.



**Table 6.3:** Offline selection requirements applied in addition to the preselection requirements (see Table 6.1).

Decay Mode	Requirement	Unit	
All modes	$m(D_s^\mp \pi^\pm)$	$\in [5300, 5800]$ MeV/c <sup>2</sup>	
	$m(h^- h^+ h^-)$	$\in [1920, 2015]$ MeV/c <sup>2</sup>	
	$t_{B_s^0}$	$> 0.4$ ps	
	$\sigma(t_{B_s^0})$	$\in [0.01, 0.1]$ ps	
	$f_{\text{BDT}}$	$> 0.475$	
	$\text{PID}_K(\pi_1^+)$	$< 0$	
	$\text{RICH}(h^\pm)$	is true	
	$\text{MUON}(h^\pm)$	is false	
	$\text{PID}_e(\pi_1^+, \pi_5^-)$	$< 5$	
$D_s^- \rightarrow K^- K^+ \pi^-$	$\text{PID}_K(K_3^-)$	$> 10$	
	or $m(K^- K^+ \pi^-)_{D^-}$	$\notin [1839, 1899]$ MeV/c <sup>2</sup>	
	$\text{PID}_K(K_3^-) - \text{PID}_p(K_3^-)$	$> 5$	
	or $m(K^- K^+ \pi^-)_{\Lambda_c^-}$	$\notin [2255, 2315]$ MeV/c <sup>2</sup>	
	$\text{PID}_p(\pi_5^-)$	$< 10$	
	$m(K_3^- K_4^+)$	$< 1800$ MeV/c <sup>2</sup>	
	$m(\pi_1^+ K_3^-)$	$\notin [1834, 1894]$ MeV/c <sup>2</sup>	
	$\rightarrow \phi \pi^-$	$\text{PID}_K(K_3^-, K_4^+)$	$> -2$
		$m(K_3^- K_4^+)$	$\in [1000, 1040]$ MeV/c <sup>2</sup>
	$\rightarrow K^{*0} K^-$	$\text{PID}_K(K_3^-)$	$> 5$
		$\text{PID}_K(K_4^+)$	$> -2$
		$m(K_3^- K_4^+)$	$\notin [1000, 1040]$ MeV/c <sup>2</sup>
		$m(K_4^+ \pi_5^-)$	$\in [842, 942]$ MeV/c <sup>2</sup>
	$\rightarrow$ non-resonant $K^- K^+ \pi^-$	$\text{PID}_K(K_3^-, K_4^+)$	$> 5$
		$\text{PID}_K(\pi_5^-)$	$< 10$
	$m(K_3^- K_4^+)$	$\notin [1000, 1040]$ MeV/c <sup>2</sup>	
	$m(K_4^+ \pi_5^-)$	$\notin [842, 942]$ MeV/c <sup>2</sup>	
$D_s^- \rightarrow \pi^- \pi^+ \pi^-$	$\text{PID}_K(\pi_3^-, \pi_4^+, \pi_5^-)$	$< 2$	
	$\text{PID}_p(\pi_3^-, \pi_4^+, \pi_5^-)$	$< 5$	
	$\text{PID}_e(\pi_3^-)$	$< 5$	
	$m(\pi_3^- \pi_4^+)$	$< 1700$ MeV/c <sup>2</sup>	
	$m(\pi_4^+ \pi_5^-)$	$< 1700$ MeV/c <sup>2</sup>	
	$m(\pi_1^+ \pi_3^-)$	$\notin [1834, 1894]$ MeV/c <sup>2</sup>	
	$m(\pi_1^+ \pi_5^-)$	$\notin [1834, 1894]$ MeV/c <sup>2</sup>	

### 6.3 Extraction of the Effective Signal Component

Thanks to the good mass resolution of the LHCb detector, and the good understanding of the particle mass parameters, the sPlot technique, described in Section 5.2.1 can be used to disentangle the different processes that contribute to the data sample. The method is implemented by performing a simultaneous maximum likelihood fit to the  $B_s^0$  and  $D_s^\mp$  invariant mass distributions, with separate parameters for each data sub-sample, i.e. 2015–2016, 2017, and 2018.

In each case the  $B_s^0$  and  $D_s^\mp$  mass distributions of the signal candidates are described with a sum of a double-sided Hypatia ( $H(m|\vec{p}^H)$ ) and a Johnson SU ( $J(m|\vec{p}^J)$ ) function [105, 106]. An extensive study of several different parametrisations is carried out by my colleague Agnieszka Dziurda, of which the aforementioned model provides the overall best fit quality.

For each data sub-sample, the signal PDF model can be written as

$$S(m_{D_s^\mp \pi^\pm} | \vec{p}_{B_s^0}^H, \vec{p}_{B_s^0}^J) = f_{B_s^0} \cdot J(m_{B_s^0} | \vec{p}_{B_s^0}^J) + (1 - f_{B_s^0}) \cdot H(m_{B_s^0} | \vec{p}_{B_s^0}^H), \quad (6.3)$$

$$S(m_{h^- h^+ h^-} | \vec{p}_{D_s^\mp}^H, \vec{p}_{D_s^\mp}^J) = f_{D_s^\mp} \cdot J(m_{h^- h^+ h^-} | \vec{p}_{D_s^\mp}^J) + (1 - f_{D_s^\mp}) \cdot H(m_{h^- h^+ h^-} | \vec{p}_{D_s^\mp}^H), \quad (6.4)$$

where the PDF shape parameters  $\vec{p}$  are mostly fixed from simulation. The set of parameters, listed in Table 6.4, includes a shared mean value  $\mu$  and the widths parameters  $\sigma$ . To allow the description of a slightly wider distribution on data with respect to simulation, the widths of the Hypatia functions,  $\sigma_{B_s^0/D_s^\mp}^H$  are floating, while the widths of the Johnson SU functions,  $\sigma_{B_s^0/D_s^\mp}^J$ , are fixed from simulation. The mean  $\mu_{B_s^0}$  and  $\mu_{D_s^\mp}$  of the combined distributions are shared between all  $D_s^\mp$  decay modes, while different values are allowed for the data sub-samples to compensate different momentum scale calibrations. The multidimensional PDF for the mass vector  $\vec{m} = (m_{D_s^\mp \pi^\pm}, m_{h^- h^+ h^-})$  is obtained by multiplying the PDFs for the individual mass dimensions:

$$S(\vec{m}) = S(m_{D_s^\mp \pi^\pm}) \cdot S(m_{h^- h^+ h^-}). \quad (6.5)$$

Combinatorial background candidates, that remain after applying the BDT selection requirements from Section 6.2.2, are described with a double exponential function in the  $m(D_s^\mp \pi^\pm)$  dimension. The invariant  $m(h^- h^+ h^-)$  mass is described with an exponential function to which the signal PDF Eq. (6.4) is added with its parameters  $\vec{p}_{D_s^\mp}^S = (\vec{p}_{D_s^\mp}^H, \vec{p}_{D_s^\mp}^J)$ . The additional term allows to describe true  $D_s^\mp$  candidates which are combined with a random pion track:

$$C(m_{D_s^\mp \pi^\pm} | f_{B_s^0}^C, c_{B_s^0}^{(1)}, c_{B_s^0}^{(2)}) = f_{B_s^0}^C \cdot e^{c_{B_s^0}^{(1)} m_{D_s^\mp \pi^\pm}} + (1 - f_{B_s^0}^C) \cdot e^{c_{B_s^0}^{(2)} m_{D_s^\mp \pi^\pm}}, \quad (6.6)$$

$$C(m_{h^- h^+ h^-} | f_{D_s^\mp}^C, c_{D_s^\mp}, \vec{p}_{D_s^\mp}^S) = f_{D_s^\mp}^C \cdot e^{c_{D_s^\mp} m_{h^- h^+ h^-}} + (1 - f_{D_s^\mp}^C) \cdot S(m_{h^- h^+ h^-} | \vec{p}_{D_s^\mp}^S). \quad (6.7)$$

Additional PDF components are used to describe the misidentified and partially reconstructed backgrounds. In the  $m(D_s^\mp \pi^\pm)$  mass distribution the  $B_d^0 \rightarrow D_s^\mp \pi^\pm$  background

### 6.3 Extraction of the Effective Signal Component

is described with the signal PDF from Eq. (6.3), shifted by the  $B_s^0$ - $B_d^0$  mass difference  $m_{B_s^0} - m_{B_d^0} = 86.8 \text{ MeV}/c^2$ :

$$B_{B_d^0 \rightarrow D_s^\mp \pi^\pm}(m_{B_s^0}) = S(m_{B_s^0} + 86.8 \text{ MeV}/c^2). \quad (6.8)$$

Furthermore, the remaining physical backgrounds from  $\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+$ ,  $B_s^0 \rightarrow D_s^\mp K^\pm$ ,  $B_d^0 \rightarrow D^- \pi^+$ , and  $B_s^0 \rightarrow D_s^{*-} \pi^+$  are described using Gaussian Kernel Estimators as implemented in `RooFit::RooKeysPdf` [83, 107]. The corresponding PDFs are extracted from simulated samples, individually for year and magnet polarities, and are added taking the recorded luminosity in the nominal data sub-sets into account.

For the  $m(h^- h^+ h^-)$  dimension, backgrounds from  $B_d^0 \rightarrow D_s^\mp \pi^\pm$ ,  $B_s^0 \rightarrow D_s^\mp K^\pm$ ,  $B_s^0 \rightarrow D_s^{*-} \pi^+$ , are reconstructed with a true  $D_s^\mp$  resonance, and are therefore described with the signal PDF from Eq. (6.4). The other backgrounds from  $\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+$  and  $B_d^0 \rightarrow D^- \pi^+$  are described with Gaussian Kernels, extracted from simulation.

As required by the sPlot method, the different PDF components are extended with a set  $\vec{N}$  of yield parameters, and added to build the total PDF

$$\begin{aligned} P(\vec{m}) = & N_{\text{Sig}} \cdot S(\vec{m}) + N_{\text{Comb}} \cdot C(\vec{m}) + N_{B_s^0 \rightarrow D_s^\mp K^\pm} \cdot B_{B_s^0 \rightarrow D_s^\mp K^\pm}(\vec{m}) + \\ & N_{\text{lowMass}} \cdot \left( f_{\text{lowMass}} \cdot B_{B_d^0 \rightarrow D_s^\mp \pi^\pm}(\vec{m}) + \right. \\ & \quad \left. (1 - f_{\text{lowMass}}) \cdot B_{B_s^0 \rightarrow D_s^{*-} \pi^+}(\vec{m}) \right) + \\ & N_{\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+} \cdot B_{\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+}(\vec{m}) + N_{B_d^0 \rightarrow D^- \pi^+} \cdot B_{B_d^0 \rightarrow D^- \pi^+}(\vec{m}). \end{aligned} \quad (6.9)$$

Due to the similar shape of the  $B_d^0 \rightarrow D_s^\mp \pi^\pm$  and partially reconstructed  $B_s^0 \rightarrow D_s^{*-} \pi^+$  backgrounds at the lower edge of the  $m(D_s^\mp \pi^\pm)$  range, the corresponding background PDFs share a yield parameter  $N_{\text{lowMass}}$ . The pseudo experiment studies shown in Section 8.4.1 indicate an overall better fit stability for a fixed fraction  $f_{\text{lowMass}}$  between the two components, while the actual value has only minor impact on the sPlot method, which is why it is fixed to  $f_{\text{lowMass}} = 1/2$ .

The fit result is shown in Fig. 6.14 in both mass dimensions. The resulting parameters are listed in Table 6.4 and 6.5. The fit results for the individual  $D_s^\mp$  decay-mode samples, and the fixed parameters of the fit are shown in the appendix.

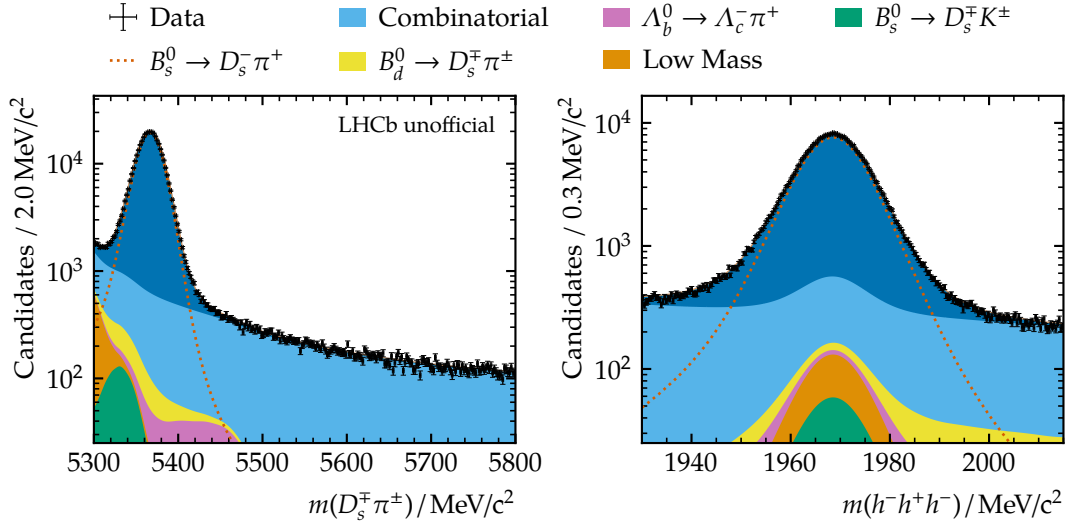
As described in Section 5.2.1, the sPlot method assumes no correlation among discriminating and control variables, which is usually hard to achieve in real world data samples. To increase the robustness of the method in the case of non-vanishing correlations, the PDF Eq. (6.9) is rephrased to

$$\begin{aligned} P'(\vec{m}) = & N_{\text{Sig}} \cdot S(\vec{m}) + N'_{\text{Bkg}} \cdot \left( \frac{N_{\text{Comb}}}{N_{\text{Bkg}}} \cdot C(\vec{m}) + \frac{N_{B_s^0 \rightarrow D_s^\mp K^\pm}}{N_{\text{Bkg}}} \cdot B_{B_s^0 \rightarrow D_s^\mp K^\pm}(\vec{m}) + \right. \\ & \frac{N_{\text{lowMass}}}{N_{\text{Bkg}}} \cdot \left( f_{\text{lowMass}} \cdot B_{B_d^0 \rightarrow D_s^\mp \pi^\pm}(\vec{m}) + (1 - f_{\text{lowMass}}) \cdot B_{B_s^0 \rightarrow D_s^{*-} \pi^+}(\vec{m}) \right) + \\ & \left. \frac{N_{\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+}}{N_{\text{Bkg}}} \cdot B_{\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+}(\vec{m}) + \frac{N_{B_d^0 \rightarrow D^- \pi^+}}{N_{\text{Bkg}}} \cdot B_{B_d^0 \rightarrow D^- \pi^+}(\vec{m}) \right), \end{aligned} \quad (6.10)$$

**Table 6.4:** Fitted shape parameters of the full, multidimensional mass fit model, Eq. (6.9).

Parameter	$D_s^-$ mode	2015–2016	2017	2018
$\mu_{B_s^0} / \text{MeV}/c^2$	All	5366.19 $\pm$ 0.05	5366.28 $\pm$ 0.05	5366.12 $\pm$ 0.05
$\sigma_{B_s^0}^J / \text{MeV}/c^2$	$\phi\pi^-$	16.13 $\pm$ 0.08	15.59 $\pm$ 0.09	15.66 $\pm$ 0.08
	$K^{*0}K^-$	15.77 $\pm$ 0.09	15.34 $\pm$ 0.09	15.46 $\pm$ 0.09
	n.r. $K^-K^+\pi^-$	15.57 $\pm$ 0.12	15.00 $\pm$ 0.12	14.89 $\pm$ 0.12
	$\pi^-\pi^+\pi^-$	15.96 $\pm$ 0.14	15.54 $\pm$ 0.15	15.36 $\pm$ 0.12
$c_{B_s^0}^{(1)} \cdot \text{GeV}/c^2$	$\phi\pi^-$	-10.1 $\pm$ 0.7	-9.3 $\pm$ 0.7	-11.1 $\pm$ 0.6
	$K^{*0}K^-$	-9.9 $\pm$ 0.7	-7.7 $\pm$ 0.7	-8.9 $\pm$ 0.7
	n.r. $K^-K^+\pi^-$	-6.5 $\pm$ 0.5	-6.4 $\pm$ 0.5	-6.0 $\pm$ 0.4
	$\pi^-\pi^+\pi^-$	-8.7 $\pm$ 0.6	-9.0 $\pm$ 0.8	-9.4 $\pm$ 0.6
$f_{B_s^0}$	$\phi\pi^-$	0.693 $\pm$ 0.026	0.736 $\pm$ 0.032	0.733 $\pm$ 0.022
	$K^{*0}K^-$	0.703 $\pm$ 0.029	0.81 $\pm$ 0.05	0.69 $\pm$ 0.04
	n.r. $K^-K^+\pi^-$	0.83 $\pm$ 0.04	0.80 $\pm$ 0.04	0.85 $\pm$ 0.04
	$\pi^-\pi^+\pi^-$	0.396 $\pm$ 0.023	0.387 $\pm$ 0.025	0.419 $\pm$ 0.021
$\mu_{D_s^-} / \text{MeV}/c^2$	All	1968.472 $\pm$ 0.021	1968.481 $\pm$ 0.022	1968.633 $\pm$ 0.020
$\sigma_{D_s^-}^J / \text{MeV}/c^2$	$\phi\pi^-$	6.923 $\pm$ 0.033	6.777 $\pm$ 0.035	6.837 $\pm$ 0.033
	$K^{*0}K^-$	6.98 $\pm$ 0.04	6.84 $\pm$ 0.04	6.89 $\pm$ 0.04
	n.r. $K^-K^+\pi^-$	6.87 $\pm$ 0.05	6.69 $\pm$ 0.05	6.64 $\pm$ 0.05
	$\pi^-\pi^+\pi^-$	9.76 $\pm$ 0.09	9.56 $\pm$ 0.09	9.37 $\pm$ 0.09
$c_{D_s^-} \cdot \text{GeV}/c^2$	$\phi\pi^-$	-7.5 $\pm$ 1.1	-6.3 $\pm$ 1.2	-7.9 $\pm$ 1.1
	$K^{*0}K^-$	-6.5 $\pm$ 0.9	-8.5 $\pm$ 1.0	-7.6 $\pm$ 0.9
	n.r. $K^-K^+\pi^-$	-4.0 $\pm$ 0.6	-5.3 $\pm$ 0.6	-5.5 $\pm$ 0.6
	$\pi^-\pi^+\pi^-$	-5.9 $\pm$ 0.4	-6.3 $\pm$ 0.5	-5.9 $\pm$ 0.4
$f_{D_s^-}$	$\phi\pi^-$	0.609 $\pm$ 0.022	0.614 $\pm$ 0.024	0.607 $\pm$ 0.022
	$K^{*0}K^-$	0.783 $\pm$ 0.022	0.829 $\pm$ 0.023	0.864 $\pm$ 0.022
	n.r. $K^-K^+\pi^-$	0.957 $\pm$ 0.012	0.958 $\pm$ 0.013	0.953 $\pm$ 0.012
	$\pi^-\pi^+\pi^-$	0.976 $\pm$ 0.011	0.978 $\pm$ 0.012	0.958 $\pm$ 0.011

### 6.3 Extraction of the Effective Signal Component



**Figure 6.14:** Result of the simultaneous fit to the invariant  $m(D_s^\mp \pi^\pm)$  and  $m(h^- h^+ h^-)$  mass distributions, combined for all  $D_s^\mp$  decay mode samples. The PDFs of the fit components are stacked, and the signal PDF is additionally shown as a dashed line.

where only the signal yield  $N_{\text{Sig}}$  and the combined background yield  $N'_{\text{Bkg}}$  float in the sPlot fit. The relative fractions of all background contributions are fixed from the aforementioned PDF, Eq. (6.9), and the combined background yield is the sum of all individual background yields:

$$N_{\text{Bkg}} = N_{\text{Comb}} + N_{B_s^0 \rightarrow D_s^\mp K^\pm} + N_{\text{lowMass}} + N_{\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+} + N_{B_d^0 \rightarrow D^- \pi^+}. \quad (6.11)$$

**Table 6.5:** Fitted yield parameters of the simultaneous extended maximum likelihood fit to the invariant  $m_{B_s^0}$  and  $m(D_s^-)$  mass distribution.

Component	$D_s^-$ Mode	2015–2016	2017	2018	Total
$B_s^0 \rightarrow D_s^- \pi^+$	$\phi\pi^-$	$49\,960 \pm 240$	$43\,370 \pm 220$	$49\,700 \pm 240$	$143\,000 \pm 400$
	$K^{*0}K^-$	$36\,140 \pm 200$	$31\,360 \pm 190$	$36\,380 \pm 200$	$103\,880 \pm 340$
	n.r. $K^-K^+\pi^-$	$22\,940 \pm 170$	$19\,900 \pm 150$	$22\,610 \pm 170$	$65\,460 \pm 280$
	$\pi^-\pi^+\pi^-$	$23\,380 \pm 180$	$20\,140 \pm 160$	$22\,770 \pm 170$	$66\,290 \pm 290$
	Total	$132\,400 \pm 400$	$114\,800 \pm 400$	$131\,500 \pm 400$	$378\,700 \pm 700$
Comb. Background	$\phi\pi^-$	$3440 \pm 120$	$2760 \pm 110$	$3490 \pm 130$	$9690 \pm 210$
	$K^{*0}K^-$	$3300 \pm 100$	$2770 \pm 90$	$3110 \pm 100$	$9180 \pm 170$
	n.r. $K^-K^+\pi^-$	$6890 \pm 120$	$5590 \pm 110$	$6820 \pm 120$	$19\,290 \pm 200$
	$\pi^-\pi^+\pi^-$	$11\,720 \pm 150$	$9210 \pm 130$	$10\,900 \pm 140$	$31\,830 \pm 250$
	Total	$25\,350 \pm 250$	$20\,330 \pm 220$	$24\,320 \pm 250$	$70\,000 \pm 400$
Low Mass	$\phi\pi^-$	$520 \pm 50$	$400 \pm 50$	$360 \pm 50$	$1270 \pm 90$
	$K^{*0}K^-$	$360 \pm 40$	$270 \pm 40$	$480 \pm 40$	$1110 \pm 70$
	n.r. $K^-K^+\pi^-$	$190 \pm 40$	$139.0 \pm 31.0$	$200 \pm 40$	$530 \pm 60$
	$\pi^-\pi^+\pi^-$	$260 \pm 50$	$250 \pm 40$	$190 \pm 40$	$690 \pm 70$
	Total	$1330 \pm 90$	$1060 \pm 80$	$1220 \pm 90$	$3610 \pm 150$

## 7 Flavour Tagging Studies

Measurements of the decay-time dependent decay rates in Eqs. (3.21) to (3.24), and the interesting processes that are parametrised with these equations requires knowledge of the initial flavour –  $B_s^0$  or  $\bar{B}_s^0$  – of the produced mesons. At the LHCb experiment, the determination of the initial flavour is referred to as flavour tagging, and described in more detail in the next sections.

A general overview about the relevant decay signatures at a hadron collider and various flavour tagging algorithms that are used at the LHCb experiment is given in Section 7.1. A study of a new, machine-learning-focused flavour tagging approach for the upcoming data taking periods is briefly described in Section 7.2. The preparations needed for the  $\Delta m_s$  measurement are finally described in Section 7.3.

