Precision tests of the Standard Model using key observables of *CP* violation and rare decays

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

vorgelegt von
Maximilian Schlupp

Fakultät Physik
Technische Universität Dortmund

Dortmund
19. Oktober 2015

Gutachter:
Prof. Dr. Bernhard Spaan
Prof. Dr. Kevin Kröninger
Dr. Matteo Palutan

Vorsitzender der Prüfungskommission:
Prof. Dr. Dmitri Yakovlev

Vertreterin der wissenschaftlichen Mitarbeiter/innen:
Dr. Bärbel Siegmann

Datum der mündlichen Prüfung:
17. Dezember 2015

Abstract

In this thesis advanced statistical methods are used for precision studies in the flavour sector of the Standard Model of particle physics. The necessary tools are developed and applied to two key measurements of the LHCb experiment: the determination of the CKM angle γ and the search for rare $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays.
The CKM angle γ is, for the first time, measured from $B_s^0 \to D_s^\mp K^\pm$ decays using a dataset corresponding to $1\,\text{fb}^{-1}$ of pp interactions at a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$. The result of $\gamma = (115^{+27}_{-43})^\circ$ is then combined with a set of γ measurements in $B \to Dh$ decays resulting in a precision on γ of $< 8^\circ$. This result improves the legacy results from the B-factories by more than a factor of two.
The rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are analysed on a dataset corresponding to $3\,\text{fb}^{-1}$ of pp interactions at a centre-of-mass energy of $\sqrt{s} = 7$ and $8\,\text{TeV}$. The branching fraction of $B_s^0 \to \mu^+\mu^-$ decays is measured to be consistent with the Standard Model prediction, while no significant excess of $B^0 \to \mu^+\mu^-$ events is found. Then, the LHCb and CMS datasets from the first run of the LHC are combined to perform a joint search for B meson decays into two muons. The decay $B_s^0 \to \mu^+\mu^-$ is observed for the first time with a statistical significance of 6σ and the first evidence for $B^0 \to \mu^+\mu^-$ decays is found with a significances of 3σ.

Zusammenfassung

Diese Arbeit beschreibt die Anwendung von statistischen Methoden auf Schlüsselmessungen des LHCb Experiments. Die dafür nötige Software wurde im Rahmen dieser Arbeit entwickelt und auf die Messung des CKM Winkels γ und die Suche der seltenen Zerfälle $B_s^0 \to \mu^+\mu^-$ und $B^0 \to \mu^+\mu^-$ angewendet. Im Verlauf der Arbeit wurde auf einem pp Kollisionsdatensatz, der einer integrierten Luminosität von $1\,\text{fb}^{-1}$ bei $\sqrt{s} = 7\,\text{TeV}$ Schwerpunktsenergie entspricht, der Winkel γ zum ersten Mal in $B_s^0 \to D_s^\mp K^\pm$ Zerfällen gemessen. Das Ergebnis von $\gamma = (115^{+27}_{-43})^\circ$ wurde anschließend mit γ Messungen aus $B \to Dh$ Zerfällen kombiniert. Die erreichte Präzision von unter 8° verbessert die Messungen der B Fabriken um mehr als einen Faktor zwei.
Die Raten von $B_s^0 \to \mu^+\mu^-$ und $B^0 \to \mu^+\mu^-$ Zerfällen wurden auf einem Datensatz von pp Kollisionen gemessen, der einer integrierten Luminosität von $3\,\text{fb}^{-1}$ bei einer Schwerpunktsenergie von $\sqrt{s} = 7$ and $8\,\text{TeV}$ entspricht. Das Ergebnis ist kompatibel mit den Vorhersagen des Standardmodells. In der Kombination der in der ersten Datennahmeperiode des LHCs aufgenommenen Datensätze der LHCb und CMS Experimente, ist es gelungen den Zerfall $B_s^0 \to \mu^+\mu^-$ zum ersten Mal mit einer statistischen Signifikanz von sechs Standardabweichungen nachzuweisen. Mit 3σ Signifikanz über dem Untergrund wurde der erste Hinweis auf den Zerfall $B^0 \to \mu^+\mu^-$ gefunden.

For Leah Christin

Contents

1 Introduction **1**

2 Basic theoretical concepts and experimental observables **5**
2.1 *CP* violation in the Standard Model 6
2.2 Flavour changing neutral currents 8
2.3 Key observables in flavour physics 10
2.3.1 The CKM angle γ from tree-level $B_s^0 \to D_s^{\mp} K^{\pm}$ decays . . . 11
2.3.2 The rare $B \to \mu^+\mu^-$ decays 15

3 Statistical methods **19**
3.1 Confidence intervals 19
3.2 Construction of confidence intervals 20
3.3 The unified frequentist method 24
3.3.1 Practical application of the unified frequentist method 24
3.3.2 The unified frequentist method with a biased estimator . . . 29
3.3.3 The unified frequentist method with nuisance parameters . . 30
3.3.4 The PLUGIN method 31
3.3.5 The Berger-Boos method 33
3.3.6 Wilks' theorem 33
3.4 Significance 34

4 The LHCb experiment at the Large Hadron Collider **37**
4.1 The LHCb detector 38
4.1.1 The tracking system 39
4.1.2 Particle identification 41
4.1.3 The trigger system 45
4.2 Flavour tagging 46

5 Measurement of the CKM angle γ with $B_s^0 \to D_s^{\mp} K^{\pm}$ decays **51**
5.1 Analysis strategy 51
5.2 Selection 52
5.3 Multivariate fit 53
5.3.1 Invariant B_s^0 and D_s^- mass templates 54
5.3.2 Particle identification shapes 56
5.3.3 Results of the multivariate fit 57
5.4 Flavour tagging 57

5.5 Determination of CP asymmetry parameters . . . 59
5.5.1 Decay time resolution and acceptance . . . 59
5.5.2 Mistag distributions . . . 60
5.5.3 The decay time distributions . . . 61
5.5.4 Results . . . 62
5.6 Measurement of the CKM angle γ . . . 64
5.6.1 Expected sensitivity . . . 64
5.6.2 Results for the CKM angle g . . . 67
5.7 Future prospects . . . 70

6 Combination of γ measurements at the LHCb experiment 73
6.1 Input measurements . . . 73
6.2 Combination method and validation . . . 75
6.3 Results . . . 78
6.4 Future Prospects . . . 80

7 Measurement of the branching fractions of $B \to \mu^+\mu^-$ decays at the LHCb experiment 81
7.1 Analysis strategy . . . 81
7.2 Selection . . . 82
7.2.1 Event selection . . . 82
7.2.2 Signal and background classification . . . 83
7.3 Decay time dependent effects . . . 85
7.4 Normalisation . . . 86
7.5 Signal shapes . . . 87
7.6 Estimation of background components . . . 89
7.6.1 Mis-identification procedure . . . 89
7.6.2 Doubly mis-identified decays . . . 90
7.6.3 Background from semi-leptonic decays . . . 93
7.6.4 Partially reconstructed decays . . . 100
7.6.5 Other potential background channels . . . 104
7.6.6 Summary of the exclusive background studies . . . 107
7.6.7 Estimation of the combinatorial background . . . 108
7.7 Expected significance . . . 109
7.8 Results . . . 110
7.9 Future prospects . . . 112

8 Combination of $B \to \mu^+\mu^-$ results from the LHCb and CMS experiments 117
8.1 Summary of the CMS analysis . . . 117
8.2 Combined maximum likelihood fit . . . 120
8.3 Expected signal significances . . . 121
8.4 Observed signal significances . . . 124
8.4.1 Significance calculation using pseudo experiments . . . 124
8.4.2 Results . . . 128

8.5 Confidence intervals 129
8.5.1 One dimensional intervals 129
8.5.2 Feldman-Cousins confidence interval 131
8.5.3 Two-dimensional confidence regions 132
8.6 Future prospects 134

9 Conclusions and prospects **135**

Bibliography **137**

A Appendix **155**
A.1 Central limit theorem 155
A.2 Coverage probabilities for a χ^2 distribution 156
A.3 Correlation tables for the $B_s^0 \to D_s^{\mp} K^{\pm}$ CP observables 157

Acknowledgements **159**

1 Introduction

The Standard Model of elementary particle physics (SM) is a description of three of the four fundamental forces in nature. The electromagnetic force, formulated in quantum electrodynamics (QED) describes electric and magnetic effects, the weak interaction allows for radioactive decays, and the strong interaction or quantum chromodynamics (QCD) holds together atomic nuclei. The SM combines QCD with the unified model of the electromagnetic and weak force, the electroweak interaction developed by S. Glashow, S. Weinberg, and G. Salam [1–3].
The SM is one of the most intensively and, in parts, precisely tested theories in physics. On the low energy frontier, for example, by the measurement of the deviation of the magnitude of the electron magnetic dipole moment from the QED prediction caused by interactions with the quantum fluctuations and polarity of the vacuum [4]. On the high energy frontier, it was confirmed that the electroweak symmetry breaking mechanism works as predicted by R. Brout, F. Englert, and P. Higgs [5–7]. The experimental tests to establish the Brout-Englert-Higgs (BEH) mechanism were performed by the ATLAS [8] and CMS [9] experiments at the Large Hadron Collider [10] operated by CERN. After the observation of the Higgs boson, which is the remnant particle of the BEH mechanism, in 2012 [11, 12], the properties of the Higgs boson have been studied successfully [13–16]. All measurements are consistent with the SM predictions.
Despite the overwhelming success of the SM, very important observations, mainly from cosmology, cannot be explained within the SM. Besides the absence of a quantum mechanical description of gravity, the possibility of neutrino masses or an explanation for the large mass hierarchy ranging from $\mathcal{O}(1\,\mathrm{eV}/c^2)$ to $\mathcal{O}(100\,\mathrm{GeV}/c^2)$, the SM is expected to provide the basic sets of particles, which are observed in nature. However, the particle content of the SM is only able to explain 5% of the energy density of the universe. The energy density of the universe is dominated by dark matter (26%) and dark energy (69%), where the latter is identified with a cosmological constant Λ. These numbers are measured by the Planck Collaboration [17] assuming a spacial flat universe and the standard model of cosmology ΛCDM (dark energy plus cold dark matter). The existence of dark matter has been shown in many cosmological observations such as in the comparison of gravitational lens effects with the visible hot gas content of colliding galaxies [18]. The SM, however, does not provide a suitable particle candidate for dark matter.
Another point is the observed matter-antimatter asymmetry on different scales of the universe. The typical fraction of antimatter consistent with the obser-

vations is $\mathcal{O}(10^{-6}$–$10^{-10})$ [19, 20]. In order to dynamically produce a matter-antimatter asymmetry, three conditions need to be satisfied, which were formulated by A. Sakharov [21]: the conservation of baryon number must be violated, the C and CP symmetry must be violated, and there must be interactions out of the thermal equilibrium. The SM provides mechanisms for all of these criteria, however the currently measured amount of matter-antimatter asymmetry within the SM is not sufficient to explain the observed asymmetry in the universe.
Despite good reasons to believe that the SM is not the ultimate theory, there is still no direct measurement that violates a SM prediction. A promising way to test the SM is the study of particle production and decay mechanisms in particle collisions. There are two main strategies to search for new physical phenomena at particle colliders. First, new heavy particles can directly be produced in collisions. If such a signature is found, this is a striking sign for physics beyond the SM. However, this approach is limited to the energy scale provided by the available particle accelerator. The second strategy is to probe SM mechanisms to a very high precision. This needs precise theoretical calculations of SM predictions, which then can be compared to the experimental measurements. The advantage of the precision tests is that this approach is not restricted to the energy scale of the collider. The possible new heavy particles can alter couplings, decay rates, or angular distributions by the occurrence in quantum loop corrections, which then result in differences between the measured and the predicted SM values. It can be shown that with precision measurements in decays of strange and beauty mesons energy scales of $\mathcal{O}(100$–$1\,000\,\text{TeV})$ can be tested [22].
All processes in particle physics have a statistical nature due to quantum mechanics. Therefore, no statement about the underlying physical mechanism can be inferred from a single process alone. As a consequence, large data samples in combination with statistical methods are used to quantify the existence of signals or precisely test SM predictions. Given the large data samples provided by the LHC experiments, statistical methods are the key to interpret the data in terms of physics models and parameters.
A focal point of this thesis is the development of a software framework that is able to statistically combine measurements, calculate confidence intervals and perform statistical tests. The implemented statistical methods are then applied to key observables of the SM flavour sector testing the CP violation mechanism and the branching fractions of very rare beauty meson decays.

This thesis begins with a short overview of the SM, CP violation and flavour changing neutral currents as well as a short introduction to the key observables of beauty decays in Chapter 2. Then Chapter 3 is an introduction to the statistical methods that are implemented and used in this thesis followed by an overview of the LHCb experiment in Chapter 4 that is used to obtain most of the analysed data. Then, this thesis presents the first measurement of the CKM angle γ in $B_s^0 \to D_s^\mp K^\pm$ decays[1]

[1] If not stated differently, charge conjugation is implied.

in Chapter 5 using data taken at the LHCb experiment with a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$. This measurement is combined with a variety of γ results from the LHCb experiment to perform the most precise measurement of the CKM angle γ by a single experiment in Chapter 6. The second part of the thesis is about the search for very rare beauty meson decays into two muons. First the measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and the search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment is presented in Chapter 7. Then the joint analysis of the combined LHCb and CMS data is presented in Chapter 7.

2 Basic theoretical concepts and experimental observables

The Standard Model of particle physics is the description of the elementary set of particles and three of the four fundamental fources in nature. The particles are grouped into a set of six quarks and six leptons all of which have a distinct property called *flavour*. Furthermore, the leptons and quarks are grouped in three *generations*, so that the full matter particle content of the SM can be described by twelve particles and the respective antiparticles:

$$\text{quark flavours: } \begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \qquad \text{lepton flavours: } \begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}.$$

Here, the up-type quarks: up (u), charm (c), and top (t) carry an electric charge of $+2/3$, while the down-type quarks: down (d), strange (s), and beauty (b) have a charge of $-1/3$. For the leptons, the electron (e), muon (μ), and tauon (τ) carry a full negative charge, while the respective electron-, muon-, and tau-neutrino are neutral.
The electroweak and strong interactions are described as the consequence of the invariance under local symmetry transformations and are, in the low energy regime, mediated by gauge bosons which are the photon, Z^0 and $W^\pm$ bosons for the electromagnetic and weak interaction, and gluons for the strong interaction. Under the strong force, multiple quarks form bound states, which are called hadrons in form of mesons and baryons, which are bound states consisting of a quark-antiquark pair and three quarks, respectively. In the year 2015, the LHCb experiment published the observation a more exotic five quark bound state [23], following a measurement of the resonant structure of a possible tetraquark candidate, the $Z(4430)^-$ [24].
Gauge boson masses are generated through the Brout-Englert-Higgs mechanism [5–7], which spontaneously breaks the electroweak symmetry to obtain the low energy electromagnetism of QED and massive Z^0 and $W^\pm$ bosons together with a massive scalar particle, the Higgs boson. The mechanism also provides a method to dynamically generate the masses of leptons and quarks.
This thesis presents precision tests of the SM in the realm of CP violation and rare decays. Hence, these two aspects are introduced in more detail in the following sections. For a more general introduction to gauge theory and the fundamental concepts of the SM see e.g. Refs. [25–27].

2.1 *CP* violation in the Standard Model

The SM is a relativistic quantum field theory with a hermitian hamiltonian that is invariant under Lorentz transformations. It conserves the simultaneous application of three discrete symmetries: *CPT* [25]. Here, C is the charge conjugation, changing all charge-like quantum numbers and therefore transforming e.g. an electron into a positron. The parity symmetry P reflects a spacial point symmetry ($\vec{x} \rightarrow -\vec{x}$) and transforms e.g. a right-handed electron into a left-handed electron. Finally T stands for the time symmetry interchanging $t \rightarrow -t$. The effect of *CP* violation was first discovered by J. Christenson, J. Cronin, V. Fitch, and R. Turlay in the decay of neutral kaons [28], which lead to the prediction of a third generation of quarks as the explanation for *CP* violation by M. Kobayashi, and T. Maskawa [29]. Together with the work of N. Cabbibo [30] *CP* violation in the weak interaction is instrumented in the complex phase of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. Since then, *CP* violation is observed in the neutral B^0 system by the BaBar and Belle experiments in the year 2001 [31, 32] and in the charged $B^{\pm}$ and neutral B_s^0 meson sectors by the LHCb experiment in 2012 [33] and 2013 [34], respectively.
In the SM, *CP* violation can also occur in the lepton sector [35] and within QCD. However, the amount of *CP* violation in QCD is negligible due to very strict constraints from the measurement of the electric dipole moment of the neutron [36, 37] and the effect of *CP* violation in the lepton sector has not been measured so far. As this thesis solely tests the mechanism of *CP* violation in the weak interaction of quarks, *CP* violation in the field of QCD and the lepton sector is not discussed any further.

Quark mixing and the CKM matrix

The basis of *CP* violation is quark mixing and the CKM matrix introduced by the dynamic mass generation of the quark fields. As mass terms for the quarks do not follow from a gauge principle, the general way to add the quark masses will introduce complex couplings and mix the various quark flavours [25]. Hence, the resulting mass matrices are not diagonal. To find the physical mass eigenstates, the mass matrix needs to be diagonalised by two unitary matrices U_u^{ij} and U_d^{ij}. However, this results in an effective rotation of the quark fields into their mass eigenstates, which, consequently, modifies the interactions of the gauge bosons with the quarks. These interactions are called neutral and charged currents for the interactions mediated by the Z^0 and $W^{\pm}$ bosons, respectively. The neutral current can be written as[1]

$$J_{Z^0}^{\rho} \propto \bar{f}_{L,R}\, \gamma^{\rho}\, a(\sin^2\theta_w)\, f_{L,R}\,, \tag{2.1}$$

[1] More details can be found in e.g. Ref. [25]

where $f_{L,R}$ can be the right- (R) or left-handed (L) component of a lepton or quark fields, γ^ρ denotes the Dirac matrices and θ_w is the weak mixing angle of the electroweak gauge fields. If in the neutral current from Eq. 2.1 the quark fields are transformed into the basis of mass eigenstates using the unitary matrices U_u and U_d the neutral current for left-handed up-type quarks u_L^i reads

$$\begin{aligned} J^\rho_{Z^0} &\propto (\bar{u}^i_L U^{\dagger,ij}_u)\gamma^\rho\, a(\sin^2\theta_w)(U^{jk}_u u^k_L) = \bar{u}^i_L\,\gamma^\rho\, a(\sin^2\theta_w)(U^\dagger_u U_u)^{ik} u^k_L \\ &\underbrace{=}_{\text{Unitarity}}\ \bar{u}^i_L\,\gamma^\rho\, a(\sin^2\theta_w)\, u^i_L\,. \end{aligned} \tag{2.2}$$

This means that the neutral current of weak interactions, due to the unitarity of the rotation matrix, gets diagonal. Hence, no flavour changing neutral currents are present at lowest order of weak interactions or *tree-level*. This is also called GIM mechanism after the publication of S. Glashow, J. Iliopoulos, and L. Maiani [38].
The picture changes, if the charged current is taken into account. The transformation into the basis of mass eigenstates affects also the tree-level couplings:

$$\begin{aligned} J^{\rho+}_W &\propto \bar{u}^i_L\,\gamma^\rho\, d^i_L \ \rightarrow\ (\bar{u}^i_L U^{\dagger,ij}_u)\,\gamma^\rho\,(U^{jk}_d d^k_L) \\ &= \bar{u}^i_L\,\gamma^\rho\,(U^\dagger_u U_d)^{ij}\, d^j_L\,. \end{aligned} \tag{2.3}$$

Now, each up-type quark u^i_L is linked to a linear combination of down-type quarks d^j_L through a rotation given by the CKM matrix

$$V_{\text{CKM}} \equiv U^\dagger_u U_d\,,$$

which is a complex unitary 3×3 matrix. Due to unitarity and the freedom of phase convention for the various quark fields, the number of free parameters in the CKM matrix reduces to four: three rotation angles and one complex phase. The single complex phase represents the only source of CP violation in the weak interaction, as the $W^\pm$ couplings for CP conjugated processes change the proportionality from V_{CKM} to $V^\dagger_{\text{CKM}}$. However, a single complex phase cannot be measured by an experiment, as the physical decay rates are proportional to squared decay amplitudes and therefore real numbers. In order to resolve CP violating effects, the interference of two different decay paths into the same final state is necessary, so that a phase difference can alter observable decay rates.
A more common parametrisation of the CKM matrix is given by the Wolfenstein parametrisation [39], which illustrates the hierarchy within the CKM matrix. The Wolfenstein parametrisation is given as

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.4}$$

$$= \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho-\mathrm{i}\eta) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\rho-\mathrm{i}\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{2.5}$$

up to fourth order of the expansion parameter λ. The parameters λ, A, ρ, and η can be written as [40]

$$\lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \qquad A\lambda = \left|\frac{V_{cb}}{V_{us}}\right|,$$
$$\text{and} \qquad A\lambda^3(\rho + \mathrm{i}\eta) = \frac{A\lambda^3(\bar{\rho} + \mathrm{i}\bar{\eta})\sqrt{1 - A^2\lambda^4}}{\sqrt{1-\lambda^2}[1 - A^2\lambda^4(\bar{\rho} + \mathrm{i}\bar{\eta})]} = V_{ub}^* ,$$

where $\bar{\rho} + \mathrm{i}\bar{\eta} = -(V_{ud}V_{ub}^*)/(V_{cd}V_{cb}^*)$. With this definition the CKM matrix, when written in terms of $\lambda \approx 0.23$, $A \approx 0.81$, $\bar{\rho} \approx 0.12$, and $\bar{\eta} \approx 0.35$, is unitary to all orders of λ [40]. The unitarity of the CKM matrix ($V_{\mathrm{CKM}}^\dagger V_{\mathrm{CKM}} = \mathbb{I}$) leads to equations such as

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

These unitarity conditions can be interpreted as triangles in the complex plain. When normalised to $V_{cd}V_{cb}^*$, the vertices of the triangle from Eq. 2.6 are (0,0), (1,0), and $(\bar{\rho},\bar{\eta})$. The three sides of the triangle are consequently given by

$$R_t = \left|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right|, \quad R_u = \left|\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right|, \quad R_c = 1, \tag{2.7}$$

where the three angles are defined by

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{2.8}$$

One goal of measuring *CP* violating observables is to over-constrain the CKM triangles in a way that inconsistencies can be detected from the comparison of indirect global fits and direct measurements. For example, in case of a fourth quark generation the CKM matrix is unitary in four dimensions. As a result, the three dimensional parametrisation can violate unitarity so that e.g. Eq. 2.6 does not hold any more and the representation in the complex plain is not a full triangle. Global fits using many flavour physics observables are provided by e.g. the CKM fitter group [41, 42]. The example of the constraints of the CKM triangle from Eq. 2.6 is given in Fig. 2.1. Providing input measurements for e.g. global SM fits is key to fundamentally test the CKM mechanism.

2.2 Flavour changing neutral currents

As shown in section 2.1 the quark mixing does not alter the weak neutral current mediated by the Z^0 boson. Hence, flavour changing neutral current are absent on tree-level. However, the charged current can change the flavour of quarks. As a consequence, flavour changing (net) neutral currents are only possible in the SM on

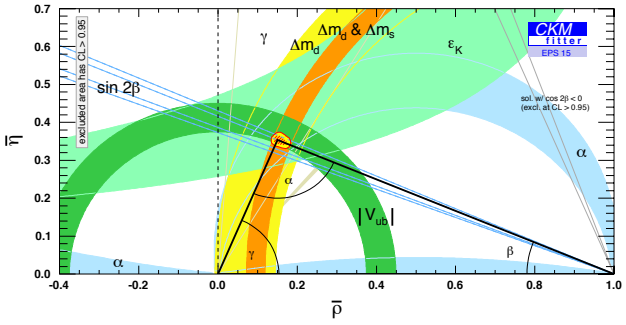


Figure 2.1: Global fit to many flavour observables provided by the CKM fitter group [42]. The red hashed area marks the 68% CL area for the apex of the triangle. A more detailed discussion of the input measurements can be found in Ref. [41].