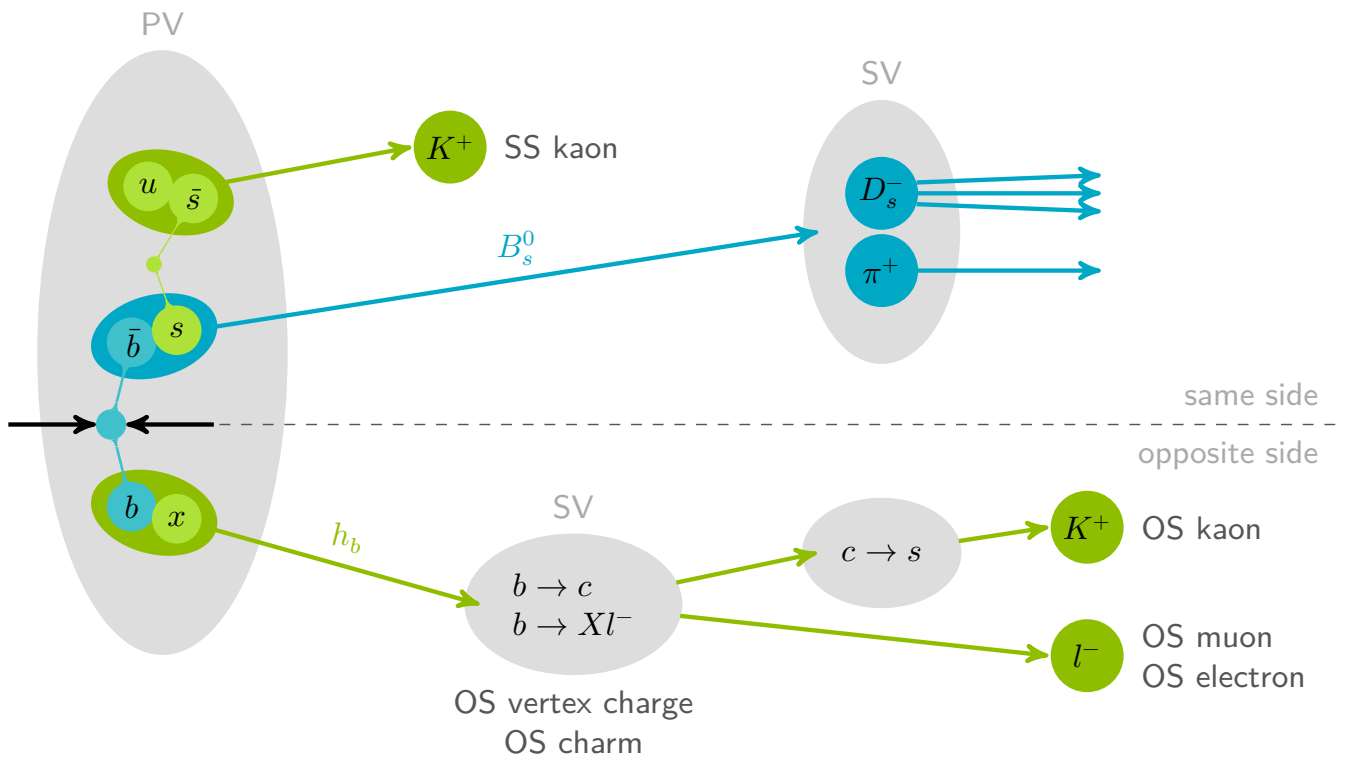
### 7.1 Beauty Flavour Tagging at LHCb

A sketch of a  $pp$  collision involving  $B$  mesons is shown in Fig. 7.1. The good vertex resolution of the LHCb detector enables the reconstruction of a majority of the tracks in an event, allowing to measure the corresponding secondary vertices and the primary vertex of the interaction. As described in Chapter 4,  $b$  quarks are usually produced in  $b\bar{b}$  quark pairs, and are often boosted along the beam direction. Therefore, a reconstructed  $B$  meson decay is usually accompanied by some opposite side (OS) process involving a  $b$  quark of opposite flavour, which hadronizes independently before it decays to lighter particles.

Similarly, the hadronization of the signal meson,  $\bar{b}q$ , leaves a  $\bar{q}$  quark which itself forms a hadron that can be reconstructed in the detector. The decay chains of these OS and same side (SS) tagging particles are thus connected to the flavour of the signal candidate. The LHCb flavour tagging algorithms therefore aim to reconstruct these particles, and measure the corresponding initial flavour of  $B$  meson candidates. The algorithms profit from the good vertex separation and the particle identification system of the LHCb detector. While naturally the same processes take place in general purpose detectors, these are sensitive to different kinematic regions and must usually handle a much larger detector occupancy. Therefore, flavour tagging is a bigger challenge at the ATLAS and CMS experiments and sometimes refers to the mere identification of particles containing a  $b$  or  $\bar{b}$  quark.

#### Key Parameters

To identify tagging particle candidates, flavour tagging algorithms implement different selection criteria, which are applied to all non-signal particles, that originate from the



**Figure 7.1:** Sketch of a  $B_s^0 \rightarrow D_s^- \pi^+$  decay topology, including the quark hadronization partners of the signal  $B_s^0$  meson. Various flavour tagging algorithms aim to identify the decay products of the OS and SS quark partners, allowing to infer the initial  $B_s^0$  flavour content. The individual algorithms are listed next to their respective tagging particles and explained in more detail in Section 7.1.2.



same primary vertex as the signal candidate. The initial flavour of the signal candidate is then defined by the charge of the tagging particle candidate, taking the particular decay chain hypothesis of the particle into account. A tag prediction is denoted with

$$d = +1 \quad \text{for } B_d^0, B_s^0, B^+ \quad \text{and} \quad d = -1 \quad \text{for } \bar{B}_d^0, \bar{B}_s^0, B^-. \quad (7.1)$$

The performance of these algorithms affects the measurable decay rates Eqs. (3.21) to (3.24) as described later in this section. It is parametrised by two key parameters, one of which describes a fraction

$$\omega = \frac{N_{\text{false}}}{N_{\text{false}} + N_{\text{true}}} \quad (7.2)$$

of  $N_{\text{false}}$  incorrectly tagged events, out of all  $N_{\text{false}} + N_{\text{true}}$  tagged events. The number of wrongly tagged events originates partly from an imperfect tagging track selection, but a lower bound is defined by neutral meson mixing: OS  $b$  quarks that hadronise to neutral  $B$  mesons can oscillate to their opposite flavour before they decay, resulting in a wrong tagging signature hypothesis. The mistag fraction Eq. (7.2) dilutes the amplitude of the trigonometric terms of the decay rates by a factor

$$D = 1 - 2\omega. \quad (7.3)$$

The second tagging performance parameter originates from a number of  $N_{\text{untagged}}$  events, which do not fulfil the selection requirements of a given flavour tagging algorithm. This reduces the fraction of tagged events by a tagging efficiency

$$\varepsilon_{\text{tag}} = \frac{N_{\text{true}} + N_{\text{false}}}{N_{\text{true}} + N_{\text{false}} + N_{\text{untagged}}}. \quad (7.4)$$

In addition, these effects potentially behave differently for different  $B$  flavours, which is why Eqs. (7.2) and (7.4) are adjusted to include small asymmetries,  $\Delta\omega$  and  $\Delta\varepsilon$ , which allow to define

$$\begin{aligned} \omega^- &= \omega - \frac{1}{2}\Delta\omega, & \varepsilon^- &= \varepsilon - \frac{1}{2}\Delta\varepsilon & \text{for } B_d^0, B_s^0, B^+, \\ \text{and } \omega^+ &= \omega + \frac{1}{2}\Delta\omega, & \varepsilon^+ &= \varepsilon + \frac{1}{2}\Delta\varepsilon & \text{for } \bar{B}_d^0, \bar{B}_s^0, B^-. \end{aligned} \quad (7.5)$$

As a consequence, each measured decay rate from some initial state  $B^0$  or  $\bar{B}^0$  is polluted with a fraction  $\omega$  of wrongly tagged events, which are reconstructed in the same final state, but originate from a different true initial state. As an example, Eq. (3.21) cannot be measured exclusively. Instead a measurement includes a fraction  $\omega^+$  of Eq. (3.22),

$$\Gamma_{\text{meas.}}(B^0(t) \rightarrow f) = \varepsilon^-(1 - \omega^-)\Gamma(B^0(t) \rightarrow f) + \varepsilon^+\omega^+\Gamma(\bar{B}^0(t) \rightarrow f). \quad (7.6)$$

The measured mixing asymmetry, which is relevant in this work, is therefore reduced by the aforementioned dilution factor,

$$A_{\text{mix,meas.}}(t) = \frac{N_{\text{u}}(t) - N_{\text{m}}(t)}{N_{\text{u}}(t) + N_{\text{m}}(t)} = \frac{D \cos(\Delta m_s t)}{\cosh\left(\frac{\Delta\Gamma_s t}{2}\right)} = DA_{\text{mix}}(t), \quad (7.7)$$

where terms describing indirect  $CP$ -violating are neglected ( $a = 0$ ).

The flavour tagging dilution directly affects the expected uncertainty of the measurement. This can be seen when assuming Gaussian uncertainties and no correlation between the measured parameters in Eq. (7.7). In this case, the uncertainty of the true mixing asymmetry, is affected by an effective tagging efficiency,

$$\sigma(A_{\text{mix}}) \propto \frac{1}{\sqrt{\varepsilon_{\text{eff}}N}}, \quad \text{with} \quad \varepsilon_{\text{eff}} \equiv \varepsilon_{\text{tag}}D^2. \quad (7.8)$$

The statistical power of the measurement corresponds to a perfectly tagged sample of  $\varepsilon_{\text{eff}}N$  signal events. The effective tagging efficiency is sometimes also referred to as “tagging power”.

### 7.1.1 Mistag Prediction, Calibration and Combination

In practice, the just introduced parameters enter a maximum likelihood fit with which physical parameters such as  $\Delta m_s$  are determined (see Section 5.2). While the likelihood can be defined using the average mistag from Eq. (7.2), better results can be achieved by estimating the tag quality of individual events. This corresponds to a per-event dilution factor, which reduces the influence of badly tagged events on the likelihood. An estimate  $\eta_i$  of the true per-event mistag  $\omega_i$  can be obtained through binary classification algorithms such as described in Section 5.1. The training labels in this case are correct or incorrect tag decisions, which can be evaluated in simulated data, or samples of flavour-specific decays. With this per-event flavour tagging information, the tagging power is defined via

$$\varepsilon_{\text{eff}} = \frac{\sum_{i=1}^{N_{\text{tag}}} s_i (1 - 2\eta_{c,i})^2}{\sum_{i=1}^N s_i}, \quad (7.9)$$

where the numerator sums all tagged events,  $N_{\text{tag}}$ , and an optional event weight  $s_i$  such as obtained from the sPlot method (see Section 5.2.1) is taken into account.

Since the distribution of the estimated mistag  $\eta_i$  does not necessarily correspond to the true mistag distribution  $\omega_i$ , a calibration is needed. A linear calibration function is often sufficient, but a distinction between both initial flavours from Eq. (7.5) must usually be taken into account. A basic, linear calibration function of the mistag predictions  $\eta_i$  is therefore given by

$$\begin{aligned} \omega^-(\eta_i) &= p_0 - \frac{1}{2}\Delta p_0 + \left(p_1 - \frac{1}{2}\Delta p_1\right)(\eta_i - \langle\eta\rangle), & \text{for } B_d^0, B_s^0, B^+ \\ \text{and } \omega^+(\eta_i) &= p_0 + \frac{1}{2}\Delta p_0 + \left(p_1 + \frac{1}{2}\Delta p_1\right)(\eta_i - \langle\eta\rangle), & \text{for } \bar{B}_d^0, \bar{B}_s^0, B^-, \end{aligned} \quad (7.10)$$

which uses the calibration parameters  $p_i$  and  $\Delta p_i$ , and the average mistag prediction  $\langle\eta\rangle$ .

With increasing statistics, more complex distributions need to be calibrated, which can be achieved through generalized calibration models. The LHCb flavour tagging working group provides and maintains the Espresso Performance Monitor (EPM) package, which implements these functions [108]. With a link function  $g$ , which is commonly chosen to

be the logistic function Eq. (5.4), the polynomial calibration functions are implemented as

$$\omega(\eta) = g\left(g^{-1}(\eta) + \sum_k \theta_k P_k(g^{-1}(\eta))\right), \quad (7.11)$$

with the calibration coefficients  $\theta_i$  for a set of polynomial basis vectors  $P_k$  of order  $k$ , which are orthogonalized to reduce the correlations of the calibration parameters. The calibration parameters vanish for a perfect calibration.

Since multiple tagging algorithms provide a tag decision  $d_i$  and mistag estimate  $\eta_i$ , their respective output features need to be combined. A probabilistic approach can be used for this, where the likelihood for an event to contain a specific flavour is defined via the per-event mistag estimates and tag predictions of the individual algorithms. For  $N$  algorithms, the likelihood  $p(b)$  to contain true  $b$  quark is defined via

$$p(b) = \prod_i \left( \frac{1 + d_i}{2} - d_i (1 - \omega_i(\eta_i)) \right), \quad (7.12)$$

$$\text{and similarly } p(\bar{b}) = \prod_i \left( \frac{1 - d_i}{2} + d_i (1 - \omega_i(\eta_i)) \right). \quad (7.13)$$

The normalized probability for the same event to contain a true  $b$  or  $\bar{b}$  quark is then obtained through

$$P(b) = \frac{p(b)}{p(b) + p(\bar{b})} \quad \text{and} \quad P(\bar{b}) = 1 - P(b), \quad (7.14)$$

respectively. This allows to define the combined tag decision and mistag estimate via

$$d_c = \text{sign}(P(b) - P(\bar{b})), \quad \text{and} \quad \eta_c = 1 - \max(P(b), P(\bar{b})). \quad (7.15)$$

It is often useful to apply an additional calibration function to a sample of combined flavour tagging results.

### 7.1.2 LHCb Flavour Tagging for LHC Run 2

During the LHCb Run 2 data taking period, a dominant fraction of the recorded data contains fully reconstructed event information. A set of various flavour tagging algorithms is applied during the data stripping, adding the required information about the initial flavour to the high-level data samples that are used for thorough offline analysis.

Since the LHC running conditions changed significantly between Run 1 and Run 2, several algorithms have been re-optimised [109]. A part of this effort was spent to refactor the flavour tagging software, streamlining the software interface and adding an abstraction layer for functionality that is commonly used by all algorithms. As a result, all stripping campaigns since spring 2018 profit from a re-optimised set of tagging algorithms.

The flavour tagging algorithms follow the general approach that has been introduced in the previous section, but differ in implementation details, which are explained in the following.

## Opposite Site Algorithms

Flavour tagging algorithms that exploit the OS  $b$  quark hadronization and decay are referred to as OS taggers. Several algorithms that are in use at LHCb are marked in Fig. 7.1, next to the particles that are being selected as tagging particles.

The OS electron and muon taggers select leptons from  $b \rightarrow X\ell^-$  decays. They apply an exclusive set of selection requirements, which have been determined from charged  $B^\pm \rightarrow J/\psi K^\pm$  decays using a BDT-based optimisation algorithm. Besides other features, the OS muon tagger requires tagging particle candidates to have a high momentum of  $p > 2.5 \text{ GeV}/c$  and transverse momentum  $p_T > 0.95 \text{ GeV}/c$ . The OS electron tagger poses more stringent requirements, taking only particles into account which fulfil  $p > 5 \text{ GeV}/c$  and transverse momentum  $p_T > 1.4 \text{ GeV}/c$ . Both algorithms use the estimated probabilities for a track to originate from a certain particle species  $X$ ,  $P_{\text{NN}}^X$ , which are obtained from the PID system, using different neural networks.

The OS kaon tagger is implemented in an analogous manner to the OS lepton taggers. A selection for kaons originating from  $b \rightarrow c \rightarrow s$  decays is applied to all tagging particle candidates. The selection requirements have also been updated for LHCb Run 2 data samples using samples of  $B^\pm \rightarrow J/\psi K^\pm$  decays. Besides different PID requirements, a larger momentum of  $p > 7.5 \text{ GeV}/c$  is required.

If multiple tagging particle candidates pass the selection requirements of any of the aforementioned taggers, the particle with highest momentum is chosen. Each algorithm afterwards uses a XGBoost BDT to estimate the per-event mistag fraction  $\eta$ . The selection requirements and BDTs of the OS electron, muon, and kaon taggers have been updated for Run 2, recovering their performance obtained on Run 1 data samples.

In addition to the single particle taggers, an OS vertex charge tagger is used to reconstruct the effective charge  $Q_{\text{vtx}}$  of a secondary decay vertex within the decay chain of the OS  $b$  hadron. A BDT is used to find combinations of two tracks that originate from a  $b$  hadron. Additional tracks that are not used by other taggers, and that do not originate from the signal primary vertex are added to the secondary vertex, if compatible. The effective charge is then evaluated as the individual track charges, weighted by the track transverse momentum,

$$Q_{\text{vtx}} = \frac{\sum_{i=1}^{N_{\text{tracks}}} Q_i p_{t,i}^\kappa}{\sum_{i=1}^{N_{\text{tracks}}} p_{t,i}^\kappa}, \quad (7.16)$$

where  $\kappa$  is chosen to maximise the tagging power on a training data sample. An estimate of the mistag rate is afterwards determined using a neural network with the average kinematic properties of the vertex tracks and the effective charge as input features.

The OS charm tagger adds additional initial flavour information by reconstructing OS charm mesons from tracks that are not used by any other tagger. Tagging particle candidates are reconstructed in seven charm meson decay modes. A BDT classifies OS charm candidates and the meson with the highest BDT response is chosen as the tagging particle. More detailed information can be found in [110].

### Same Side Algorithms

Depending on the flavour of the SS hadronization partner quark, different processes can be exploited to infer SS flavour tagging information. As shown in Fig. 7.1, strange  $B$  mesons are often accompanied by SS kaons, which are selected by the SS kaon tagger.

A combination of two BDTs is trained on simulated  $B_s^0 \rightarrow D_s^- \pi^+$  decays, first of which selects good SS kaon candidates. The second BDT evaluates several event features and the response of the first BDT for up to three candidates to estimate the mistag probability  $\eta$ . The tag decision is then determined as described in Eq. (7.14). Similarly to the OS electron, muon, and kaon taggers, the SS kaon tagger has been re-optimised for LHCb Run 2 conditions by the LHCb WG, recovering and exceeding the Run 1 performance. As part of this, the BDTs superseded a combination of two neural networks. The initial algorithm is described in more detail in [111].

For signal candidates of  $B_d^0$  flavour, the SS pion and SS proton algorithms are used, each exploiting the hadronization of the light  $u$  and  $d$  quark partners of the signal meson. Both algorithms use a BDT to select the best tagging particle candidate. The BDTs are trained on a data sample of flavour specific decays of  $B_d^0 \rightarrow D^- \pi^+$ , from which only particles with small decay-times  $t < 2.2$  ps are considered. The mistag estimate is obtained by transforming the BDT response of the selected tagging particle [112].

## 7.2 Early Studies of an Inclusive Flavour Tagging Algorithm

The increased target luminosity during LHC Run 3 will pose new challenges to the flavour tagging algorithms. The performance of the current taggers is expected to decrease, given an increasing number of reconstructed PVs and associated tracks per event. Moreover, the LHCb trigger strategy poses strict requirements to the detector output rate. While the majority of recorded events was stored with the full event information during Run 2, this will only be possible for a small fraction of events during Run 3. As a consequence the flavour tagging algorithms are discussed to run in the online trigger, such that only the tag information for the selected signal candidates is stored, while the underlying non-signal tracks are discarded. This would make later re-evaluation of the tag information impossible, while adding strict CPU time constraints to each flavour tagging algorithm.

Therefore, an inclusive flavour tagging approach was proposed, taking into account the information of all underlying tracks and vertices of an event [113]. Within this algorithm, the individual track and vertex features are evaluated using two BDTs, to estimate the charge relation between signal candidate and track or signal candidate and vertex, respectively.

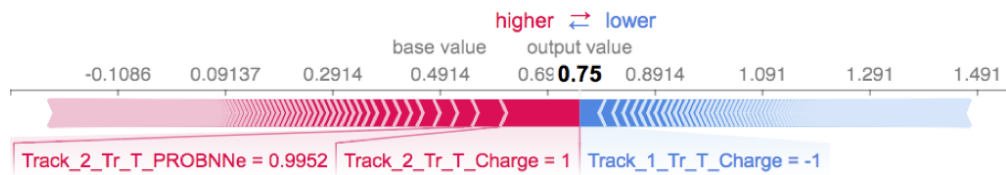
Based on this idea, the LHCb flavour tagging group is actively developing an inclusive tagging algorithm that is using a deep neural network with recurrent units (RNN, see e.g. [91, 114]) to directly predict the initial flavour of a signal candidate. This network allows to evaluate a variable number of tracks. It is planned to supersede the existing algorithm, while matching the time constraints posed by the LHCb Run 3 trigger system. While a significant increase of the tagging power has been seen in various preliminary

studies, the robustness of the algorithm must be proven. As of today, the exact flavour tagging strategy is still under discussion, but in an extreme case the inclusive tagging algorithm will be the only source of initial tag information and offline re-evaluation might be impossible for a large fraction of the Run 3 data sample.

Therefore, the SHAP values introduced in Section 5.1.3 are studied for the inclusive flavour tagging algorithm, to gain some insight into the algorithm's decision making. The results of this study are briefly discussed in the following. The study has been performed by Augustinas Šukys during a RISE internship, which was supervised by the author of this thesis. As a baseline, a neural network has been trained on 3 M simulated decays of  $B^\pm \rightarrow J/\psi K^\pm$ , using 18 features of up to 60 tracks per event, which are ordered by descending transverse momentum. A fraction of 400 k data points is used as a validation sample during the training. The network structure feeds the input vector through one dense layer, one recurrent layer using gated recurrent units, another dense layer, and into the final output node. The SHAP values are calculated using the DeepSHAP method [93].

Additionally, a XGBoost BDT has been trained on the same data sample, allowing for significantly reduced training times, which are a factor  $\sim 30$  smaller than a training of the baseline deep neural network. As explained in Section 5.1.3, the SHAP values can furthermore be calculated efficiently for the BDT.