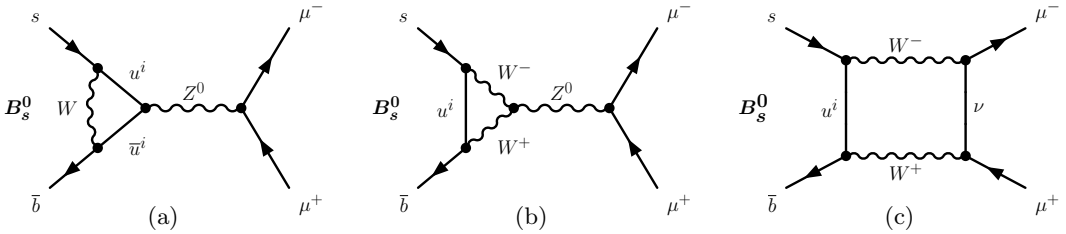


Figure 2.2: Leading order SM Feynman diagrams for the process of $B_s^0 \to \mu^+\mu^-$ as an example for FCNCs. Introducing at least two more vertices compared to the tree-level allows for the charged weak current to actually change the quark flavour so that the decay $B_s^0 \to \mu^+\mu^-$ can occur. (a) and (b) show the Z^0 penguin diagrams and (c) the $W^\pm$ box diagram.

loop-level, which are processes with at least four vertices. For example, Feynman diagrams for the flavour changing neutral current decay $B_s^0 \to \mu^+\mu^-$ in the SM are shown in Fig. 2.2. Within the SM $B_s^0 \to \mu^+\mu^-$ decays can occur via a W loop prior to a neutral current $Z^0 \to \mu^+\mu^-$ decay in Fig. 2.2(a) and Fig. 2.2(b) or via a W box diagram in Fig. 2.2(c).
The calculation of such FCNC processes is often parametrised in terms of an effective hamiltonian [43–45]

$$\mathcal{H}_{\text{eff}} = \frac{\mathrm{G}_F}{\sqrt{2}} \sum_i c_i(\mu) Q_i(\mu) \,, \tag{2.9}$$

where G_F is the Fermi constant, μ the renormalisation scale, and c_i are the *Wilson coefficients* [43, 46], which are the expansion coefficients and can be interpreted as effective coupling strength for each operator Q_i. The operators describe *long distance effects*, which means that they depend on the incoming and outgoing particle fields and

the structure of the interaction. For a specific decay with initial state $|M\rangle$ and final state $|F\rangle$ the decay amplitude $\mathcal{A}$ can be calculated as

$$\mathcal{A}(M \to F) = \langle F|\mathcal{H}_{\text{eff}}|M\rangle = \mathcal{H}_{\text{eff}} = \frac{\mathrm{G}_F}{\sqrt{2}} \sum_i c_i(\mu)\langle F|Q_i(\mu)|M\rangle \,, \tag{2.10}$$

so that the sum in Eq. 2.10 only contains relevant operators Q_i for the decay $M \to F$. The operators are evaluated at the renormalisation scale μ, which separates the long distance (Q_i) from the short distance (c_i) effects [45]. Hence, all physical interactions of the scale $m > \mu$ are within the Wilson coefficients c_i, especially the weak interaction or potential new particles that have a mass larger than the renormalisation scale. As the Wilson coefficients and operators can be precisely calculated from theory, the operator product expansion from Eq. 2.9 provides a powerful tool for searches for physics beyond the SM.

2.3 Key observables in flavour physics

After the successful operations of the B-factories BaBar and Belle, which outperformed expectations and delivered an overwhelming and rich physics legacy [47], the upcoming LHCb experiment formulated a set of key measurements [48] in which the LHCb collaboration can have a great impact within the first years of data-taking. The first measurement regards many different decay channels, all of which aim to measure the least well known CKM angle γ (cf. Section 2.1). The B-factories BaBar and Belle measured $\gamma = (69^{+17}_{-16})$ [49] and $\gamma = (68^{+15}_{-14})^\circ$ [50], respectively. One goal of the LHCb physics programme is to perform a precision measurement of the CKM angle γ and use this as a test for the CKM mechanism in the SM and as a possible search for physics beyond the SM. For the latter the comparison from a precise tree-level baseline measurement (not sensitive to most new models) and measurements from loop processes are needed. The SM measurement can be performed with $B \to Dh$ decays, where B denotes either a B^0, B^+, or B_s^0 meson, D denotes e.g. a D^0 or D_s^- meson, and h could either be a π^+ or K^+ meson. The loop contributions to γ can then be measured in e.g. $B^0 \to \pi^+\pi^-$ or $B_s^0 \to K^+K^-$.
The second key measurement in the regime of CP violation is the measurement of mixing induced CP violation in the B_s^0 sector using $B_s^0 \to J/\psi\,\phi$ decays.
The last key observables are coming from the regime of flavour changing neutral currents (cf. Section 2.2) making use of the high luminosity at the LHC collider. There are the classes of radiative decays focussing on probing the photon polarisation and CP violating observables in $B_s^0 \to \phi\gamma$ decays. Then the decay $B^0 \to K^{*0}\mu^+\mu^-$ is given as a theoretical clean test for new physics models that have different Lorentz structures compared to the SM. But most importantly, for the test of flavour changing neutral currents, the search for very rare $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays. As the two decays have very similar decay properties, $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are

treated as signal simultaneously and are denoted $B \to \mu^+\mu^-$. The importance of $B \to \mu^+\mu^-$ decays is due to the theoretically clean environment of a heavy meson decaying into two leptons and the vast variety of physics models beyond the SM which affects the decay $B \to \mu^+\mu^-$.
This thesis studies two of the five key measurements mentioned above: the measurement of the CKM angle γ from tree-level decays and the search for $B \to \mu^+\mu^-$ decays.

2.3.1 The CKM angle γ from tree-level $B_s^0 \to D_s^\mp K^\pm$ decays

At the B factories, the traditional way to measure the CKM angle γ was to use decay-time integrated or *time-independent* methods [51–55]. However, it is also possible to perform *time-dependent* γ measurements from tree-level $B^0_{(s)} \to D^\mp_{(s)}\pi^\pm, K^\pm$ decays [56–59], which involves resolving the B–$\overline{B}$ oscillations in dependence of the decay-time of the B meson. First time-dependent measurements have been performed by BaBar [60, 61] and Belle [62, 63].
Most of the time-independent methods get the sensitivity on the CKM angle γ by utilising external measurements. As the interfering diagrams are of the form $B^- \to D^0 K^- \leftrightarrow B^- \to \overline{D}^0 K^-$, the decay amplitudes depend on decay parameters of the D meson system such as the ratio of interfering amplitudes $r_D^f = \left|\overline{D}^0 \to f / D^0 \to f\right|$ and the strong phase difference δ_D^f, where f denotes the D decay final state. The strong phase difference δ_D arises from e.g. final state scatterings [64] and is invariant under *CP* transformations. The final state scattering includes effects of the form $D \to f' \to f$, where f' marks an intermediate resonance. To increase the sensitivity on the CKM angle γ, the D system parameters are taken from dedicated studies of, for example, the CLEO collaboration (e.g. from Ref. [65] for $D^0 \to K^0_S K^\pm \pi^\mp$ decays). For the analysis of $B_s^0 \to D_s^\mp K^\pm$ decays, however, the sensitivity to γ stems from the interference of tree-level amplitudes, illustrated by the leading order Feynman diagrams in Fig. 2.3. Here, the final state of the D_s^- meson is — besides experimental considerations — not of greater interest and therefore the CKM angle γ can be measured without any input for the D meson decay system. Further, it can be seen from Fig. 2.3 that the charge-explicit decays $B_s^0 \to D_s^- K^+$ in Fig. 2.3(a) and $\overline{B}_s^0 \to D_s^- K^+$ in Fig. 2.3(b) occur at tree-level and have approximately the same amplitude. In terms of the Wolfenstein parameter λ (cf. Eq. 2.4), the decay amplitudes for $B_s^0 \to D_s^- K^+$ (A_f) and $\overline{B}_s^0 \to D_s^- K^+$ ($\bar{A}_f$) are proportional to [59]

$$\begin{aligned} A_f &\propto V_{cb}^* V_{us} = A\lambda^3 + \mathcal{O}(\lambda^4)\,, \\ \bar{A}_f &\propto V_{cs}^* V_{ub} = A\lambda^3(\rho - \mathrm{i}\eta) + \mathcal{O}(\lambda^4) = A\lambda^3\sqrt{\bar{\rho}^2 + \bar{\eta}^2}\, e^{-\mathrm{i}\gamma}\,, \end{aligned} \tag{2.11}$$

respectively. The same statement holds for the charge-conjugated amplitudes $\bar{A}_{\bar{f}}$ and $A_{\bar{f}}$. Indeed, a measure for the sensitivity of a decay to *CP* observables is the ratio of the interfering diagrams, which in case of $B_s^0 \to D_s^\mp K^\pm$ decays is

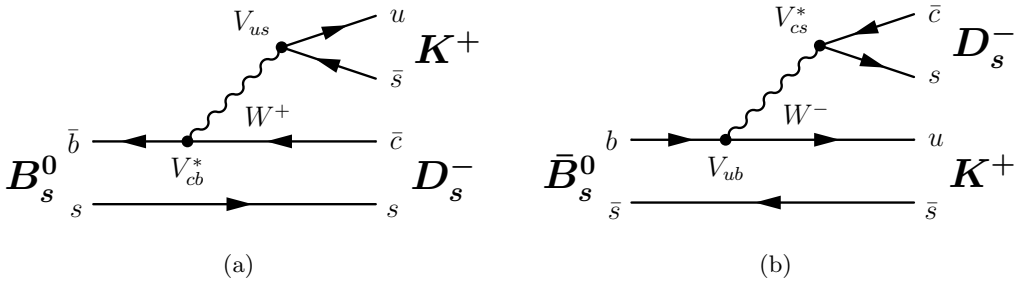


Figure 2.3: Feynman diagrams for the decay $B_s^0 \to D_s^- K^+$ (a) and $\overline{B}_s^0 \to D_s^- K^+$ (b). Both decays are of the same order of magnitude in the Wolfenstein parameter λ: $A\lambda^3$. The decay $\overline{B}_s^0 \to D_s^- K^+$ through the proportionality to V_{ub} introduces a dependence on the CKM angle γ. In the interference of the two decay paths through $B_s^0 - \overline{B}_s^0$ mixing, the weak phase difference $\gamma - 2\beta_s$ can be measured.

$r_B^{D_s K} = \left|A(\overline{B}_s^0 \to D_s^- K^+)/A(B_s^0 \to D_s^- K^+)\right| \approx |V_{ub}V_{cs}^*/V_{cb}^*V_{us}| \approx 0.4$, while for previous time-dependent measurements in the B^0 sector this ratio was an order of magnitude smaller $r_B^{D^*\pi} \approx 0.02$ [66, 67].
The sensitivity to γ in decays such as $B_s^0 \to D_s^- K^+$ arises from the interference of a $B_s^0 \to D_s^- K^+$ and a mixed $B_s^0 \to \overline{B}_s^0 \to D_s^- K^+$ decay. Hence, the *CP* violating observables depend not solely on γ, but on $\gamma - 2\beta_s$, where $\beta_s \equiv \arg(-V_{ts}V_{tb}^*/V_{cs}V_{cb}^*)$ is the mixing angle in the $B_s^0 - \overline{B}_s^0$ system.
The analysis of $B_s^0 \to D_s^\mp K^\pm$ decays can therefore be interpreted in two different ways. Either β_s with an external input for γ can be constrained or vice versa. Both interpretations have appealing outcomes. With the tree-level measurement of β_s with an external input of γ it is possible to estimate the loop (or penguin) contributions $\Delta_{\phi_s}^P$ to $\phi_s = -2\beta_s + \Delta_{\phi_s}^P$ the *CP* violating phase in $B_s^0 \to J/\psi K^+K^-$ decays [68]. This directly constrains contributions from physics beyond the SM.
On the other hand improved accuracy on the CKM angle γ from tree-level serves as an important baseline measurement in order to test the CKM mechanism of the SM. The LHCb collaboration measured ϕ_s to be small $\phi_s = 0.010 \pm 0.039$ rad [68, 69], which corresponds to an uncertainty of $\approx 2^\circ$, the uncertainties on the world averaged for the CKM angle γ from direct measurements is $\mathcal{O}(8^\circ)$ with an absolute value of approximately 70° [41,42]. Given that a weak phase difference is measured for the first time in $B_s^0 \to D_s^\mp K^\pm$ decays and, therefore, the expected uncertainty on $\gamma - 2\beta_s$ is large compared to the world average, $-2\beta_s$ is substituted with ϕ_s and constrained to the LHCb measurement [69] to extract the CKM angle γ.

Notation and $B_s^0 \to D_s^\mp K^\pm$ decay rates

In order to measure the CKM angle γ the decay time distribution for the $B_s^0 \to D_s^\mp K^\pm$ decays has to be known. Therefore, the theoretical formulae are given in Eq. 2.14. The following notation is adopted from [66]. A general more formal derivation of the formulae can be found in e.g. Chapter 9 of Ref. [64] with a slightly different notation. The B_s^0 meson mass eigenstates are written as

$$|B_{s,L}\rangle = p|B_s^0\rangle + q|\overline{B}_s^0\rangle, \tag{2.12}$$

$$|B_{s,H}\rangle = p|B_s^0\rangle - q|\overline{B}_s^0\rangle, \tag{2.13}$$

with $|p|^2 + |q|^2 = 1$. The mass difference $\Delta m_d = m_H - m_L$ and the decay width difference $\Delta\Gamma_s = \Gamma_{s,L} - \Gamma_{s,H}$ of the heavy (B_H) and light (B_L) B_s^0 mass eigenstates are defined to be positive throughout this thesis, following the definition in Ref. [69]. The decay amplitude for a B_s^0 meson to decay into a $D_s^- K^+$ final state is defined as A_f. The respective conjugated decays are denoted $\bar{A}_f$ ($\overline{B}_s^0 \to D_s^- K^+$), $A_{\bar{f}}$ ($B_s^0 \to D_s^+ K^-$), and $\bar{A}_{\bar{f}}$ ($\overline{B}_s^0 \to D_s^+ K^-$). In order to achieve time dependent decay rates that are independent from global phase transformations, all quantities are written in terms of $|q/p|$, $|A_f|$ and $|\bar{A}_f|$, or $|\bar{A}_{\bar{f}}|$ and $|A_{\bar{f}}|$, $\lambda_f \equiv (q/p)(\bar{A}_f/A_f)$, and $\lambda_{\bar{f}} \equiv (q/p)(\bar{A}_{\bar{f}}/A_{\bar{f}})$. Here, the parameters λ_f and $\lambda_{\bar{f}}$ hold the dependence on the weak phase and the strong phase differences $\gamma + \phi_s$ and $\delta_B^{D_s K}$, respectively. The four time dependent decay rates can be written in terms of the parameters that are invariant under global phase transformations as

$$\begin{aligned}\frac{d\Gamma_{B_s^0 \to f}(t)}{dt} = e^{-\Gamma_s t}\frac{1}{2}|A_f|^2(1+|\lambda_f|^2)\Bigg[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \mathcal{A}_f^{\Delta\Gamma}\sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ + C_f\cos(\Delta m_s t) - S_f\sin(\Delta m_s t)\Bigg],\end{aligned} \tag{2.14}$$

$$\begin{aligned}\frac{d\Gamma_{\bar{B}_s^0 \to f}(t)}{dt} = e^{-\Gamma_s t}\frac{1}{2}|A_f|^2\left|\frac{p}{q}\right|^2(1+|\lambda_f|^2)\Bigg[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \mathcal{A}_f^{\Delta\Gamma}\sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ - C_f\cos(\Delta m_s t) + S_f\sin(\Delta m_s t)\Bigg],\end{aligned} \tag{2.15}$$

$$\begin{aligned}\frac{d\Gamma_{\bar{B}_s^0 \to \bar{f}}(t)}{dt} = e^{-\Gamma_s t}\frac{1}{2}|\bar{A}_{\bar{f}}|^2(1+|\bar{\lambda}_{\bar{f}}|^2)\Bigg[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \mathcal{A}_{\bar{f}}^{\Delta\Gamma}\sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ + C_{\bar{f}}\cos(\Delta m_s t) - S_{\bar{f}}\sin(\Delta m_s t)\Bigg],\end{aligned} \tag{2.16}$$

$$\begin{aligned}\frac{d\Gamma_{B_s^0 \to \bar{f}}(t)}{dt} = e^{-\Gamma_s t}\frac{1}{2}|\bar{A}_{\bar{f}}|^2\left|\frac{q}{p}\right|^2(1+|\bar{\lambda}_{\bar{f}}|^2)\Bigg[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \mathcal{A}_{\bar{f}}^{\Delta\Gamma}\sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \\ - C_{\bar{f}}\cos(\Delta m_s t) + S_{\bar{f}}\sin(\Delta m_s t)\Bigg],\end{aligned} \tag{2.17}$$

where $\bar{\lambda}_{\bar{f}} = 1/\lambda_{\bar{f}}$ and the dependence on the weak and strong phase differences is in the *CP* asymmetry observables C_f, $C_{\bar{f}}$, S_f, $S_{\bar{f}}$, $\mathcal{A}_f^{\Delta\Gamma}$, and $\mathcal{A}_{\bar{f}}^{\Delta\Gamma}$. These are given as

$$\begin{aligned} C_f &= \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2} = -C_{\bar{f}} = -\frac{1-|\lambda_{\bar{f}}|^2}{1+|\lambda_{\bar{f}}|^2}\,, \\ \mathcal{A}_f^{\Delta\Gamma} &= \frac{-2\mathcal{R}e(\lambda_f)}{1+|\lambda_f|^2}\,, \quad \mathcal{A}_{\bar{f}}^{\Delta\Gamma} = \frac{-2\mathcal{R}e(\lambda_{\bar{f}})}{1+|\lambda_{\bar{f}}|^2}\,, \\ S_f &= \frac{2\mathcal{I}m(\lambda_f)}{1+|\lambda_f|^2}\,, \quad S_{\bar{f}} = \frac{2\mathcal{I}m(\lambda_{\bar{f}})}{1+|\lambda_{\bar{f}}|^2}\,, \end{aligned} \tag{2.18}$$

where it can be seen that the *CP* observables satisfy

$$C_f{}^2 + S_f{}^2 + \mathcal{A}_f^{\Delta\Gamma 2} = \frac{1}{(1+|\lambda_f|^2)^2}\left((1-|\lambda_f|^2)^2 + 4|\lambda_f|^2\right) = 1\,. \tag{2.19}$$

The equality of $C_f = -C_{\bar{f}}$ in Eq. 2.18 comes from the relation $\left|\lambda_{\bar{f}}\right| = 1/\left|\lambda_f\right|$. This holds when $|q/p| = 1$, $|A_f| = \left|\bar{A}_{\bar{f}}\right|$, and $\left|A_{\bar{f}}\right| = \left|\bar{A}_f\right|$, i.e. the pure B_s^0–$\overline{B}_s^0$ mixing and the pure decay processes are *CP* conserving. The *CP* asymmetry observables in Eq. 2.18 depend on the physical parameters $\gamma + \phi_s$, $\delta_B^{D_sK}$, and the interfering amplitude ratio, $r_B^{D_sK}$ as follows:

$$\begin{aligned} C_f &= \frac{1-\left(r_B^{D_sK}\right)^2}{1+\left(r_B^{D_sK}\right)^2}\,, \\ \mathcal{A}_f^{\Delta\Gamma} &= \frac{-2r_B^{D_sK}\cos(\delta_B^{D_sK}-(\gamma+\phi_s))}{1+\left(r_B^{D_sK}\right)^2}\,, \quad \mathcal{A}_{\bar{f}}^{\Delta\Gamma} = \frac{-2r_B^{D_sK}\cos(\delta_B^{D_sK}+(\gamma+\phi_s))}{1+\left(r_B^{D_sK}\right)^2}\,, \\ S_f &= \frac{2r_B^{D_sK}\sin(\delta_B^{D_sK}-(\gamma+\phi_s))}{1+\left(r_B^{D_sK}\right)^2}\,, \quad S_{\bar{f}} = \frac{-2r_B^{D_sK}\sin(\delta_B^{D_sK}+(\gamma+\phi_s))}{1+\left(r_B^{D_sK}\right)^2}\,. \end{aligned} \tag{2.20}$$

As the weak phase difference arises from the complex couplings introduced by the CKM matrix, the sign changes under *CP* transformations. This allows to separately measure the weak and strong phase differences from the *CP* observables $\mathcal{A}_f^{\Delta\Gamma}$, $\mathcal{A}_{\bar{f}}^{\Delta\Gamma}$, S_f, and $S_{\bar{f}}$.

2.3.2 The rare $B \to \mu^+\mu^-$ decays

The wide interest in $B \to \mu^+\mu^-$ decays does not solely arise from the fact that FCNCs are suppressed in the SM. In addition, the branching fractions of $B \to \mu^+\mu^-$ decays can precisely be calculated from theory, as $B \to \mu^+\mu^-$ processes are one of the theoretically cleanest rare B decays [70]. The $B \to \mu^+\mu^-$ decays are dominated by the Z^0 penguin or loop diagrams in Fig. 2.2(a) and Fig. 2.2(b), and the box diagram in Fig. 2.2(c). In principle, contributions of the scalar Higgs boson must be included in the calculation of the branching fractions. However, due to the coupling of the Higgs boson being proportional to the fermion masses, Higgs diagrams are further suppressed compared to the Z^0 decays. Hence, scalar contributions are negligible in the SM. For $B \to \mu^+\mu^-$ decays Eq. 2.9 reduces to

$$\mathcal{A}(B \to \mu^+\mu^-) = -\frac{\mathrm{G}_F}{\sqrt{2}} \frac{\alpha}{\pi \sin^2 \theta_w} V^*_{td(s)} V_{tb} c_{10} \times Q_{10} + h.c.\,, \tag{2.21}$$

where

$$Q_{10} = (\bar{q}_L \gamma^\rho b_L)(\bar{l}\gamma_\rho \gamma_5 l) \tag{2.22}$$

with α being the fine-structure constant, θ_w the weak mixing angle, γ^ρ the Dirac matrices, and $\gamma_5 = \mathrm{i}\gamma_0\gamma_1\gamma_2\gamma_3$. Using the Eq. 2.21 to calculate the branching fractions of $B \to \mu^+\mu^-$ decays, it is crucial to know the Wilson coefficient c_{10}. For FCNC processes, the coefficients c_i can be written as linear combinations of CKM factors and basic functions $\mathcal{F}(x_i)$ [71], where $x_i = m_i^2/m_W^2$. The basic functions $\mathcal{F}(x_i)$ are called *Inami-Lim* functions for a set of effective vertices. For FCNC processes involving a B_q decay, with $q = s,d$ in case of a B_s^0 or B^0 meson, the c_i coefficients are proportional to

$$c_i \propto \sum_{j=u,c,t} V^*_{jb} V_{jq} \times \mathcal{F}_i(x_j)\,. \tag{2.23}$$

Due to unitarity of the CKM matrix the sum of CKM elements in Eq. 2.23 cancels if $\mathcal{F}_i(x_j) = \text{const.}$ (GIM cancellation). However, as the $\mathcal{F}_i(x_j)$ depend on the quark masses, the GIM cancellation is not exact on quark-loop level. For $B_q \to \mu^+\mu^-$ decays Eq. 2.21 in the SM becomes [72]

$$\mathcal{B}(B_q \to \mu^+\mu^-)_{\mathrm{SM}} = \frac{\mathrm{G}_F}{\pi} \left(\frac{\alpha}{4\pi \sin^2 \theta_w} \right)^2 \tau_{B_q} F^2_{B_q} m_{B_q} m^2_\mu \sqrt{1 - \frac{4m^2_\mu}{m^2_{B_q}}} \left| V^*_{tb} V_{tq} \right|^2 \times \mathcal{Y}^2(x_t)\,. \tag{2.24}$$

Here, τ_{B_q} is the lifetime of the B_q meson, $F^2_{B_q}$ is the hadronic decay constant, and $\mathcal{Y}(x_t)$ the combined Z^0 penguin and $W^\pm$ box Inami-Lim function. The m^2_μ term marks another suppression mechanism present in $B \to \mu^+\mu^-$ decays. Due to the pseudo-scalar nature of B mesons and angular momentum conservation, the dimuon

system needs to be in a spin zero state [73]. This can only be achieved when one of the muons has the incorrect helicity, e.g. the anti-muon is left-handed. Given the latest calculations for the hadronic decay constant [74–76], the latest combined top mass estimate [77], the decay-time integrated SM branching fractions including next-to-leading order electroweak and next-to-next-to-leading order QCD corrections are [78]

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.66 \pm 0.23) \times 10^{-9} \,, \quad \text{and} \tag{2.25}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10} \,, \tag{2.26}$$

where the leading contribution to the uncertainties are coming the CKM elements and the decay constants.