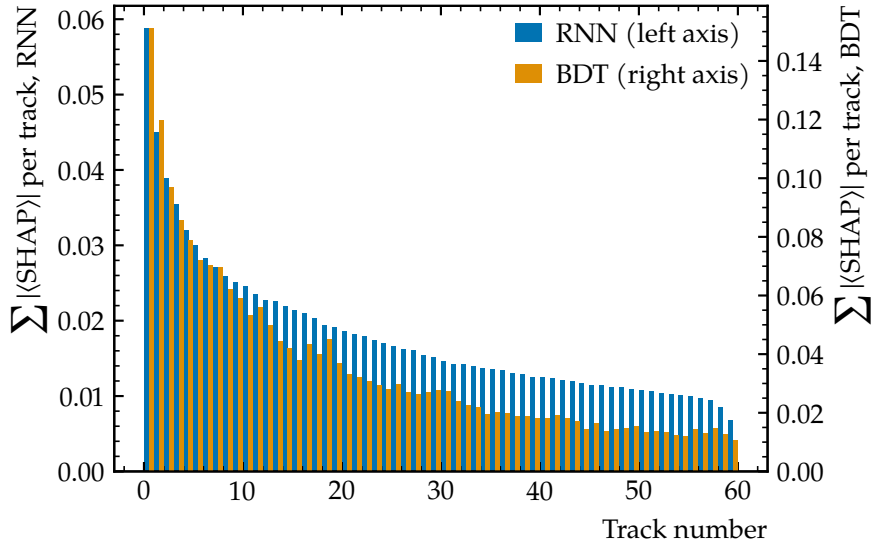
The neural network SHAP values for a single event are exemplary shown in Fig. 7.2. In this example, the second track can be considered to originate from an electron due to the value of  $P_{\text{NN}}^e \sim 1$ . The charge therefore dominates the tag decision as a  $B^-$  meson and for this single event, the algorithm behaves like an OS electron tagger. A more general



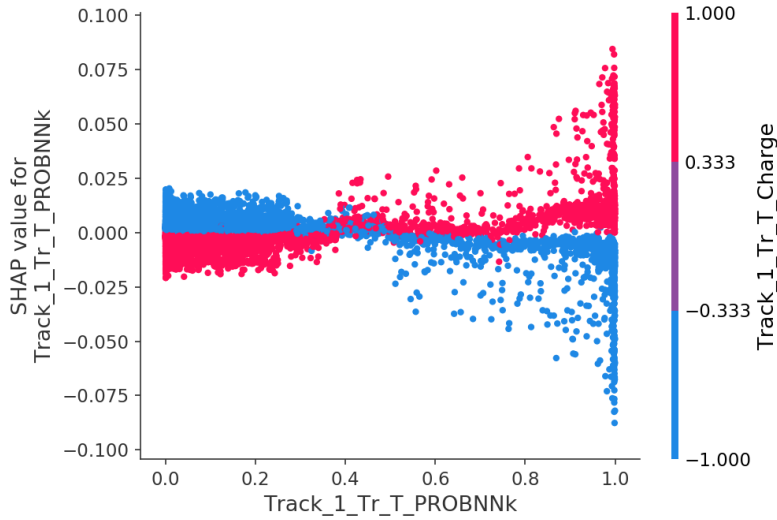
**Figure 7.2:** Contributions of individual features to a single tag prediction, according to the SHAP values. The decision is dominated by the track electron probability and charge of the second track.

judgement of the algorithms can be obtained by considering a larger number of events. The sum of the absolute SHAP values of each track is shown in Fig. 7.3, for 10 000 events. The distributions are similar for the neural network and BDT algorithms, confirming that the SHAP values provide an algorithm-agnostic measure of feature importance for the considered data sample. It is especially interesting to identify interactions of features. A scatter plot of the SHAP value for the track kaon probability is shown in Fig. 7.4 in dependence of the track kaon probability. In addition, the track charge is colored differently. At a track kaon probability of  $\sim 0.4$ , the SHAP values start to increase while the charge distribution is inverted, indicating that the algorithm identifies kaons and adjusts the tag prediction accordingly.

These insights suggest that the unintuitive machine learning algorithm are able to



**Figure 7.3:** Sum of absolute SHAP values per track for a RNN (blue bars) and a BDT (orange bars). While the absolute scale of the contributions is different, the values decrease for lower transverse momentum tracks in both cases, and the slope is similar for 10 the most important tracks.

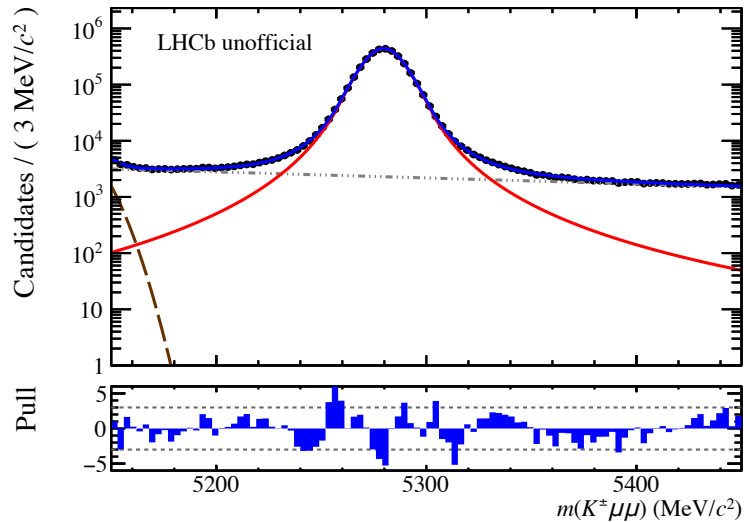


**Figure 7.4:** SHAP value of the first track kaon probability  $P_{NN}^K$  for 1000 events, depending on the value of  $P_{NN}^K$ . Tracks with positive charge are marked red, negatively charged tracks are marked blue. The dispersion of the SHAP values originates from higher-dimensional feature interaction, indicating that the RNN is able to identify kaon tracks and assign appropriate tag predictions.

exploit sensible correlations. A more thorough analysis of the Shapley values could therefore lead to a better confidence for the future usage of the algorithms or reveal potential problems. Instead of deep neural networks, BDTs could furthermore be used as proxy algorithm with faster training times, but similar predictive behaviour.

### 7.3 Opposite Side Tagging Precalibration for $B_s^0 \rightarrow D_s^- \pi^+$ Data

The decay-time dependent measurement of  $\Delta m_s$  presented in this thesis uses the combination of all available Run 2 flavour tagging algorithms. The individual OS tagger mistag estimates are precalibrated, before both OS combination and SS kaon tagging information are used in the decay time fit in Chapter 8. To prevent over-fitting effects, caused by reusing the  $B_s^0 \rightarrow D_s^- \pi^+$  data sample, the OS precalibration parameters are extracted from a sWeighted sample of  $B^\pm \rightarrow J/\psi K^\pm$  decays. The data samples are preprocessed as described in [115]. It is provided by my colleagues Vukan Jevtic, et al. as part of the measurement of the CKM angle  $\sin(2\beta)$  in decays of  $B_d^0 \rightarrow J/\psi K_S^0$ . The invariant mass distribution of the  $B^+$  candidates, with a fitted PDF to extract sWeights is shown in Figure 7.5 as an example.



**Figure 7.5:** Fit to the invariant mass of a sample of  $B^\pm \rightarrow J/\psi K^\pm$  decays, used to extract sWeights [115]. The sample is further reweighted to match some kinematic distributions to the  $B_s^0 \rightarrow D_s^- \pi^+$  data sample, and precalibration parameters for several OS taggers are extracted, as explained in the text.

The sWeighted  $B^\pm \rightarrow J/\psi K^\pm$  sample is then reweighted to match the kinematic distributions of the  $B_s^0 \rightarrow D_s^- \pi^+$  data sample, as described in the following section. Afterwards, individual calibration parameters are extracted for each OS tagging algorithm and the algorithms are combined. Finally this information is used to determine an estimated tagging power for the data sample.

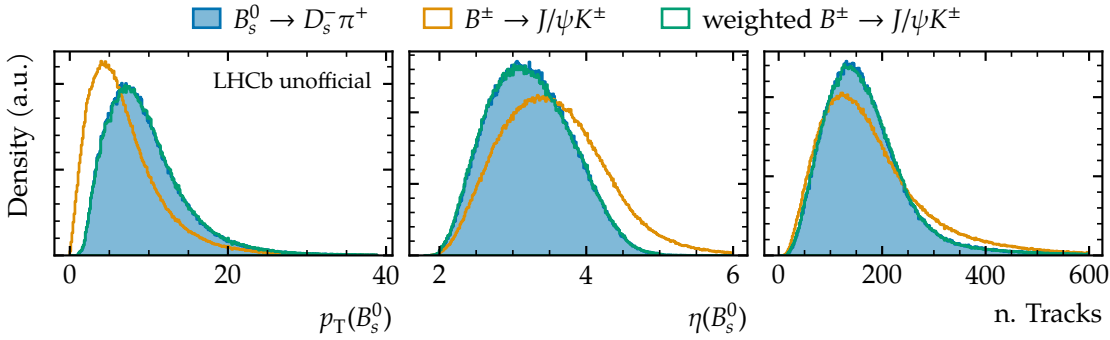


### 7.3.1 Multivariate Reweighting of $B^\pm \rightarrow J/\psi K^\pm$ Data

Due to the different decay topology of the calibration sample of  $B^\pm \rightarrow J/\psi K^\pm$  decays, weights are extracted to match three key distributions of the calibration sample to the sample of  $B_s^0 \rightarrow D_s^- \pi^+$  decays. While the OS tagging algorithms are expected to be independent of the signal  $B$  meson decay, differences in the decay-specific trigger and selection requirements translate into different OS tagging results.

The distributions of the pseudorapidity  $\eta$ , the number of tracks per event  $N_{\text{tracks}}$ , and the transversal momentum  $p_T$  are displayed in Fig. 7.6. They are used to extract weights to match the  $B^\pm \rightarrow J/\psi K^\pm$  data sample to the  $B_s^0 \rightarrow D_s^- \pi^+$  data sample, using the GBREWEIGHTING method, as implemented in the `hep_ml` package [116].

As shown in Fig. 7.7, the effect is most dominant for the OS kaon tagger. A later study, using a naive combination of uncalibrated OS taggers suggests that the overall effect of the precalibration is negligible. The effect on  $CP$  parameters in measurements, such as  $B_s^0 \rightarrow D_s^\mp K^\pm$ , is expected to be larger [104].

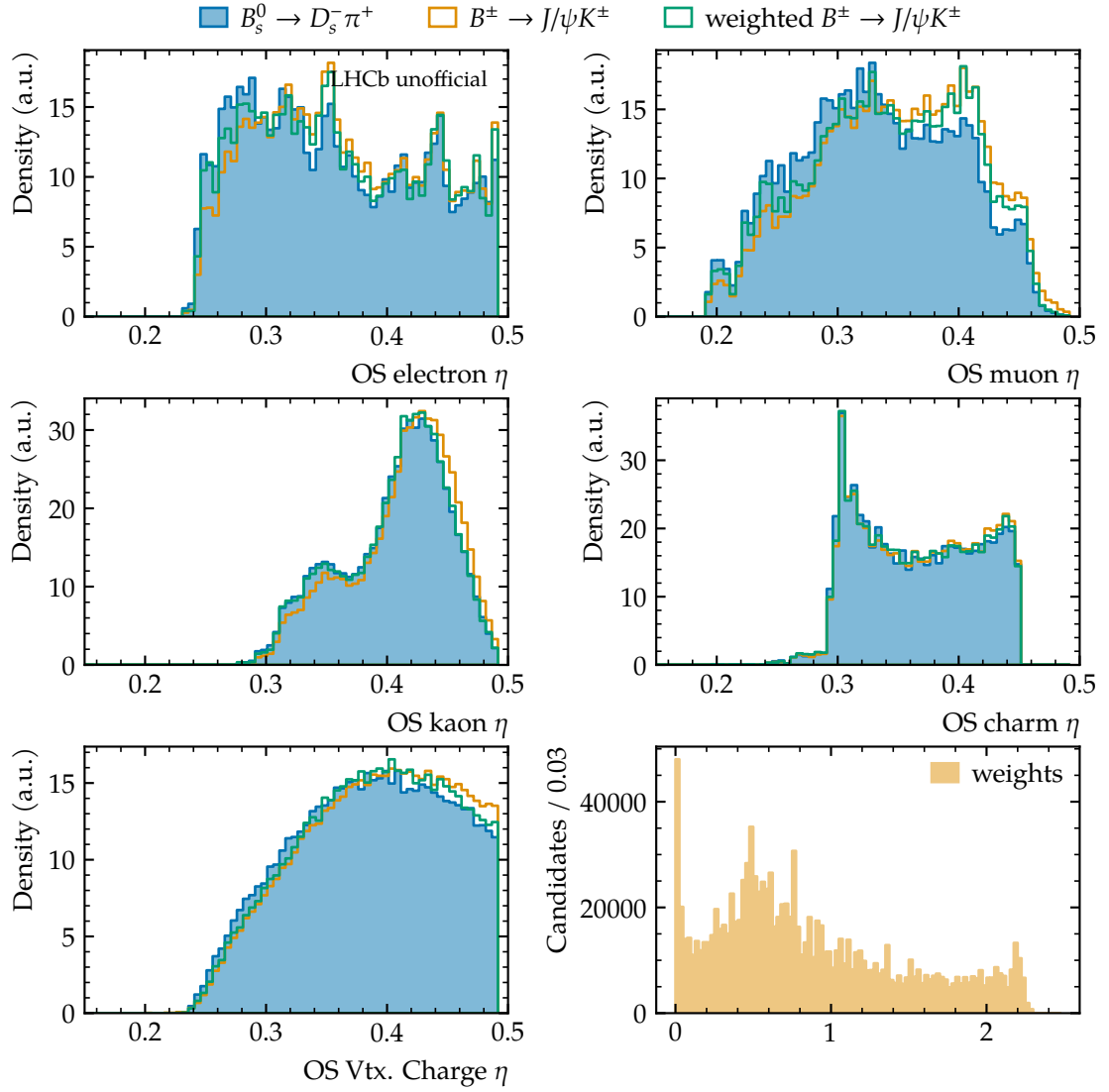


**Figure 7.6:** sPlots of transverse momentum  $p_T$ , pseudorapidity  $\eta$  and number of tracks,  $N_{\text{tracks}}$ , used to reweight  $B^\pm \rightarrow J/\psi K^\pm$  data to the  $B_s^0 \rightarrow D_s^- \pi^+$  data sample. The per-year samples are combined in the plots.

### 7.3.2 Calibration and Combination of Opposite Side Tagging Algorithms

Primarily to reduce the number of fit parameters, all OS taggers need to be combined into a single tag decision  $d^{\text{OS}}$  and mistag estimate  $\eta^{\text{OS}}$ . To decorrelate the individual OS taggers, and to linearise and thus simplify the calibration of the OS combination, each OS tagger is first precalibrated using the previously described sample of  $B^\pm \rightarrow J/\psi K^\pm$  decays. The sample is weighted to match the signal  $B_s^0 \rightarrow D_s^- \pi^+$  kinematic distributions, as described in subsection 7.3.1.

A second order polynomial with a logistic link function, as described in subsection 7.1.1 is used for the precalibration. Non-linearities such as seen for the OS electron tagger in e.g. the OS electron tagger are well-modelled resulting in a OS combination that can be described with a linear function such as Eq. (7.10), reducing the number of free parameters in the final decay-time fit.

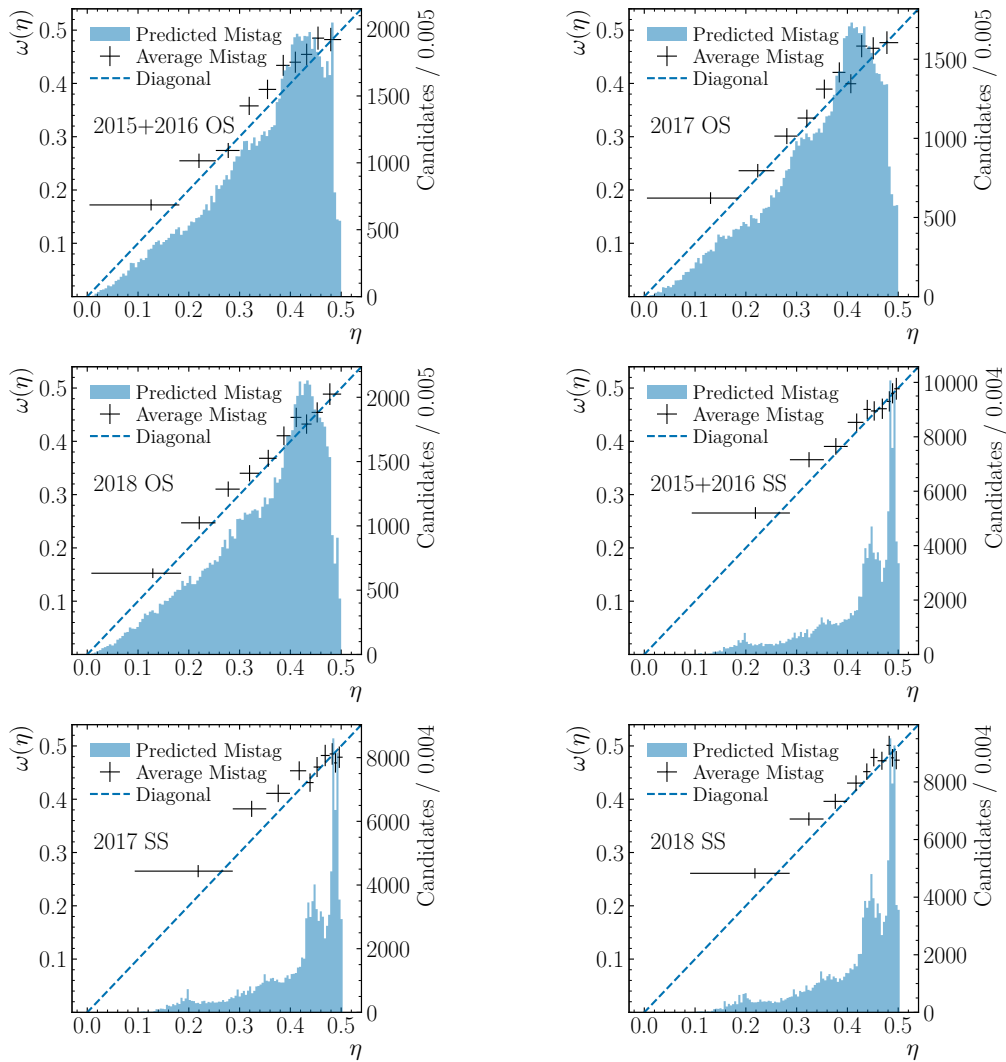


**Figure 7.7:** Distribution of the predicted mistag estimates of the  $B^\pm \rightarrow J/\psi K^\pm$ ,  $B_s^0 \rightarrow D_s^- \pi^+$ , and reweighted  $B^\pm \rightarrow J/\psi K^\pm$  data samples. The weight distribution, which is applied to the  $B^\pm \rightarrow J/\psi K^\pm$  samples, is shown in the lower right plot.

### 7.3 Opposite Side Tagging Precalibration for $B_s^0 \rightarrow D_s^- \pi^+$ Data

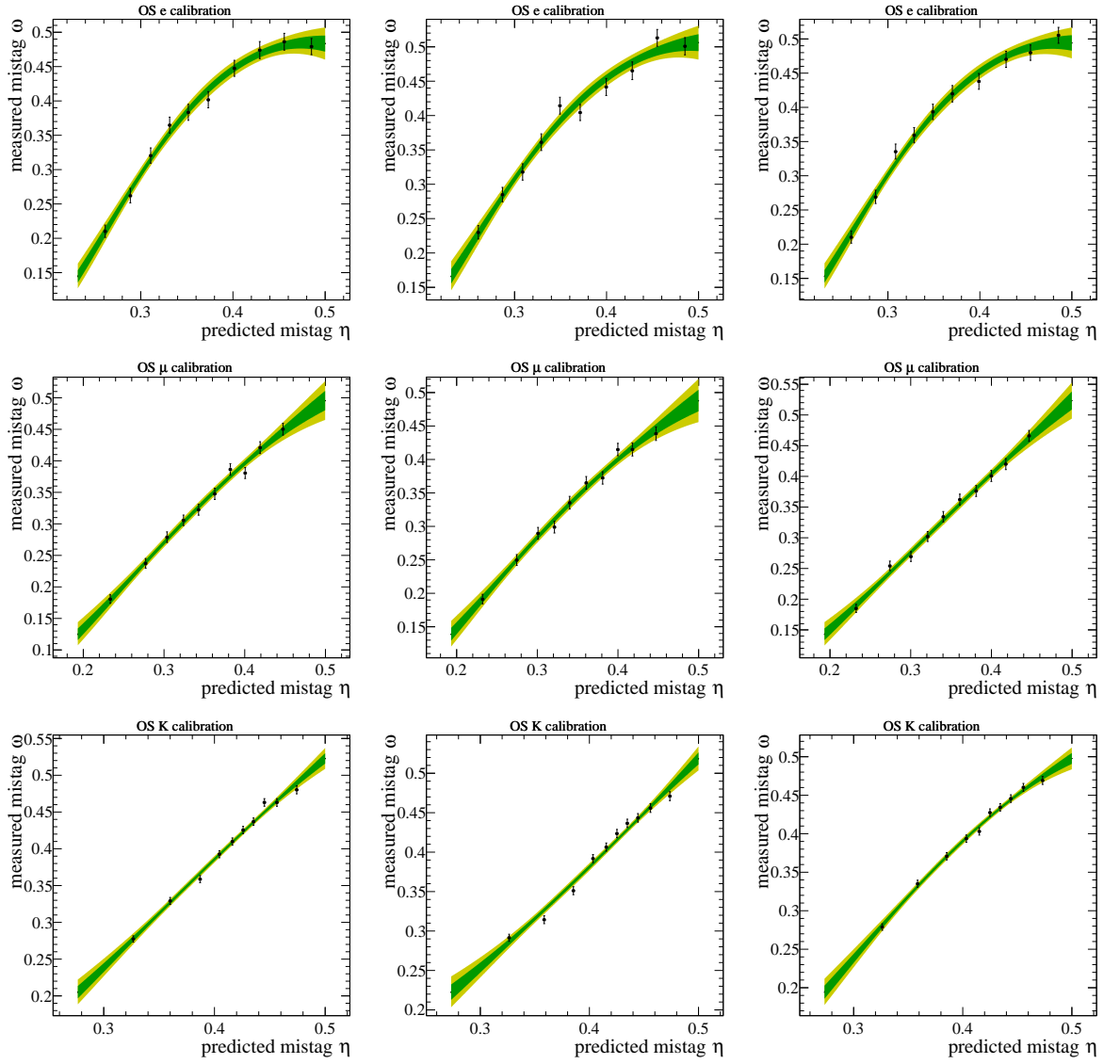
The precalibration of the OS tagging algorithms is performed using the Espresso Performance Monitor, extracting a set of  $i = 1 \dots 3$  tagging parameters  $p_i^t$  per OS tagging algorithm  $t$ . The distribution of the combined OS mistag is shown in the first three graphs of Figure 7.8, together with the average mistag in bins of the predicted mistag. The calibration curves, extracted from the  $B^\pm \rightarrow J/\psi K^\pm$  data sample are shown in Figs. 7.9 and 7.10. In both cases, a value of  $\Delta m_s = 17.757 \text{ ps}^{-1}$  is assumed. As shown in the remaining graphs of Figure 7.8, the SS kaon mistag distribution can also be well described using a linear function such that the mistag estimates are directly used in the decay time fit in Chapter 8.

An estimate of the final tagging power can be extracted in a third step, again assuming the aforementioned value for  $\Delta m_s$ . This yields an expected tagging power of  $\varepsilon_{\text{tag}}^* \approx 7\%$ . No cross-validation is performed for this estimate, explaining an overestimated tagging power with respect to the actual tagging power reported in chapter 8.