$B \to \mu^+\mu^-$ decays beyond the Standard Model

Another important point why $B \to \mu^+\mu^-$ decays are an excellent probe for the validity of the SM is the dependence of the branching fractions on potential effects of physics beyond the SM. The $B \to \mu^+\mu^-$ decays are especially sensitive to new models that modify scalar and pseudoscalar couplings or have an enhanced (pseudo-)scalar sector. The prediction for the branching fractions of $B_s^0 \to \mu^+\mu^-$ decays in a generic beyond SM physics model can be written as [79]

$$\frac{\mathcal{B}(B_s^0 \to \mu^+\mu^-)}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{\mathrm{SM}}} = |S|^2 \left(1 - \frac{4m_\mu^2}{m_{B_s^0}^2}\right) + |P|^2 \,, \tag{2.27}$$

where

$$S = \frac{m_{B_s^0}^2}{2m_\mu} \frac{(C_S - C_S')}{\left|C_{10}^{\mathrm{SM}}\right|} \,, \quad \text{and} \quad P = \frac{m_{B_s^0}^2}{2m_\mu} \frac{(C_P - C_P')}{C_{10}^{\mathrm{SM}}} + \frac{(C_{10} - C_{10}')}{C_{10}^{\mathrm{SM}}} \,. \tag{2.28}$$

Here, $C_S^{(\prime)}$ and $C_P^{(\prime)}$ are the Wilson coefficients corresponding to scalar and pseudoscalar contributions, which are either neglected (due to Higgs couplings) or absent in the SM. The coefficients C' are the chirality flipped operators. In the SM limit this translates to $S \to 0$ and $P \to 1$. Furthermore, the parametrisation in Eq. 2.27 clearly shows the sensitivity to an enhanced (pseudo-)scalar sector. This makes $B \to \mu^+\mu^-$ decays an excellent probe for physics models beyond the SM that generate quark masses from two different Higgs fields, e.g. the *two Higgs doublets type II* (2HDM-II) theories or Supersymmetry. It can be shown (e.g. in Ref. [80]) that the decay amplitude can be proportional to $\tan^3\beta$, where $\tan\beta$ is governed by the ratio of the two Higgs vacuum expectation values.
Certain physics models beyond the SM are grouped into a class of *minimal flavour violating* (MFV) models [81, 82]. In these theories all flavour changing and *CP* violating effects are based on the CKM matrix. However, the Wilson coefficients can

be altered by new particles. For these types of models the ratio of branching fractions of $B \to \mu^+\mu^-$ decays provides a strict observable [83]

$$\mathcal{R} \equiv \frac{\mathcal{B}(B^0 \to \mu^+\mu^-)}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)} = \frac{\tau(B^0)}{\tau(B_s^0)} \frac{m_{B^0}}{m_{B_s^0}} \frac{F_{B^0}^2}{F_{B_s^0}^2} \frac{|V_{td}|^2}{|V_{ts}|^2} , \tag{2.29}$$

which is predicted in the SM as [84]

$$\mathcal{R}_{\text{SM}} = \left(2.95^{+0.28}_{-0.25}\right) \times 10^{-2} . \tag{2.30}$$

Experimental status of $B \to \mu^+\mu^-$

The search for $B \to \mu^+\mu^-$ decays has a long history. Starting back in 1984 with a search performed by the CLEO collaboration [85] setting a limit on $\mathcal{B}(B^0 \to \mu^+\mu^-) < 0.02\%$. Over 30 years numerous experiments measured upper limits for the branching fractions of $B \to \mu^+\mu^-$ decays until the LHCb collaboration found the first evidence for $B_s^0 \to \mu^+\mu^-$ decays in 2012 [86] measuring $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9}$ with a significance of 3.5σ. The history of the searches is illustrated in Fig. 2.4. Prior to

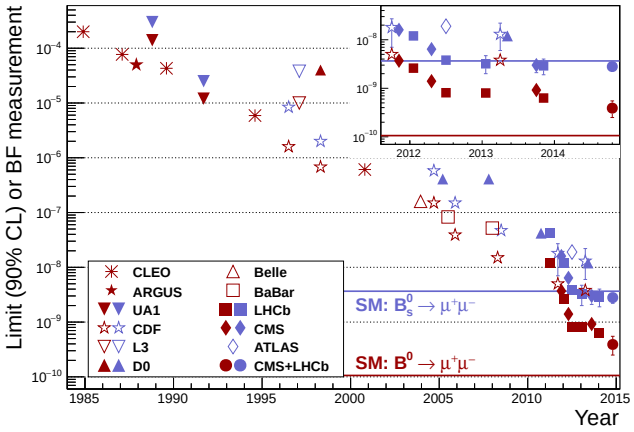


Figure 2.4: Searches for $B_s^0 \to \mu^+\mu^-$ (red) and $B^0 \to \mu^+\mu^-$ (blue) decays over a span of 30 years. The markers represent the various experiments. When a marker is shown without an error bar the measured result is a 90% upper limit, while a measured branching fraction is shown with the corresponding 68% confidence interval. The shown measurements are from Refs. [85–116]. Figure taken from [84].

the LHC era the limits for the $B \to \mu^+\mu^-$ decays were about an order of magnitude over the SM predictions. The LHC experiments LHCb and CMS were able to reach the SM prediction within two years of data-taking. This theses describes the updated LHCb analysis based on the full RunI dataset and the joint analysis of $B \to \mu^+\mu^-$ decays based on the combined LHCb and CMS datasets.

3 Statistical methods

In this thesis statistical methods are used to measure physics parameters out of ensembles of measurements and to combine results to achieve the highest possible precision. Hence, a major task is the implementation, testing, and application of the statistical methods. Thus, a short introduction to the methods used in the physics analyses in the Chapters 5 to 8 is given. It is, however, beyond the scope of this thesis to give a full introduction into statistics. A more fundamental and detailed introduction to statistical methods in high energy physics can be found e.g. in Refs. [117–120].
The main purpose of the statistical methods within the presented work is to evaluate confidence intervals and signal significances. Therefore, the concept of confidence intervals is introduced in Section 3.1 and different ways of constructing confidence intervals are reviewed in Sections 3.2 and 3.3. The definition and methods to calculate the significance of a signal are presented in Section 3.4.

3.1 Confidence intervals

A confidence interval is a generalised way to display the level of uncertainty. Consider a gaussian random variable x distributed as

$$g(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} , \tag{3.1}$$

where μ is the mean and $\sigma^2 \equiv V$ ($\sigma = \sqrt{V}$) the variance (standard deviation) of the underlying probability density distribution (PDF). Often, it is the goal of a physics measurement to estimate a parameter of the underlying distribution, e.g. the mean parameter μ. The relevant uncertainty is the variance of the distribution of the best estimate $\hat{\mu}$. For the given example it can be shown (e.g. in Chapter 5, Ref. [118]) that the variance can be expressed by $\sigma^2_{\hat{\mu}} = \sigma^2/N$ or in terms of the standard deviation $\sigma_{\hat{\mu}} = \sigma/\sqrt{N}$, where N is the number of measurements. Due to the central limit theorem (cf. Appendix A.1) this statement holds also for more general cases so that in the large sample limit ($N \to \infty$) $\sigma_{\hat{\mu}} \propto \frac{1}{\sqrt{N}}$.

The classical interpretation of the result $\mu = \hat{\mu} \pm \sigma_{\hat{\mu}}$ is obvious: if the estimation of $\hat{\mu}$ is repeated under same conditions (e.g. $N = const$), then $\sigma_{\hat{\mu}}$ is the standard deviation

of the resulting $\hat{\mu}$ distribution. Or equivalently, for a sufficiently large number of estimations, the interval $\mathcal{I}^i = [\hat{\mu}^i - \sigma_{\hat{\mu}}, \hat{\mu}^i + \sigma_{\hat{\mu}}] = [\mu_1, \mu_2]$ covers the true value μ_t in 68.27% of the times. Note, that this is not a probability statement about the true value, as — in classical statistics — the true value is fixed and unknown. There is no concept of a probability distribution for the true value. Instead, the statement is about the *coverage* of a set of confidence intervals. An illustration is given in Fig. 3.1, where the mean of the *known* underlying gaussian distribution from Eq. 3.1 with $\mu_t = 0$ and $\sigma_t = 1$ is estimated from a set of nine independent measurements. The estimation of the mean (and $\sigma_{\hat{\mu}} = 1/\sqrt{9}$) is repeated $\log_{10}(n) \in \{1,2,3,4\}$ times. It can be seen that the confidence intervals are subjects of statistical fluctuations and that for a large number of estimations the interpretation of the standard deviation of the $\hat{\mu}$ distribution and the coverage probability of a confidence interval with a coverage probability of 68.27% are equivalent[1]. In general non-gaussian cases, the standard deviation is still a measure for the spread of the distribution, but the interpretation depends on the distribution. This can easily be seen with a Poisson distribution $p(k|\lambda) = \lambda^k/k! \exp(-\lambda)$ with mean λ and standard deviation $\sqrt{\lambda}$. If a single event is observed this yields $\lambda = 1 \pm 1$. However the probability integral of this interval is 91.97%. Therefore, the more universal form of interpreting an uncertainty is the definition based on coverage.

3.2 Construction of confidence intervals

In the previous example to illustrate the properties of a confidence interval the parameters of the underlying PDF are fixed and therefore exactly known. Now, using the gaussian example from Eq. 3.1, the true mean μ_t is unknown (but the width is still fixed $\sigma = 1$) and the task is to construct a valid confidence interval $[\mu_1, \mu_2]$ for all possible measurements x of the experiment so that

$$\mathcal{P}(\mu \in [\mu_1, \mu_2]) = \alpha \tag{3.2}$$

holds for all possible μ. Here, $\mathcal{P}(\mu \in [\mu_1, \mu_2])$ denotes the probability[2] for μ to be in the interval $[\mu_1, \mu_2]$ and α is the confidence level (CL). The classical method to construct such intervals is called *Neyman construction* based on the solution of J. Neyman [122]. Here, for each *fixed* value of μ the probability to measure x given the value for μ, $\mathcal{P}(x|\mu)$, is examined to construct an interval so that

$$\mathcal{P}(x \in [x_1, x_2]|\mu) = \alpha \,. \tag{3.3}$$

[1] Indeed, this is true for a single confidence interval, but a probability statement cannot be illustrated with a single estimation.

[2] This notation follows Ref. [121].

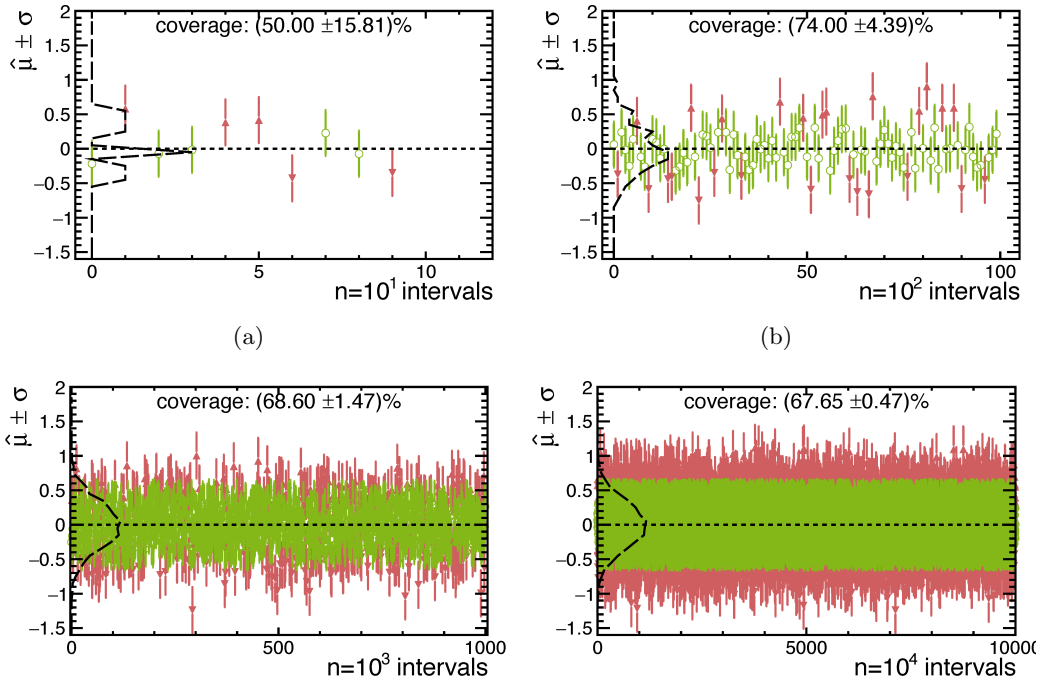


Figure 3.1: Results of a series of confidence interval estimations for the mean of a standard gaussian distribution. The single estimates are based on a measured set of $N = 9$ independent variables x, hence the standard deviation is $1/\sqrt{9}$. The results are shown for $\log_{10}(n) \in \{1, 2, 3, 4\}$ in Fig. 3.1(a), 3.1(b), 3.1(c), and 3.1(d), respectively. All intervals covering the true value of $\mu_t = 0$ (black dashed line) are shown in green, while all intervals that fail to cover the true value are shown in red. The resulting distribution of $\hat{\mu}$ values is indicated as a long dashed black line.

However, there is an implicit freedom on how the actual x are chosen to construct the interval $\mathcal{I}_x = [x_1, x_2]$. As long as Eq. 3.3 is satisfied an arbitrary x can be accepted by $\mathcal{I}_x$. The most common choices are

$$\mathcal{P}(x < x_1) = \mathcal{P}(x > x_2) = \frac{1-\alpha}{2}\,, \tag{3.4}$$

which constructs a *two-sided* confidence interval, and

$$\mathcal{P}(x < x_l) = 1 - \alpha\,, \tag{3.5}$$

which corresponds to a *one-sided* confidence interval or — in this case — an upper limit. The conditions in Eq. 3.4 and Eq. 3.5 are two possibilities for the *ordering* principle, which reflects the freedom on how to construct the confidence level. An

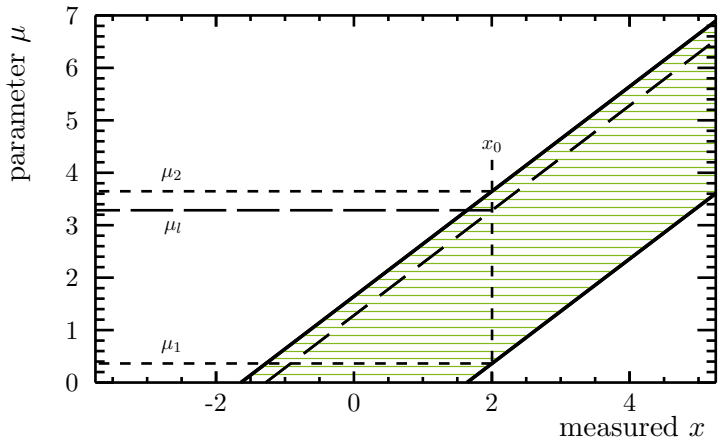


Figure 3.2: Result of a Neyman construction for a restricted mean parameter ($\mu \geq 0$) of a gaussian PDF with fixed width $\sigma = 1$. The 90% CL two-sided confidence belt boundaries are the solid lines while the one-sided upper limit is shown as a dashed line. For a measurement of x_0, the resulting confidence interval and the upper limit for the physics parameter μ can be read off on the μ-axis as $[\mu_1, \mu_2]$ for the two-sided interval and $\mu < \mu_l$ for the upper limit as the intersections with a vertical line drawn at $x = x_0$.

example of this classical construction for a gaussian PDF with fixed width parameter σ is shown in Fig. 3.2. The confidence intervals are constructed horizontally for each possible fixed physics parameter value μ, the interval $\mathcal{I}_x$ is filled with values of x until either Eq. 3.4 or Eq. 3.5 is true. When the measurement is performed and yields a particular result x_0, the confidence interval for the parameter μ is the union of all μ values where the corresponding confidence intervals intersect a vertical line drawn at x_0 or simply $\mu_1(x_0)$ and $\mu_2(x_0)$. In the example of Fig. 3.2 the parameter μ is restricted to positive values (e.g. a measurement of a branching fraction, mass, or decay time), thus confidence intervals do not exist for $\mu < 0$. However, the measurement can observe negative x. In fact, one drawback of the classical construction is that for negative values of $x < -1.64$ ($x < -1.28$) the two-sided (one-sided) 90% confidence interval becomes the empty set. This is a possible correct yet unfortunate result, which indicates that the measurement is part of the 10% of the times where the confidence interval does not cover the true value. One could either allow unphysical values for μ, that implies to know the PDF $\mathcal{P}(x|\mu < 0)$ in the unphysical range, which is conceptionally difficult. Or whenever negative x are observed, zero is taken as the best estimate for μ at the cost of over-coverage (i.e. an unnecessary high upper limit).
Another issue arises with the classical construction if the choice of ordering is made based on the observed data, e.g. a two-sided interval is quoted when the observation is 3σ above zero, while for signals below 3σ an upper limit is published. This is called

flip-flopping in Ref. [121]. The resulting confidence region for the flip-flopping case is shown in Fig. 3.3 for a standard gaussian with fixed width and the mean parameter constrained to positive values. The Neyman construction gets the exact coverage from the fact that $\mathcal{P}(x \in [x_1, x_2]|\mu) = \alpha$ for all fixed μ. However, when the choice of ordering is based on data this condition is not fulfilled for all μ. In case of a standard gaussian distribution with fixed width parameter $\sigma = 1$ the region of $\mu \in [1.36, 4.64]$ has a coverage which is below 90%, e.g. $\mathcal{P}(x \in [x_1, x_2]|\mu \in [1.36, 4.28]) = 85\%$. Hence, the resulting confidence intervals do not guarantee the correct coverage probability.

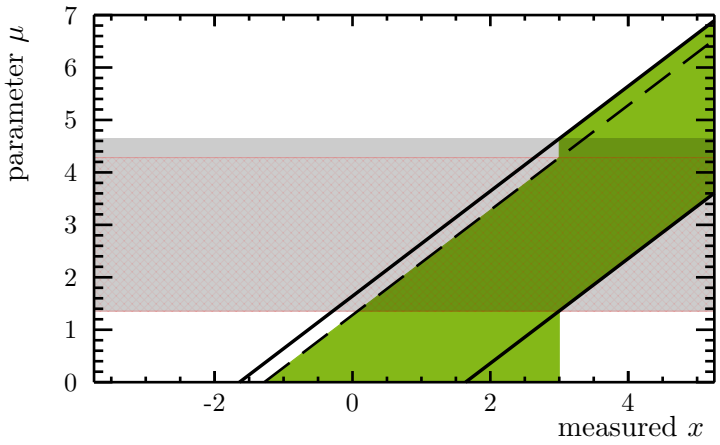


Figure 3.3: Illustration of the effect when the choice of the ordering principle is made based on the observed data. Here, for a standard gaussian example, an upper limit (dashed line) is constructed for all measured values less than 3σ and a two-sided confidence level (solid lines) is calculated for all measurements above 3σ. The resulting confidence region (green) does not provide proper coverage for all μ. The light shaded area has incorrect coverage probability, e.g. for the red hatched it is found: $\mathcal{P}(x \in [x_1, x_2]) = 85\%$.

3.3 The unified frequentist method

A method that solves the issues of empty sets and flip-flopping is the unified frequentist method, based on M. Kendall and A. Stuart [123, 124] and made well-known to the high energy physics community by G. Feldman and R. Cousins [121]. At the heart of the unified frequentist method (also referred to as FC method) is the ordering principle based on a probability ratio:

$$\lambda \equiv \frac{\mathcal{P}(x|\mu)}{\mathcal{P}(x|\hat{\mu})} \,, \tag{3.6}$$

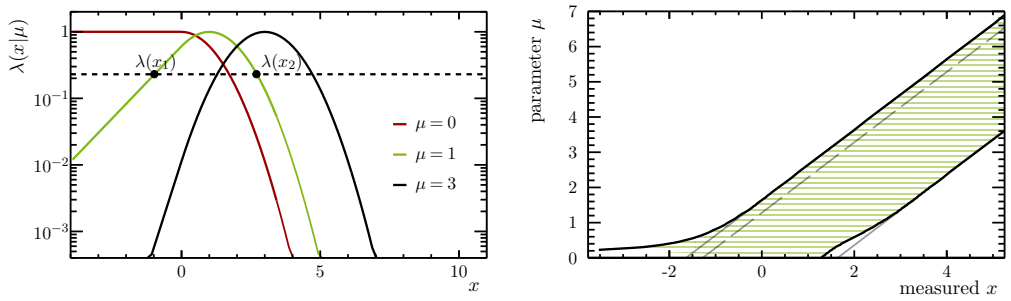
where $\hat{\mu}$ is the best estimate for μ. The motivation for λ will be given in Section 3.3.1. The confidence intervals for a given fixed μ in a full Neyman construction are filled according to decreasing values of $\lambda(x)$ until the coverage condition from Eq. 3.3 is satisfied.
For the example of a standard gaussian with fixed width and constrained mean parameter $\mu \geq 0$ the probability ratio from Eq. 3.6 becomes

$$\lambda(x|\mu) = \begin{cases} \exp\left(x\mu - \frac{1}{2}\mu^2\right) & \text{for } x < 0 \\ \exp\left(-\frac{1}{2}(x-\mu)^2\right) & \text{for } x \geq 0 \,, \end{cases} \tag{3.7}$$

where $\hat{\mu}$ from Eq. 3.6 is replaced by $\hat{\mu} = \max(0, x)$ respecting the constraint of $\mu \geq 0$. The distribution of $\lambda(x|\mu)$ is shown in Fig. 3.4(a) for $\mu = 0$, 1, and 3. A clear asymmetry is visible from the applied boundary leading to a constant function $\lambda(x|\mu = 0) = 1$ for $x < 0$. In this way the default two-sided interval construction automatically turns into a one-sided limit for x and due to the symmetry of the Neyman construction this is also true for μ. The automatic transition from a two-sided into a one-sided interval is the reason for the name *unified* method. The union of all intervals for the mean μ form the confidence belt constructed with the FC ordering principle, which is shown in Fig. 3.4(b). It can be seen that for large observed x, due to the distance to the boundary of $\mu \geq 0$, the confidence intervals are identically to a standard two-sided gaussian intervals. For measurements around $x_0 \approx 0$ the, now one-sided, confidence interval does not drop to the conventional upper-limit, because the conventional interval is only valid when the decision of ordering is not based on the data. So the slightly larger upper-limit is part of avoiding the flip-flopping issue [121].

3.3.1 Practical application of the unified frequentist method

One of the main features of the unified method discussed in Section 3.3 is that the interpretation of the resulting interval is purely determined by the observed



(a) (b)

Figure 3.4: Construction of FC confidence intervals for various gaussian mean parameters μ. (a) FC ordering function λ for $\mu = 0$ (red), $\mu = 1$ (green), and $\mu = 3$ (black). The intersections $\lambda(x_1)$ and $\lambda(x_2)$ with a line of constant λ define the boundaries of the FC confidence interval when Eq. 3.3 is satisfied simultaneously. (b) Comparison of the FC 90% intervals (hatched green area) with the standard Neyman construction of a two- (solid grey) and one-sided (dashed grey) confidence interval.

data. Hence, the interval for a measured parameter can be estimated after the measurement solely from the observed data. Another way to describe the FC procedure is that for a defined confidence level (e.g. $\alpha = 90\%$) a critical value, λ_c, for $\lambda(x|\mu)$ is calculated so that a fraction α of the experiments satisfy $\lambda(x|\mu) > \lambda_c$ or equivalently

$$\mathcal{P}(\lambda(x|\mu) > \lambda_c) = \alpha \,. \tag{3.8}$$

If now the observation is fixed $x = x_0$, the question is how likely a value of μ is, given the observed data[3]. The answer to this question is provided by hypothesis tests[4], which is indeed the motivation for the ordering principle in the unified method. The Neyman-Pearson lemma (see e.g. Refs. [117,118,120]) states that the likelihood ratio λ provides the most powerful test to distinguish between two simple hypotheses. Here, simple hypotheses are those which completely define the probability functions. Now the interpretation of Eq. 3.6 is slightly different and therefore the probability statement is replaced by the likelihood function $\mathcal{L}(\mu|x_0)$. For a measurement with one observable the likelihood function has the same analytic form as the underlying PDF (cf. Eq. 3.7 for $x \geq 0$), but is now a function of μ given the observed data x_0. The likelihood ratio

$$\Delta\chi^2 \equiv -2\ln\lambda = -2\ln\left(\frac{\mathcal{L}(\mu|x_0)}{\mathcal{L}(\hat{\mu}|x_0)}\right) \tag{3.9}$$

[3]Importantly, this is not a statement about the true value μ_t.
[4]For a detailed introduction to hypothesis tests in experimental physics, see e.g. Refs. [117–120].

for a given measured $x = x_0$ is therefore a measure for how likely the hypothesis for μ (referred to as *null* hypothesis) is compared to the alternative $\hat{\mu}$ hypothesis. The notation follows Ref. [121] and is due to the fact that $-2\ln\mathcal{L}$ for gaussian cases and without boundaries becomes $(x-\mu)^2$, which follows a χ^2 distribution with one degree of freedom, as x is a gaussian random variable.
The function $\Delta\chi^2$ is called a test statistic, which is a scalar function of the data and physics parameters. Based on the test statistic, decisions are inferred to either accept or reject the null hypothesis. Therefore, the p-value of the data is calculated from the test statistic. The p-value is the probability to observe a dataset that is at least as extreme as the measured one under the given test statistic and hypothesis. Mathematically this is the integral over the PDF of the test statistic $f(\Delta\chi^2)$ from the observed test statistic value $\Delta\chi^2_{\text{obs}}$ to infinity:

$$\mathrm{p} = \mathcal{P}(\Delta\chi^2 \geq \Delta\chi^2_{\text{obs}}) = \int_{\Delta\chi^2_{\text{obs}}}^{\infty} f\left(\widetilde{\Delta\chi^2}\right) \mathrm{d}\left(\widetilde{\Delta\chi^2}\right) . \tag{3.10}$$

This is the exact inverse of the statement provided by the original FC method in Eq. 3.8. Hence, a specific confidence level with coverage α can be determined by solving $\mathrm{p}(\mu|x_0) = 1-\alpha = 1-\mathrm{CL}$. In practice this is done numerically. For the example of a standard gaussian with fixed mean and an observed value of $x = x_0$, this means that a p-value for many values of μ can be calculated the following way[5]:

1.) Choose a null hypothesis for μ, say $\mu = \mu_i$.

2.) Calculate the observed test statistic value for the given hypothesis $\mu = \mu_i$ as

$$\Delta\chi^2_{\text{obs}} = \chi^2(\mu_i|x_0) - \chi^2(\hat{\mu}|x_0) = -2\ln(\mathcal{L}(\mu_i|x_0)) - (-2\ln\mathcal{L}(\hat{\mu}|x_0)) ,$$

where $\hat{\mu} = \max(0, x_0)$.

3.) Generate a random number x' from the underlying gaussian PDF under the null hypothesis $\mu = \mu_i$, $g(x|\mu_i)$, and calculate the test statistic value for the new pseudo-dataset x' as

$$\Delta\chi^2_{\text{toy}} = \chi^2(\mu_i|x') - \chi^2(\hat{\mu}'|x') = -2\ln(\mathcal{L}(\mu_i|x')) - (-2\ln\mathcal{L}(\hat{\mu}'|x')) ,$$

with $\hat{\mu}' = \max(0, x')$.

4.) Repeat 3.) until the p-value can be calculated with sufficient uncertainty from

$$\mathrm{p} = \frac{N(\Delta\chi^2_{\text{toy}} \geq \Delta\chi^2_{\text{obs}})}{N_{\text{toy}}} \pm \sqrt{\frac{\mathrm{p}(1-\mathrm{p})}{N_{\text{toy}}}} , \tag{3.11}$$

where $N(\Delta\chi^2_{\text{toy}} \geq \Delta\chi^2_{\text{obs}})$ is the number of x' which result in $\Delta\chi^2_{\text{toy}} \geq \Delta\chi^2_{\text{obs}}$ and N_{toy} denotes the total number of random numbers generated in step 3.).

[5] A slightly different formulation based on a different argumentation can be found in Ref. [125].

5.) Repeat 1.) − 4.) for a sufficiently large number of μ_i hypotheses to reach the desired $1-\mathrm{CL}$ confidence interval.

In step 3.) it is necessary to generate random numbers from the underlying PDF. These random numbers are called *pseudo-experiment* or *pseudo-dataset* as they are treated as an actual measurement. Sometimes the prefix *pseudo* is substituted by *toy* to indicate that it is no actual data being measured by a detector. In step 4.) the p-value is calculated by approximating the integral from Eq. 3.10 with the fraction of pseudo-experiments resulting in a test statistic value equal or greater than $\Delta\chi^2_{\mathrm{obs}}$. The uncertainty is the standard error of a binomial distribution. This uncertainty only has the correct coverage in the large sample limit, so that in cases where it is important to prevent under-coverage Clopper-Pearson intervals [126] are used. These intervals use the full binomial PDF to construct intervals.
The result of the procedure 1.) − 5.) is commonly presented in $1-\mathrm{CL}$ plots, which show the $\mathrm{p}=1-\mathrm{CL}$ values in dependence of the null hypotheses μ_i. There are some distinct features of the $1-\mathrm{CL}$ plot: the $1-\mathrm{CL}$ curve reaches the maximum at $\mu_i=\hat{\mu}$. As the maximum likelihood with fixed parameter μ_i can only be as large as the maximum likelihood with the best estimate for μ if $\mu_i=\hat{\mu}$ and thus $\Delta\chi^2_{\mathrm{obs}}=0$. As the test statistic $\Delta\chi^2$ is positive semi-definite, the p-value gets unity for $\mu_i=\hat{\mu}$. The confidence interval with a given confidence level α can be read off by the intersections of a horizontal line at $\mathrm{p}=1-\mathrm{CL}=1-\alpha$.
The outcome of the construction of FC confidence intervals is shown in Fig. 3.5 for four different possible measured values $x_0\in\{-2,0,1.4,4\}$. Each figure shows the distribution of the $\Delta\chi^2$ test statistic for various values of μ where the pseudo-datasets that satisfy $\Delta\chi^2_{\mathrm{toy}}\geq\Delta\chi^2_{\mathrm{obs}}$ are highlighted. The $1-\mathrm{CL}$ plots shown as an insertion are constructed from the integral over the highlighted area for a fixed value of μ and from the pure χ^2 approximation for $f(\Delta\chi^2|\mu)$. It can be seen from the comparison of the four scenarios how the unified frequentist method evolves from a 90% CL upper limit for $x_0=-2$ and $x_0=0$ to a two-sided 90% confidence interval for $x_0=4$. The boundary curve of the acceptance region ($\Delta\chi^2=\Delta\chi^2_{\mathrm{obs}}$) can be taken as a guideline for this phenomenon as this directly translates into a $1-\mathrm{CL}$ curve given the χ^2 approximation. If the χ^2 approximation is true, the confidence intervals corresponding to a coverage probability of standard two-sided gaussian $n\sigma$ intervals is given by $\Delta\chi^2(\mu)=n^2$. For the purely two-sided interval for $x_0=4$ the values for μ resulting in $\Delta\chi^2=1$ correspond to $\mu=4\pm1$, which is the exact gaussian standard error $\sigma_{\hat{\mu}}=\sigma/\sqrt{N}$ with $\sigma=1$ and $N=1$ (cf. Section 3.1). However, the χ^2 approximation only provides correct confidence intervals (in this example), if the observed x_0 is exactly at the physical boundary or the observed x_0 is largely separated from the boundary so that the confidence interval is not affected by the boundary. While the latter case is trivial, the first case is true, because the effect of the boundary is integrated out by calculating the p-value. The distribution $f(\Delta\chi^2,\mu)$ is shown in Fig. 3.6 for $\Delta\chi^2\leq3$ and $0\leq\mu\leq2$. For small values of μ, i.e. $\mu\lesssim1$, deviations from the χ^2 distribution are evident. In turn, this means that for all observations where the acceptance boundary curve ($\Delta\chi^2=\Delta\chi^2_{\mathrm{obs}}$) is close to the

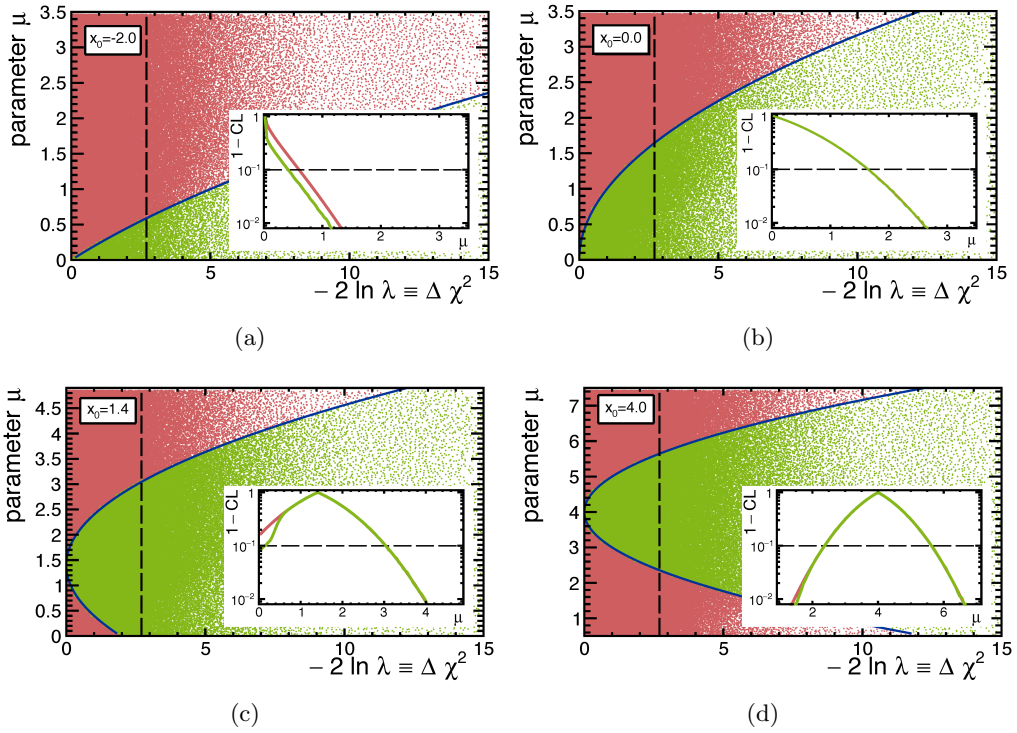


Figure 3.5: Example construction of Feldman-Cousins intervals given a single measurement x_0. The distribution $f(\Delta\chi^2|\mu)$ (points) is based on the test statistic $\Delta\chi^2$ from Eq. 3.7 and 10^5 random gaussian numbers for each mean hypothesis μ_i. The blue curve is the boundary of the acceptance region, $\Delta\chi^2 = \Delta\chi^2_{\text{obs}}$, which marks the lower value for the p-value calculation (Eq. 3.11). The insertion plots show the resulting $1 - \text{CL}$ curves for the χ^2 approximation (red) and the p-values from Eq. 3.11 (green) in dependence of the μ hypotheses.

boundary region the coverage of χ^2 approximated confidence intervals is incorrect. The observed value of $x_0 = 1.4$ in Fig. 3.5(c) is chosen as an example where the correct 90% interval is two-sided while the χ^2 approximation returns a one-sided upper limit. This is a good example where an approximation can lead to a critical change in the interpretation of the measured result.

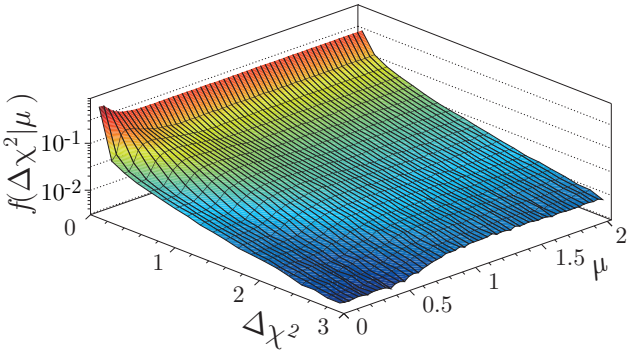


Figure 3.6: The distribution of the test statistic $\Delta\chi^2$, $f(\Delta\chi^2|\mu)$, in dependence of the parameter μ. For large enough parameter values $\mu \gtrsim 1$, $f(\Delta\chi^2|\mu)$ follows approximately a χ^2 distribution with one degree of freedom, while for parameter values close to the boundary of $\mu \geq 0$ a clear deviation from the χ^2 distribution is evident. Thus, the coverage probability is wrong for the χ^2 approximation close to a strict boundary.

3.3.2 The unified frequentist method with a biased estimator

For finite sample sizes the likelihood estimator is expected to be biased, i.e. the expectation value of the estimator is not the true value ($\langle\hat{\mu}\rangle \neq \mu_t$). The bias is often evaluated using pseudo-experiments. If such a bias is present on the likelihood estimator, the confidence intervals from the FC method are widened to maintain the correct coverage. This could in principle be interpreted as a systematic uncertainty folded into the confidence interval calculation. The effect is illustrated in Fig. 3.7 where intervals from a measurement of $x_0 = 1.5$ and $x_0 = 2.0$ are shown and compared to an estimator that has a bias of $x_0 + b = 1.0 + 0.5$ and $x_0 + b = 1.0 + 1.0$, where b is the bias. If the estimator is biased, the null hypotheses do not correspond to the best estimates and hence the sampling distribution becomes a non-central χ^2 distribution. The resulting confidence intervals are broadened to ensure correct coverage.

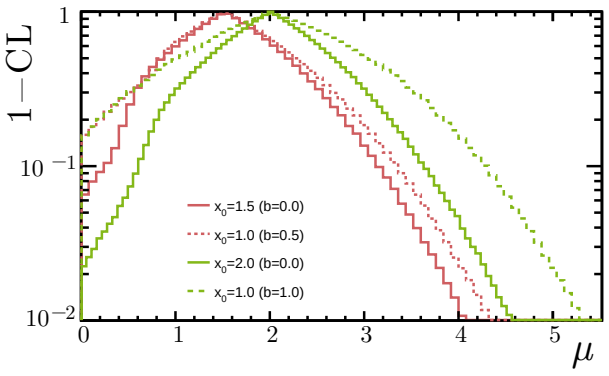


Figure 3.7: Feldman-Cousins intervals for a biased likelihood estimator (dotted lines). Biases of $b = 0.5$ and $b = 1.0$ are tested along with the corresponding unbiased measurements $x_0 = 1.5$ and $x_0 = 2.0$ (solid lines). The confidence intervals from the biased measurements are clearly widened.

3.3.3 The unified frequentist method with nuisance parameters

In order to clarify that the following discussion is independent of the gaussian example a slight change in notation is useful. In general, the parameter of interest (POI) is not the only parameter which is unknown and on which the underlying PDF depends. In fact, the PDF can depend on an r-dimensional set of parameters of interest $\vartheta \equiv \vec{\vartheta} = (\vartheta_1, \ldots, \vartheta_r)$ and an s-dimensional set of *nuisance* parameters $\theta \equiv \vec{\theta} = (\theta_1, \ldots, \theta_s)$ that need to be determined, but are not of prime interest from the physics point of view. These nuisance parameters can be *constrained* from e.g. external measurements. It is furthermore possible to estimate the same physics parameters in different channels. This could be separate particle decays or subsets of the data, which need a different probability description. The total likelihood function can then be written as[6]

$$\mathcal{L}(\vartheta, \theta | \mathbb{X}_c, X_q) = \prod_{q \in \mathbb{G}} g_q(X_q|\theta_q) \times \prod_c \left[P(N_c|\nu_c) \prod_i^{N_c} f_c(x_{c,i}|\vartheta, \theta) \right], \tag{3.12}$$

where the likelihood function now is a function of the POIs and the nuisance parameters, given the measured datasets in the categories c, $\mathbb{X}_c = (x_{c,1}; \ldots, x_{c,i}; \ldots; x_{c,N_c})$, and a set of external measurements X_q. The term $P(N_c|\nu_c)$ marks the poisson probability to observe N_c events in each category c, when ν_c events are expected. The constraints are incorporated by additional terms in the likelihood denoted $g_q(X_q|\theta_q)$, as usually these constraints are gaussian PDFs for the measurement X_q, given the nuisance parameter θ_q. The index q is from a subset of indices $\mathbb{G}$ that includes all constrained

[6] Modified notation from [127].

nuisance parameters. The terminology for estimating the most probable values $\hat{\vartheta}$ or *best fit* value is to perform a simultaneous extended maximum likelihood fit. In case of a single category, the "simultaneous" and in absence of a poisson probability term the "extended" is dropped. In practice, the PDFs are built with RooFit [128] implemented in the Root [129] framework. Furthermore instead of maximising the likelihood from Eq. 3.12, $-\ln\mathcal{L}(\vartheta,\theta|\mathbb{X}_c,X_q)$ is minimised using the programme Minuit [130, 131], because sums can be derived better and are numerically more stable than products.

The full unified frequentist method relies on the power of the Neyman construction to ensure the correct coverage for all possible values of ϑ. This directly implies to construct an $r+s$-dimensional confidence belt or to sample an $r+s$-dimensional parameter space. For more than a couple of parameters, this is computationally not feasible and therefore a treatment for the nuisance parameters is necessary.
In the construction of a $1-\mathrm{CL}$ curve, p-values are evaluated for different hypothesised values for the POIs (cf. Section 3.3.1). In the general case, the question arises how to define the p-value when only the parameters of interest (ϑ) are specified in a given hypothesis. When it is not possible to quote the p-value as a function of the nuisance parameters, the supremum p-value over the nuisance parameters space needs to be found.

3.3.4 The plugin method

A way to handle the nuisance parameters was suggested by M. Kendall and A. Stuart [123] and uses a special choice of nuisance parameters for each null hypothesis. The test statistic can be written as

$$\Delta\chi^2(\vartheta) = -2\ln\lambda = -2\ln\frac{\mathcal{L}(\vartheta,\hat{\hat{\theta}}(\vartheta)|\mathbb{X}_c,X_q)}{\mathcal{L}(\hat{\vartheta},\hat{\theta}|\mathbb{X}_c,X_q)}\,, \tag{3.13}$$

where $\hat{\vartheta}$ and $\hat{\theta}$ are the parameters that maximise the likelihood unconditionally, and $\hat{\hat{\theta}}(\vartheta)$ is the set of nuisance parameters that maximise the likelihood under the given fixed ϑ hypothesis. The test statistic is supposed to approximate the supremum as large p-values indicate a better agreement between the dataset and the hypothesis. The set $\hat{\hat{\theta}}(\vartheta)$ provides the highest probability for the dataset to come from the underlying model. The ratio in Eq. 3.13 is also called *profile likelihood ratio*, as the nuisance parameters are *profiled* for each ϑ hypothesis. The method sometimes is also called plugin method [132], because the nuisance parameters are *plugged-in* at the respective best fit values under the POI hypothesis.
The recipe to construct confidence intervals only differs slightly from Section 3.3.1, however, as the notation and techniques have changed and the plugin method is the nominal method when statistical interpretations are made in this thesis, it is

explicitly recalled. The notation of Eq. 3.12 and Eq. 3.13 is used. Let $\mathbb{X}_c^0$ be the nominal measured datasets and X_q^0 the nominal set of external measurements, then the procedure to follow is:

0.) Perform the nominal maximum likelihood fit to the dataset to get $\mathcal{L}(\hat{\vartheta},\hat{\theta}|\mathbb{X}_c^0,X_q^0)$.

1.) Choose a null hypothesis for ϑ, say $\vartheta = \vartheta_i$.

2.) Perform a maximum likelihood fit with fixed POI ϑ_i to find $\mathcal{L}(\vartheta_i,\hat{\hat{\theta}}(\vartheta_i)|\mathbb{X}_c^0,X_q^0)$.

3.) Calculate the observed test statistic value, $\Delta\chi^2_{\text{obs}}$ for the given hypothesis $\vartheta = \vartheta_i$ according to Eq. 3.13.

4.) Generate a pseudo-dataset $\mathbb{X}_c'$ for each category c from the underlying PDFs under the null hypothesis with plugged-in nuisance parameters $f_c(\mathbb{X}_c|\vartheta_i,\hat{\hat{\theta}}(\vartheta_i))$. Here, the event yield fluctuates according to the poisson term in Eq. 3.12.

5.) Generate a pseudo-dataset X_q' from $g_q(X_q|\hat{\hat{\theta}}_q(\vartheta_i))$.

6.) Perform a maximum likelihood fit with floating POIs and nuisance parameters to find $\mathcal{L}(\hat{\vartheta},\hat{\theta}(\hat{\vartheta})|\mathbb{X}_c',X_q')$. Here, any boundary condition can be applied, e.g. $\vartheta \geq 0$.

7.) Perform a maximum likelihood fit with fixed POIs ϑ_i and floating nuisance parameters to find $\mathcal{L}(\vartheta_i,\hat{\hat{\theta}}(\vartheta_i)|\mathbb{X}_c',X_q')$.

8.) Calculate the test statistic value for the new pseudo-datasets $\mathbb{X}_c'$ and X_q', $\Delta\chi^2_{\text{toy}}$.

9.) Repeat 4.)–8.) until the p-value can be calculated with sufficient uncertainty from

$$\mathrm{p} = \frac{N(\Delta\chi^2_{\text{toy}} \geq \Delta\chi^2_{\text{obs}})}{N_{\text{toy}}} \pm \sqrt{\frac{\mathrm{p}(1-\mathrm{p})}{N_{\text{toy}}}}\,.$$

where $N(\Delta\chi^2_{\text{toy}} \geq \Delta\chi^2_{\text{obs}})$ is the number of pseudo-experiments which result in $\Delta\chi^2_{\text{toy}} \geq \Delta\chi^2_{\text{obs}}$, and N_{toy} denotes the total number of pseudo-datasets generated in step 4.).

10.) Repeat 1.)–9.) for a sufficiently large number of ϑ_i hypotheses to reach the desired $1-\mathrm{CL}$ confidence interval.

3.3.5 The Berger-Boos method

A different method to handle nuisance parameters is based on a p-value correction proposed by R. Berger and D. Boos [133]. The idea is not to choose a specific set of nuisance parameters, but to calculate a p-value in dependence of the nuisance parameters. Instead of searching for the supremum p-value over the complete nuisance parameter space, a subspace of the nuisance parameters with known confidence level α is considered. Then, to obtain a valid p-value a correction is needed:

$$\tilde{\mathrm{p}} = \sup_{\theta \in C_\alpha} \mathrm{p}(\vartheta, \theta) + (1 - \alpha) \,, \tag{3.14}$$

where C_α is the confidence region with confidence level α. In practice, the correction must be small compared to the desired precision of the p-value. This means to sample large parts of the nuisance parameter space. Given finite computing resources the p-values in dependence of the nuisance parameters are expected to fluctuate significantly. Hence, the practical application of the Berger-Boos method results in over-coverage.

3.3.6 Wilks' theorem

Wilks' theorem [134] is a generalisation of the result of Section 3.2, where for a gaussian random variable with fixed known width the distribution of the test statistic becomes a χ^2 distribution with one degree of freedom. The theorem from Ref. [134] in the notation used in this thesis with $\vec{\zeta} \equiv (\vartheta, \theta)^T$ reads:

Theorem — If the measurement of the variable x is distributed according to the PDF $f(x|\zeta)$, such that best estimates $\hat{\zeta}$ of $\zeta = (\vartheta_1, \ldots, \vartheta_r, \theta_1, \ldots, \theta_s)$ exist which are distributed in large samples according to

$$\mathcal{L}(\zeta|x) = \frac{|V|^{-1/2}}{\sqrt{(2\pi)^{r+s}}} \exp\left(-\frac{1}{2} \sqrt{N}(\hat{\zeta} - \zeta) V^{-1} \sqrt{N} (\hat{\zeta} - \zeta)^T \right) \times \left(1 + \mathcal{O}\left(\frac{1}{\sqrt{N}} \right) \right), \tag{3.15}$$

then, when the null hypothesis is true that $\vartheta_{1,\ldots,r} = \vartheta^0_{1,\ldots,r}$, the distribution of $\Delta\chi^2 = -2 \ln \lambda$ (cf. Eq. 3.13) is, except for terms of $\mathcal{O}(1/\sqrt{N})$, a χ^2 distribution with r degrees of freedom.

Wilks' theorem needs, in order to be valid, an approximated multivariate gaussian (Eq. 3.15) behaviour for the likelihood function $\mathcal{L}(\zeta)$ or in other words, Wilks' theorem works on samples where the central limit theorem (cf. Appendix A.1) is valid. In most physics analyses the parameter ranges for ζ are finite and as a consequence the variances of the resulting underlying distribution functions are finite. Hence, for

$N \to \infty$ a gaussian behaviour is expected. However, from the general form of the likelihood function in Eq. 3.12 it is not evident when the regime of large samples is reached.
The more severe restriction to the applicability of Wilks' theorem is the absence of physical boundaries. As illustrated in Section 3.3.1 the presence of a boundary breaks the gaussian symmetry in the distribution of the likelihood estimates. As a consequence, whenever a single parameter of ζ converges close to a physical boundary the assumption in Eq. 3.15 is not valid. Here, *close* is defined by the desired confidence level, e.g. in Fig. 3.5(c) the 90% confidence interval includes the boundary and is therefore incorrect, while for the same measured dataset Wilks' theorem provides a 68.27% confidence level with good approximate coverage probability. Wilks' theorem provides a powerful tool to get an idea of how the confidence intervals are shaped, or to construct high dimensional confidence intervals, where even the Kendall-Stuart method is computationally too expensive. For complicated likelihood functions this could already be the case for $r > 1$. A connection between gaussian standard deviation, coverage probability and $\Delta\chi^2$ values for $r = 1$ and $r = 2$ is given in Table A.1 in Appendix A.2.