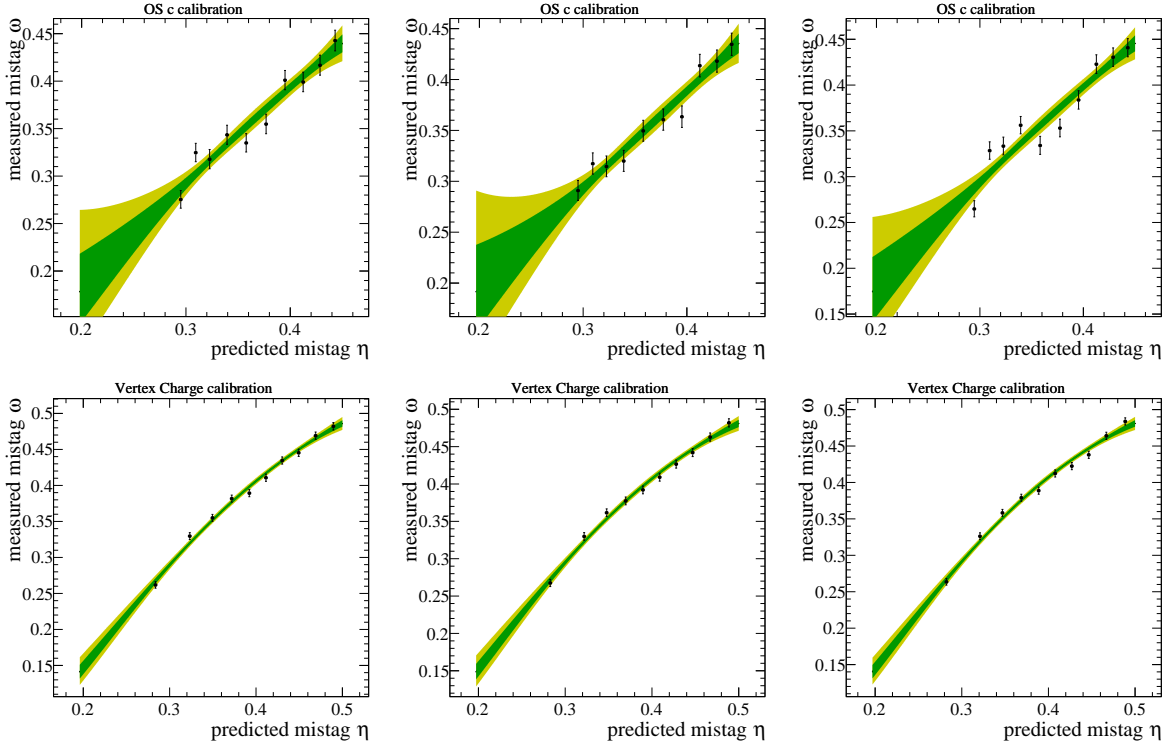
**Figure 7.8:** Mistag distributions and average mistag  $\omega$  per bin of the estimated mistag  $\eta$ , assuming  $\Delta m_s = 17.757 \text{ ps}^{-1}$ . From top left to bottom right: OS combination, separately for the per-year data sub-samples, followed by SS kaon mistag distributions.

## 7 Flavour Tagging Studies



**Figure 7.9:** Individual OS tagger calibration curves, described with a second order polynomial and a logistic link function, extracted from weighted  $B^\pm \rightarrow J/\psi K^\pm$  samples. From top to bottom: OS electron, OS muon, and OS kaon taggers. From left to right: 2015–2016, 2017, and 2018 data samples.

### 7.3 Opposite Side Tagging Precalibration for $B_s^0 \rightarrow D_s^- \pi^+$ Data



**Figure 7.10:** Individual OS tagger calibration curves, described with a second order polynomial and a logistic link function, extracted from weighted  $B^\pm \rightarrow J/\psi K^\pm$  samples. From top to bottom: OS charm and OS vertex charge taggers. From left to right: 2015–2016, 2017, and 2018 data samples.



# 8 Measurement of the Oscillation Frequency $\Delta m_s$

The thus far described studies allow to perform a likelihood fit of the PDF described in Eqs. (3.21) to (3.24) to a statistically pure sample of  $B_s^0 \rightarrow D_s^- \pi^+$  decays, determining the value of  $\Delta m_s$ .

Detector effects are taken into account with several adjustments to the theoretical PDF: The detector resolution is taken into account by convolving the PDF with a resolution function  $R(t|\delta_t)$  and the detector acceptance is described as a decay-time dependent function  $\varepsilon(t)$ . Most importantly, the effective  $CP$  coefficients  $C_{\text{eff}}^{\text{cosh}}(\vec{d}, \vec{\eta})$  and  $C_{\text{eff}}^{\text{cos}}(\vec{d}, \vec{\eta})$  parametrise the flavour tagging information from the previous Chapter as well as production and detection asymmetries,  $a_{\text{prod}}$  and  $a_{\text{det}}$ , respectively.

In the following sections, more details are described for the acceptance function (Section 8.1) and the decay-time uncertainty calibration, which is also used to correct a decay-time bias (Section 8.2). The final fit to data, using the adjusted decay-time PDF and the fit result are described in Section 8.3. Several fit validation tests are shown in Section 8.4

## 8.1 Description of the Detector Acceptance with Cubic Splines

As described in Section 4.2.1, the tracking system, and especially the VELO detector, provide very high vertex resolution, allowing to distinguish primary and secondary vertex. The particle decay-time  $t$  can be deduced with high precision by combining the vertex information with momentum information. At very low decay-times  $t \lesssim 1$  ps however, the vertex separation becomes ambiguous, effectively reducing the reconstruction efficiency for these events. This effect is further amplified with the BDT selection, shown in Section 6.2.2.

The effect is modelled by a decay-time dependent acceptance function  $\varepsilon_{\text{acc}}(t)$ , for which a cubic spline function is used. The spline is built from  $n_{\text{acc}} = 8$  cubic B-splines  $B_i(t|\vec{k})$ , with knots positioned at  $k_i \in [0.4, 0.5, 1.0, 1.5, 2.0, 3.0, 12.0, 15.0]$ ps, where first and last knot position are repeated threefold:

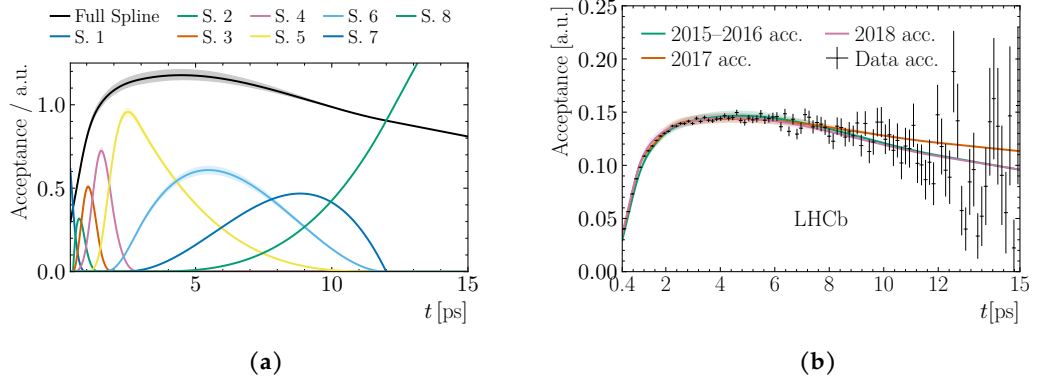
$$\varepsilon_{\text{acc}}(t) = \sum_{i=0}^7 v_i B_i(t|\vec{k}). \quad (8.1)$$

The B-splines are implemented similarly to [117], but linearly extrapolated beyond the first and last knot. To improve the fit stability, especially at large decay-times  $t \gtrsim 8$  ps, where the statistics is very limited, the coefficient at knot  $k_6 = 12.0$  ps is fixed to  $v_6 = 1$  and

coefficient  $v_7$  is calculated as in in the previous  $CP$  violation measurement in  $B_s^0 \rightarrow D_s^\mp K^\pm$  decays [104]:

$$v_7 = v_6 + \frac{v_5 - v_6}{t_5 - t_6} (t_7 - t_6). \quad (8.2)$$

Similarly to the Run 1  $CP$  violation measurement, this function could be used to model the detector acceptance in the planned update of the  $B_s^0 \rightarrow D_s^\mp K^\pm$  measurement with Run 2 data.



**Figure 8.1:** Cubic spline functions to model a decay-time dependent efficiency. The individual spline components for the 2018 sample are shown in (a). The fitted acceptance functions for all data samples, together with the combined acceptance data and decay-time distribution is shown in (b).

Figure 8.1a shows the individual spline components with a common, arbitrary normalization, as well as the full spline function. The decay-time histogram can be normalized to a decay-time PDF with no flavour tagging and fixed decay constants  $\Gamma_s$  and  $\Delta\Gamma_s$  to visualize the acceptance in data. It is shown in Fig. 8.1b for all data samples simultaneously, using the splines of the nominal fit result described later in this chapter.

Studies with pseudo experiments (see Section 8.4.1) show a significant bias of the knot coefficient  $v_1$ , which is why several alternative parametrisations have been tested. Since no effect on the measurement of  $\Delta m_s$  is observed and a significant amount of computing time needs to be invested to recheck for biases, the initial set of knot positions is unchanged.

## 8.2 Correction of the Decay-time and Decay-time Uncertainty

Similarly to the imperfect knowledge of the tag information, the imperfect estimation of the decay-time uncertainty  $\delta_t$  can affect the measurement of  $\Delta m_s$ . Therefore, the decay-time uncertainty, which is estimated by the vertex fitter (see Section 6.2.1), needs to be calibrated. While determining the calibration procedure, a decay-time bias has been observed that induces a significant shift of  $\Delta m_s$  and hence need to be corrected. The decay-time calibration and decay-time bias correction are described further in the following section. These studies have been performed by my co-author of [18], Michele



Veronesi. Therefore, the description will only be as detailed as needed to follow the general procedure. The figures shown in this section are based on the original work and only visually adjusted.

### 8.2.1 Decay-time Uncertainty Calibration

A sample of prompt  $D_s^-$  candidates is used to construct fake, prompt  $D_s^- \pi^+$  candidates by combining the  $D_s^-$  candidate with a random  $\pi^+$  track from the primary vertex. The decay-time distribution of these fake candidates is centered at  $t = 0$ , by definition. The mass distribution of the prompt  $D_s^-$  candidates is used to construct sPlots of the decay-time distribution. A maximum likelihood fit of a double crystal ball function to describe the signal component, and an exponential function, to describe the combinatorial background, is used to extract the sWeights.

This procedure is repeated in multiple bins of the decay-time uncertainty, which is obtained from the decay-tree fitter (see Section 6.2.1). The widths of the decay-time sPlots are then compared to the average decay-time uncertainty in each bin. In each decay-time uncertainty bin, the width is extracted by fitting three Gaussian PDFs to the decay-time distribution, with a shared mean, individual widths  $\sigma_i$ , and recursively defined fractions between the Gauss components,  $f_i$ . A combined dilution is computed via

$$D_{\text{res}} = \sum_{i=0}^1 f_i \cdot e^{\sigma_i^2 \frac{\Delta m_s^2}{2}},$$

taking only the narrow widths,  $\sigma_0$  and  $\sigma_1$  into account. The wide gauss component is assumed to originate from background candidates and therefore not included. Afterwards, an average width is calculated via

$$\sigma_{\text{eff}} = \sqrt{-\frac{2}{\Delta m_s^2} \ln D_{\text{res}}}.$$

A linear function is fitted to the  $(\sigma_{\text{eff}}, \bar{\delta}_t)$  pairs, with the average decay-time uncertainty  $\bar{\delta}_t$  in each bin, and calibration parameters  $r_0$  and  $r_1$ :

$$\sigma_t(\delta_t) = r_0 + r_1 \cdot \delta_t. \quad (8.3)$$

The fit result is shown in Fig. 8.2, exemplary for the 2017 data sample. The resulting calibration parameters are listed in Table 8.1.

### 8.2.2 Decay-time Bias Correction

The decay-time uncertainty calibration studies reveal a decay-time bias, i.e. a non-vanishing mean of the prompt decay-time distribution. As described in Section 6.2.4, this

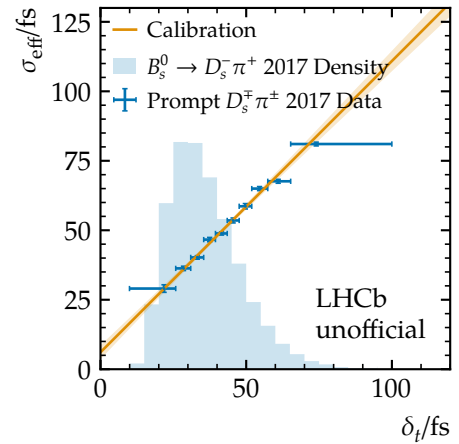


Figure 8.2: Calibration function for the decay-time uncertainty, estimated by the decay-tree fitter.

**Table 8.1:** Decay-time uncertainty calibration parameters. A linear function is used to calibrate the estimated decay-time uncertainty in the final decay-time fit.

Sample	$r_0/\text{fs}$	$r_1$	$\langle\delta_t\rangle/\text{fs}$
20152016	$8.4 \pm 2.6$	$1.002 \pm 0.059$	46.97
2017	$6.1 \pm 2.5$	$1.048 \pm 0.059$	45.90
2018	$5.6 \pm 2.5$	$1.052 \pm 0.059$	45.74

is partly reduced by an optimised offline selection. A remaining bias of several fs has a significant effect on the  $\Delta m_s$  measurement, shifting the central value by more than  $3\sigma$ . To prevent a systematic limitation of the measurement, this effect needs to be corrected.

The bias originates from the misalignment of the VELO detector. As described in Section 4.2.1, the detector consists of two movable halves on which the detector components are mounted. Studies with simulated events and an artificially misaligned VELO detector show that a misalignment in the  $x$ -direction, i.e. the direction of motion of the halves, or the rotation around the  $y$ -axis, can induce aforementioned decay-time bias.

**Table 8.2:** Decay-time bias of the  $B_s^0 \rightarrow D_s^- \pi^+$  data samples.

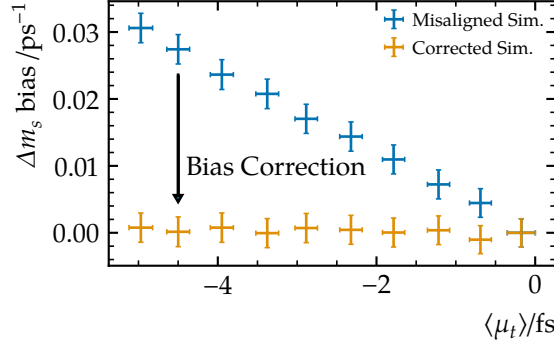
Sample	$\varphi_{\text{sig}}/\text{fs}$
20152016	$-2.25 \pm 0.11$
2017	$-3.05 \pm 0.11$
2018	$-2.39 \pm 0.11$

The bias can be effectively corrected by shifting the mean of the resolution function  $R(t|\delta_t)$ . The correction factor can only be measured in prompt  $D_s^\mp \pi^\pm$  samples and needs to be calibrated for the signal sample of  $B_s^0 \rightarrow D_s^- \pi^+$  decays. The calibration function is extracted from simulated samples, for which ten different misalignments of the VELO halves'  $x$ -alignment are induced, ranging from between  $\Delta T_x = 0 \mu\text{m}$  to  $\Delta T_x = 9 \mu\text{m}$ . For each simulated sample, the prompt and signal decay-time bias,  $\varphi_{\text{pr}}$  and  $\varphi_{\text{sig}}$  respectively, are determined and

a linear function is fitted to these pairs. This yields the calibration function

$$\varphi_{\text{sig}}(\varphi_{\text{pr}}) = (0.031 \pm 0.197) \text{fs} + (0.496 \pm 0.032) \text{fs}^{-1} \cdot \varphi_{\text{pr}}. \quad (8.4)$$

The decay-time bias is afterwards determined individually for each data sample and listed in Table 8.2. The effect of the calibration, applied to the differently misaligned, simulated samples is shown in Fig. 8.3.



**Figure 8.3:** Fitted values of  $\Delta m_s$  for different misaligned, simulated samples, before (blue) and after (orange) decay time bias correction.

### 8.3 Decay-time Fit to Flavour-tagged Data

Finally, to extract  $\Delta m_s$ , the tagging information needs to be parametrised. This is incorporated into effective  $CP$  coefficients which also allow for production and detection asymmetries,  $a_{\text{prod}}$  and  $a_{\text{det}}$ , respectively:

$$\begin{aligned} C_{\text{eff}}^{\text{cos}}(\vec{d}, \vec{\eta}) &= q(1 - q \cdot a_{\text{det}}) \cdot (\Delta^- - a_{\text{prod}}\Delta^+) \cdot C_{\text{phys}}^{\text{cos}} \\ \text{and} \quad C_{\text{eff}}^{\text{cosh}}(\vec{d}, \vec{\eta}) &= (1 + q \cdot a_{\text{det}}) \cdot (\Delta^+ - a_{\text{prod}}\Delta^-) \cdot C_{\text{phys}}^{\text{cosh}}. \end{aligned} \quad (8.5)$$

Here,  $q$  is the final state charge, which is given by the leftover pion charge. The detection and production asymmetries are defined as

$$a_{\text{prod}} = \frac{\sigma(\bar{B}_s^0) - \sigma(B_s^0)}{\sigma(\bar{B}_s^0) + \sigma(B_s^0)} \quad \text{and} \quad a_{\text{det}} = \frac{\varepsilon(D_s^+ \pi^-) - \varepsilon(D_s^- \pi^+)}{\varepsilon(D_s^+ \pi^-) + \varepsilon(D_s^- \pi^+)}, \quad (8.6)$$

with the production cross-section  $\sigma(B_s^0)$  of initial states, and the reconstruction and detection efficiencies  $\varepsilon(D_s^\mp \pi^\pm)$  of the final states. These effects are incorporated into the decay-time PDFs similarly to Eq. (7.6), adding a factor  $1 - a_{\text{prod}}$  to the decay rates with initial  $B_s^0$  mesons, a factor  $1 - a_{\text{det}}$  to the decay rates with final  $D_s^+ \pi^-$  states, and inverting the signs for the respective opposing flavours. The coefficients of Eq. (8.5) are then obtained by expanding all terms and comparing the coefficients of the trigonometric and hyperbolic functions. The flavour tagging information of both OS and SS taggers are taken into account simultaneously in the  $\Delta^\pm$  parameters. They include the tag decisions  $d_i$ , the flavour-specific tagging efficiencies

$$\varepsilon_b^i = \varepsilon_{\text{tag}}^i - \frac{1}{2}\Delta \varepsilon_{\text{tag}}^i, \quad (8.7)$$

$$\text{and} \quad \varepsilon_{\bar{b}}^i = \varepsilon_{\text{tag}}^i + \frac{1}{2}\Delta \varepsilon_{\text{tag}}^i, \quad (8.8)$$

and the flavour-specific calibrated mistag  $\omega(\eta)$  as defined in Eq. (7.10). For signal candidates that have been tagged by both taggers, i.e.  $d^{\text{OS}} = \pm 1$ ,  $d^{\text{SS}} = \pm 1$ , the  $\Delta^\pm$  parameters

are defined as

$$\begin{aligned} \Delta^\pm &= \frac{1}{4} \varepsilon_b^{\text{OS}} \varepsilon_b^{\text{SS}} \left[ 1 + d^{\text{SS}} (1 - 2\omega_b^{\text{SS}}) + d^{\text{OS}} (1 - 2\omega_b^{\text{OS}}) \right. \\ &\quad \left. + d^{\text{OS}} d^{\text{SS}} (1 - 2\omega_b^{\text{SS}} - 2\omega_b^{\text{OS}} + 4\omega_b^{\text{OS}} \omega_b^{\text{SS}}) \right] \\ &\quad \pm \frac{1}{4} \varepsilon_b^{\text{OS}} \varepsilon_b^{\text{SS}} \left[ 1 - d^{\text{SS}} (1 - 2\omega_b^{\text{SS}}) - d^{\text{OS}} (1 - 2\omega_b^{\text{OS}}) \right. \\ &\quad \left. + d^{\text{OS}} d^{\text{SS}} (1 - 2\omega_b^{\text{SS}} - 2\omega_b^{\text{OS}} + 4\omega_b^{\text{OS}} \omega_b^{\text{SS}}) \right], \end{aligned} \quad (8.9)$$

for candidates tagged by only one tagger,  $d^i \neq d^j = 0$ , they are defined as

$$\begin{aligned} \Delta^\pm &= \frac{1}{2} \varepsilon_b^i \left[ 1 - \varepsilon_b^j + d_i (1 - \varepsilon_b^j - 2\omega_b^i + 2\omega_b^i \varepsilon_b^j) \right] \\ &\quad \pm \frac{1}{2} \varepsilon_b^i \left[ 1 - \varepsilon_b^j + d_i (1 - \varepsilon_b^j - 2\omega_b^i + 2\omega_b^i \varepsilon_b^j) \right], \end{aligned} \quad (8.10)$$

and for untagged signal candidates,  $d^{\text{OS}} = d^{\text{SS}} = 0$ , they are defined as

$$\Delta^\pm = (1 - \varepsilon_b^{\text{OS}} - \varepsilon_b^{\text{SS}} + \varepsilon_b^{\text{OS}} \varepsilon_b^{\text{SS}}) \pm (1 - \varepsilon_b^{\text{OS}} - \varepsilon_b^{\text{SS}} + \varepsilon_b^{\text{OS}} \varepsilon_b^{\text{SS}}). \quad (8.11)$$

Using the effective  $CP$  coefficients of Eq. (8.5), the effective decay-time PDF follows to be

$$P_{\text{eff}}(t|q, \vec{d}, \vec{\eta}) \sim e^{-\Gamma_s t} \left[ C_{\text{eff}}^{\text{cosh}}(\vec{d}, \vec{\eta}) \cdot \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + C_{\text{eff}}^{\text{cos}}(\vec{d}, \vec{\eta}) \cdot \cos(\Delta m_s t) \right]. \quad (8.12)$$

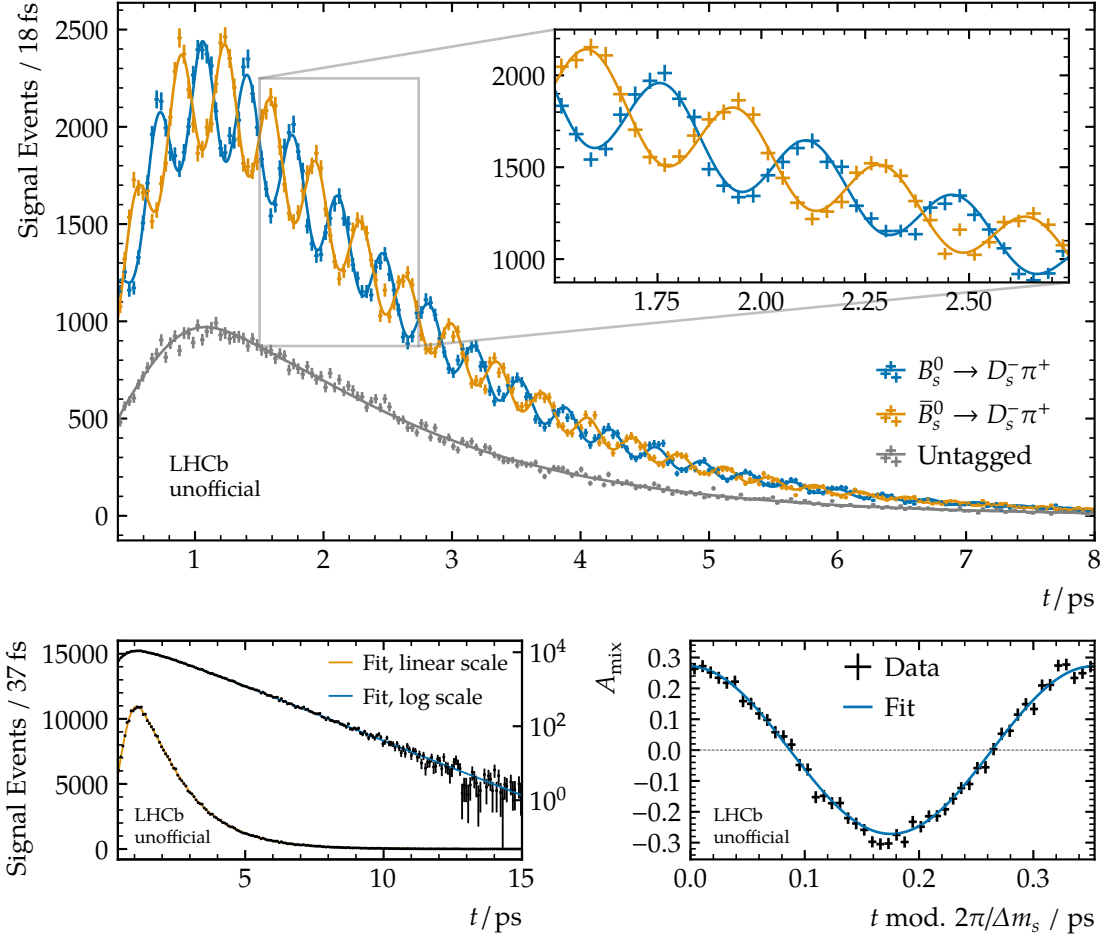
The full decay-time PDF includes the PDFs for both tagging categories,  $P(\eta^{\text{OS}})$  and  $P(\eta^{\text{SS}})$ , as well as the decay-time uncertainty distribution  $P(\delta_t)$ , all of which are described with a corresponding data histogram. The decay-time PDF is conditional on the final state  $q$ , the flavour tagging information  $\vec{d}$  and  $\vec{\eta}$ , as well as the decay-time resolution  $\delta_t$ .

Neglecting a normalization term, the PDF is

$$\mathcal{P}(t; \delta_t, q, \vec{d}, \vec{\eta}) \sim \varepsilon(t) \cdot P(\eta^{\text{OS}}) \cdot P(\eta^{\text{SS}}) \cdot P(\delta_t) \int R(t - t' | \delta_t) \cdot P_{\text{eff}}(t' | q, \vec{d}, \vec{\eta}) dt'. \quad (8.13)$$

To describe variations between data samples, the PDF is fitted simultaneously to all data samples, but with individual sets of parameters. Only the oscillation frequency  $\Delta m_s$ , and the production and detection asymmetries  $a_{\text{prod}}$  and  $a_{\text{det}}$  are shared across all samples. The estimated statistic uncertainties of the fitted parameters are corrected with the asymptotically correct approach [118]. Projections of the decay-time fit result are shown in Fig. 8.4, the fitted parameters are listed in Tables 8.3 and 8.4, and fixed fit parameters are listed in Tables 8.5 and 8.5b.

The tagging power of the sample is re-evaluated with these results. The combined mistag rate of the sample is  $\omega = (36.21 \pm 0.17)\%$ , with a tagging efficiency of  $\varepsilon_{\text{tag}} = (80.30 \pm 0.07)\%$ . This corresponds to an effective tagging efficiency of  $\varepsilon_{\text{eff}} = (6.10 \pm 0.15)\%$ .



**Figure 8.4:** Projections of the decay-time fit result. The top plot displays the fitted unmixed ( $B_s^0 \rightarrow D_s^- \pi^+$ ), unmixed ( $\bar{B}_s^0 \rightarrow D_s^- \pi^+$ ), and untagged decay-rates and corresponding data samples. The lower left plot shows the combined PDF projects and correspond data sample in linear scale and logarithmic scale. The lower right plot shows a projection of the mixing asymmetry Eq. (3.41) into one first oscillation period.

**Table 8.3:** Free parameters which are shared among all data samples. The parameters include the oscillation frequency  $\Delta m_s$  and the production asymmetry  $a_{\text{prod}}$ .

Param.	Value	Unit
$\Delta m_s$	$17.7683 \pm 0.0051$	$\text{ps}^{-1}$
$a_{\text{prod}}$	$-0.0031 \pm 0.0032$	

**Table 8.4:** Floating, per-dataset parameters of the decay-time fit, including the spline acceptance coefficients  $v_i$ , the flavour tagging efficiencies  $\varepsilon_{\text{tag}}$ , calibration parameters  $p_i$ , and their asymmetries  $\Delta\varepsilon_{\text{tag}}$  and  $\Delta p_i$ , respectively.

Param.	2015–2016	2017	2018
$p_0^{\text{OS}}$	$0.385 \pm 0.004$	$0.376 \pm 0.004$	$0.374 \pm 0.004$
$p_0^{\text{SS}}$	$0.4345 \pm 0.0032$	$0.4373 \pm 0.0032$	$0.4373 \pm 0.0030$
$p_1^{\text{OS}}$	$0.99 \pm 0.04$	$0.88 \pm 0.04$	$0.882 \pm 0.035$
$p_1^{\text{SS}}$	$0.747 \pm 0.035$	$0.71 \pm 0.04$	$0.783 \pm 0.033$
$\Delta p_0^{\text{OS}}$	$0.008 \pm 0.004$	$0.004 \pm 0.005$	$0.012 \pm 0.004$
$\Delta p_0^{\text{SS}}$	$-0.0163 \pm 0.0033$	$-0.0220 \pm 0.0035$	$-0.0123 \pm 0.0033$
$\Delta p_1^{\text{OS}}$	$0.00 \pm 0.04$	$0.07 \pm 0.04$	$0.02 \pm 0.04$
$\Delta p_1^{\text{SS}}$	$0.01 \pm 0.04$	$0.06 \pm 0.04$	$0.05 \pm 0.04$
$\Delta\varepsilon_{\text{tag}}^{\text{OS}}$	$0.007 \pm 0.005$	$0.003 \pm 0.005$	$-0.003 \pm 0.005$
$\Delta\varepsilon_{\text{tag}}^{\text{SS}}$	$-0.001 \pm 0.005$	$0.002 \pm 0.005$	$-0.007 \pm 0.005$
$v_1$	$0.325 \pm 0.010$	$0.345 \pm 0.011$	$0.382 \pm 0.011$
$v_2$	$0.468 \pm 0.015$	$0.477 \pm 0.017$	$0.533 \pm 0.018$
$v_3$	$0.779 \pm 0.024$	$0.830 \pm 0.026$	$0.860 \pm 0.027$
$v_4$	$0.958 \pm 0.029$	$0.906 \pm 0.029$	$1.024 \pm 0.031$
$v_5$	$1.107 \pm 0.029$	$1.078 \pm 0.030$	$1.132 \pm 0.030$
$v_6$	$1.33 \pm 0.05$	$1.18 \pm 0.05$	$1.31 \pm 0.05$

**Table 8.5:** Constant per-data-sample parameters (a), and shared parameters (b), including the average mistag rates  $\langle\eta\rangle$ , the decay-time bias correction  $\hat{\mu}_t$ , the decay-time resolution calibration parameters  $r_i$ , the flavour tagging efficiency  $\varepsilon_{\text{tag}}^{\text{OS,SS}}$ , and the fixed acceptance spline coefficient  $v_7$ . The shared parameters include the  $CP$  coefficient  $C$ , the decay constants  $\Gamma_s$  and  $\Delta\Gamma_s$ , and the detection asymmetry  $a_{\text{det}}$ , which is fixed to prevent spoiling a potential future measurement.

(a)					(b)		
Param.	2015–2016	2017	2018	Unit	Param.	Value	Unit
$\langle\eta\rangle^{\text{OS}}$	0.3562	0.3463	0.3464		$C$	1.0	
$\langle\eta\rangle^{\text{SS}}$	0.4162	0.4164	0.4156		$\Gamma_s$	0.6600	$\text{ps}^{-1}$
$\varepsilon_{\text{tag}}^{\text{OS}}$	0.4126	0.4084	0.4123		$\Delta\Gamma_s$	0.085	$\text{ps}^{-1}$
$\varepsilon_{\text{tag}}^{\text{SS}}$	0.6918	0.6992	0.6973		$a_{\text{det}}$	0	
$\hat{\mu}_t$	-2.254	-3.047	-2.394	fs			
$r_0$	8.377	6.092	5.5521	fs			
$r_1$	1.002	1.048	1.052				
$v_7$	1.0	1.0	1.0				

## 8.4 Model Validation

The robustness of the fit procedure is tested in multiple ways, which are explained in more detail in the following sections. Two versions of tests using pseudo data samples are described in Sections 8.4.1 and 8.4.2. In addition, the fit procedure is repeated using different, independent sub-samples of the data, shown in Section 8.4.3.

### 8.4.1 Pseudo Experiments

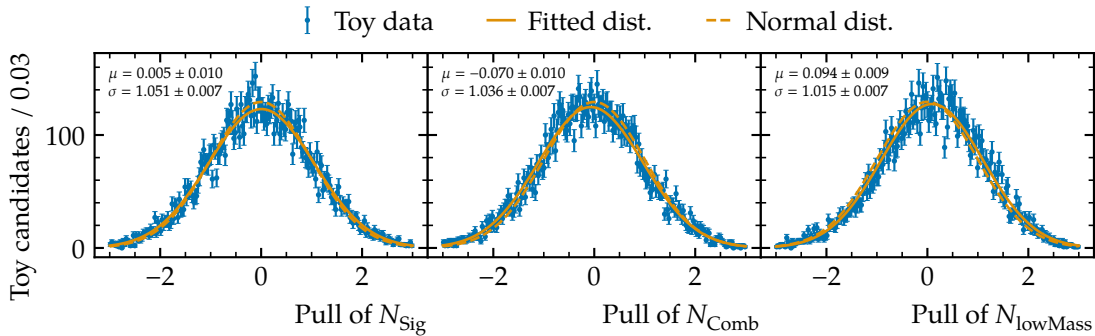
One set of tests uses simulated pseudo experiments, or “toy” experiment, for which the data samples are randomly drawn in 8 uncorrelated data sample dimension from the fitted PDFs from Eqs. (6.9) and (8.13). For each toy experiment, individual sub-samples are generated based on the different data samples, i.e.  $D_s^-$  final states and year of data taking. The number of generated events per toy sample is drawn from a Poisson distribution with the number of occurrences matching the nominally fitted yields from Table 6.5. The generated sub-samples are then combined to be technically indistinguishable from the data samples.

Both likelihood fits, the mass fit and the decay-time fit, are repeated 1500 times for each sample and the distributions of fitted parameters is tested for dominant systematic deviations from the generated values. This is commonly parametrised with the “pull” of a parameter  $x$ ,

$$g_i = \frac{x_i - \mu}{\sigma_i}, \quad (8.14)$$

with the fitted value  $x_i$  of the  $i$ -th pseudo experiment, the generated value  $\mu$  and the parameter’s statistic uncertainty  $\sigma_i$ . The pull distribution follows a normal distribution, if the fit model is unbiased.

The mass fit parameters agree well with normal distributions, and the deviation of the yield parameters from the generated values is below 10%. The combined pull distribution of the yield parameters is shown in Fig. 8.5. Most importantly, the  $\Delta m_s$



**Figure 8.5:** Pull distribution of combined yield parameters from 1500 pseudo experiments. The distributions are well described with a normal distribution, such that the extended likelihood fit is considered stable.

parameter pull distribution is well described with a normal distribution, hence the fit

model introduces no bias for the physical parameter of interest. Similarly to the mass fit, most other decay-time fit parameters are unbiased up to 10%, while the SS calibration asymmetry parameters  $\Delta p_{0/1}$  are shifted by up to 18%. As mentioned in Section 8.1, the acceptance coefficients  $v_1$  and  $v_2$  show a significant bias, which has no effect on  $\Delta m_s$ . The pull distributions of the decay-time fit are shown in Fig. 8.6, where the distributions of per-year parameters are combined.

In a previous set of toy experiments, the statistical uncertainties of the decay-time fits were corrected by a scaling factor of

$$s = \frac{\sum_i w_i^s}{\sum_i (w_i^s)^2}, \quad (8.15)$$

using the sWeights  $w_i^s$ . While this approach can be numerically evaluated efficiently, the uncertainty estimates are underestimated by up to 10%. Therefore, the asymptotically correct sWeight correction is used in the nominal fit procedure [118].

#### 8.4.2 sPlot Validation: Correlation among Observables

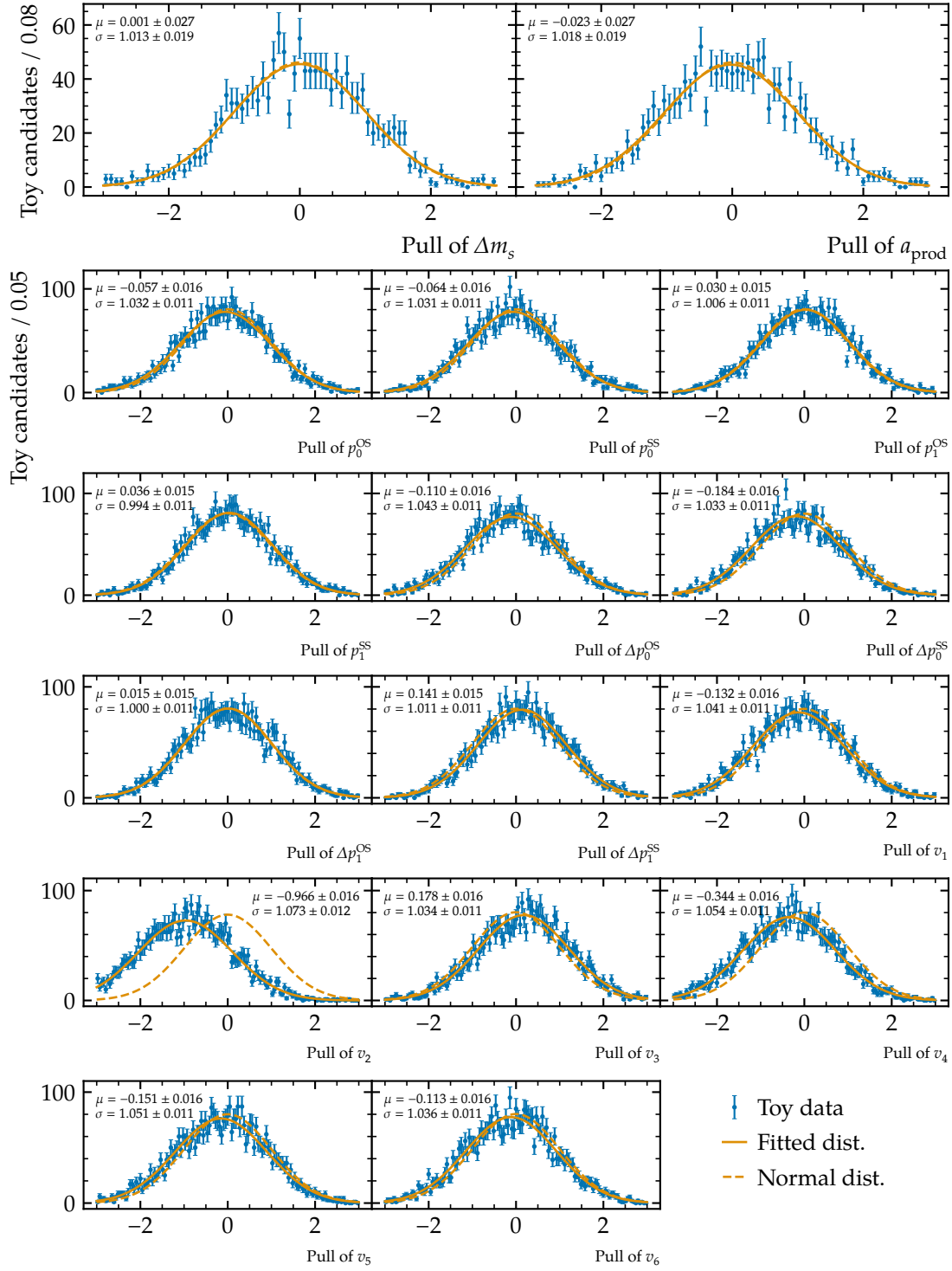
As explained in Section 5.2.1, the sPlot method assumes uncorrelated discriminating and test variables, which are the mass variables  $m_{B_s^0}$ ,  $m(D_s^-)$  and decay-time variable  $t$ , in this case. The data samples however show a correlation coefficient of up to  $\rho(m_{B_s^0}, t) = (12.5 \pm 0.1)\%$  in case of the  $D_s^- \rightarrow K^{*0}K^-$  data sub-sample. The effect of this on the measurement of  $\Delta m_s$  has been studied by my colleague Quentin Fühling, and is described in this section. A systematic uncertainty is assigned, as described in Section 9.2.

To verify that the observed correlation has no effect on the applicability of the sPlot method, two sets of bootstrapped data samples are generated. Fully simulated signal events are therefore combined with data points from the upper  $B_s^0$  invariant mass band,  $m_{B_s^0} \in [5600, 6800] \text{ MeV}/c^2$ , which is mapped into the nominal fit range,  $m_{B_s^0} \in [5300, 5800] \text{ MeV}/c^2$ . The peaking background contributions listed in Section 6.3 are added from simulated samples, to which the non-resonant  $K^-K^+\pi^-$  selection requirements are applied. As in the previously described toy samples, the admixture of dataset components resembles the nominal mass fit yields, drawn from a Poisson distribution.

In one of the bootstrapped data samples, the invariant  $B_s^0$  and  $D_s^\mp$  mass variables are shuffled, removing any correlation with the decay-time variable  $t$ . The decay-time likelihood fit is then repeated 300 times with each of the correlated and uncorrelated data samples. For each fit the difference  $\Delta_C = \Delta m_s^{\text{corr}} - \Delta m_s^{\text{uncorr}}$  of the correlated (superscript “corr”) and uncorrelated (superscript “uncorr”) fit results is calculated. The distribution of  $\Delta_C$  is checked to be compatible with zero and the remainder is assigned as a systematic uncertainty, as explained in Section 9.2.

In the course of this study, a significant bias of  $0.01 \text{ ps}^{-1}$  can be observed in the correlated samples as soon as the  $B_s^0 \rightarrow D_s^\mp K^\pm$  background component is described in the mass fit. This effect originates from the knowingly wrongly applied sPlot technique, but can be significantly reduced by reformulating the mass fit PDF Eq. (6.10) as described in Section 6.3.





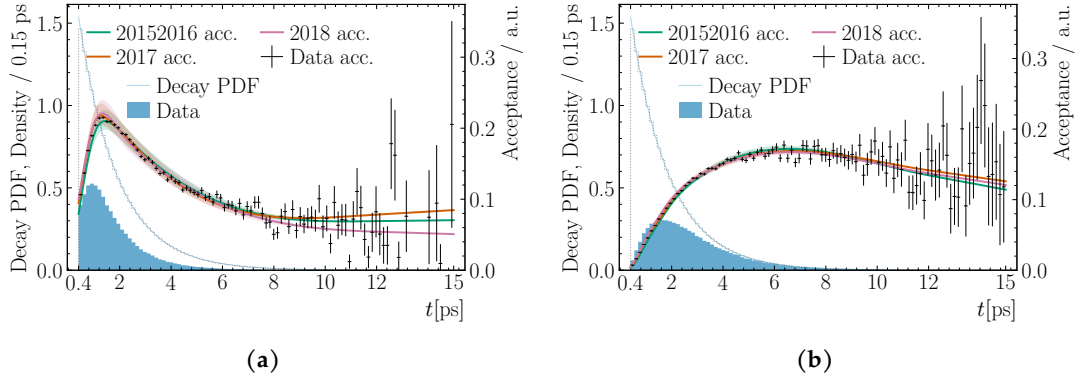
**Figure 8.6:** Pull distributions of the decay-time fit parameters, obtained from 1500 pseudo experiments. The distribution of per-data-sample parameters (see Table 8.4) are combined. No significant bias is seen for the parameter of interest,  $\Delta m_s$ .

### 8.4.3 Data Sample Splits

To test for possible effects of several analysis steps on the value of  $\Delta m_s$ , the mass and decay-time likelihood fits are repeated on different sub-samples of the data.

Different run conditions are tested by repeating the analysis within each year sub-sample. The final-state specific offline selection is tested by repeating the analysis in each  $D_s^\mp$  final state sample. The flavour tagging strategy is tested by performing decay-time fits using only OS or SS tagging information, and by splitting the data sample into low ( $p(B_s^0) < 130 \text{ GeV}/c^2$ ) and high ( $p(B_s^0) \geq 130 \text{ GeV}/c^2$ )  $B_s^0$  momentum regions. Finally, the BDT selection and acceptance modelling is verified by repeating the analysis in a low ( $0.475 < f_{\text{BDT}} < 0.94$ ) and high ( $f_{\text{BDT}} \geq 0.94$ ) BDT region.

Large deviations of the result for  $\Delta m_s$  within the different splits could hint towards significant systematic effects induced by the corresponding method. The results are listed in Table 8.6, and a systematic uncertainty is assigned as explained in Chapter 9. The effect of the different BDT requirements on the acceptance is shown in Fig. 8.7.



**Figure 8.7:** Decay-time dependent efficiency, obtained in the low BDT bin (a) and the high BDT bin (b).

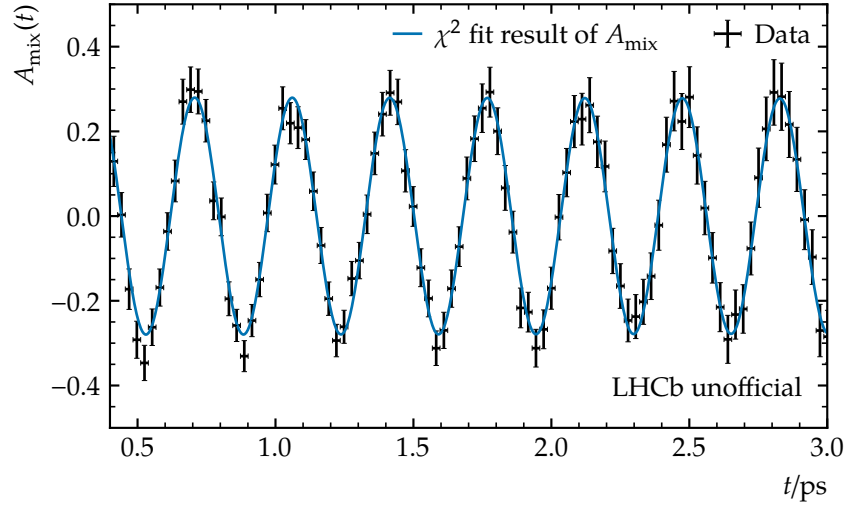
**Table 8.6:** Decay-time fit results for  $\Delta m_s$  from different data sub-samples. If more than two sub-samples are present, the weighted averages (avg.) of all sub-samples, leaving one out are also compared with the leftover result. For each split, the average of the individual samples' results is compared to the nominal fit result, assuming full correlation.