3.4 Significance

The evaluation of signal significances is closely related to hypothesis tests. In order to quantify if the observation is a statistical fluctuation of the background or not, the p-value of the dataset given the background-only hypothesis is calculated. For a measurement of a branching fraction, the background-only hypothesis corresponds to the signal branching fraction equal to zero. The commonly quoted significance z is the translation of the p-value into a number of gaussian standard deviations. The gaussian standard deviations can be calculated as the lower boundary of a one-sided integral over the tail of a standard gaussian distribution, where the integral is equal to the p-Value:

$$\mathrm{p} = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-\frac{\widetilde{x}^2}{2}} d\,\widetilde{x}\,. \tag{3.16}$$

This definition implies a significance of $z = 0$ for a p-value of $\mathrm{p} = 0.5$ and for a signal significance of 3σ (5σ) the corresponding p-value is 0.135% (2.87×10^{-7}).
For an accurate calculation of the significance the distribution of the test statistic $f(\Delta\chi^2, \mu{=}0)$ has to be constructed from pseudo-experiments. However, if FC intervals are constructed for the outcome of the experiment that include the background-only hypothesis as one null hypothesis, the p-value for the significance calculation can directly be read off from the $1 - \mathrm{CL}$ curve at $\mu = 0$. This is true for the case where the parameter μ is constrained to positive values.
There are also approximative methods to calculate the significance based on the χ^2

approximations. If the measured test statistic value is $\Delta\chi^2_0 = 4$, the p-value calculated with Wilks' theorem, as the integral over a χ^2 distribution with one degree of freedom (for a single POI), results in $\mathrm{p} = 4.55\%$. Solving Eq. 3.16 for z, the significance is $z = 1.69\sigma$. However, as for the significance of a signal the interesting quantity is the one-sided deviation from zero ($\mu > 0$), the significance from Wilks' theorem is underestimated, because the χ^2 approximation results in a two-sided test. In the gaussian regime and at the null hypothesis of $\mu = 0$ it is evident from Eq. 3.7 that the test statistic is $\Delta\chi^2 = (x - \mu)^2$ for all $x \geq 0$ and $\Delta\chi^2 = 0$ for all $x < 0$. Hence, the distribution $f(\Delta\chi^2|\mu = 0)$ follows

$$f(\Delta\chi^2|\mu = 0) = \frac{1}{2} \times \delta(0) + \frac{1}{2} \times \chi^2_1 \,,$$

where $\delta(0)$ is the delta distribution and χ^2_1 the χ^2 distribution with one degree of freedom. As a consequence, when testing a significance of $\mu > 0$ and μ is constrained to positive values, Wilks' theorem results in a p-value that is a factor of two too large. The correction of the p-value results in a significance of $z = 2\sigma$, which is equivalent to the calculation of the significance using the square root of $\Delta\chi^2$: $z = \sqrt{\Delta\chi^2}\sigma = 2\sigma$.

4 The LHCb experiment at the Large Hadron Collider

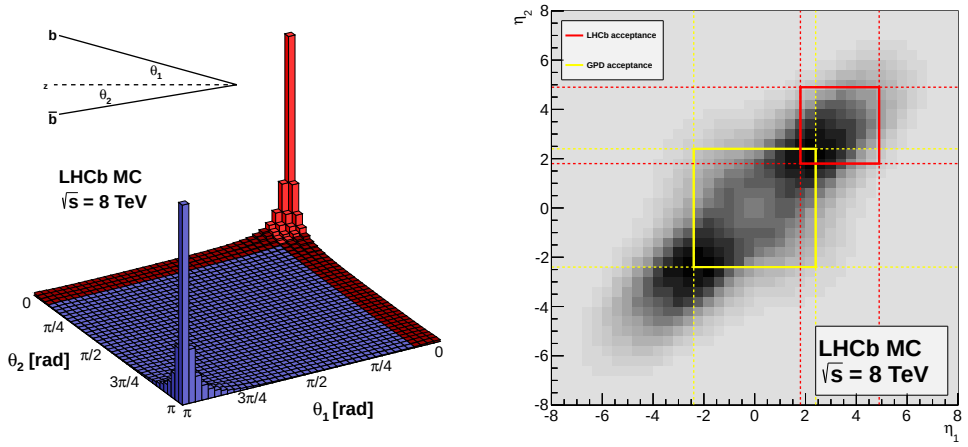
The datasets analysed in this thesis are mainly recorded by the LHCb experiment [135] and partly by the CMS experiment [9], both located at the Large Hadron Collider (LHC) [10] near Geneva, Switzerland. The LHC is the largest and most powerful particle accelerator and collider operated by the European Organization for Nuclear Research (CERN). The core of the physics programme at the LHC is the study of proton-proton (pp) interactions. Therefore, two high intensity proton beams are brought to collision at four interaction points with a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ during the years 2010 and 2011 and $\sqrt{s} = 8\,\text{TeV}$ in 2012. The proton beams are organised in up to 2808 bunches containing about 10^{11} protons per bunch. With a crossing rate of filled bunches of 20 MHz a peak instantaneous luminosity of $L = 8 \times 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$ is reached in Run I[1] by the LHC. Despite the fact that the pp interaction rate is half of the design rate, the instantaneous luminosity has approximately reached the design luminosity of $L = 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. This has been achieved by increasing the average number of pp interactions per bunch crossing.
Four main experiments are situated at the pp interaction points: ALICE [136], ATLAS [8], CMS [9], and LHCb [135]. The ALICE experiment focuses on the investigation of heavy ion collisions, which are provided by the LHC in dedicated lead-lead or lead-proton runs. The ATLAS and CMS experiments are high luminosity general-purpose detectors. To reach maximum luminosities ATLAS and CMS are built to reconstruct large numbers of visible pp interactions per bunch crossing (up to 40 in Run I). In addition, the focus of the general purpose detectors is on the decays of high mass particles, which primarily decay into particles that propagate perpendicular to the beam axis. Therefore, the pp interaction point is centred in the detector acceptance. Finally, the LHCb detector is a high precision experiment specialised on physics involving beauty and charm quarks.

In this thesis, most of the analysed data originates from the LHCb experiment. Hence, the LHCb detector is described in more detail in Section 4.1. Although one part of the presented work is a combination of results partly obtained with data from the CMS experiment, the CMS detector is not described in this thesis. A detailed description of the CMS experiment can be found in Refs [9, 137, 138].

[1] Run I includes the first data taking periods up to the long shutdown in 2013.

4.1 The LHCb detector

In order to achieve maximum precision in the field of beauty and charm physics, the LHCb detector has a distinct geometry. Instead of having the pp interaction point centred in the detector, LHCb is a single-arm forward spectrometer covering a pseudorapidity of $2 < \eta < 5$, where pseudorapidity is defined as $\eta = -\ln[\tan(\theta/2)]$. Here, θ is the polar angle between the momentum of the particle and the beam axis. This design concept is motivated by the fact that, at the LHC, b quark pairs are highly boosted into the forward (and backward) direction at the LHC. The distribution of production angles for b and $\bar{b}$ quarks and the acceptance regions of the LHCb experiment in comparison with general-purpose detectors (GPD) are shown in Fig. 4.1(a) and 4.1(b), respectively. For the GPDs the acceptance is assumed to be $|\eta| < 2.4$. With an acceptance for angles up to approximately 300 mrad LHCb is able to measure $\approx 25\%$ of the produced $b\bar{b}$ pairs.



(a) (b)

Figure 4.1: Distribution of production angles for b and $\bar{b}$ quarks at the LHC in pp collisions at a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$. (a) The distribution of the production angles θ_1 and θ_2 with respect to the beam axis. The LHCb acceptance is highlighted in red. (b) Pseudorapidity distribution of the b and $\bar{b}$ quarks with the comparison of the LHCb accpetance (red) and the acceptance of the general-purpose detectors (yellow). Figures from [139].

The LHCb experiment aims for highest precision and efficiencies in the reconstruction of b hadrons. Therefore, the instantaneous luminosity is artificially reduced by a disalignment of the proton beams to achieve constant running conditions throughout an LHC fill. This controlling of the luminosity [140, 141] leads to an instantaneous luminosity of $L = 4 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. This is twice the design luminosity of the LHCb experiment. The schematic layout of the LHCb detector is shown in Fig. 4.2. The pp interaction point is 11.25 m displaced from the centre of the LHCb cavern at the origin of a right handed cartesian coordinate system. The positive z-direction is defined as the clockwise circulating proton beam axis, while the positive y-direction points vertically towards the cavern ceiling. Completing the right handed coordinate system, the x-direction points away from the centre of the LHC ring.
The pp interaction point is situated in the vertex locator, which is part of the LHCb tracking system (cf. Section 4.1.1). The tracking system additionally consists of the tracker turicensis located upstream and three tracking stations located downstream of a normal conducting dipole magnet. The magnet provides an integrated field of 4 Tm for tracks with a length of 10 m. The polarity of the LHCb magnet can be changes to minimise detection asymmetries due to the detector geometry.
Particle identification (cf. Section 4.1.2) is provided by two Ring Imaging Cherenkov detectors in combination with the calorimeter system, which is composed of a scintillating pad and preshower detector, an electromagnetic, and a hadronic calorimeter. For the identification of muons a dedicated detector system consisting of five stations of multi-wire proportional chambers is installed.

4.1.1 The tracking system

The vertex locator

The vertex locator (VELO) [143] is a silicon strip detector surrounding the pp interaction point. As the VELO is the first tracking detector the main purpose is to reconstruct primary and displaced secondary vertices, which are key in reconstruction of particle decay times and to resolve the impact parameter (IP) of reconstructed tracks.
The VELO consists of 21 detector modules measuring the radial distance with respect to the beam axis and the azimuthal angle in the x-y plane. Each module is built out of two half-disks situated on opposite sides of the proton beam. In order to achieve the best IP resolution the half-disks are instrumented as close to the proton beams as permitted by safety considerations. Therefore, each VELO half is mounted on a movable support structure, which optimises the position around the proton beam for each LHC fill. As a result, the VELO sensors can be moved as close as 7 mm to the proton beam [144], which is less than the aperture during the proton injection phase of the LHC.

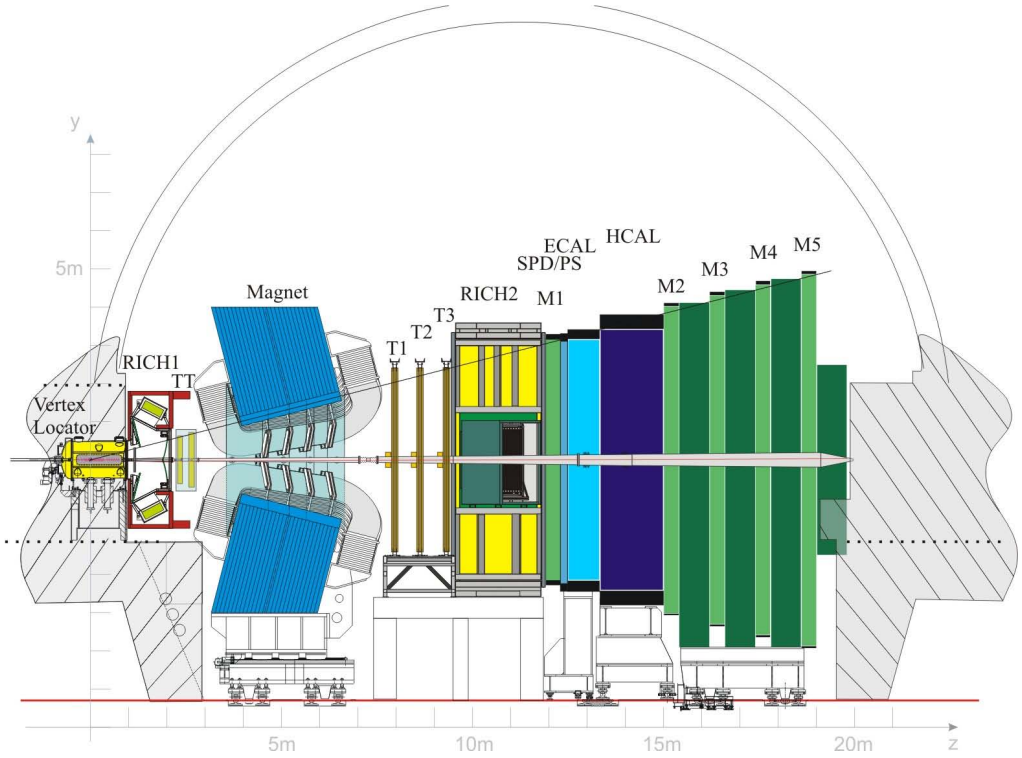


Figure 4.2: Schematic cross section of the LHCb detector. Starting with the *pp* interaction point situated in the vertex locator (VELO) at the origin of a cartesian coordinate system. Along the positive z-direction the first Ring Imaging Cherenkov detector (RICH1), the tracker turicensis (TT), the dipole magnet, the tracking stations (T1–T3), and the second Ring Imaging Cherenkov detector (RICH2) follow the interaction point. Then the calorimeter system follows, which is composed of a scintillator pad and preshower detector (SPD/PS), an electromagnetic (ECAL), and a hadronic calorimeter (HCAL). Lastly the muon stations (M1-M5) complete the LHCb detector. Figure from Ref. [142].

The silicon trackers and the outer tracker

The first of the silicon trackers [145], the tracker turicensis (TT), is located right upstream of the dipole magnet and covers the full LHCb acceptance. The TT serves as a tracking station for low energy particles that are bent out of the LHCb detector acceptance by the dipole magnet. Additionally, the TT helps to reconstruct long-lived neutral particles such as the K^0_S meson that can decay outside of the VELO. The TT consists of four layers of silicon micro-strip sensors, which are arranged in an "x-u-v-x" pattern. Here, the first and last layer (the x-layer) has vertical strips, while in the inner u- and v-layers the strips are rotated by $\pm 5^\circ$ with respect to the x-layers to increase the resolution in y-direction. The other silicon trackers are located in the high multiplicity[2] inner parts of the tracking stations T1–T3 downstream of the magnet. The inner tracker (IT) incorporates the same x-u-v-x pattern as the TT. The outer tracker (OT) covering the remaining outer parts of the LHCb acceptance in the T1–T3 stations is made of gaseous straw tubes. Each of the OT stations consists of four detector layers, arranged in the x-u-v-x pattern.

Performance of the tracking system

The full LHCb tracking systems in combination with the dipole magnet provides an excellent momentum resolution of $\delta p/p = 0.4\%$ for particles with momenta below $20\,\text{GeV}/c$ and 0.6% for particles with momenta around $100\,\text{GeV}/c$ [146]. This translates into a relative invariant mass resolution of $\approx 0.5\%$ up to masses of $b\bar{b}$ resonances such as the $\Upsilon(1S)$ meson for two-body muon decays.
The primary vertex resolution, which is mainly driven by the VELO, correlates with the number of tracks originating from the primary vertex. For vertices with more than 25 tracks, a primary vertex resolution of $13\,\mu\text{m}$ in the x-y plane and $71\,\mu\text{m}$ along the beam axis is achieved. The IP resolution with respect to these vertices is less than $35\,\mu\text{m}$ for particle tracks with transverse momenta greater than $1\,\text{GeV}/c$.

4.1.2 Particle identification

The RICH detectors

Two ring imaging Cherenkov (RICH) detectors [135] are installed between the VELO and the TT and downstream of the dipole magnet following the tracking stations.

[2] Charged particle fluxes of $5 \times 10^5\,\text{cm}^{-2}\text{s}^{-1}$ are expected in the inner most parts of the tracking stations [145].

Both detectors utilise the Cherenkov effect [147, 148], which describes the photon emission from charged particles traversing dielectric material at speeds greater than the phase speed of light in the respective medium. The photon emission angle θ_C is defined by the speed of a particle v and the refractive index n of the dielectric medium:

$$\cos\theta_C = \frac{c}{nv},$$

where c is the speed of light in vacuum. In combination with the momentum measurement of the tracking system, particle hypotheses can be tested and distinguished by assigning masses to the reconstructed particle tracks.
Different radiators are used in the RICH1 and RICH2 detectors to ensure an excellent PID performance in a momentum range of 2–100 GeV/c. The RICH1 detector is filled with C_4F_{10} gas at room temperature and pressure together with a 5 cm thick layer of aerogel, which is added to maintain the PID performance at low particle momenta and to increase the kaon-proton separation. Due to the choice of radiators the RICH1 detector differentiates light charged hadrons effectively in a momentum range of 2–40 GeV/c over the full LHCb acceptance. The distribution of Cherenkov angles for isolated tracks in dependence of the momentum is shown in Fig. 4.3 for the C_4F_{10} radiator. Isolated tracks are defined as tracks where the reconstructed Cherenkov ring is not intersected with another ring from the same radiator. These are about 2% of all tracks in data [149].

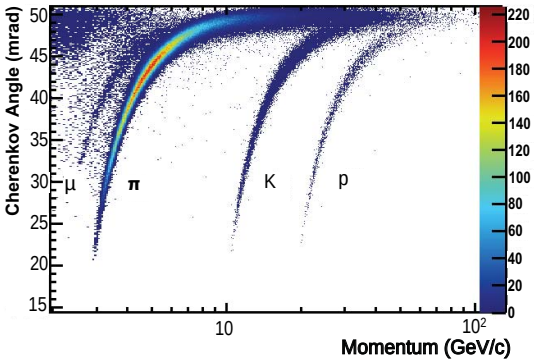


Figure 4.3: Cherenkov angles of isolated tracks in the C_4F_{10} radiator [149]. Clear bands of muons, pions, kaons, and protons are visible as well as respective momentum thresholds. PID information can be extracted from Cherenkov angle or the absence of Cherenkov light.

Clear differences in the bands of Cherenkov angles can be seen for muons, pions, kaons, and protons. PID information can be extracted from Cherenkov angle or from the absence of Cherenkov light due to momentum thresholds. For example, the momentum threshold for kaons to emit Cherenkov photons in C_4F_{10} is 9.3 GeV/c, therefore reconstructed rings with radii corresponding to momenta less than 9.3 GeV/c can only originate from lighter pions or muons.
The RICH2 detector provides PID information in the momentum range of 15–100 GeV/c as it is filled with CF_4, which has a lower refractive index compared to C_4F_{10}. High momentum tracks are mainly produced under small opening angles with respect to the beam axis, therefore the RICH2 acceptance is limited to 15–120 mrad (100 mrad) in the bending (non-bending) plane of the magnet.

The calorimeter system

The calorimeter system [150] is composed of a scintillating pad and preshower detector (SPD/PS) prior to an electromagnetic (ECAL) and hadronic calorimeter (HCAL) located between the muon stations M1 and M2. Due to fast response times the calorimeter system plays a crucial role in the hardware trigger stage (cf. Section 4.1.3) by measuring the transverse energy and position of electrons, photons, and hadrons. All calorimeter detectors follow the same principle of alternating shower and detection layers. As detection layers scintillating tiles are used with increasing granularity towards the beam axis to ensure approximately constant angular resolution.
The SPD and PS scintillating pads are separated by a 12 mm thick lead layer (2.5 radiation lengths). Hence, the SPD detector can distinguish between neutral and charged particles (predominantly photons and electrons) and the PS pads detect first showers of electrons while hadrons are unlikely to shower in 12 mm of lead. The subsequent ECAL structure adds 25 radiation lengths of alternating lead and scintillating pad layer to optimise the energy resolution. Finally, the hadronic calorimeter consists of alternating layers of iron shower material and scintillating pad detectors corresponding to a hadronic interaction lenth of $\lambda = 5.6$.

The muon system

The muon system [151, 152] is essential for triggering and identifying muons at the LHCb experiment. The first of the five muon stations is placed in front of the calorimeter system in order to improve the transverse momentum (p_T) resolution in the LHCb trigger. The other four muon stations are located after the HCAL, where each of the stations is separated by an 80 cm thick iron absorber, which is

supposed to stop remaining hadrons. All muon stations are multi-wired proportional chambers (MWPC) with the exception of the inner region of M1. Here, the particle rate exceeds MWPC safety limits. Hence, triple gas electron multiplier (GEM) [153] are used [152].
The MWPC chambers are made out of cathode pads that run along the x-direction and anode wires that are parallel to the y-axis. By applying a logical `AND` on the cathode and anode read-out sensitivity in the x-y plane is achieved within a single detector.

The particle identification algorithms and performances

The analysis of CP observables in $B_s^0 \to D_s^{\mp} K^{\pm}$ decays as well as the search for very rare $B \to \mu^+\mu^-$ decays rely on the correct identification of particles. The final PID information is provided by a combined likelihood difference between two particle hypotheses [146]. Therefore, each PID detector constructs a likelihood that aims to separate kaons, protons, electrons, and muons from pions by using reconstructed tracks. Then each sub-detector matches these tracks to Cherenkov angles and expected number of photoelectrons (RICH1/2), calorimeter cluster positions (E/HCAL), and muon station hits. As the predominantly produced particles in pp interactions are pions, the null hypothesis for each track is that of a pion, where the hypothesis is tested by assigning the π mass to the reconstructed track. For every track in an event the likelihoods are evaluated under the K, p, e, and μ mass hypotheses in order to construct the likelihood differences: $\Delta \ln \mathcal{L}_{X\pi} = \ln \mathcal{L}_X - \ln \mathcal{L}_\pi$, where X denotes the respective particle mass hypothesis. The single likelihoods from the RICH, calorimeter, and muon systems are added to get the global combined PID information:

$$\mathrm{DLL}_{X\pi} = \sum_i \Delta \ln \mathcal{L}_{X\pi} \,, \tag{4.1}$$

where i denotes the three PID subsystems.
A typical cut in order to separate kaons from pions is $\mathrm{DLL}_{K\pi} > 5$, which has an efficiency of $\approx 85\%$ with a $\pi \to K$ mis-identification rate of $\approx 3\%$.
In the search for $B \to \mu^+\mu^-$ decays the muon identification relies on a fast binary classifier called `IsMuon` [154], while the $\mathrm{DLL}_{\mu\pi}$ likelihood is used to suppress the $B \to h^+h'^-$ background. The binary `IsMuon` decision is based on existing hits in the muon station in a field of interest around an extrapolated track from the tracking system. In order for `IsMuon` to be true, hits need to be present in M2 and M3 for tracks with momentum of $3 < p < 6\,\mathrm{GeV}/c$, in M2, M3, and either M4 or M5 for tracks with $6 < p < 10\,\mathrm{GeV}/c$, and in M2, M3, M4, and M5 for tracks with momentum greater $10\,\mathrm{GeV}/c$. The classifier is up to 99% efficient for muons from $B_s^0 \to \mu^+\mu^-$ decays with an integrated hadron mis-identification rate of about 1% for tracks with $p > 3\,\mathrm{GeV}/c$.

4.1.3 The trigger system

The LHCb trigger system is built to reduce the rate of incoming data to a manageable level, while keeping the output of interesting physics events at a maximum. Therefore, the trigger system consists of a hardware trigger stage, the Level 0 trigger (L0), and a subsequent software trigger, the High Level Trigger (HLT). The overall trigger system reduces the rate of physics events to about 5 kHz. A more detailed description of the LHCb trigger system can be found in [155–157].

The L0 hardware trigger reduces the bunch crossing rate of 40 MHz or the effective inelastic pp collision rate of 13 MHz during RunI to an L0 output rate of 1 MHz. Therefore, the latency for the L0 decisions is fixed to a maximum of 4 µs, where parallel processing and a dedicated event buffer allows for the synchronous running at 40 MHz. The L0 trigger uses information of the calorimeter and muon systems to form a decision. Here, the calorimeter systems provide information about potential hadron, photon, and electron candidates: high transverse energy deposits are searched for in the HCAL and ECAL in combination with information from the SPD and PS detectors to separate electrons from photons.
Muons are identified in L0 from tracks reconstructed in the muon stations. From the track position in M1 and M2 the transverse momentum (p_T) of a track can be measured with a resolution of $\approx 25\%$ compared to the full muon track reconstruction. The L0 single and dimuon triggers search the muons with high transverse momenta, where the single muon L0 sets a threshold on the highest transverse momentum of a track, while the L0 dimuon trigger sets a threshold on the product of the highest and next-to-highest transverse momentum track.