Split	Difference	Result / ps <sup>-1</sup>	
Polarity	Up - Down	-0.019	± 0.010
	avg{individual} - nominal	0.000 031	± 0.000 020
$D_s^\mp$ final state	PhiPi - KstK	-0.015	± 0.012
	PhiPi - NonRes	0.001	± 0.014
	PhiPi - PiPiPi	0.017	± 0.014
	KstK - NonRes	0.016	± 0.014
	KstK - PiPiPi	0.032	± 0.015
	NonRes - PiPiPi	0.016	± 0.016
	avg{KstK, NonRes, PiPiPi} - PhiPi	0.001	± 0.010
	avg{PhiPi, NonRes, PiPiPi} - KstK	-0.019	± 0.011
	avg{PhiPi, KstK, PiPiPi} - NonRes	0.003	± 0.012
	avg{PhiPi, KstK, NonRes} - PiPiPi	0.021	± 0.013
	avg{individual} - nominal	-0.000 152	± 0.000 006
Years	20152016 - 2017	-0.003	± 0.012
	20152016 - 2018	0.010	± 0.012
	2017 - 2018	0.013	± 0.012
	avg{2017, 2018} - 20152016	-0.004	± 0.010
	avg{20152016, 2018} - 2017	-0.008	± 0.011
	avg{20152016, 2017} - 2018	0.011	± 0.010
	avg{individual} - nominal	-0.000 022	± 0.000 004
	Tagging	0S-SS	0.020
avg{0S,SS} - nominal		0.000 24	± 0.000 07
$B_s^0$ momentum	BsP1 - BsP2	0.023	± 0.010
	avg{BsP1,BsP2} - nominal	-0.000 324	± 0.000 015
BDT splits	BDT <sub>Low</sub> - BDT <sub>High</sub>	0.005	± 0.011
	avg{individual} - nominal	0.000 339	± 0.000 033

### 8.5 Perspective of Indirect $CP$ Violation in $B_s^0 \rightarrow D_s^- \pi^+$

In the measurement presented so far, indirect  $CP$  violation has been neglected, such that the normalization factors of Eqs. (3.21) to (3.24) were identical. In this case,  $a = 0$  and the mixing asymmetry is well described with the leading term of Eq. (3.41).

To test for a possible sensitivity on indirect  $CP$  violation, an ad-hoc least-squares fit of the mixing asymmetry from Eq. (3.40) is performed. The fit result is shown in



**Figure 8.8:** Least squares fit of the mixing asymmetry Eq. (3.40) to  $B_s^0 \rightarrow D_s^- \pi^+$  data.

Fig. 8.8, and yields  $\Delta m_s^{A_{\text{mix}}} = (17.773 \pm 0.005) \text{ ps}^{-1}$ , which is close to the nominal fit result. Furthermore, the dilution factor which combines effects of flavour tagging and decay-time resolution, is obtained as  $D^{A_{\text{mix}}} = 0.283 \pm 0.009$ . No significance on indirect  $CP$  violation is found,  $a^{A_{\text{mix}}} = (1 \pm 3)\%$ , and the uncertainty is one order of magnitude above latest single measurements [49]. A large correlation of  $\rho(D, a) = 91\%$  between  $D$  and  $a$ , and potential differences of the sPlot method between mixed and unmixed data samples indicate that a comprehensive study of experimental effects are needed for a future measurement. Moreover, the usage of Eq. (3.27) could provide a better numerical sensitivity. No correlation between  $\Delta m_s$  and  $a$  is seen, justifying the negligence of indirect  $CP$  violation in Section 8.3.

In the perspective of significantly increased statistics in LHC Run 3 and beyond, a measurement of indirect  $CP$  violation could be feasible, using decays of  $B_s^0 \rightarrow D_s^- \pi^+$ .

## 9 Systematic Uncertainties

The measurement shown in this thesis represents one of the most precise LHCb measurement to date. While frequency measurements can generally obtain a high precision, the huge data sample that is presented here reveals several new challenges: As shown in the previous chapters, notably Sections 6.2.4, 6.3 and 8.2, effects like small discrepancies between data and simulation, an insufficient knowledge of the detector, or a simplification in the data model can add up to a significant discrepancy between underlying and measured physical parameters.

Systematic uncertainties are assigned for these effects, as described in more detail in the following sections. A summary of all non-negligible systematic uncertainties is given in Table 9.1.

**Table 9.1:** Non-negligible systematic uncertainties of the  $\Delta m_s$  measurement.

Systematic	$\sigma_{\text{syst}}^{\Delta m_s}/\text{ps}^{-1}$
Reconstruction effects:	
VELO alignment	0.0020
VELO z-scale	0.0018
Momentum scale uncertainty	0.0007
Multidimensional fit model:	
background parametrization	0.0002
$f_{\text{lowMass}}$	0.0005
Decay-time fit model:	
decay-time resolution model	0.0011
correlation among observables	0.0011
data sample splits	0.0003
Total systematic uncertainty	0.0032
Statistical uncertainty	0.0051

### 9.1 Reconstruction Effects

As described in Section 4.2.1, the tracking system allows to reconstruct particle momenta and decay vertices with high precision. This is challenged by the precision of the  $\Delta m_s$  measurement, which is why several reconstruction uncertainties contribute to the dominating systematic uncertainty of this measurement.

## 9 Systematic Uncertainties

The decay time  $t$  is obtained from the measured decay-length  $l$ , momentum  $p$  and mass  $m$  of the reconstructed particles via

$$t = \frac{m \cdot l}{p}. \quad (9.1)$$

A systematic offset of any of these parameters translates into a shift of the decay-time  $t$  which directly translates into a shift of the measured oscillation frequency  $\Delta m_s$ . Potential effects of this are therefore taken into account as described in the following.

### Momentum Scale Uncertainty

During Run 2, LHCb achieved a precision of  $3 \cdot 10^{-4}$  of the momentum scale [119]. The effect of this on  $\Delta m_s$  is studied by scaling the momenta in a sample of simulated events by  $\pm 3 \cdot 10^{-4}$  of their generated values. For each variation, a decay-time fit is performed and  $\Delta m_s$  is extracted. The difference of the measurements,  $\Delta(\Delta m_s) = 0.0014 \text{ ps}^{-1}$  is used to assign a systematic of  $\sigma_{p\text{-scale}} = 0.0007 \text{ ps}^{-1}$ .

### VELO $z$ Scale

The precision of the VELO module positions was evaluated during assembly, before first data taking [120, 121]. This is unchanged during Run 2 and the positions along the beam-axis ( $z$ -direction) is known with a relative precision of  $100 \mu\text{m}/\text{m} = 10^{-6}$ . The resulting uncertainty linearly translates to a systematic uncertainty of  $\Delta m_s$  of  $\sigma_{z\text{-scale}} = 0.0018 \text{ ps}^{-1}$ .

### VELO Alignment and Decay-time Bias Correction

An additional uncertainty originates from the VELO alignment procedure, for which a correction is described in Section 8.2.2. For each misalignment scenario (see Fig. 8.3), the difference of the corrected fit results for  $\Delta m_s$  and the perfectly aligned fit result is considered. The largest deviation is taken as a systematic uncertainty of the measurement, which yields the dominant systematic uncertainty of  $\sigma_t = 0.002 \text{ ps}^{-1}$ .

## 9.2 Analysis Methods

The methods used in this analysis can induce additional systematic errors to the measurement of  $\Delta m_s$ . Therefore, various aspects of their application, as described in Chapters 6 to 8, are studied to assign appropriate systematic uncertainties:

### Mass Fit Model

The influence of the mass fit model on the sPlot method and the value of  $\Delta m_s$  is tested by varying the fit parametrisation.

The data-to-simulation correction that is used to extract the signal and physical background shapes is changed to use an alternative set of weighting variables. For the signal

parametrisation this results in a worse fit quality, while the difference for  $\Delta m_s$  is negligible. The resulting physical background shapes induce a difference of below  $2 \text{ ns}^{-1}$ , which is assigned as a systematic uncertainty.

To test the combinatorial background parametrisation, two adjusted PDFs are compared: The second exponential parameter,  $c_{B_s^0}^{(2)} m_{D_s^\mp \pi^\pm}$ , is floating and a double exponential function is used to describe the  $m(h^- h^+ h^-)$  invariant mass distribution. The effect on  $\Delta m_s$  is negligible in both cases.

Since the low mass background fraction is arbitrarily fixed to  $f_{\text{lowMass}} = 0.5$ , the mass fits are repeated using values of  $f_{\text{lowMass}} = 0$  and  $f_{\text{lowMass}} = 1$ . The largest difference of the decay-time fit result to the nominal fit result is found to be below  $5 \text{ ns}^{-1}$  and assigned as a systematic uncertainty.

Finally the fixed yields of the physical background components are varied between  $\pm 50\%$  of their nominal value, which has a negligible effect on the measured value of  $\Delta m_s$ . No additional systematic uncertainty is assigned.

### Correlation of discriminating sPlot variables

In the studies shown in Section 8.4.2, a small deviation of the measured value of  $\Delta m_s$  between correlated and uncorrelated data samples of  $\Delta_C = (-0.00112 \pm 0.00008) \text{ ps}^{-1}$  is measured. The central value and uncertainty of this deviation is added in quadrature and assigned as a systematic uncertainty of the sPlot procedure.

### Decay-time Fit Model

The pseudo experiment studies (Section 8.4.1) are able to uncover any possible systematic error induced by the decay-time fit model. While some acceptance parameters are significantly biased, this has no effect on  $\Delta m_s$ , which is reproduced unbiased as shown in Fig. 8.6 and no systematic uncertainty is assigned.

### Decay-time Acceptance Parametrisation

Two alternative decay-time acceptances are studied, one of which is using the double number of knots,  $k_i \in [0.45, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 4.0, 6.0, 10, 14.9] \text{ ps}$ . The other parametrisation uses knot positions at  $k_i \in [0.4, 0.41, 1.0, 1.5, 2.0, 6.0, 14.9] \text{ ps}$  and fixed knot coefficients,  $v_6 = v_7 = 1$ . Both parametrisation yield a negligible difference of the fitted value for  $\Delta m_s$ .

Additionally, the fit is repeated with the coefficients  $v_1, v_2$ , and  $v_3$  fixed to  $\pm 10\%$  of their nominally fitted values. All resulting fits yield negligible differences for the fitted value of  $\Delta m_s$ , with respect to the nominal fit result and no systematic uncertainty is assigned.

### Flavour Tagging

Since all flavour tagging parameters are unconstrained in the decay-time fit, and the mistag distribution is well described with a linear function (see Fig. 7.8), a effect is

## 9 Systematic Uncertainties

expected to be small. To test this, the linear calibration function is changed for a linear function with a logistic link function that is possibly improving the fit convergence around  $\omega = 0$  and  $\omega = 0.5$ .

Additionally, the decay-time fit is repeated, while the OS precalibration (see Section 7.1.1) is completely neglected. Instead a naive combination of the OS tagging information with Eq. (7.15) is used in the fit.

In both scenarios the resulting differences of  $\Delta m_s$  with respect to the nominal fit result is found negligible and no systematic uncertainty is assigned.

### Correlation of fixed Parameters $\Gamma_s$ and $\Delta\Gamma_s$

While the correlation between  $\Gamma_s$  and  $\Delta\Gamma_s$  is known to be  $\rho(\Gamma_s, \Delta\Gamma_s) = -12.4\%$  [22], the decay-time fit assumes uncorrelated parameters. The effect of this on the value of  $\Delta m_s$  is tested by rerunning 1000 toy experiments as described in Section 8.4.1, but with the values of  $\Gamma_s$  and  $\Delta\Gamma_s$  drawn from a correlated Gaussian distribution. For each random seed, the fit result is compared to the result of the nominal pseudo experiment. The mean and widths of these differences are then added in quadrature to evaluate a systematic error, which is found to be negligible.

### Decay-time Resolution Model

A possible systematic error due to the decay-time resolution model is evaluated by repeating the decay-time fit with two alternative resolution calibration models. One alternative includes also the wide Gauss components,  $\sigma_2$  of the decay-time fits on prompt  $D_s^\mp \pi^\pm$  samples, which are described in Section 8.2.1. The second alternative uses only the narrow Gaussian component, with width  $\sigma_0$ . Half of the difference of the fitted values of  $\Delta m_s$  for both alternatives is used to assign a systematic uncertainty of  $\sigma_{t\text{-res.}} = 0.0011 \text{ ps}^{-1}$ .

### Data Sample Splits

For each data sample split that is described in Section 8.4.3, the weighted average of  $\Delta m_s$  from the split samples is compared with the nominal fit result. The resulting deviations are listed in Table 8.6, where full correlation is assumed between both values. The largest deviation is taken as a systematic uncertainty,  $\sigma_{\text{split}} = 0.0003 \text{ ps}^{-1}$ .



## 10 Conclusions & Outlook

The first two data taking periods of the Large Hadron Collider have produced a variety of measurements that continue the success story of the Standard Model of particle physics. In addition to the discovery of the Higgs particle, new milestones have been reached, including the discovery of pentaquarks and the observation of  $CP$  violation in the charm system.

These continued successes, however, increase the pressure of the still open questions, such as the nature of dark matter or the process of baryogenesis. First hints of physics beyond the Standard Model might guide the way towards an updated theory [17, 58], allowing hopes for exciting times ahead for particle physics. Regardless of these exciting results, fundamentally new developments of the theory might only be possible through even more accurate knowledge of as many SM parameters as possible. This will allow for global parameter fits of the SM and effective theories, which will shine light into the last remaining hideouts for New Physics.

With these efforts, the field of particle physics enters an era of high precision. The work described in this thesis represents an early example of this, as one of the most precise measurements performed by the LHCb experiment so far.

The full LHCb Run 2 data sample, corresponding to an integrated luminosity of  $\mathcal{L}_{\text{int}} = 6 \text{ fb}^{-1}$  has been used to extract a sample of 378 700  $B_s^0 \rightarrow D_s^- \pi^+$  decays. The sample selection includes a powerful BDT to suppress combinatorial background, and an extensive study of additional potential background processes has been performed. The flavour tagging information from several tagging algorithms has been calibrated and used to extract the oscillation frequency  $\Delta m_s$  from the decay-time distribution of these particles. Several additional detector effects have been taken into account and respective systematic uncertainties were assigned to obtain the most precise single measurement of

$$\Delta m_s = (17.7683 \pm 0.0051 \pm 0.0032) \text{ ps}^{-1}.$$

If combined with other LHCb measurement, this yields the most precise single experiment measurement of  $\Delta m_s^{\text{LHCb}} = (17.7656 \pm 0.0057) \text{ ps}^{-1}$  [18].

Following Ref. [56], using the current world averages for the  $B$  meson masses [22], and assuming symmetric uncertainties for all parameters, this result enables an estimate of the ratio of CKM matrix elements,  $|V_{td}/V_{ts}| = 0.2043 \pm 0.0011$ , which is an important ingredient to the CKM picture.

Considering the increased statistics to be collected during LHC Run 3 [122], and assuming similar selection efficiencies as in this measurement, the statistic uncertainty of  $\Delta m_s$  will decrease to  $\sim 1/3$  of the current value. The Vertex Locator, which gives rise to the leading systematic uncertainty, receives a major upgrade before LHCb Run 3. To

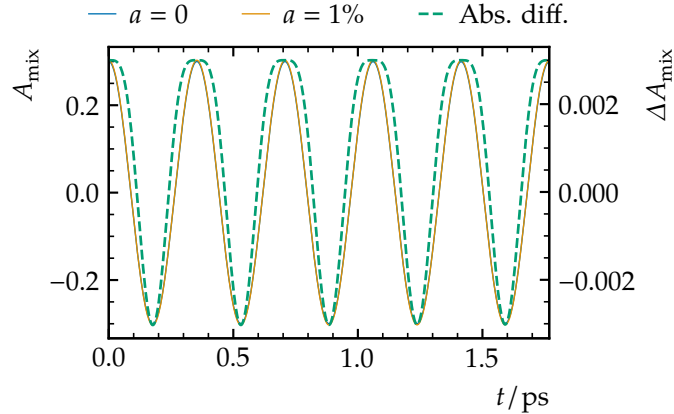
prevent being a limiting factor for a future  $\Delta m_s$  measurement, its alignment and scale uncertainty would need to improve significantly, which will be challenging.

Within the broader experimental landscape, the Belle II experiment plans to collect an integrated luminosity of  $\mathcal{L}_{\text{int}} \approx 5 \text{ ab}^{-1}$  at the  $\Upsilon(5S)$  resonance, which decays to  $B_s^0$  mesons around 20% of the time [59, 123]. While the tagging power of the  $b$ -factory is significantly higher when compared to the LHCb experiment, a potential Belle II measurement of tagged, flavour specific  $B_s^0$  decays is not expected to be competitive.

Therefore, the measurement presented in this thesis might turn out to be a long-lasting reference for future detector and software developments. It will furthermore be an essential external parameter for different time-dependent measurements of  $CP$  violation using  $B_s^0$  mesons.

Moreover, it is an important ingredient for theory models, even though all uncertainties are currently dominated by QCD factors or the relative uncertainty of  $\Delta m_d$ . The uncertainties of the current best theory estimation of  $\Delta m_s^{\text{th.}} = 18.4_{-1.2}^{+0.7} \text{ ps}^{-1}$  is expected to improve to  $\pm 0.5 \text{ ps}^{-1}$  by 2025, which would indicate a tension of  $\sim 1.2\sigma$ , if central values stay unchanged. With this, constraints on the scales of different New Physics scenarios can be determined, which might explain currently discussed lepton flavour anomalies. These constraints include upper bounds on a leptoquark mass,  $m_{\text{LQ}} \lesssim 30 \text{ TeV}$ , and a  $Z'$  mass,  $m_{Z'} \lesssim \sim 9 \text{ TeV}$  [56], which would be in reach of future collider experiments [124].

A potential measurements of indirect  $CP$  violation with decays of  $B_s^0 \rightarrow D_s^- \pi^+$  could furthermore complement semi-leptonic measurements of  $CP$  violation, and potentially uncover additional effects of New Physics. To illustrate this, a projection of the mixing asymmetry Eq. (3.40) for an assumed indirect  $CP$  violation, corresponding to  $a = 1\%$ , is shown in Fig. 10.1.



**Figure 10.1:** Mixing asymmetry (left axis), assuming a dilution  $D = 0.3$ , the new value of  $\Delta m_s$ , and no indirect  $CP$  violation,  $a = 0$  (blue line). A “New Physics” scenario with  $a = 1\%$  (orange line) is also shown. The difference of both is at the permille-level and shown as a green, dashed line on the right axis.

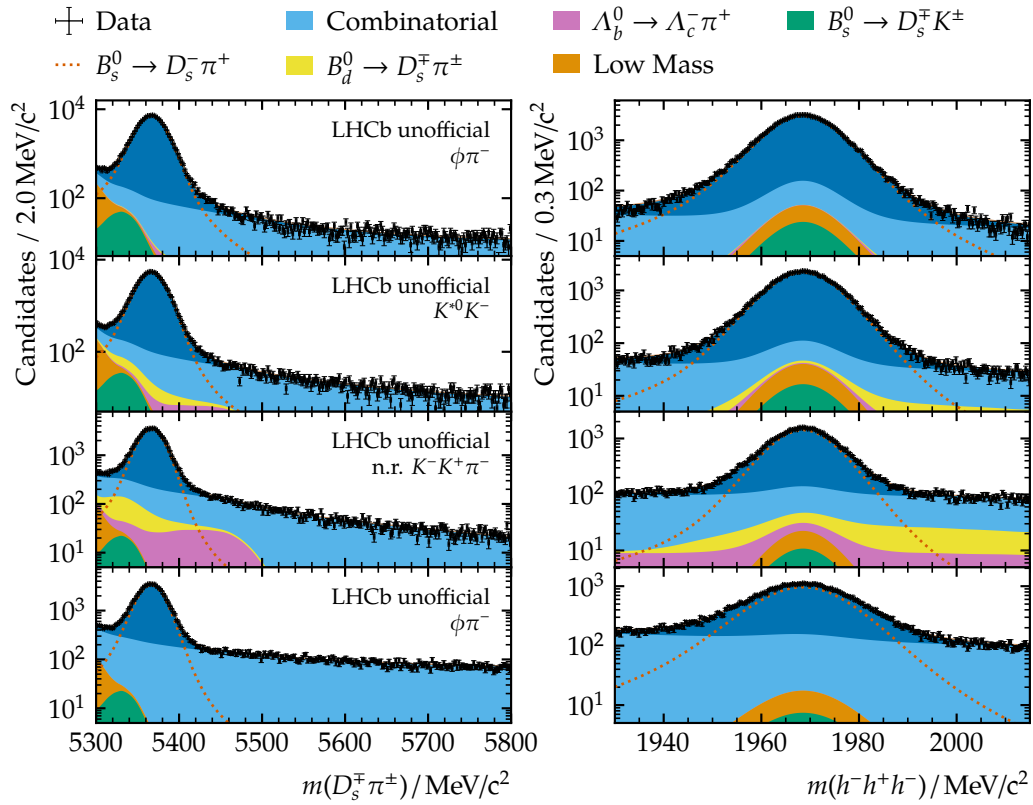
Ultimately, decays of  $B_s^0 \rightarrow D_s^- \pi^+$  not only allowed us to measure  $\Delta m_s$  and contribute to the evolution of the SM, but are a wonderful example of the quantum nature of strange oscillating beauty-mesons.

# Appendix

## Fixed Parameters of the Multidimensional Mass Fit

An extensive description of the fit model, used in Section 6.3, is given in [100]. For reference, all fixed parameters, needed to evaluate the PDF Eq. (6.9), are listed in Tables A.1 to A.3. The yield parameters of background components that are less than 2% of the signal yield are fixed from known branching fractions and efficiencies, which are measured on simulated events.

The fit results are shown separately for each  $D_s^\mp$  final state in Fig. A.1.



**Figure A.1:** Result of the simultaneous fit of Eq. (6.9) to the invariant  $m_{B_0}$  and  $m(D_s^-)$  mass distributions, separately for each individual  $D_s^\mp$  final state sample. From top to bottom:  $D_s^- \rightarrow \phi\pi^-$ ,  $D_s^- \rightarrow K^{*0}K^-$ ,  $D_s^- \rightarrow \text{n.r. } K^-K^+\pi^-$ , and  $D_s^- \rightarrow \pi^-\pi^+\pi^-$ .

**Table A.1:** Parameters of the double-sided Hypatia, describing the invariant  $m_{B_s^0}$  mass, obtained from fits to simulated data samples.

Parameter	year	$D_s^- \rightarrow \phi\pi^-$	$D_s^- \rightarrow K^{*0}K^-$	$D_s^- \rightarrow \text{n.r. } K^-K^+\pi^-$	$D_s^- \rightarrow \pi^-\pi^+\pi^-$
$\mu_{D_s^\mp} / \text{MeV}/c^2$	2015–2016	$5367.72 \pm 0.04$	$5367.77 \pm 0.05$	$5367.73 \pm 0.04$	$5367.68 \pm 0.11$
	2017	$5367.95 \pm 0.05$	$5367.82 \pm 0.05$	$5367.82 \pm 0.06$	$5367.77 \pm 0.24$
	2018	$5367.84 \pm 0.05$	$5367.73 \pm 0.04$	$5367.81 \pm 0.04$	$5467.30 \pm 0.08$
$f_{D_s^\mp}$	2015–2016	$0.124 \pm 0.011$	$0.121 \pm 0.013$	$0.115 \pm 0.010$	$0.199 \pm 0.029$
	2017	$0.162 \pm 0.015$	$0.126 \pm 0.014$	$0.142 \pm 0.017$	$0.247 \pm 0.077$
	2018	$0.141 \pm 0.012$	$0.126 \pm 0.012$	$0.163 \pm 0.012$	$0.142 \pm 0.036$
$\sigma_{D_s^H}^H / \text{MeV}/c^2$	2015–2016	$39.69 \pm 2.46$	$39.23 \pm 3.01$	$37.49 \pm 2.49$	$33.32 \pm 2.50$
	2017	$39.26 \pm 2.52$	$35.71 \pm 3.22$	$33.82 \pm 2.68$	$30.81 \pm 5.80$
	2018	$39.89 \pm 4.34$	$34.72 \pm 1.94$	$39.97 \pm 1.26$	$28.69 \pm 2.62$
$\tau_{D_s^\mp}$	2015–2016	$0.360 \pm 0.010$	$0.343 \pm 0.012$	$0.347 \pm 0.011$	$0.353 \pm 0.021$
	2017	$0.328 \pm 0.013$	$0.345 \pm 0.013$	$0.319 \pm 0.017$	$0.311 \pm 0.062$
	2018	$0.346 \pm 0.011$	$0.338 \pm 0.012$	$0.294 \pm 0.012$	$0.338 \pm 0.035$
$\nu_{D_s^\mp}$	2015–2016	$-0.203 \pm 0.022$	$-0.141 \pm 0.029$	$-0.192 \pm 0.026$	$-0.048 \pm 0.062$
	2017	$-0.125 \pm 0.033$	$-0.169 \pm 0.031$	$-0.162 \pm 0.045$	$-0.035 \pm 0.022$
	2018	$-0.160 \pm 0.035$	$-0.179 \pm 0.026$	$-0.211 \pm 0.037$	$-0.285 \pm 0.081$
$\sigma_{D_s^J}^J / \text{MeV}/c^2$	2015–2016	$14.39 \pm 0.096$	$14.22 \pm 0.102$	$14.19 \pm 0.086$	$14.56 \pm 0.157$
	2017	$13.82 \pm 0.096$	$13.88 \pm 0.107$	$13.63 \pm 0.129$	$13.92 \pm 0.338$
	2018	$13.99 \pm 0.088$	$13.85 \pm 0.096$	$13.49 \pm 0.081$	$14.23 \pm 0.340$
$a_{D_s^\mp}^{(1)}$	2015–2016	$0.273 \pm 0.033$	$0.233 \pm 0.033$	$0.266 \pm 0.033$	$0.282 \pm 0.036$
	2017	$0.266 \pm 0.033$	$0.262 \pm 0.047$	$0.315 \pm 0.054$	$0.271 \pm 0.083$
	2018	$0.239 \pm 0.049$	$0.307 \pm 0.036$	$0.381 \pm 0.035$	$0.538 \pm 0.148$
$a_{D_s^\mp}^{(2)}$	2015–2016	$2.643 \pm 0.142$	$2.772 \pm 0.196$	$1.952 \pm 0.168$	$3.022 \pm 0.431$
	2017	$2.960 \pm 0.188$	$2.725 \pm 0.186$	$2.762 \pm 0.218$	$3.484 \pm 0.600$
	2018	$2.795 \pm 0.168$	$2.712 \pm 0.154$	$2.407 \pm 0.159$	$2.399 \pm 0.244$
$\lambda_{D_s^\mp}$	2015–2016	$-1.476 \pm 0.113$	$-1.403 \pm 0.105$	$-1.542 \pm 0.143$	$-1.390 \pm 0.089$
	2017	$-1.367 \pm 0.083$	$-1.463 \pm 0.153$	$-1.594 \pm 0.196$	$-1.339 \pm 0.168$
	2018	$-1.337 \pm 0.133$	$-1.615 \pm 0.138$	$-1.769 \pm 0.148$	$-3.319 \pm 1.772$
$n_{D_s^\mp}^{(2)}$	2015–2016	$1.070 \pm 0.183$	$0.974 \pm 0.207$	$2.029 \pm 0.261$	$1.158 \pm 0.464$
	2017	$0.966 \pm 0.199$	$0.999 \pm 0.218$	$1.278 \pm 0.267$	$0.980 \pm 0.426$
	2018	$0.986 \pm 0.250$	$1.175 \pm 0.185$	$1.687 \pm 0.207$	$1.517 \pm 0.278$
$\beta_{D_s^\mp}$	All			0.0 (fixed)	
$\zeta_{D_s^\mp}$	All			0.0 (fixed)	
$n_{D_s^\mp}^{(1)}$	All			50.0 (fixed)	

**Table A.2:** Parameters of the double-sided Hypatia, describing the invariant  $m(D_s^-)$  mass, obtained from fits to simulated data samples.

Parameter	year	$D_s^- \rightarrow \phi\pi^-$	$D_s^- \rightarrow K^{*0}K^-$	$D_s^- \rightarrow \text{n.r. } K^-K^+\pi^-$	$D_s^- \rightarrow \pi^-\pi^+\pi^-$
$\mu_{D_s^-} / \text{MeV}/c^2$	2015–2016	$1968.75 \pm 0.01$	$1968.84 \pm 0.01$	$1968.75 \pm 0.01$	$1969.13 \pm 0.03$
	2017	$1968.75 \pm 0.01$	$1968.84 \pm 0.01$	$1968.77 \pm 0.01$	$1969.11 \pm 0.03$
	2018	$1968.75 \pm 0.01$	$1968.84 \pm 0.01$	$1968.76 \pm 0.01$	$1969.11 \pm 0.03$
$f_{D_s^-}$	2015–2016	$0.129 \pm 0.010$	$0.087 \pm 0.010$	$0.086 \pm 0.010$	$0.313 \pm 0.048$
	2017	$0.136 \pm 0.010$	$0.068 \pm 0.008$	$0.080 \pm 0.008$	$0.272 \pm 0.053$
	2018	$0.121 \pm 0.009$	$0.069 \pm 0.006$	$0.095 \pm 0.011$	$0.308 \pm 0.048$
$\sigma_{D_s^-}^H / \text{MeV}/c^2$	2015–2016	$26.73 \pm 0.92$	$26.27 \pm 1.35$	$29.92 \pm 1.77$	$30.48 \pm 1.16$
	2017	$27.20 \pm 0.74$	$26.86 \pm 1.89$	$26.66 \pm 2.32$	$30.35 \pm 1.51$
	2018	$24.31 \pm 1.18$	$23.85 \pm 2.19$	$27.78 \pm 1.50$	$29.30 \pm 1.15$
$\tau_{D_s^-}$	2015–2016	$0.485 \pm 0.007$	$0.446 \pm 0.008$	$0.462 \pm 0.010$	$0.380 \pm 0.041$
	2017	$0.468 \pm 0.008$	$0.454 \pm 0.008$	$0.454 \pm 0.009$	$0.386 \pm 0.041$
	2018	$0.490 \pm 0.007$	$0.448 \pm 0.006$	$0.465 \pm 0.009$	$0.388 \pm 0.041$
$\nu_{D_s^-}$	All	0.0 (fixed)			
$\sigma_{D_s^-}^J / \text{MeV}/c^2$	2015–2016	$6.420 \pm 0.037$	$6.551 \pm 0.034$	$6.436 \pm 0.047$	$9.206 \pm 0.173$
	2017	$6.218 \pm 0.037$	$6.496 \pm 0.033$	$6.322 \pm 0.039$	$9.039 \pm 0.171$
	2018	$6.338 \pm 0.031$	$6.494 \pm 0.022$	$6.313 \pm 0.043$	$9.083 \pm 0.167$
$a_{D_s^-}^{(1)}$	2015–2016	$0.190 \pm 0.010$	$0.173 \pm 0.013$	$0.161 \pm 0.015$	$0.183 \pm 0.015$
	2017	$0.203 \pm 0.007$	$0.142 \pm 0.013$	$0.127 \pm 0.013$	$0.173 \pm 0.018$
	2018	$0.165 \pm 0.009$	$0.136 \pm 0.013$	$0.162 \pm 0.014$	$0.184 \pm 0.014$
$a_{D_s^-}^{(2)}$	2015–2016	$0.375 \pm 0.040$	$0.379 \pm 0.057$	$0.395 \pm 0.097$	$0.844 \pm 0.230$
	2017	$0.494 \pm 0.079$	$0.310 \pm 0.048$	$0.261 \pm 0.041$	$0.805 \pm 0.216$
	2018	$0.298 \pm 0.028$	$0.272 \pm 0.043$	$0.378 \pm 0.071$	$0.888 \pm 0.199$
$\beta_{D_s^-}$	All	0.0 (fixed)			
$\zeta_{D_s^-}$	All	0.0 (fixed)			
$\lambda_{D_s^-}$	All	-1.1 (fixed)			
$n_{D_s^-}^{(1)}$	All	8.0 (fixed)			
$n_{D_s^-}^{(2)}$	All	8.0 (fixed)			

**Table A.3:** Parameters of the double exponential function, Eq. (6.6), describing the combinatorial background invariant mass  $m_{B_s^0}$ . The parameters are obtained from a fit to the  $B_s^0$  invariant mass sideband,  $m_{B_s^0} \in [5600, 5800] \text{ MeV}/c^2$ . The parameters marked with “\*” are floating in the nominal fit.

Parameter	$D_s^- \rightarrow \phi\pi^-$	$D_s^- \rightarrow K^{*0}K^-$	$D_s^- \rightarrow \text{n.r. } K^-K^+\pi^-$	$D_s^- \rightarrow \pi^-\pi^+\pi^-$
$c_{B_s^0}^{(1)} \cdot \text{GeV}/c^2$	$-0.26 \pm 0.23$	$-0.65 \pm 0.13$	$-0.55 \pm 0.10$	$-0.47 \pm 0.08$
$c_{B_s^0}^{(2)} \cdot \text{GeV}/c^2$	$-4.79 \pm 2.52$	$-7.26 \pm 2.32$	$-6.83 \pm 1.11$	$-4.93 \pm 1.35$ *
$f_{B_s^0}^C$	$0.86 \pm 0.10$	$0.87 \pm 0.05$	$0.82 \pm 0.03$	$0.91 \pm 0.04$ *

**Table A.4:** Fixed yield parameters of the multidimensional fit PDF Eq. (6.9).

Background	Year	$D_s^- \rightarrow \phi\pi^-$	$D_s^- \rightarrow K^{*0}K^-$	$D_s^- \rightarrow \text{n.r. } K^-K^+\pi^-$	$D_s^- \rightarrow \pi^-\pi^+\pi^-$
$B_d^0 \rightarrow D^-\pi^+$	2015	3.41	56.45	192.0	-
	2016	15.7	237.1	818.5	-
	2017	22.1	265.2	931.5	-
	2018	21.6	319.4	1122.7	-
$\Lambda_b^0 \rightarrow \Lambda_c^-\pi^+$	2015	6.22	26.5	116.5	-
	2016	62.3	142.3	664.8	-
	2017	52.0	133.1	650.2	-
	2018	74.4	159.3	765.8	-
$B_s^0 \rightarrow D_s^\mp K^\pm$	2015	65.9	45.7	29.5	30.1
	2016	339.4	235.6	151.5	155.1
	2017	323.1	227.0	145.1	148.9
	2018	391.4	277.3	173.5	176.8

# Bibliography

- [1] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. [DOI: 10.1103/PhysRev.105.1413](https://doi.org/10.1103/PhysRev.105.1413).
- [2] J. H. Christenson et al. “Evidence for the  $2\pi$  Decay of the  $K_2^0$  Meson”. *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. [DOI: 10.1103/PhysRevLett.13.138](https://doi.org/10.1103/PhysRevLett.13.138).
- [3] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. [DOI: 10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531).
- [4] Makoto Kobayashi and Toshihide Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. [DOI: 10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [5] S. W. Herb et al. “Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions”. *Phys. Rev. Lett.* 39 (5 Aug. 1977), pp. 252–255. [DOI: 10.1103/PhysRevLett.39.252](https://doi.org/10.1103/PhysRevLett.39.252).
- [6] The D0 collaboration. “Search for high mass top quark production in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV”. *Phys. Rev. Lett.* 74 (1995), pp. 2422–2426. [DOI: 10.1103/PhysRevLett.74.2422](https://doi.org/10.1103/PhysRevLett.74.2422). [arXiv: hep-ex/9411001](https://arxiv.org/abs/hep-ex/9411001).
- [7] The CDF collaboration. “Observation of top quark production in  $p\bar{p}$  collisions”. *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. [DOI: 10.1103/PhysRevLett.74.2626](https://doi.org/10.1103/PhysRevLett.74.2626). [arXiv: hep-ex/9503002](https://arxiv.org/abs/hep-ex/9503002).
- [8] F.J. Hasert et al. “Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment”. *Physics Letters B* 46.1 (1973), pp. 138–140. ISSN: 0370-2693. [DOI: 10.1016/0370-2693\(73\)90499-1](https://doi.org/10.1016/0370-2693(73)90499-1).
- [9] The UA1 collaboration. “Experimental observation of isolated large transverse energy electrons with associated missing energy at  $s = 540$  GeV”. *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. [DOI: 10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2).
- [10] The UA2 collaboration. “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN  $pp$  collider”. *Physics Letters B* 122.5 (1983), pp. 476–485. ISSN: 0370-2693. [DOI: 10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2).
- [11] The UA1 collaboration. “Experimental observation of lepton pairs of invariant mass around  $95 \text{ GeV}/c^2$  at the CERN SPS collider”. *Physics Letters B* 126.5 (1983), pp. 398–410. ISSN: 0370-2693. [DOI: 10.1016/0370-2693\(83\)90188-0](https://doi.org/10.1016/0370-2693(83)90188-0).
- [12] The UA2 collaboration. “Evidence for  $Z^0 \rightarrow e^+e^-$  at the CERN  $pp$  collider”. *Physics Letters B* 129.1 (1983), pp. 130–140. ISSN: 0370-2693. [DOI: 10.1016/0370-2693\(83\)90744-X](https://doi.org/10.1016/0370-2693(83)90744-X).

## Bibliography

- [13] The DONUT collaboration. “Observation of tau neutrino interactions”. *Phys. Lett. B* 504 (2001), pp. 218–224. [doi: 10.1016/S0370-2693\(01\)00307-0](https://doi.org/10.1016/S0370-2693(01)00307-0). [arXiv: hep-ex/0012035](https://arxiv.org/abs/hep-ex/0012035).
- [14] The ATLAS collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. *Phys. Lett. B* 716 (2012), pp. 1–29. [doi: 10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). [arXiv: 1207.7214](https://arxiv.org/abs/1207.7214).
- [15] The CMS collaboration. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. *Phys. Lett. B* 716 (2012), pp. 30–61. [doi: 10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). [arXiv: 1207.7235](https://arxiv.org/abs/1207.7235).
- [16] N. Aghanim et al. “Planck 2018 results”. *Astronomy & Astrophysics* 641 (Sept. 2020), A6. ISSN: 1432-0746. [doi: 10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910).
- [17] The LHCb collaboration. “Test of lepton universality in beauty-quark decays”. *Submitted to Nature Physics* (Mar. 2021). [arXiv: 2103.11769](https://arxiv.org/abs/2103.11769).
- [18] The LHCb collaboration. “Precise Determination of the  $B_s^0$ - $\bar{B}_s^0$  oscillation frequency”. *Submitted to Nature Physics* (Apr. 2021). [arXiv: 2104.04421](https://arxiv.org/abs/2104.04421).
- [19] Robin Eichhorn. “Sensitivity study of the oscillation frequency  $\Delta m_s$  in  $B_s^0 \rightarrow D_s^- \pi^+$  with Run 2 data of the LHCb experiment”. Master’s Thesis. Technische Universität Dortmund, 2018.
- [20] Julian Jung. “Optimierung eines multivarianten Klassifizierers für Messungen von  $CP$ -Verletzung mit dem LHCb-Experiment”. Bachelor’s Thesis. Technische Universität Dortmund, 2019.
- [21] David Griffiths. *Introduction to elementary particles*. 2008. ISBN: 978-3-527-40601-2.
- [22] The Particle Data Group collaboration. “Review of Particle Physics”. *PTEP* 2020.8 (2020), p. 083C01. [doi: 10.1093/ptep/ptaa104](https://doi.org/10.1093/ptep/ptaa104).
- [23] Hausbrandt Trieste 1892 S.p.A. Adapted from image licensed via CC BY-SA 3.0. 2011. URL: [https://en.wikipedia.org/wiki/File:Hausbrandt\\_Espresso.JPG](https://en.wikipedia.org/wiki/File:Hausbrandt_Espresso.JPG) (visited on 05/24/2021).
- [24] The LHCb collaboration. “Search for violations of Lorentz invariance and  $CPT$  symmetry in  $B_{(s)}^0$  mixing”. *Phys. Rev. Lett.* 116.24 (2016), p. 241601. [doi: 10.1103/PhysRevLett.116.241601](https://doi.org/10.1103/PhysRevLett.116.241601). [arXiv: 1603.04804](https://arxiv.org/abs/1603.04804).
- [25] J. M. Link et al. “Charm System Tests of  $CPT$  and Lorentz Invariance with FOCUS”. *Phys. Lett. B* 556 (2003), pp. 7–13. [doi: 10.1016/S0370-2693\(03\)00103-5](https://doi.org/10.1016/S0370-2693(03)00103-5). [arXiv: hep-ex/0208034](https://arxiv.org/abs/hep-ex/0208034).
- [26] The KTeV collaboration. “Precise Measurements of Direct  $CP$  Violation,  $CPT$  Symmetry, and Other Parameters in the Neutral Kaon System”. *Phys. Rev. D* 83 (2011), p. 092001. [doi: 10.1103/PhysRevD.83.092001](https://doi.org/10.1103/PhysRevD.83.092001). [arXiv: 1011.0127](https://arxiv.org/abs/1011.0127).
- [27] The CMS collaboration. “Measurement of the mass difference between top and antitop quarks”. *JHEP* 06 (2012), p. 109. [doi: 10.1007/JHEP06\(2012\)109](https://doi.org/10.1007/JHEP06(2012)109). [arXiv: 1204.2807](https://arxiv.org/abs/1204.2807).



- [28] J. H. Christenson et al. “Evidence for the  $2\pi$  Decay of the  $K_2^0$  Meson”. *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. [DOI: 10.1103/PhysRevLett.13.138](https://doi.org/10.1103/PhysRevLett.13.138).
- [29] The BABAR collaboration. “Observation of  $CP$  Violation in the  $B^0$  Meson System”. *Phys. Rev. Lett.* 87 (9 Aug. 2001), p. 091801. [DOI: 10.1103/PhysRevLett.87.091801](https://doi.org/10.1103/PhysRevLett.87.091801).
- [30] The Belle collaboration. “Observation of Large  $CP$  Violation in the Neutral  $B$  Meson System”. *Phys. Rev. Lett.* 87 (9 Aug. 2001), p. 091802. [DOI: 10.1103/PhysRevLett.87.091802](https://doi.org/10.1103/PhysRevLett.87.091802).
- [31] The LHCb collaboration. “Observation of  $CP$  Violation in Charm Decays”. *Phys. Rev. Lett.* 122.21 (2019), p. 211803. [DOI: 10.1103/PhysRevLett.122.211803](https://doi.org/10.1103/PhysRevLett.122.211803). [arXiv: 1903.08726](https://arxiv.org/abs/1903.08726).
- [32] Martinus Veltman. *Diagrammatica: The Path to Feynman Diagrams*. Cambridge Lecture Notes in Physics. Cambridge University Press, 1994. [DOI: 10.1017/CB09780511564079](https://doi.org/10.1017/CB09780511564079).
- [33] W. N. Cottingham and D. A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. 2nd ed. Cambridge University Press, 2007. [DOI: 10.1017/CB09780511791406](https://doi.org/10.1017/CB09780511791406).
- [34] Michael Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Boca Raton, FL: CRC Press, 2018. ISBN: 9780813350196.
- [35] R. P. Feynman. “Space-Time Approach to Quantum Electrodynamics”. *Phys. Rev.* 76 (6 Sept. 1949), pp. 769–789. [DOI: 10.1103/PhysRev.76.769](https://doi.org/10.1103/PhysRev.76.769).
- [36] Lincoln Wolfenstein. “Parametrization of the Kobayashi-Maskawa Matrix”. *Phys. Rev. Lett.* 51 (21 Nov. 1983), pp. 1945–1947. [DOI: 10.1103/PhysRevLett.51.1945](https://doi.org/10.1103/PhysRevLett.51.1945).
- [37] J. Charles et al. “Current status of the Standard Model CKM fit and constraints on  $\Delta F = 2$  New Physics”. *Phys. Rev. D* 91.7 (2015), p. 073007. [DOI: 10.1103/PhysRevD.91.073007](https://doi.org/10.1103/PhysRevD.91.073007). [arXiv: 1501.05013](https://arxiv.org/abs/1501.05013).
- [38] C. Jarlskog. “Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal  $CP$  Nonconservation”. *Phys. Rev. Lett.* 55 (10 Sept. 1985), pp. 1039–1042. [DOI: 10.1103/PhysRevLett.55.1039](https://doi.org/10.1103/PhysRevLett.55.1039).
- [39] Neta A. Bahcall et al. “The Cosmic triangle: Assessing the state of the universe”. *Science* 284 (1999), pp. 1481–1488. [DOI: 10.1126/science.284.5419.1481](https://doi.org/10.1126/science.284.5419.1481). [arXiv: astro-ph/9906463](https://arxiv.org/abs/astro-ph/9906463).
- [40] John M. O’Meara et al. “The Deuterium to Hydrogen Abundance Ratio toward a Fourth QSO: HS 0105+1619”. *The Astrophysical Journal* 552.2 (May 2001), pp. 718–730. ISSN: 1538-4357. [DOI: 10.1086/320579](https://doi.org/10.1086/320579).
- [41] Andrei D Sakharov. “Violation of  $CP$  in variance,  $C$  asymmetry, and baryon asymmetry of the universe”. *Soviet Physics Uspekhi* 34.5 (May 1991), pp. 392–393. [DOI: 10.1070/pu1991v034n05abeh002497](https://doi.org/10.1070/pu1991v034n05abeh002497).