The High Level Trigger is split into two stages, HLT1 and HLT2 and runs synchronously to the full detector read-out rate of 1 MHz. The two HLT stages perform a partial (HLT1) and full (HLT2) event reconstruction on a event filter farm consisting of 29.000 logical CPU cores [157].
The HLT1 stage uses high IP track segments from the VELO detector and tracks that can be matched to hits in the muon stations together with forward tracking [158] to form a trigger decision. The HLT1 triggers on good quality tracks that have a displacement from the primary vertex and have a transverse momentum of $p_T > 1.0\,\text{GeV}/c$ ($p_T > 1.6\,\text{GeV}/c$) for muon (non-muon) track candidates. Furthermore, the HLT1 has dedicated trigger selections for single muon candidates with $p > 4.8\,\text{GeV}/c$ dropping the IP criterion, dimuon candidates, high p_T electrons, photons, diprotons, displaced vertices and jets with hight transverse energies [157].
The HLT2 trigger stage performs a full event reconstruction for all tracks with $p_T > 0.3\,\text{GeV}/c$ reducing the HLT1 output rate of ≈ 80 kHz to the final 5 kHz permanently written to storage. The important trigger selections used for the analyses of $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_s \to D^\mp_s K^\pm$ decays are the muon and generic beauty triggers. The generic beauty triggers select multi-body B meson decays using close high quality

tracks that are separated from but share the same associated PV[3]. From these tracks inclusive vertices, i.e. without any vertex quality criterion, are reconstructed, which are separated from minimum-bias events (almost unbiased random pp interactions) by the output of a boosted decision tree (BDT) [159, 160] that uses kinematic event properties as an input [161, 162].
The HLT2 stage selects single muons as well as dimuon events. However, the output rate requirements for the HLT2 stage are much stricter compared the HLT1 rate. Therefore, single muons are only triggered when the candidate track has a very good quality and is significantly separated from the PV or has a large transverse momentum ($p_T > 10\,\text{GeV}/c$). The HLT2 dimuon decisions are based on good quality tracks, which form a good common vertex. The single decision mainly relies on two discriminating variables: the invariant mass of the reconstructed dimuon system and the vertex separation of the PV with respect to to the dimuon origin vertex. The invariant mass allows for dedicated resonance selections such as J/ψ or $\psi(2S)$ or inclusive high dimuon mass events ($m_{\mu^+\mu^-} > 4.7\,\text{GeV}/c$) without a vertex separation requirement. Due to the rate budget of the HLT2, low invariant dimuon masses can only be triggered when the reconstructed SV is significantly separated from the PV. The set of muon triggers in the HLT are discussed in Ref. [163] in more detail.

4.2 Flavour tagging

The measurement of CP observables, e.g. in $B_s^0 \to D_s^{\mp} K^{\pm}$ decays (cf. Section 2.3.1), relies on the knowledge of the production flavour of the B_s^0 meson, i.e. whether a B_s^0 or $\overline{B}_s^0$ meson is initially produced. The set of algorithms and methods used to determine the production flavour of the decaying meson is called flavour tagging.
As B_s^0 mesons oscillate, it is not trivial to determine the production flavour of the B_s^0 particle. Therefore, a short introduction to flavour tagging, its algorithms, and the calibration of the flavour tagging output is given in next sections.

The flavour tagging algorithms

The flavour tagging at the LHCb experiment utilises characteristics of $b\bar{b}$ pair production at the LHC. Therefore, an event is split into two *sides*. The *same side* (SS) is defined as the set of particles involved in processes associated with the signal b quark. In contrast, the *opposite side* (OS) is the set of particles associated to the non-signal b quark.
The algorithms (or taggers) operating on the same side make use of the quark

[3] This is the PV to which the reconstructed track has the smallest impact parameter.

produced in the hadronisation process of the signal b quark. For example, in association with the hadronisation of a b quark into a $\overline{B}{}^0_s$ meson an additional s quark is produced, which can handronise into a K^- meson. It is evident that the charge of the associated kaon depends on the production flavour of the B^0_s meson. This charged kaon is searched for by the *same side kaon* (SSK) tagger. An illustration of the SSK tagger for a $\overline{B}{}^0_s \to D^-_s K^+$ decay is shown in the top half of Fig. 4.4. More details on the SS taggers can be found in [164, 165].
The taggers operating on the opposite side [166] pursue two different strategies. The first class of taggers aims to identify single final state particles of non-signal b hadron decays. Here, leptons (either an electron or a muon, l) from semileptonic decays as well as charged kaons from $b \to c \to s$ decay chains are used to determine the production flavour of the signal b hadron. The second way is to assign an effective charge to the decay vertex of the OS b hadron. In analogy to an OS B^+ decay a positive vertex charge results in a $\overline{B}{}^0_s$ tag in case of a $\overline{B}{}^0_s \to D^-_s K^+$ signal decay. The OS taggers are illustrated in the lower half of Fig. 4.4.

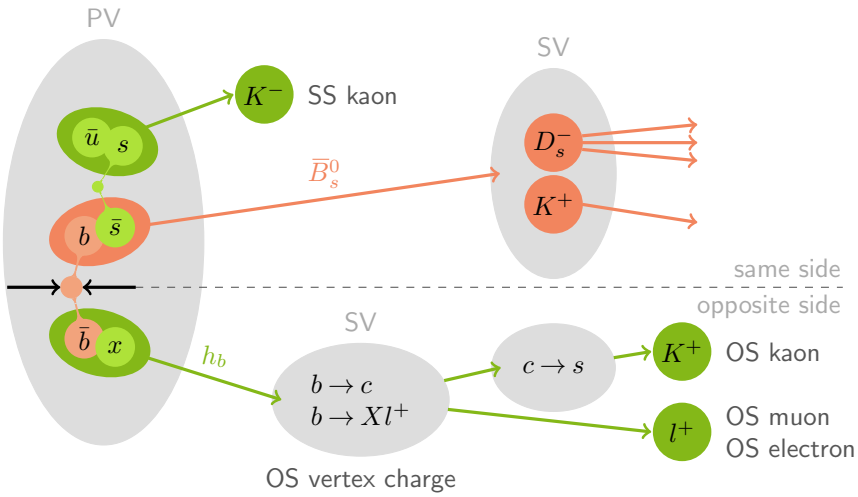


Figure 4.4: Schematic view of the same side (SS) kaon, opposite side (OS) vertex charge, and the single particle OS kaon, electron and muon taggers. The signal particles are shown in red and the particles associated with taggers are shown in green. The primary (PV), secondary (SV) and tertiary vertices are indicated as grey ellipses. Figure based on [167].

The calibration of flavour tagging

The performance of the flavour taggers is quantified by three measures: the tagging efficiency ε_{tag}, the mistag rate ω, and the effective tagging efficiency or tagging power ε_{eff}. The quantities are defined as follows:

$$\varepsilon_{\text{tag}} = \frac{\text{tagged candidates}}{\text{all candidates}}\,, \quad \omega = \frac{\text{wrong tags}}{\text{all tags}}\,, \quad \text{and}$$
$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2\,. \tag{4.2}$$

Each tagging algorithm suffers from an intrinsic mistag rate ω, which is due to wrong identification of tagging particles, decays of high mass B resonances (OS and SS taggers), or the oscillation of neutral tagging particles (OS taggers). The tagging power is of particular importance as the statistical sensitivity of *CP* asymmetry measurements such as in $B^0 \to J/\psi K^0_{\text{S}}$ [168] decays scale with $1/\sqrt{\varepsilon_{\text{eff}}}$. For the example of a measured *CP* asymmetry

$$A_{CP,m} = \frac{\Gamma_m(\overline{B}\to f) - \Gamma_m(B\to f)}{\Gamma_m(\overline{B}\to f) + \Gamma_m(B\to f)}\,, \tag{4.3}$$

where Γ_m denotes the measured decay rates, this can qualitatively be explained as follows[4]: statistical precision scales with $1/\sqrt{N}$. However, as only tagged events can be used to calculate the *CP* asymmetry in Eq. 4.3, the number of events N is effectively reduced by ε_{tag} ($\sigma \propto 1/\sqrt{\varepsilon_{\text{tag}}}$). Furthermore, the measured decay rates are diluted by the mistag rate ω:

$$\Gamma_m(\overline{B}\to f) = \omega\Gamma_t(B\to f) + (1-\omega)\Gamma_t(\overline{B}\to f)$$
$$\Gamma_m(B\to f) = (1-\omega)\Gamma_t(B\to f) + \omega\Gamma_t(\overline{B}\to f)\,,$$

where Γ_t denotes the actual decay rates. This leads to a dilution factor of $(1-2\omega)$ for the measured *CP* asymmetry: $A_{CP,m} = (1-2\omega)A_{CP,t}$, where $A_{CP,t}$ is the actual *CP* asymmetry defined by Eq. 4.3 interchanging the measured decay rates with the actual decay rates Γ_t. As a results, the uncertainties on physics parameters scale according to $1/(1-2\omega)$ and together with the tagging efficiency this results in

$$\sigma(A_{CP,t}) \propto \frac{\sigma(A_{CP,m})}{\sqrt{\varepsilon_{\text{tag}}(1-2\omega)^2}} = \frac{\sigma(A_{CP,m})}{\sqrt{\varepsilon_{\text{eff}}}}.$$

The effective tagging efficiency, in other words, is a measure for the overall quality of the tagging algorithms and is therefore optimised in the development of tagging algorithms.

Each of the previously mentioned tagging algorithms provides a mistag probability estimate η, which is calculated by a multilayer perceptron [169] artificial neural

[4] A more quantitatively discussen can e.g. be found in Ref. [164].

network (MLP), implemented in the ROOT package [129], on a candidate-by-candidate basis. The MLPs for the taggers used in the analysis of $B_s^0 \to D_s^{\mp} K^{\pm}$ decays are trained on simulated events[5] using geometric, kinematic, and PID information. Hence, the output of the MLPs need to be calibrated on data, which is done using flavour-specific decays, which determine the flavour of the B meson at decay time. For non-oscillating charged mesons the mistag rate ω can be measured by counting the charge of final state particles. For oscillating neutral B mesons a full decay time analysis is performed to extrapolate the flavour at decay time back to the production flavour. The linear calibration function $\omega(\eta)$ is a parametrised as

$$\omega(\eta) = p_0 + p_1(\eta - \langle\eta\rangle), \tag{4.4}$$

where $\langle\eta\rangle$ is the average mistag estimate. For an ideally calibrated tagger the parameters would result in $p_0 = \langle\eta\rangle$ and $p_1 = 1$. Potential tagging asymmetries can be accounted for with two different calibration function for B_s^0 and $\overline{B}_s^0$ mesons.

[5] Recently, the MLPs are also trained on large samples of $B^+ \to J/\psi K^+$ decays.

5 Measurement of the CKM angle γ with $B_s^0 \to D_s^\mp K^\pm$ decays

This chapter describes the measurement of the CKM angle γ from $B_s^0 \to D_s^\mp K^\pm$ decays. The first step is to measure the *CP* violating observables from Eq. 2.18, then in a second step the CKM angle γ is extracted using the relations in Eq. 2.20 connecting the measured observables to the physics parameters. For the published analysis [66] the work presented in this thesis is the latter of the two steps. Prior to the measurement of the *CP* observables, the branching fraction of $B_s^0 \to D_s^\mp K^\pm$ decays was measured in Ref. [170], for which the fit model was validated on sets of pseudo-experiments. The basic outcome of this work is summarised in Ref. [171] and is reflected in the way the constrained terms are handled in the PLUGIN method described in Section 3.3.4.

5.1 Analysis strategy

In order to measure the *CP* observables from Eq. 2.18, a fit to the decay-time distributions from Eq. 2.14 of selected $B_s^0 \to D_s^\mp K^\pm$ decays is performed. Two different approaches are used to fit the decay time distributions. The first approach makes use of a full description of the decay time distributions of the signal and the various background components. This method is the nominal one and is called *cFit*. The second approach uses the *sPlot* technique [172], which is a way to perform a statistical background subtraction.
Both methods are based on a three-dimensional maximum likelihood fit, in the following referred to as *multivariate fit*, to the B_s^0 and D_s^- invariant mass, and the PID information of the companion hadron, $\mathrm{DLL}_{K\pi}$. The multivariate fit is performed to constrain the number of events for the signal and background components in the cFit or to extract *sWeight*s, which are then used in a weighted maximum likelihood fit [173] to statistically subtract the background components in the sFit. Throughout the analysis, the kinematically similar decay mode $B_s^0 \to D_s^- \pi^+$ is used as a control channel to evaluate signal shapes, correct for differences between simulation and data, and to provide an independent cross-check of the cFit by measuring the B_s^0 oscillation frequency Δm_s. The CKM angle γ is measured from the *CP* observables

determined by the cFit.
An overview of the selection of $B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^- \pi^+$ candidates is given in Section 5.2 followed by the description of the multivariate fit in Section 5.3. Then the necessary cFit inputs of flavour tagging and decay times resolution and acceptance are described in Section 5.4 and Section 5.5.1, respectively. The results of the measurement of the *CP* observables are given in Section 5.5 and finally the expected sensitivity on the CKM angle γ and the final γ result are presented in Section 5.6.

5.2 Selection

The selection for the signal decay $B_s^0 \to D_s^\mp K^\pm$ and the control channel $B_s^0 \to D_s^- \pi^+$ is equal with the exception of a particle identification criterion on the companion hadron (the kaon or pion). Hence, whenever it is referred to the signal and control channel the notation $B_s^0 \to D_s^\mp h^\pm$ is used.
At the hardware trigger stage most events are selected due to high momentum signal tracks depositing large transverse energy in the HCAL and ECAL. Then the software trigger selects generic B decays based on inclusive vertices, which are separated from the primary vertex (cf. Section 4.1.3). From the set of triggered events a first loose selection of $B_s^0 \to D_s^\mp h^\pm$ decays is performed. The selection exploits the fact that the lifetime of B_s^0 and D_s^- mesons is large enough that the particles decay after a sizeable flight distance. This leads to two distinct vertices separated from the associated primary vertex. In a first step good quality D_s^- candidates are reconstructed in five different decay modes given in Table 5.1. To reconstruct a single $B_s^0 \to D_s^\mp h^\pm$ candidate a fourth track is added to the D_s^- candidate. After fitting for the common vertex a boosted decision tree (vertex BDT) trained on simulated events is used to separate random particle combinations from signal decays. The discriminating variables are the transverse momentum and the vertex separation quality with respect to the nearest PV of the B_s^0 candidate, and the sum of the vertex qualities of the fits to the B_s^0 and the D_s^- decay vertex.

Table 5.1: List of different D_s^- final states. The $D_s^- \to K^- K^+ \pi^-$ mode is split into resonant $D_s^- \to \phi\pi^-$ and $D_s^- \to K^{*0} K^-$, and the non-resonant component $D_s^- \to (K^- K^+ \pi^-)_{\text{nonres}}$.

$\boldsymbol{D_s^-}$ **decays products**	**reconstructed as**
$\phi\pi^-$ $K^{*0}\ K^-$ $(K^-K^+\pi^-)_{\text{nonres}}$	$K^-K^+\pi^-$
$K^-\ \pi^+\ \pi^-$ $\pi^-\ \pi^+\ \pi^-$	$K^-\ \pi^+\ \pi^-$ $\pi^-\ \pi^+\ \pi^-$

The classification of signal and combinatorial background is done by a BDT (called BDTG) trained on data from $B_s^0 \to D_s^- \pi^+$ with $D_s^- \to K^- K^+ \pi^-$. As a signal proxy *sWeights* are used, extract by a simplified fit to the $B_s^0 \to D_s^- \pi^+$ data and the high mass sideband, $m_{B_s^0} > 5.445\,\mathrm{GeV}/c^2$, is used as combinatorial background. Besides transverse momentum and impact parameter quality of all signal tracks, the main separating variables are the decay vertex qualities, the flight distances, and the radial distances to the beam axis of the D_s^- and B_s^0 candidates. The other variables are pointing parameters that measure the agreement between the distance vector of the origin and end vertex and the reconstructed momentum vector of the decaying particles.
The samples are further purified by particle identification cuts on the daughters of the D_s^- meson. For each subchannel the cuts are tightened until no resonant structure is observed in the D_s^- mass spectrum. Additionally, many background sources from falsely reconstructed D_s^- mesons are suppressed by explicitly vetoing mis-identified $\Lambda_c^+ \to p K^- \pi^+$ and $D^+ \to K^+ \pi^- \pi^+$ decays, 2-body D^0 decays from $B^0 \to D^{*-} \pi^+$ processes, and inclusive $B \to J/\psi X$ decays. Possible non-charm final states such as $B_{(s)}^0 \to K^+ K^- \pi^+ \pi^-$ are reduced by a requirement on the vertex separation quality and the decay time of the reconstructed D_s^- candidates.
Finally, the cut on the BDTG classifier is optimised in dependence of the particle identification criterion of the companion hadron, which forms the B_s^0 candidate together with the D_s^- meson. The target function, $S/\sqrt{S+B}$, which is maximised in the optimisation, is an approximation of the signal significance. The number of $B_s^0 \to D_s^\mp K^\pm$ signal events, S, and the number of background events, B, are evaluated from a simplified one-dimensional fit to the invariant mass distribution of the full dataset. The optimal cuts are BDTG > 0.3 and $\mathrm{DLL}_{K\pi} > 5$. To separate the $B_s^0 \to D_s^- \pi^+$ control channel more strictly from the signal decay the particle identification on the companion pion is set to $\mathrm{DLL}_{K\pi} < 0$.

5.3 Multivariate fit

The multivariate fit is a three dimensional simultaneous fit to the B_s^0 invariant mass, the D_s^- invariant mass, and the companion hadron PID, $\mathrm{DLL}_{K\pi}$, in five categories of the D_s^- final state. Hence, the signal and background distributions need to be known in all dimensions and categories. The distributions are mainly determined from simulated events, where the two LHCb magnet polarities are merged. However, a large number of $B^0 \to D^- \pi^+$ events is used to calibrate the invariant mass scale and to correct for differences between the simulation and data.

5.3.1 Invariant B_s^0 and D_s^- mass templates

Signal invariant mass shapes

A set of simulated events is used to determine the invariant mass distribution of the B_s^0 and D_s^- signal candidates. The model for both distributions is a gaussian core function with two power law tails in either direction of the invariant mass mean in order to account for reconstruction and final state radiation effects. The evaluation of the B_s^0 and D_s^- invariant mass shapes is performed separately for each D_s^- final state.

Combinatorial background mass shapes

The invariant B_s^0 mass distributions of the combinatorial background component for the different D_s^- final states are obtained from the high B_s^0 mass sideband ($m_{B_s^0} > 5.8\,\mathrm{GeV}/c^2$). The lower B_s^0 mass sideband is dominated by partially and mis-reconstructed background decays, so that no unbiased estimation for the combinatorial background is possible. For all reconstruction modes a single exponential is sufficient to describe the combinatorial background, except for the $D_s^- \to K^-K^+\pi^-$ mode in the fit to $B_s^0 \to D_s^-\pi^+$ decays. Here, an exponential together with a constant offset is used to model the combinatorial background in the B_s^0 invariant mass spectrum. The combinatorial background distributions is primarily defined with respect to the $B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^-\pi^+$ signals. However, this does not imply that the combinatorial component has a non-peaking structure in the invariant D_s^- mass. In fact, the combinatorial background distribution in the reconstructed D_s^- mass contains random combinations of three tracks, which have a non-peaking structure in the D_s^- candidates mass, but also a peaking D_s^- contribution, which is combined with a random hadron track to form combinatorial background for the B_s^0 mass distribution. The only D_s^- final state where the peaking invariant D_s^- mass in the combinatorial background is present, are the resonant $D_s^- \to K^-K^+\pi^-$ channels. Here, the invariant D_s^- mass signal distribution from Section 5.3.1 is added to the single exponential function that is used otherwise.

Fully reconstructed background mass shapes

Fully reconstructed background decays are processes where no final state particle has been missed during the reconstruction. These occur either when the decay has the same final state composition as the signal or tracks are falsely identified. The list of fully reconstructed background components considered in this analysis is shown in Table 5.2. The invariant B_s^0 mass shapes of the most important backgrounds, $B^0 \to D^-\pi^+$ to the $B_s^0 \to D_s^-\pi^+$ signal and $B_s^0 \to D_s^-\pi^+$ to the $B_s^0 \to D_s^\mp K^\pm$ signal, are taken from pure data samples of $B^0 \to D^-\pi^+$ and $B_s^0 \to D_s^-\pi^+$ decays. The

Table 5.2: The list of considered fully reconstructed backgrounds in the nominal fits to $B_s^0 \to D_s^{\mp} K^{\pm}$ and $B_s^0 \to D_s^- \pi^+$. A short description is given how the background can contribute to the respective signal channel. The signal final state particle (D_s^- or the companion hadron h) that is involved in the mis-identification process is given in brackets.

Channel	**Background to**	**Mechanism**
	$\boldsymbol{B_s^0 \to D_s^- \pi^+}$	
$B^0 \to D_s^- \pi^+$	$B_s^0 \to D_s^- \pi^+$	Same final state
$B_s^0 \to D_s^- K^+$	$B_s^0 \to D_s^- \pi^+$	$\pi^+ \to K^+$ (h)
$\Lambda_b^0 \to \Lambda_c^- \pi^+ \to (\overline{p} K^+ \pi^-) \pi^+$	$B_s^0 \to D_s^- \pi^+ \to (K^+ K^- \pi^-) \pi^+$	$\overline{p} \to K^-$ (D_s^-)
$B^0 \to D^- \pi^+ \to (K^+ \pi^- \pi^-) \pi^+$	$B_s^0 \to D_s^- \pi^+ \to (K^+ K^- \pi^-) \pi^+$	$\pi^- \to K^-$ (D_s^-)
$B^0 \to D^- \pi^+ \to (K^+ \pi^- \pi^-) \pi^+$	$B_s^0 \to D_s^- \pi^+ \to (\pi^+ K^- \pi^-) \pi^+$	$K^+ \to \pi^+$ and $\pi^- \to K^-$ (D_s^-)
	$\boldsymbol{B_s^0 \to D_s^{\mp} K^{\pm}}$	
$B^0 \to D_s^- K^+$	$B_s^0 \to D_s^- K^+$	Same final state
$B_s^0 \to D_s^- \pi^+$	$B_s^0 \to D_s^- K^+$	$\pi^+ \to K^+$ (h)
$B^0 \to D^- K^+ \to (K^+ \pi^- \pi^-) K^+$	$B_s^0 \to D_s^- K^+ \to (K^+ K^- \pi^-) K^+$	$\pi^- \to K^-$ (D_s^-)
$\Lambda_b^0 \to \Lambda_c^- K^+ \to (\overline{p} K^+ \pi^-) K^+$	$B_s^0 \to D_s^- K^+ \to (K^+ K^- \pi^-) K^+$	$\overline{p} \to K^-$ (D_s^-)
$\Lambda_b^0 \to D_s^- p$	$B_s^0 \to D_s^- K^+$	$p \to K^+$ (h)

resulting distributions are fit with a non-parametric model using gaussian kernel functions [174].

The invariant B_s^0 mass shapes for the backgrounds with the same final state, $B^0 \to D_s^- h^+$, are assumed to have the same distribution as the signal, a gaussian core function with two power law tails. The shape is taken from simulated events, with the exception of the gaussian mean, which is taken from the calibrated signal mass shape. As the signal shapes are calibrated with $B_s^0 \to D_s^- \pi^+$ decays, the gaussian mean for the B^0 backgrounds is shifted by the mass difference of B_s^0 and B^0 mesons, $\Delta m = 86.6\,\mathrm{MeV}/c^2$ [175]. The widths of the gaussian core functions are taken from the signal shapes and are corrected for kinematic differences between B^0 background and B_s^0 signal decays.

The other invariant mass templates for the background channels in Table 5.2 are determined solely from simulated events. For all background channels that have a correctly reconstructed D_s^- meson, the invariant mass distribution of the D_s^- is assumed to be the same as for the signal.

The invariant D_s^- mass distribution for the $B^0 \to D^- \pi^+$ can be obtained from the same pure data sample used to extract the distribution of the invariant B_s^0 mass. In this case only the D^- meson has to be re-reconstructed under the D_s^- mass hypothesis.

For the remaining background channels where one or more particles are mis-identified, the distributions are taken from simulated events. Here, the different channels are reconstructed under the corresponding signal hypotheses.

Partially reconstructed background mass shapes

If one or more particles are missed in the event reconstruction, the decay is considered as partially reconstructed background to the $B_s^0 \to D_s^\mp K^\pm$ or $B_s^0 \to D_s^- \pi^+$ signals. Due to the absent particle, the reconstructed invariant B_s^0 mass distribution of partially reconstructed backgrounds is shifted to lower mass ranges. A list of the considered partially reconstructed background is given in Table 5.3. For all partially reconstructed channels the invariant B_s^0 mass distribution is taken from simulated events. The templates for the invariant D_s^- mass distributions are assumed to be equal to the distributions of the $B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^- \pi^+$ signal distributions, because all the considered modes are decaying via a D_s^- meson.

Table 5.3: The list of considered partially reconstructed backgrounds in the nominal fits to $B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^- \pi^+$. A short description is given how the background can contribute to the respective signal channel. The particle that is partially reconstructed or mis-identified is given in brackets.

Channel	Background to	Mechanism
$B_s^0 \to D_s^{*\pm} \pi^\mp$	$B_s^0 \to D_s^- \pi^+$	missed γ/π^0 ($D_s^{*\pm}$)
$B_s^0 \to D_s^- \rho^+$	$B_s^0 \to D_s^- K^+$	missed π^+ and $\pi^+ \to K^+$ (ρ^+)
$B_s^0 \to D_s^{*-} \pi^+$	$B_s^0 \to D_s^- K^+$	missed γ/π^0 (D_s^{*-}) and $\pi^+ \to K^+$ (π^+)
$\Lambda_b^0 \to D_s^{*-} p$	$B_s^0 \to D_s^- K^+$	missed γ/π^0 (D_s^{*-}) and $p \to K^+$ (p)

5.3.2 Particle identification shapes

The particle identification probability distributions are extracted from large data samples used for the calibration of the particle identification system [176] of the LHCb detector. The $\mathrm{DLL}_{K\pi}$ distribution for kaons, pions, and protons are evaluated for the $B_s^0 \to D_s^\mp K^\pm$ ($\mathrm{DLL}_{K\pi} > 5$) and $B_s^0 \to D_s^- \pi^+$ ($\mathrm{DLL}_{K\pi} < 0$) separately. All shapes are corrected for kinematic differences between the calibration channel and the respective signal or background decays. For the $B^0 \to D^- \pi^+$, $B_s^0 \to D_s^- \pi^+$, and the combinatorial background channels, the correction factors are directly calculated from data. All other correction factors for the background sources are extracted using simulated events, where the simulated events are weighted to better describe the distributions measured in data.

5.3.3 Results of the multivariate fit

The multivariate fit is performed simultaneously in all five D_s^- final state categories for merged magnet polarities. It yields a total number of

$$N(B_s^0 \to D_s^\mp K^\pm) = 1768 \pm 49 \text{ and}$$
$$N(B_s^0 \to D_s^- \pi^+) = 28264 \pm 182$$

$B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^- \pi^+$ candidates, respectively. The single most important background component to both decay modes is the combinatorial background with 2430 ± 62 and 9030 ± 127 events for the $B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^- \pi^+$ channels, respectively. The fits to invariant B_s^0 mass, the invariant D_s^- mass, and the $\mathrm{DLL}_{K\pi}$ distribution of the companion particle are shown for the $B_s^0 \to D_s^\mp K^\pm$ signal in Fig. 5.1. To increase the clarity of the figures background components with similar final states are grouped into a single component.

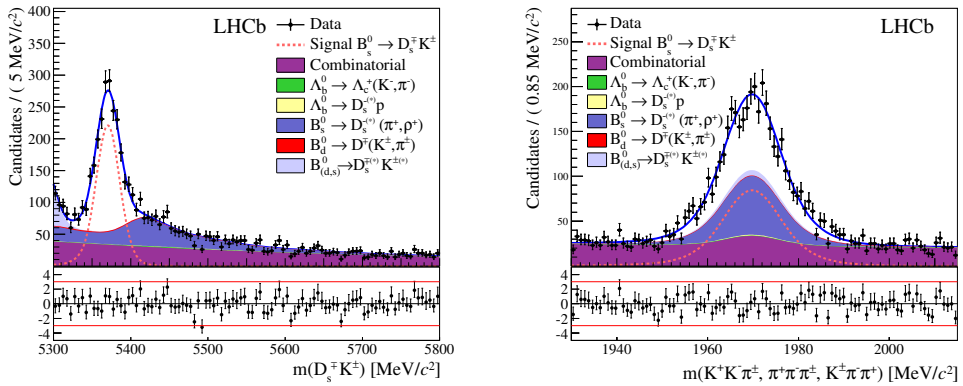
5.4 Flavour tagging

In order to fit the $B_s^0 \to D_s^\mp K^\pm$ time-dependent decay rates from Eq. 2.14, the initial flavour of the B_s^0 mesons needs to be known. Therefore the $B_s^0 \to D_s^\mp K^\pm$ analysis relies on the methods of flavour tagging, cf. Section 4.2. To determine the initial flavour of a B_s^0 candidate a combination of all opposite side (OS) taggers and the same side kaon (SSK) tagger are used. As described in Section 4.2 the mistag distribution must be calibrated for each signal channel. For the calibration of the combined OS tagger sets of flavour specific decays are used as a signal proxy: $B^+ \to J/\psi K^+$, $B^+ \to \overline{D}^0 \pi^+$, $B^0 \to D^{*-} \mu^+ \nu_\mu$, $B^0 \to J/\psi K^{*0}$, and $B_s^0 \to D_s^- \pi^+$. For each of these flavour specific channels a set of calibration parameters (cf. Eq. 4.4) is determined and averaged. The combined OS tagger calibration parameters for the $B_s^0 \to D_s^\mp K^\pm$ signal are

$$p_0^{\mathrm{OS}} = 0.3834 \pm 0.0014\,(\mathrm{stat}) \pm 0.0040\,(\mathrm{syst}) \text{ and} \tag{5.1}$$
$$p_1^{\mathrm{OS}} = 0.972 \pm 0.012\,(\mathrm{stat}) \pm 0.035\,(\mathrm{syst})\,.$$

The systematic uncertainties account for possible dependence on the final states of the calibration channel, the kinematic properties of the B_s^0 signal, and the event properties [66].
The SSK tagger is calibrated with $B_s^0 \to D_s^- \pi^+$ events accounting for systematic effects from the $B_s^0 \to D_s^- \pi^+$ fit model, the uncertainty of the decay time resolution, and the background components in the $B_s^0 \to D_s^- \pi^+$ fit. The calibration parameters for the SSK tagger are determined to be

$$p_0^{\mathrm{SSK}} = 0.4244 \pm 0.0086\,(\mathrm{stat}) \pm 0.0071\,(\mathrm{syst}) \text{ and} \tag{5.2}$$
$$p_1^{\mathrm{SSK}} = 1.255 \pm 0.140\,(\mathrm{stat}) \pm 0.104\,(\mathrm{syst})\,.$$



(a) (b)

(c)

Figure 5.1: Results of the multivariate fit to the B_s^0 invariant mass (a), the D_s^- invariant mass (b), and the companion hadron $\mathrm{DLL}_{K\pi}$ distributions(c). The distributions are the sum of all D_s^- final states. The $B_s^0 \to D_s^{\mp} K^{\pm}$ signal component is given as a dotted red line, while various similar background components are grouped to increase the readability [66].

In order to use the calibration parameters from $B_s^0 \to D_s^- \pi^+$ for the $B_s^0 \to D_s^{\mp} K^{\pm}$ analysis, the mistag distributions and the calibration parameters should be equal. The calibration is cross-checked on simulated events, while the mistag distributions are compared on data and on simulated events. No systematic difference is found. Finally, a possible tagging asymmetry is taken into account, which is determined by studying the calibration separately for B^0 and $\overline{B}^0$ mesons. While the tagging asymmetry can be measured by using $B^+ \to J/\psi K^+$ for the combined OS tagger, it is more difficult for the same side kaon tagger. Due to the fast B_s^0 oscillations it is not possible to determine the asymmetry directly from $B_s^0 \to D_s^- \pi^+$ data. Instead

charged kaons associated with prompt D_s^- mesons, i.e. mesons that originate from the PV, are used to measure the tagging asymmetry. Here, the D_s^- momentum distribution is reweighted to match the momentum distribution of $B_s^0 \to D_s^- \pi^+$ candidates.
The dataset is split into three disjunct tagging samples. One where the events are only tagged by the combined OS tagger, one where the events are only tagged by the SSK tagger and one where both taggers made a tag decision. In the latter case the outputs of both taggers are combined to give a combined tag decision and a combined mistag probability ω. The final tagging power (cf. Eq. 4.2) for the three samples is listed in Table 5.4.

Table 5.4: Summary of the tagging performance for $B_s^0 \to D_s^- \pi^+$ events. The data is split into samples where only the OS, only the SSK or combined tagger (OS+SSK) tagged the event.

Sample	$\epsilon_{\mathbf{eff}}$ [%]
OS-only	$1.61 \pm 0.03 \pm 0.08$
SSK-only	$1.31 \pm 0.22 \pm 0.17$
OS+SSK	$2.15 \pm 0.05 \pm 0.09$

5.5 Determination of CP asymmetry parameters

The general strategy is to fit the $B_s^0 \to D_s^\mp K^\pm$ samples to measure the CP asymmetry parameters from Eq. 2.20 and use the fit to $B_s^0 \to D_s^- \pi^+$ data, which measures the oscillation frequency of B_s^0 mesons Δm_s, as a cross-check for the CP asymmetry fit. The dataset is split into the five final states of the D_s^- meson (cf. Table 5.1) considered in the $m_{B_s^0}$, $m_{D_s^-}$, and $\mathrm{DLL}_{K\pi}$ dimensions, the output of the three tagging categories, and whether the initial state meson is tagged as a B_s^0 or $\overline{B}_s^0$ meson. In order to measure the CP observables, the decay time distributions for the signal and backgrounds, the decay time acceptance and the decay time resolution need be to known. As the per candidate estimated decay time uncertainties and the per candidate estimated mistag probabilities are two additional fit dimensions, these distributions have to be modelled, too.

5.5.1 Decay time resolution and acceptance

A single gaussian resolution function is estimated from simulated signal events and convoluted with the decay time distributions for the signal and background. However, the width of the resolution function needs to be calibrated on data to reflect the

measured decay time resolution. The calibration is taken from the measurement of the B_s^0–$\overline{B}_s^0$ oscillation frequency Δm_s on $B_s^0 \to D_s^- \pi^+$ data [177]. The distributions of the per candidate decay time uncertainty are taken from $B_s^0 \to D_s^- \pi^+$ data for the signal, from the high B_s^0 mass sidebands for the combinatorial background and from simulated events for the other backgrounds, where the simulation is weighted to account for discrepancies with respect to the measured $B_s^0 \to D_s^- \pi^+$ data. The decay time acceptance, i.e. the biasing effect on the decay time distribution induced by the detector acceptance, trigger, reconstruction, and selection requirements, is correlated with the *CP* observables and hence must be determined prior to the final fit. An effective acceptance description, using segments of polynomial functions, is extracted from a fit to the decay time distribution of the $B_s^0 \to D_s^- \pi^+$ control channel. In this fit, the oscillation frequency Δm_s is measured along with the decay time acceptance parameters as a sanity check for the acceptance function. The result of $\Delta m_s = 17.772 \pm 0.022\,\mathrm{ps}^{-1}$, where the uncertainty is only statistical, is fully consistent with the published result of the dedicated analysis [177] $\Delta m_s = 17.768 \pm 0.023\,(\mathrm{stat}) \pm 0.006\,(\mathrm{syst})\,\mathrm{ps}^{-1}$. Hence, the decay time acceptance is able to correctly describe the $B_s^0 \to D_s^- \pi^+$ data. The resulting decay time acceptance is corrected for the differences between the control channel and the $B_s^0 \to D_s^{\mp} K^{\pm}$ signal obtained from simulated events. It is assumed that the acceptance is independent of the actual decay mode and does not need to be evaluated for each background channel separately. This assumption is cross-checked using large simulated samples of $B_s^0 \to D^{*\pm} \pi^{\mp}$ and $B_s^0 \to D_s^- \rho^+$ events. Here, the discrepancy between the signal and the control channels is expected to be large, due to the mis-reconstruction and mis-identification of particles, which are necessary for the control channels to be categorised as background for the $B_s^0 \to D_s^{\mp} K^{\pm}$ signal. No difference in the acceptance is observed.

5.5.2 Mistag distributions

To take the mistag probabilities into account as an additional observable in the cFit, the shape of the mistag distribution has to be modelled for each tagging decision (OS, SSK, and OS+SSK) and for all signal and background components. For all channels involving a B_s^0 decay, the distributions are evaluated with a weighted data sample of $B_s^0 \to D_s^- \pi^+$ events. For the B^0 and Λ_b^0 backgrounds a weighted set of $B^0 \to D_s^- \pi^+$ decays is used to determine the mistag distribution for the OS taggers. No explicit treatment of the Λ_b^0 is necessary, due to similar kinematic properties in the event. The mistag model for the combinatorial background is taken from the high-mass sideband of $B_s^0 \to D_s^- \pi^+$ decays.

5.5.3 The decay time distributions

The decay time description of the signal follows the time dependent decay rates given in Eq. 2.14. However, the background components do not share the same decay time dependence, but can be combined into groups of decays that show similar behaviour to $B_s^0 \to D_s^{\mp} K^{\pm}$ or $B_s^0 \to D_s^- \pi^+$, or are non-oscillating decays, or contribute to the combinatorial background. The list of exclusive background channels considered in each category is shown in Table 5.5. The background components that have a similar lifetime behaviour to $B_s^0 \to D_s^{\mp} K^{\pm}$ events are those that are able to decay into the finale state f and its CP conjugate $\bar{f}$ at any given decay time independent from the flavour of the b hadron. These backgrounds are also sensitive to CP asymmetries and have therefore a separate set of CP observables. These CP parameters are effectively constrained to remain within physical limitations in contrast to the signal CP observables, which are freely floating.
The second group of background channels behaves like $B_s^0 \to D_s^- \pi^+$ events, which decay only into one CP final state. In other words, the B meson can only decay into the final state f, while the anti-meson $\overline{B}$ can only decay into the CP conjugate final state $\bar{f}$. In this case, the CP parameters are fixed to $C_f = C_{\bar{f}} = 1$ and $S_f = S_{\bar{f}} = \mathcal{A}_f^{\Delta\Gamma} = \mathcal{A}_{\bar{f}}^{\Delta\Gamma} = 0$, due to the missing interference of decay amplitudes.
For the non-oscillating background, the parametrisation is a single exponential where the lifetimes are constrained to the corresponding world averages [175].
The combinatorial background is composed of random tracks forming a so-called *fake* B_s^0 candidate. Hence, the kinematics are not determined by any distinct particle decay time. As the tagging decisions show a dependences on the decay kinematics, the decay time distribution for the combinatorial background is evaluated separately for each tagging category (untagged, OS, SSK, OS+SSK) using the high B_s^0 mass sideband ($m_{B_s^0} > 5.7\,\mathrm{GeV}/c^2$) of $B_s^0 \to D_s^{\mp} K^{\pm}$ and $B_s^0 \to D_s^- \pi^+$ data. A double exponential function is used to model the decay time distributions for the combinatorial background.

Table 5.5: List of background channels categorised in groups of decay time behaviour. The explanation is given in the text. The $B_s^0 \to D_s^{\mp} K^{\pm}$ and $B_s^0 \to D_s^- \pi^+$ decays are listed as background, because in the fit to $B_s^0 \to D_s^{\mp} K^{\pm}$ candidates the $B_s^0 \to D_s^- \pi^+$ component is considered as a background and vice versa.

like $B_s^0 \to D_s^- \pi^+$	**like $B_s^0 \to D_s^{\mp} K^{\pm}$**	**non-oscillating**
$B_s^0 \to D_s^- \pi^+$	$B_s^0 \to D_s^{\mp} K^{\pm}$	$\Lambda_b^0 \to D_s^- p$
$B_s^0 \to D_s^{*-} \pi^+$	$B^0 \to D^{\pm} \pi^{\mp}$	$\Lambda_b^0 \to D_s^{*-} p$
$B_s^0 \to D_s^- \rho^+$		$\Lambda_b^0 \to \Lambda_c^+ K^-$
$B^0 \to D_s^- K^+$		$\Lambda_b^0 \to \Lambda_c^+ \pi^-$
$B^0 \to D^- K^+$		
$B^0 \to D_s^- \pi^+$		

5.5.4 Results

For the final fit the decay time distributions including the flavour tagging are convoluted with the single gaussian decay time resolution model and multiplied by the decay time acceptance, per-event decay time resolution, mistag probability functions, and the templates for the B_s^0 mass, D_s^- mass, and the $\mathrm{DLL}_{K\pi}$ probability distributions. In order to increase the stability of the fit to $B_s^0 \to D_s^\mp K^\pm$ events the oscillation frequency Δm_s, the decay rate Γ_s, the decay rate difference $\Delta\Gamma_s$, and the correlation between the latter two $\rho(\Gamma_s,\Delta\Gamma_s)$ for the B_s^0 meson are fixed in the final fit. Additionally, the decay widths for the B^0 and Λ_b^0 hadron are fixed to values obtained from independent measurements [69, 175, 177, 178]:

$$\begin{aligned}
\Gamma_s &= 0.661 \pm 0.007\,\mathrm{ps}^{-1}\,, \quad \Delta\Gamma_s = 0.106 \pm 0.013\,\mathrm{ps}^{-1}\,,\\
\Gamma_d &= 0.658 \pm 0.003\,\mathrm{ps}^{-1}\,, \quad \Gamma_{\Lambda_b^0} = 0.676 \pm 0.003\,\mathrm{ps}^{-1}\,,\\
\Delta m_s &= 17.768 \pm 0.024\,\mathrm{ps}^{-1}\,, \quad \rho(\Gamma_s,\Delta\Gamma_s) = -0.39\,.
\end{aligned}$$

The final fit is restricted to the mass range around the B_s^0 and D_s^- mass, while the $\mathrm{DLL}_{K\pi}$ ranges differ for the $B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^- \pi^+$ fit due to the difference in the $\mathrm{DLL}_{K\pi}$ selection. The resulting decay time distribution is fit within $0.4 - 15.0\,\mathrm{ps}$. The number of background events per background channel is fixed to the value obtained by the multivariate fit. The fit projection in the decay time of the B_s^0 candidate is shown in Fig. 5.2. Various background components are grouped to increase the clarity of the plots. Furthermore, the cross-check fit to fully selected $B_s^0 \to D_s^- \pi^+$ events with fixed *CP* observables $C_f = C_{\bar{f}} = 1$ and $S_f = S_{\bar{f}} = \mathcal{A}_f^{\Delta\Gamma} = \mathcal{A}_{\bar{f}}^{\Delta\Gamma} = 0$ measures $\Delta m_s = 17.762 \pm 0.021\,\mathrm{ps}^{-1}$, which is in excellent agreement with the world average [40], the LHCb dedicated analysis [177], and the result previously obtained in the decay time acceptance study 5.5.1. The results of the *CP* observables for the nominal cFit are

$$\begin{aligned}
C_f &= 0.53 \pm 0.25\,(\mathrm{stat}) \pm 0.04\,(\mathrm{syst})\,,\\
S_f &= -1.09 \pm 0.33\,(\mathrm{stat}) \pm 0.08\,(\mathrm{syst})\,,\\
S_{\bar{f}} &= -0.36 \pm 0.34\,(\mathrm{stat}) \pm 0.08\,(\mathrm{syst})\,,\\
\mathcal{A}_f^{\Delta\Gamma} &= 0.37 \pm 0.42\,(\mathrm{stat}) \pm 0.20\,(\mathrm{syst})\,,\\
\mathcal{A}_{\bar{f}}^{\Delta\Gamma} &= 0.20 \pm 0.41\,(\mathrm{stat}) \pm 0.20\,(\mathrm{syst})\,,
\end{aligned} \tag{5.3}$$

where the quoted uncertainties are the statistical and the systematic uncertainties. In order to validate the results, the data is split into two samples for each of the following quantities: the LHCb magnet polarity, the trigger decision and the BDTG response. Furthermore, a large number of pseudo-experiments is generated taking into account production, detection and *CP* asymmetries as well as oscillations of signal and background components. The data splits and the pseudo-experiments are used to test the result for consistency and to evaluated the systematic uncertainties.

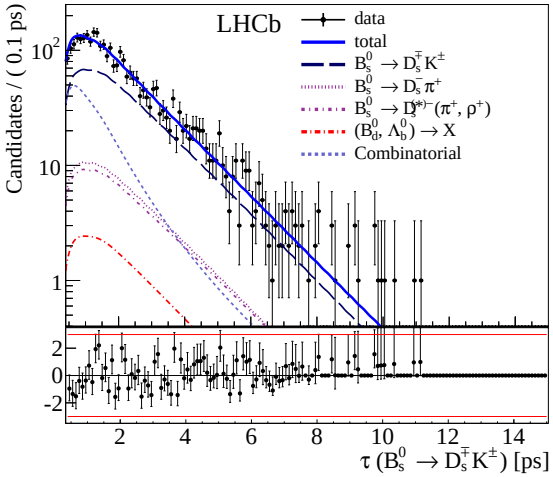


Figure 5.2: Distribution of the decay time of selected $B_s^0 \to D_s^\mp K^\pm$ candidates (black points). The result of the total fit is shown as a blue solid line, where the $B_s^0 \to D_s^\mp K^\pm$ signal component is dashed black and the $B_s^0 \to D_s^- \pi^+$ in dotted violet. For clarity, the $B_s^0 \to D^{(*)-}(\pi^+, \rho^+)$ (dashed dotted violet) component is the sum of $B_s^0 \to D_s^{*-}\pi^+$ and $B_s^0 \to D_s^- \rho^+$, while $(B^0, \Lambda_b^0) \to X$ (dashed dotted red) combines all B^0 and Λ_b^0 hadron background components. The combinatorial background is shown as a dotted blue line [66].

The latter arise from fitting with fixed values of Δm_s, Γ_s, $\Delta\Gamma_s$ and combinatorial background parameters, and the imperfect knowledge of the decay time acceptance and resolution. In the results of Eq. 5.3 only the total systematic uncertainty is quoted. The full list of systematic effects and the correlations with the CP observables can be found in [66]. Furthermore, the CP observables show no bias when compared to the input values of the pseudo-experiments.

Comparison with the cross-check sFit

The cross-check sFit uses the same selection and multivariate fit as the nominal cFit. However, the multivariate fit extracts *sWeights*, which are then used to only fit the statistically background subtracted signal decay time distribution. The results for the CP observables are compared to the cFit in Table 5.6. A difference is observed in the S_f parameter, which is not covered by the systematic uncertainty estimate. However, studies with large sets of pseudo-experiments show that both fitters yield unbiased results. Hence, no additional systematic is introduced on the S_f result.

Table 5.6: Results of the fits to the *CP* asymmetry parameters for the nominal cFit and the cross-check sFit. The listed uncertainties are the statistical (first) and systematic (second) uncertainties.

Observable	**cFit result**	**sFit result**
C_f	$0.53 \pm 0.25 \pm 0.04$	$0.52 \pm 0.25 \pm 0.04$
S_f	$-1.09 \pm 0.33 \pm 0.08$	$-0.90 \pm 0.31 \pm 0.06$
$S_{\bar{f}}$	$-0.36 \pm 0.34 \pm 0.08$	$-0.36 \pm 0.34 \pm 0.06$
$\mathcal{A}_f^{\Delta\Gamma}$	$0.37 \pm 0.42 \pm 0.20$	$0.29 \pm 0.42 \pm 0.17$
$\mathcal{A}_{\bar{f}}^{\Delta\Gamma}$	$0.20 \pm 0.41 \pm 0.20$	$0.14 \pm 0.41 \pm 0.18$

5.6 Measurement of the CKM angle γ

The *CP* asymmetry observables C_f, S_f, $S_{\bar{f}}$, $\mathcal{A}_f^{\Delta\Gamma}$, and $\mathcal{A}_{\bar{f}}^{\Delta\Gamma}$ determined in Sections 5.5.4 are used to measure the physics parameters related to the $B_s^0 \to D_s^{\mp} K^{\pm}$ system defined in Eq. 2.20: the CKM angle γ, the ratio of decay amplitudes $r_B^{D_s K}$, and the strong phase difference $\delta_B^{D_s K}$. The phase ϕ_s is taken from [69].