## Bibliography

- [42] Ulrich Nierste. “Three Lectures on Meson Mixing and CKM phenomenology”. *Helmholtz International Summer School on Heavy Quark Physics*. Mar. 2009. [arXiv: 0904.1869](#).
- [43] The HFLAV collaboration. “Averages of b-hadron, c-hadron, and  $\tau$ -lepton properties as of 2018”. *Eur. Phys. J. C* 81.3 (2021), p. 226. [DOI: 10.1140/epjc/s10052-020-8156-7](#). [arXiv: 1909.12524](#).
- [44] V. Weisskopf and E. Wigner. “Berechnung der natürlichen Linienbreite auf Grund der Diracschen Lichttheorie”. *Z. Phys.* 63 (1930), pp. 54–73. [DOI: 10.1007/BF01336768](#).
- [45] The LHCb collaboration. “Observation of  $CP$  violation in two-body  $B_s^0$ -meson decays to charged pions and kaons”. *JHEP* 03 (2021), p. 075. [DOI: 10.1007/JHEP03\(2021\)075](#). [arXiv: 2012.05319](#).
- [46] The LHCb collaboration. “Observation of  $CP$  Violation in Charm Decays”. *Phys. Rev. Lett.* 122.21 (2019), p. 211803. [DOI: 10.1103/PhysRevLett.122.211803](#). [arXiv: 1903.08726](#).
- [47] Ulrich Nierste. “ $CP$  asymmetry in flavor-specific B decays”. *39th Rencontres de Moriond on Electroweak Interactions and Unified Theories*. June 2004. [arXiv: hep-ph/0406300](#).
- [48] The LHCb collaboration. “Measurement of the semileptonic  $CP$  asymmetry in  $B^0 - \bar{B}^0$  mixing”. *Phys. Rev. Lett.* 114 (2015), p. 041601. [DOI: 10.1103/PhysRevLett.114.041601](#). [arXiv: 1409.8586](#).
- [49] The LHCb collaboration. “Measurement of the  $CP$  asymmetry in  $B_s^0 - \bar{B}_s^0$  mixing”. *Phys. Rev. Lett.* 117.6 (2016). [Addendum: *Phys.Rev.Lett.* 118, 129903 (2017)], p. 061803. [DOI: 10.1103/PhysRevLett.117.061803](#). [arXiv: 1605.09768](#).
- [50] The HFLAV collaboration. “Averages of  $b$ -hadron,  $c$ -hadron, and  $\tau$ -lepton properties as of summer 2016”. *Eur. Phys. J. C* 77.12 (2017), p. 895. [DOI: 10.1140/epjc/s10052-017-5058-4](#). [arXiv: 1612.07233](#).
- [51] The LHCb collaboration. “Measurement of  $CP$  violation in  $B^0 \rightarrow J/\psi K_S^0$  and  $B^0 \rightarrow \psi(2S)K_S^0$  decays”. *JHEP* 11 (2017), p. 170. [DOI: 10.1007/JHEP11\(2017\)170](#). [arXiv: 1709.03944](#).
- [52] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. [DOI: 10.1103/PhysRevD.2.1285](#).
- [53] Takeo Inami and C. S. Lim. “Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes  $KL \rightarrow \mu\mu$ ,  $K^+ \rightarrow \pi + \nu\nu$  and  $K^0 \leftrightarrow \bar{K}^0$ ”. *Progress of Theoretical Physics* 65.1 (Jan. 1981), pp. 297–314. ISSN: 0033-068X. [DOI: 10.1143/PTP.65.297](#). eprint: <https://academic.oup.com/ptp/article-pdf/65/1/297/5252099/65-1-297.pdf>.

- [54] Takeo Inami and C. S. Lim. “Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes  $KL \rightarrow \mu\mu$ ,  $K^+ \rightarrow \pi + \nu\nu$  and  $K^0 \leftrightarrow \bar{K}^0$ ”. *Progress of Theoretical Physics* 65.5 (May 1981), pp. 1772–1772. ISSN: 0033-068X. DOI: [10.1143/PTP.65.1772](https://doi.org/10.1143/PTP.65.1772). eprint: <https://academic.oup.com/ptp/article-pdf/65/5/1772/5255913/65-5-1772.pdf>.
- [55] Andrzej J. Buras, Matthias Jamin, and Peter H. Weisz. “Leading and next-to-leading QCD corrections to  $\alpha$ -parameter and  $B^0 - \bar{B}^0$  mixing in the presence of a heavy top quark”. *Nuclear Physics B* 347.3 (1990), pp. 491–536. ISSN: 0550-3213. DOI: [10.1016/0550-3213\(90\)90373-L](https://doi.org/10.1016/0550-3213(90)90373-L).
- [56] Luca Di Luzio et al. “ $\Delta M_s$  theory precision confronts flavour anomalies”. *JHEP* 12 (2019), p. 009. DOI: [10.1007/JHEP12\(2019\)009](https://doi.org/10.1007/JHEP12(2019)009). arXiv: [1909.11087](https://arxiv.org/abs/1909.11087).
- [57] J. Charles et al. *CKM Fitter, Preliminary results as of Summer 2019*. 2019. URL: [http://ckmfitter.in2p3.fr/www/results/plots\\_summer19/ckm\\_res\\_summer19.html](http://ckmfitter.in2p3.fr/www/results/plots_summer19/ckm_res_summer19.html).
- [58] The Muon  $g-2$  collaboration. “Measurement of the anomalous precession frequency of the muon in the Fermilab Muon  $g - 2$  Experiment”. *Phys. Rev. D* 103.7 (2021), p. 072002. DOI: [10.1103/PhysRevD.103.072002](https://doi.org/10.1103/PhysRevD.103.072002). arXiv: [2104.03247](https://arxiv.org/abs/2104.03247).
- [59] T. Aushev et al. “Physics at Super B Factory” (Feb. 2010). arXiv: [1002.5012](https://arxiv.org/abs/1002.5012).
- [60] The LHCb collaboration. “The LHCb Detector at the LHC”. *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [61] “LHC Machine”. *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [62] The ATLAS collaboration. “Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC” (June 2019). CDS: [2677054](https://cds.cern.ch/record/2677054)
- [63] The LHCb collaboration. “Framework TDR for the LHCb Upgrade: Technical Design Report” (Apr. 2012). CDS: [1443882](https://cds.cern.ch/record/1443882)
- [64] The LHCb collaboration. “LHCb VELO Upgrade Technical Design Report” (Nov. 2013). CDS: [1624070](https://cds.cern.ch/record/1624070)
- [65] The LHCb collaboration. “LHCb PID Upgrade Technical Design Report” (Nov. 2013). CDS: [1624074](https://cds.cern.ch/record/1624074)
- [66] The LHCb collaboration. “LHCb Tracker Upgrade Technical Design Report” (Feb. 2014). CDS: [1647400](https://cds.cern.ch/record/1647400)
- [67] The LHCf collaboration. “Technical design report of the LHCf experiment: Measurement of photons and neutral pions in the very forward region of LHC” (Feb. 2006).
- [68] The TOTEM collaboration. “TOTEM: Technical design report. Total cross section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN” (Jan. 2004).

## Bibliography

- [69] The FASER collaboration. “Technical Proposal for FASER: ForwArd Search Experiment at the LHC” (Dec. 2018). [arXiv: 1812.09139](#).
- [70] The LHCb collaboration.  *$\bar{b}b$  production angle plots*. URL: [https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb\\_ProductionAngles.html](https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html) (visited on 05/02/2021).
- [71] F Alessio, R Jacobsson, and S Schleich. “The LHCb Online Framework for Experiment Protection, and Global Operational Control and Monitoring”. *Journal of Physics: Conference Series* 331.2 (Dec. 2011), p. 022002. [DOI: 10.1088/1742-6596/331/2/022002](#).
- [72] The LHCb collaboration. “LHCb reoptimized detector design and performance: Technical Design Report” (2003). [CDS: 630827](#)
- [73] The LHCb collaboration. URL: <https://lbtwiki.cern.ch/bin/view/VELO/VELOConferencePlots> (visited on 05/03/2021).
- [74] The LHCb collaboration. “LHCb VELO (Vertex Locator): Technical Design Report”. Technical design report. LHCb (2001). [CDS: 504321](#)
- [75] P. A. Čerenkov. “Visible emission of clean liquids by action of  $\gamma$  radiation”. *Doklady Akademii Nauk SSSR* 451.2 (1934).
- [76] T. Blake. “Ring Imaging Cherenkov Detectors and the Rare Decay  $Bd \rightarrow K^0 u^+ u^-$  at LHCb”. PhD thesis. London U., 2008.
- [77] The LHCb collaboration. “Measurement of the  $B^\pm$  production cross-section in pp collisions at  $\sqrt{s} = 7$  and 13 TeV”. *JHEP* 12 (2017), p. 026. [DOI: 10.1007/JHEP12\(2017\)026](#). [arXiv: 1710.04921](#).
- [78] The LHCb collaboration. “Measurements of prompt charm production cross-sections in pp collisions at  $\sqrt{s} = 13$  TeV”. *JHEP* 03 (2016). [Erratum: *JHEP* 09, 013 (2016), Erratum: *JHEP* 05, 074 (2017)], p. 159. [DOI: 10.1007/JHEP03\(2016\)159](#). [arXiv: 1510.01707](#).
- [79] R. Aaij et al. “Tesla : an application for real-time data analysis in High Energy Physics”. *Comput. Phys. Commun.* 208 (2016), pp. 35–42. [DOI: 10.1016/j.cpc.2016.07.022](#). [arXiv: 1604.05596](#).
- [80] The LHCb collaboration. “Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC”. *JINST* 14.04 (2019), P04013. [DOI: 10.1088/1748-0221/14/04/P04013](#). [arXiv: 1812.10790](#).
- [81] I Bird et al. “Update of the Computing Models of the WLCG and the LHC Experiments” (Apr. 2014). [CDS: 1695401](#)
- [82] Concezio Bozzi. “LHCb Computing Resource usage in 2018” (Feb. 2019). [CDS: 2657833](#)
- [83] Rene Brun et al. *root-project/root: v6.18/02*. Version v6-18-02. Aug. 2019. [DOI: 10.5281/zenodo.3895860](#).

- [84] G. Barrand et al. “GAUDI — A software architecture and framework for building HEP data processing applications”. *Computer Physics Communications* 140.1 (2001). CHEP2000, pp. 45–55. ISSN: 0010-4655. DOI: [10.1016/S0010-4655\(01\)00254-5](https://doi.org/10.1016/S0010-4655(01)00254-5).
- [85] The LHCb collaboration. *LHCb Starterkit Lessons » First Analysis Steps » The LHCb data flow*. Version c22d51d. URL: <https://lhcb.github.io/starterkit-lessons/first-analysis-steps/dataflow.html> (visited on 05/02/2021).
- [86] Jim Pivarvski. *Python Data Science, HL-LHC Analysis Mini-Workshop*. URL: <https://github.com/jpivarvski-talks/2021-05-04-hllhc-workshop> (visited on 05/04/2020).
- [87] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024). arXiv: [1410.3012](https://arxiv.org/abs/1410.3012).
- [88] D. J. Lange. “The EvtGen particle decay simulation package”. *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: [10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4).
- [89] J. Allison et al. “Geant4 developments and applications”. *IEEE Transactions on Nuclear Science* 53.1 (2006), pp. 270–278. DOI: [10.1109/TNS.2006.869826](https://doi.org/10.1109/TNS.2006.869826).
- [90] Tianqi Chen and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [91] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016. ISBN: 0262035618.
- [92] L. S. Shapley, K. J. Arrow, E. W. Barankin, et al. “A VALUE FOR n-PERSON GAMES”. *Contributions to the Theory of Games (AM-28), Volume II*. Princeton University Press, 1953, pp. 307–318. ISBN: 9780691079356. URL: <http://www.jstor.org/stable/j.ctt1b9x1zv.24>.
- [93] Scott M Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions”. *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>.
- [94] F. James and M. Roos. “Minuit - a system for function minimization and analysis of the parameter errors and correlations”. *Computer Physics Communications* 10.6 (1975), pp. 343–367. ISSN: 0010-4655. DOI: [10.1016/0010-4655\(75\)90039-9](https://doi.org/10.1016/0010-4655(75)90039-9).
- [95] Hans Dembinski and Piti Ongmongkolkul et al. “scikit-hep/iminuit” (Dec. 2020). DOI: [10.5281/zenodo.4310361](https://doi.org/10.5281/zenodo.4310361).
- [96] Muriel Pivk and Francois R. Le Diberder. “SPlot: A Statistical tool to unfold data distributions”. *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. DOI: [10.1016/j.nima.2005.08.106](https://doi.org/10.1016/j.nima.2005.08.106). arXiv: [physics/0402083](https://arxiv.org/abs/physics/0402083).

## Bibliography

- [97] The HEP Software Foundation collaboration. “A Roadmap for HEP Software and Computing R and D for the 2020s”. *Comput. Softw. Big Sci.* 3.1 (2019), p. 7. DOI: [10.1007/s41781-018-0018-8](https://doi.org/10.1007/s41781-018-0018-8). arXiv: [1712.06982](https://arxiv.org/abs/1712.06982).
- [98] F Mölder et al. “Sustainable data analysis with Snakemake [version 1; peer review: 1 approved, 1 approved with reservations]”. *F1000Research* 10.33 (2021). DOI: [10.12688/f1000research.29032.1](https://doi.org/10.12688/f1000research.29032.1).
- [99] The LHCb collaboration. “Precise measurement of the  $f_s/f_d$  ratio of fragmentation fractions and of  $B_s^0$  decay branching fractions” (Mar. 2021). arXiv: [2103.06810](https://arxiv.org/abs/2103.06810).
- [100] Alessandro Bertolin et al. “Precision measurement of the  $B_s^0 - \bar{B}_s^0$  oscillation frequency with the  $B_s^0 \rightarrow D_s^- \pi^+$  decays” (May 2020). CDS: [2718032](https://cds.cern.ch/record/2718032)
- [101] The LHCb collaboration. “The PIDCalib package” (July 2016). CDS: [2202412](https://cds.cern.ch/record/2202412)
- [102] R. E. Kalman. “A New Approach to Linear Filtering and Prediction Problems”. *Journal of Basic Engineering* 82.1 (Mar. 1960), pp. 35–45. ISSN: 0021-9223. DOI: [10.1115/1.3662552](https://doi.org/10.1115/1.3662552).
- [103] The LHCb collaboration. “Study of beauty hadron decays into pairs of charm hadrons”. *Phys. Rev. Lett.* 112 (2014), p. 202001. DOI: [10.1103/PhysRevLett.112.202001](https://doi.org/10.1103/PhysRevLett.112.202001). arXiv: [1403.3606](https://arxiv.org/abs/1403.3606).
- [104] The LHCb collaboration. “Measurement of CP asymmetry in  $B_s^0 \rightarrow D_s^\mp K^\pm$  decays”. *JHEP* 03 (2018), p. 059. DOI: [10.1007/JHEP03\(2018\)059](https://doi.org/10.1007/JHEP03(2018)059). arXiv: [1712.07428](https://arxiv.org/abs/1712.07428).
- [105] Diego Martínez Santos and Frederic Dupertuis. “Mass distributions marginalized over per-event errors”. *Nucl. Instrum. Meth.* A764 (2014), pp. 150–155. DOI: [10.1016/j.nima.2014.06.081](https://doi.org/10.1016/j.nima.2014.06.081). arXiv: [1312.5000](https://arxiv.org/abs/1312.5000).
- [106] N. L. Johnson. “Systems of frequency curves generated by methods of translation”. *Biometrika* 36 (1949), pp. 149–176. DOI: [10.1093/biomet/36.1-2.149](https://doi.org/10.1093/biomet/36.1-2.149).
- [107] Kyle S. Cranmer. “Kernel estimation in high-energy physics”. *Comput. Phys. Commun.* 136 (2001), pp. 198–207. DOI: [10.1016/S0010-4655\(00\)00243-5](https://doi.org/10.1016/S0010-4655(00)00243-5). arXiv: [hep-ex/0011057](https://arxiv.org/abs/hep-ex/0011057).
- [108] Jack Wimberley et al. *Espresso Performance Monitor*. URL: <https://gitlab.cern.ch/lhcb-ft/EspressoPerformanceMonitor/>.
- [109] Kevin Heinicke. “Optimization of Flavour Tagging Algorithms for the LHCb Experiment”. MA thesis. Tech. U., Dortmund, Sept. 2016. CDS: [2229990](https://cds.cern.ch/record/2229990)
- [110] The LHCb collaboration. “B flavour tagging using charm decays at the LHCb experiment”. *JINST* 10.10 (2015), P10005. DOI: [10.1088/1748-0221/10/10/P10005](https://doi.org/10.1088/1748-0221/10/10/P10005). arXiv: [1507.07892](https://arxiv.org/abs/1507.07892).
- [111] The LHCb collaboration. “A new algorithm for identifying the flavour of  $B_s^0$  mesons at LHCb”. *JINST* 11.05 (2016), P05010. DOI: [10.1088/1748-0221/11/05/P05010](https://doi.org/10.1088/1748-0221/11/05/P05010). arXiv: [1602.07252](https://arxiv.org/abs/1602.07252).
- [112] The LHCb collaboration. “New algorithms for identifying the flavour of  $B^0$  mesons using pions and protons”. *Eur. Phys. J. C* 77.4 (2017), p. 238. DOI: [10.1140/epjc/s10052-017-4731-y](https://doi.org/10.1140/epjc/s10052-017-4731-y). arXiv: [1610.06019](https://arxiv.org/abs/1610.06019).

- [113] Tatiana Likhomanenko, Denis Derkach, and Alex Rogozhnikov. “Inclusive Flavour Tagging Algorithm”. *J. Phys. Conf. Ser.* 762.1 (2016). Ed. by Luis Salinas and Claudio Torres, p. 012045. DOI: [10.1088/1742-6596/762/1/012045](https://doi.org/10.1088/1742-6596/762/1/012045). arXiv: [1705.08707](https://arxiv.org/abs/1705.08707).
- [114] D.E. Rumelhart, G.E. Hintont, and R.J. Williams. “Learning representations by back-propagating errors”. *Nature* 323 (1986), pp. 533–536.
- [115] Vukan Jevtic, Patrick Mackowiak, and Gerwin Meier. “Measurement of  $CP$  Violation in  $B^0 \rightarrow \psi K_S^0$  and  $B_s^0 \rightarrow J/\psi K_S^0$  decays” (Aug. 2020). CDS: [2707700](https://cds.cern.ch/record/2707700)
- [116] Alex Rogozhnikov et al. *hep\_ml*. Version v0.5.2. May 2018. DOI: [10.5281/zenodo.1247391](https://doi.org/10.5281/zenodo.1247391).
- [117] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. *Nature Methods* 17 (2020), pp. 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [118] Christoph Langenbruch. “Parameter uncertainties in weighted unbinned maximum likelihood fits” (Nov. 2019). arXiv: [1911.01303](https://arxiv.org/abs/1911.01303).
- [119] Matthew David Needham. “Momentum scale calibration of the LHCb spectrometer”. LHCb-INT-2017-008 (Apr. 2017). CDS: [2260848](https://cds.cern.ch/record/2260848)
- [120] The LHCb collaboration. “Precision measurement of the  $B_s^0$ - $\bar{B}_s^0$  oscillation frequency with the decay  $B_s^0 \rightarrow D_s^- \pi^+$ ”. *New J. Phys.* 15 (2013), p. 053021. DOI: [10.1088/1367-2630/15/5/053021](https://doi.org/10.1088/1367-2630/15/5/053021). arXiv: [1304.4741](https://arxiv.org/abs/1304.4741).
- [121] Stephanie Hansmann-Menzemer, Georg Krocker, and Sebastian Wandernoth. “Measurement of  $\Delta m_s$  in the decay  $B_s^0 \rightarrow D_s^- \pi^+$  using  $1\text{fb}^{-1}$ ” (Apr. 2014). CDS: [1445564](https://cds.cern.ch/record/1445564)
- [122] The LHCb collaboration. “Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era” (Aug. 2018). arXiv: [1808.08865](https://arxiv.org/abs/1808.08865).
- [123] The Belle-II collaboration. “The Belle II Physics Book”. *PTEP* 2019.12 (2019). Ed. by E. Kou and P. Urquijo. [Erratum: *PTEP* 2020, 029201 (2020)], p. 123C01. DOI: [10.1093/ptep/ptz106](https://doi.org/10.1093/ptep/ptz106). arXiv: [1808.10567](https://arxiv.org/abs/1808.10567).
- [124] The FCC collaboration. “FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3”. *Eur. Phys. J. ST* 228.4 (2019), pp. 755–1107. DOI: [10.1140/epjst/e2019-900087-0](https://doi.org/10.1140/epjst/e2019-900087-0).





# Danksagung

Für die Möglichkeit, in den letzten viereinhalb Jahren noch tiefer in einige aktuelle Forschungsschwerpunkte der Teilchenphysik einzutauchen, möchte ich mich zuerst bei meinem Doktorvater, Herrn Prof. Dr. Spaan bedanken. Die gewährten Freiheiten, die Unterstützung, zu Reisen und eigene Ideen umzusetzen und das Vertrauen in die eigene Arbeit haben die Promotion zu einer besonderen Zeit werden lassen, in der ich viele neue Erfahrungen sammeln konnte.

Mein Dank gilt zudem meinem Zweitgutachter, Herrn Prof. Dr. Dr. Rhode, für die Zeit, meine Arbeit zu beurteilen. Ich möchte mich außerdem beim Bundesministerium für Bildung und Forschung für die Finanzierung meiner Arbeit bedanken, sowie beim Sonderforschungsbereich 876, für die Möglichkeit für fachübergreifende Diskussionen und Arbeiten.

Meine Promotionsarbeit gründet auf der Arbeit von vielen anderen Doktoranden und Wissenschaftlerinnen, die Detektorkomponenten entwickeln, Software schreiben, und Informationen zur Verfügung stellen und denen ich meinen Dank aussprechen möchte. Die Zusammenarbeit mit meinen Kollegen vom CERN war eine herausragende Erfahrung, wofür ich dem  $\Delta m_s$ -Team danken möchte. Vor allem gilt dies Agnieszka, dafür, dass Sie sich während der letzten zweieinhalb Jahren sehr viel Zeit für Diskussionen (und Debugging) genommen hat und eine wunderbare Balance zwischen Profession und Privatem gehalten hat – dzięki!

Die Zeit der Promotion (und vieles was dazu geführt hat) wird für mich immer mit dem Lehrstuhlleben bei E5 verbunden sein. Für die wunderbare Arbeitsatmosphäre vor Ort und viele schöne Gespräche, die nicht immer mit dem täglichen Geschäft zu tun hatten, muss vielen Generationen von Lehrstuhlmitgliedern gedankt werden. Um nur einige zu nennen: Philipp und Patrick für die Diskussionen zu den "richtigen" Themen; Quentin für die gemeinsame Arbeit an der Analyse und Zeit im Büro (wenn nicht gerade eine globale Pandemie ausgebrochen ist); Louis, Antje, Vanessa und Ramon für die guten Bürogespräche; Vukan und Alex für allerlei Diskussionen zum Flavour Tagging; Timon und Moritz für die schönen Erinnerungen an die Arbeit im und um den LHCb-Kontrollraum; Julian für die offenen Ohren und Diskussionen; unter anderen Sophie und Holger für die Korrektur von Teilen meiner Arbeit; Max, Matthias und dem Adminteam für die viele Zeit, die in unsere Infrastruktur fließt; Britta für die gute Stimmung im Sekretariat; und natürlich dem Ministry of Coffee Affairs, für die exzellente Versorgung mit Qualitätsgetränken, nach der Machtübernahme.

Schließlich gilt mein Dank meinen Freunden und meiner Familie, für den Rückhalt, die Rückzugsorte und die Ablenkungen. Die intensiven Gespräche mit Jonas und der Blick von außen haben enorm geholfen, motiviert zu bleiben. Die Unterstützung meiner Eltern Knut und Ulrike, meiner Schwiegereltern Ans und Michael, und meines Bruders Sebastian und seiner Familie, war für mich von unschätzbarem Wert – vor allem in der letzten Phase meiner Arbeit.

Zuletzt danke ich meiner Verlobten, Kristina. Für dein Verständnis, deine Nachsicht, und deine Liebe. Ebenso meinen Söhnen Jona und Levi für die regelmäßige Neufokussierung und den schönen Ausblick in die Zukunft (in der ich hoffentlich noch verstehe, was in dieser Arbeit geschrieben ist und ihr es vielleicht eines Tages wissen möchtet).