Starting from the measured *CP* observables in Table 5.6, the strategy to extract γ is to maximise a multivariate gaussian likelihood

$$\mathcal{L}(\vec{\alpha}) \propto \exp\left(-\frac{1}{2}\left(\vec{A}(\vec{\alpha}) - \vec{A}_{\text{obs}}\right)^T V^{-1} \left(\vec{A}(\vec{\alpha}) - \vec{A}_{\text{obs}}\right)\right). \tag{5.4}$$

Here, the physics parameters define the vector $\vec{\alpha} = (\gamma, r_B^{D_s K}, \delta_B^{D_s K})$, $\vec{A}(\vec{\alpha})$ marks the relations of the *CP* observables as a function of the physics parameters from Eq. 2.20, and $\vec{A}_{\text{obs}} = (C_f, S_f, S_{\bar{f}}, \mathcal{A}_f^{\Delta\Gamma}, \mathcal{A}_{\bar{f}}^{\Delta\Gamma})$ is the measured result of the *CP* observables from Table 5.6. The matrix V is the full experimental covariance matrix built from the statistical and systematic correlation matrices given in Table A.2 and Table A.3 in the Appendix A.3. The maximum value of the likelihood is reached, when the difference between the measured *CP* observables $\vec{A}_{\text{obs}}$ and the theoretical calculations $\vec{A}(\vec{\alpha})$ is minimal.

5.6.1 Expected sensitivity

A set of pseudo-experiments is used to evaluate the expected sensitivity on the CKM angle γ. These pseudo-datasets accounting for the production asymmetry and oscillation of B_s^0 and $\overline{B}_s^0$ mesons, and the detection asymmetry for oppositely charged kaons are equal to those used to validate the two decay time fitters. Therefore, the sensitivity on the CKM angle γ can be calculated from the set of *CP* observables, $\vec{A}_{\text{toy}}$, coming from the results of the nominal cFit applied to the validation toys. The

strategy is to determine the expected sensitivity from pseudo-experiments generated at $\gamma = 70^\circ$ and perform the study again at the best fit value coming from data. In this way, it can be checked whether the expectation depends on the CKM angle γ. The reason this study is performed a posteriori is that the time the cFit needs to converge is in the order of hours. This prevents the toy studies to be repeated for each single change in the analysis, which would result in fitting $\mathcal{O}(10^3)$ pseudo-datasets with the nominal fitter.
Due to the unconstrained determination of the *CP* observables, which are physically constrained (cf. Eq. 2.19) and the intrinsic ambiguity in the evaluation of γ, a few selection criteria are applied in order to perform and stabilise the minimisation process of the likelihood in Eq. 5.4. Results from the nominal fit to the pseudo-datasets that show too small or too large uncertainties on the *CP* observables A^i_{toy}, are discarded:

$$\frac{\sigma(A^i_{\text{toy}})}{|A^i_{\text{toy}}|} < 1\% \quad \text{and} \quad |\sigma(A^i_{\text{toy}})| > 10\,.$$

For each set of $\vec{A}_{\text{toy}}$ a profile likelihood scan is performed, where the test statistic from Eq. 3.13 is used together with Wilks' theorem (cf. Section 3.3.6) to translate the $\Delta\chi^2$ value into a $\text{p} = 1 - \text{CL}$ value. The CKM angle γ is scanned between 0–360° to find two equivalent solutions. Two examples of resulting $1 - \text{CL}$ curves from these profile likelihood scans are shown in Fig. 5.3.

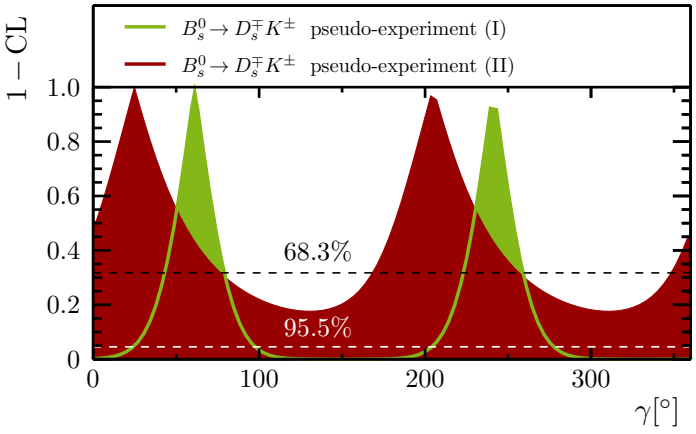
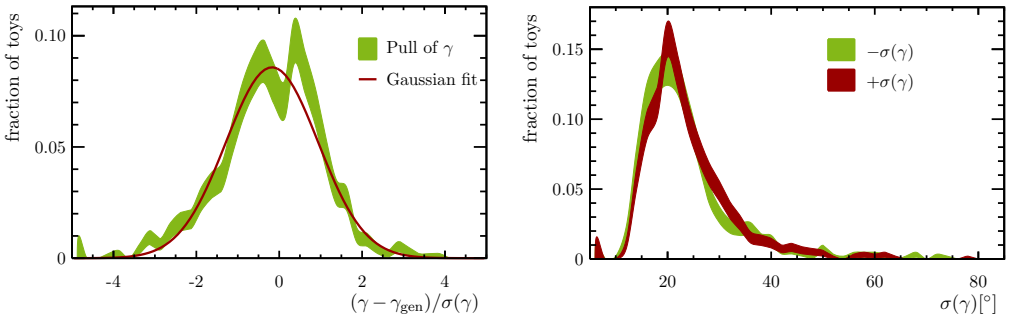


Figure 5.3: $1 - \text{CL}$ curves for the CKM angle γ from profile likelihood scans of $B_s^0 \to D_s^\mp K^\pm$ pseudo-experiments. (I) A case in light green where the γ measurement works as expected and (II) a case in red where the 68.3% confidence interval is interrupted by the scan range boundary. In this case periodicity conditions are used to determine the confidence interval.

It can be seen that the extraction of the CKM angle γ has an intrinsic periodicity of 180°. Therefore, the results are presented modulo 180°, where the nominal result is chosen as the solution closest to the value used to generate the respective pseudo-experiment. Whenever the nominal determination of the 68.3% confidence interval is interrupted by a scan range boundary, 0° or 360°, periodicity conditions are used to fold the physics parameter back into the fit range (red curve of pseudo-experiment II in Fig. 5.3). Each scan results in the best fit value and the confidence interval boundaries, which are translated into asymmetric uncertainties, $\pm\sigma(\gamma)$, by calculating the difference of the best fit value and the confidence interval boundaries. For example, the $1-\mathrm{CL}$ curve of pseudo-experiment II in Fig. 5.3 has a slightly larger positive than negative uncertainty. The distributions of the asymmetric uncertainties and the relative difference between best fit value and the value used to generate the pseudo-experiment normalised to the asymmetric uncertainty (pull) are studied. The results are shown in Fig. 5.4. The distributions of the pulls does not exactly follow an unbiased overall gaussian distribution. A gaussian fit to the pull distribution of γ yields $\mu = -0.17 \pm 0.04$ and $s = 1.11 \pm 0.03$ for the mean and width of the gaussian model, respectively. As the pull is expected to be distributed as a gaussian with mean of zero and a width of unity, it can be concluded that the estimation of the CKM angle γ is biased at $\gamma = 70°$ and the profile likelihood scan method used to estimate the uncertainties results in under-coverage, i.e. too small uncertainties. As a consequence, the final estimation of the CKM angle γ from data needs to be performed with the PLUGIN method (cf. Section 3.3.4), as it is expected to have better coverage and accounts for biases in the parameter estimation.



(a) Pull for $\gamma = 70°$ (b) Asymmetric uncertainties for $\gamma = 70°$

Figure 5.4: Distributions for pseudo-experiments generated with $\gamma = 70°$. No periodicity folding is applied. (a) The pull distribution of γ with a gaussian fit and (b) Comparison of positive (red) and negative (green) uncertainties computed as the difference of the best fit value and the 1σ confidence interval boundaries.

The expected sensitivity on the CKM angle γ in $B_s^0 \to D_s^{\mp} K^{\pm}$ decays is defined as the maximum of the expectation values of the two asymmetric uncertainty distributions in Fig. 5.4(b). The maximum is chosen, as the profile likelihood methods is expected to under-cover. The expectation values are

$$\langle -\sigma(\gamma) \rangle \approx 25^\circ \text{ and}$$
$$\langle +\sigma(\gamma) \rangle \approx 27^\circ .$$

Thus, the expected sensitivity on the CKM angle γ is

$$\langle \sigma(\gamma) \rangle \equiv \langle +\sigma(\gamma) \rangle \approx 27^\circ . \tag{5.5}$$

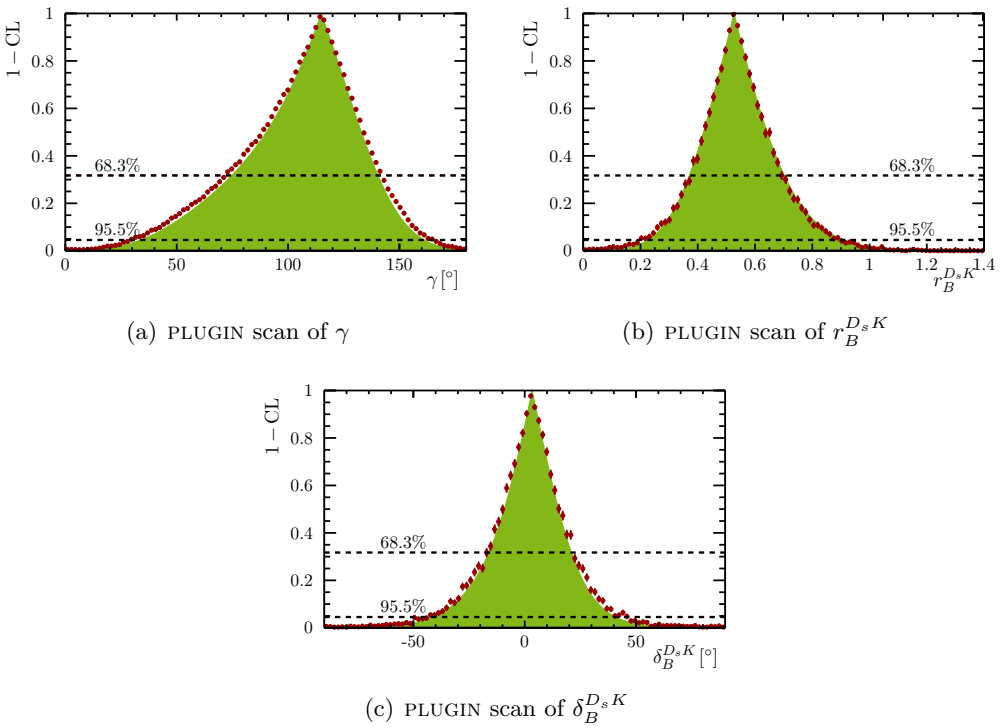
5.6.2 Results for the CKM angle γ

For the determination of the nominal γ result, the vector of observables $\vec{A}_{\text{obs}}$ from the multivariate likelihood in Eq. 5.4 is fixed to the values measured in data (cf. Table 5.6). The first step is an auxiliary profile likelihood scan with 100 scan points between zero and 360° in order to check, if two solutions are found and how the shape of the profile likelihood behaves with respect to the scan boundaries. Both solutions are found and result in identical confidence intervals $\gamma = (115^{+25}_{-40})^\circ$ modulo 180°. The PLUGIN method (cf. Section 3.3.4) is used to get a better estimate of the uncertainties. Now the fit range is restricted to $0 - 180^\circ$. A profile likelihood scan with 100 scan points provides the starting values for the nuisance parameters $\delta_B^{D_s K}$ and $r_B^{D_s K}$ at which the pseudo-experiments are generated. As before, ϕ_s is fixed to the measurement in Ref. [69]. For each scan point $\mathcal{O}(10^3)$ sets of random measurements are generated. This procedure is done for the CKM angle γ, the amplitude ratio $r_B^{D_s K}$, and the strong phase $\delta_B^{D_s K}$ separately resulting in

$$\gamma = (115^{+27}_{-43})^\circ , \tag{5.6}$$
$$r_B^{D_s K} = 0.53^{+17}_{-16} , \text{ and} \tag{5.7}$$
$$\delta_B^{D_s K} = (3^{+19}_{-20})^\circ , \tag{5.8}$$

where angles are given modulo 180° and the uncertainties combine the statistical and systematic uncertainty. In order to break down the respective uncertainty contributions the statistical uncertainty is evaluated using a scan of the profile likelihood ratio where all systematic uncertainties including the uncertainty on ϕ_s are fixed to zero. The ratio of statistical over total uncertainty is applied to the total uncertainty from the nominal PLUGIN result. This assumes that the ratio is not affected by the change to the faster profile likelihood ratio method. The uncertainties are then the respective differences in squares between total, statistical and systematic uncertainty and the uncertainty from ϕ_s yielding

$$\gamma = \left(115^{+26}_{-35}\,(\text{stat})^{+8}_{-25}\,(\text{syst}) \pm 4(\phi_s)\right)^\circ . \tag{5.9}$$



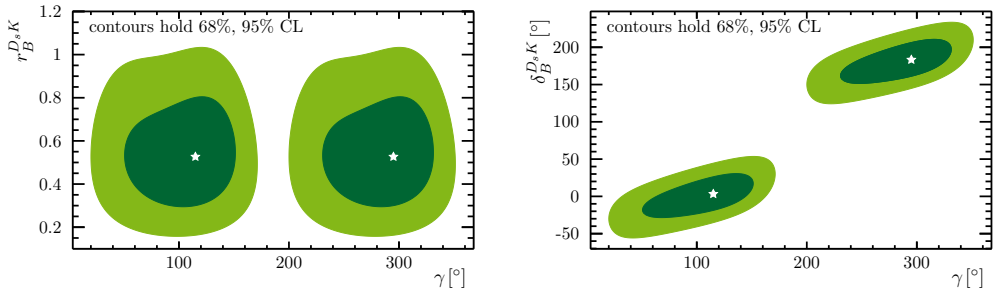
(a) PLUGIN scan of γ

(b) PLUGIN scan of $r_B^{D_sK}$

(c) PLUGIN scan of $\delta_B^{D_sK}$

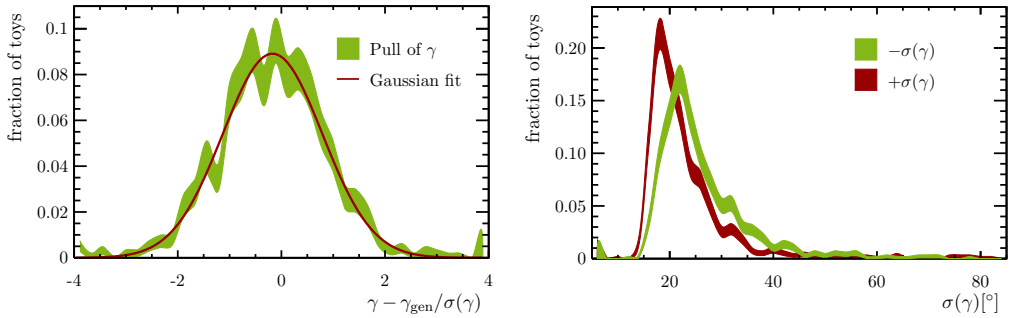
Figure 5.5: $1-\mathrm{CL}$ distributions of the PLUGIN results (red points) and profile likelihood scans (green). (a) The PLUGIN result for the CKM angle γ. (b) Result for the amplitude ratio $r_B^{D_sK}$ and (c) the plugin result for the strong phase $\delta_B^{D_sK}$.

The $1-\mathrm{CL}$ curves are shown for the profile likelihood ratio scan and for the PLUGIN approach in Fig. 5.5. The overall agreement between the profile likelihood ratio scan and the PLUGIN method is excellent. Therefore, the two-dimensional scans in the dimensions of γ, $r_B^{D_sK}$, and $\delta_B^{D_sK}$ are performed with the faster profile likelihood ratio method. The 68.3% and 95.5% contours are shown in Fig 5.6. The ambiguities in the estimations are purposely shown by extending the scan range for γ and $\delta_B^{D_sK}$. As shown in Section 5.6.1 the expected sensitivity on the CKM angle γ is $\approx 27°$ and nearly symmetrical. Hence, it is remarkable that the measurement of the CKM angle γ yields such asymmetric uncertainties of $+\sigma = +27°$ and $-\sigma = -43°$. A set of pseudo-experiments is generated with $\gamma = 116°$ in order to check for any distinct features in this phase space region. Analogously to Section 5.6.1, the distributions of the pull and asymmetric uncertainties are studied. The pull distribution in Fig. 5.7(a) is biased towards negative values. The systematic shift is calculated as $+3.2 \pm 0.9°$. However, as shown in Section 3.3.2, when constructing confidence intervals based on the FC ordering principle, the resulting intervals account for the intrinsic bias of the estimator.



(a) profile likelihood scan for $r_B^{D_sK}$ versus γ (b) profile likelihood scan for $\delta_B^{D_sK}$ versus γ

Figure 5.6: Two-dimensional profile likelihood ratio scans of the two nuisance parameters and the CKM angle γ in $B_s^0 \to D_s^{\mp} K^{\pm}$ decays. The contours are constructed to have 68.3% and 95.5% coverage (cf. Table A.1). (a) Amplitude ratio $r_B^{D_sK}$ versus CKM angle γ. (b) Strong phase $\delta_B^{D_sK}$ versus γ.



(a) Pull for $\gamma = 116°$ (b) Asymmetric uncertainties for $\gamma = 116°$

Figure 5.7: Distributions for pseudo-experiments generated with $\gamma = 116°$. No periodicity folding is applied. (b) The pull distribution of γ with a gaussian fit, and (c) Comparison of positive (red) and negative (green) asymmetric uncertainties computed as the difference of the best fit value and the 1σ confidence interval boundaries.

In addition, the measurement of the *CP* observables is unbiased so that the final γ result is not corrected for the observed bias. The expectation values for the two one-sided uncertainties (cf. Fig. 5.7(b)) are $\langle +\sigma(\gamma) \rangle = 24°$ and $\langle -\sigma(\gamma) \rangle = 28°$. The p-values for the measured uncertainties on γ given the distributions from the pseudo-experiments are 18% for $\langle +\sigma(\gamma) \rangle$ and 9% for $\langle -\sigma(\gamma) \rangle$. This shows that the measured uncertainties can be explained by a statistical fluctuation.
The precision on γ obtained from $B_s^0 \to D_s^{\mp} K^{\pm}$ decays alone is promising. The comparison to other measurements of γ on the same dataset, e.g. from $B^+ \to DK^{\pm}$ with $D \to K_S^0 h^+ h^-$ decays which results in $\gamma = (44^{+43}_{-38})°$ [179], shows that the time-dependent analysis of $B_s^0 \to D_s^{\mp} K^{\pm}$ decays is competitive. Additionally, in

$B_s^0 \to D_s^\mp K^\pm$ decays the CKM angle γ can be determined independently of the D meson system parameters, which makes the analysis an important input for the γ combination, where the set of γ measurements performed at the LHCb experiment is combined to achieve higher precision.

5.7 Future prospects

The RunII of the LHC collider will help to increase the precision on the CKM angle γ. In order to estimate the expected sensitivity on γ from $B_s^0 \to D_s^\mp K^\pm$ decays two scenarios in terms of integrated luminosities are considered. For the first scenario a conservative additional $5\,\text{fb}^{-1}$ at centre-of-mass energies of $\sqrt{s} = 13\,\text{TeV}$ is assumed, while the second scenario assumes an additional integrated luminosity of $10\,\text{fb}^{-1}$. However, as e.g. measured by the LHCb collaboration [180, 181], the $b\bar{b}$ production cross-section increases approximately linear with higher centre-of-mass energies. Hence, the measured number of events per channel scales according to

$$\mathcal{N} = \frac{L_{13} \times 13\,\text{TeV}/8\,\text{TeV}}{L}\,, \tag{5.10}$$

where L_{13} is the assumed additional integrated luminosity at $\sqrt{s} = 13\,\text{TeV}$ and L is the integrated luminosity from the analysed dataset scaled to $\sqrt{s} = 8\,\text{TeV}$ centre-of-mass energies.
For the integrated luminosities of $L_{13} = 5\,\text{fb}^{-1}$ and $L_{13} = 10\,\text{fb}^{-1}$ the statistical uncertainties of the measured CP observables in $B_s^0 \to D_s^\mp K^\pm$ decays from Eq. 5.3) are scaled according to $1/\sqrt{\mathcal{N}}$. Then, the nominal method to determine the CKM angle γ (cf. Section 5.6) is performed on the CP observables with scaled uncertainties. The central values, the statistical and the systematic correlations are assumed to be the same as in Section 5.6. The systematic uncertainties are assumed to scale in the same way as the statistical uncertainties. The results of the profile likelihood scans are shown in Fig. 5.8. The expected sensitivity on the CKM angle γ from $B_s^0 \to D_s^\mp K^\pm$ decays is $\sigma(\gamma) \approx 10^\circ$ and $\sigma(\gamma) \approx 8^\circ$ for an estimated total integrated luminosity of $L_{13} = 5\,\text{fb}^{-1}$ and $L_{13} = 10\,\text{fb}^{-1}$, respectively. It is evident that the sensitivity on the CKM angle γ from $B_s^0 \to D_s^\mp K^\pm$ decays alone is not sufficient to set a precise benchmark SM measurement, which can be compared to results from loop processes.

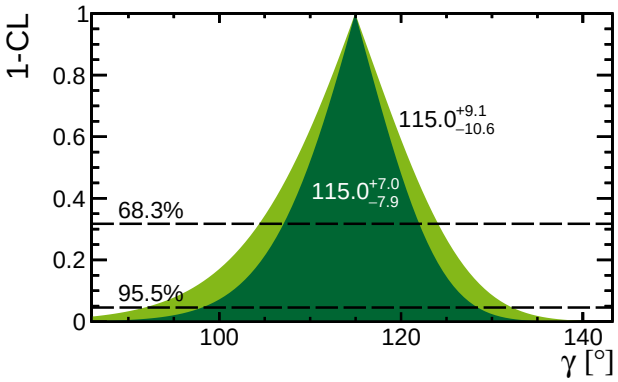


Figure 5.8: Extrapolations of the uncertainties on the CKM angle γ from $B_s^0 \to D_s^{\mp} K^{\pm}$ decays for total integrated luminosities of $L_{13} = 5\,\text{fb}^{-1}$ (light green) and $L_{13} = 10\,\text{fb}^{-1}$ (dark green). Here, the systematic uncertainties are assumed to scale along with the statistical uncertainties.

6 Combination of γ measurements at the LHCb experiment

One of the main aspects of the physics programme of the LHCb experiment is the precise determination of the CKM angle γ. To get the best precision from the available datasets, measurements from a variety of decay channels are combined. For this purpose the GammaCombo framework [182] is developed with the author of this thesis playing a major role in the development and testing of the framework. Although the origin of the GammaCombo framework was the first LHCb γ combination [132], it has been used in other analyses such as the analysis of $B_s^0 \to D_s^{\mp} K^{\pm}$ decays (see Chapter 5), the full likelihood combination of the $B \to \mu^+\mu^-$ results from the LHCb and CMS experiments (see Chapter 8) or in the measurement of V_{ub} [183] to constrain the right-handed coupling component.
For the first LHCb γ combination the method proposed by R. Berger and D. Boos (cf. Section 3.3.5) is implemented as a test of the PLUGIN method. For the input measurement of $B^+ \to DK^+$ with $D \to K^0_S K\pi$ decays, the nominal PLUGIN method is cross-checked with the Berger-Boos method and yield consistent results. Applying the Berger-Boos method to the full combination showed that the computing resources needed to calculate p-values that are not affected by either large statistical fluctuations (controlled by the number of pseudo-experiments) or by a large Berger-Boos correction (controlled by the confidence region of the nuisance parameters) are enormous. Hence, the coverage of the full combination is cross-checked in an explicit study.

6.1 Input measurements

The first update of a combined LHCb result for the CKM angle γ where the time-dependent measurement from $B_s^0 \to D_s^{\mp} K^{\pm}$ events is included, was presented at the CKM conference in 2014 [184]. The measurements that are additionally included are coming from $B \to Dh$ decays, where h denotes either a pion or a kaon. In order to summarise many decay channels in one decay chain symbol, explicit flavour and charge indications are dropped from the illustrations of the decay chains. The analyses are separated mainly by the different final states of the D meson, where the D meson can decay into CP eigenstates such as $D \to K^+K^-$ or $D \to \pi^+\pi^-$ [51, 52].

In other analyses, the D meson can decay either into suppressed non-CP eigenstates such as $D \to K^+\pi^-$ or $D \to K^+\pi^-\pi^+\pi^-$ [53, 55], or into three-body self-conjugate states such as $D \to K^0_S\pi^+\pi^-$ or $D \to K^0_S K^+K^-$ [54]. The analyses considered and the size of the datasets used in the combination are

- $B^+ \to Dh^+$, $D \to hh$, $1\,\text{fb}^{-1}$ [185],
- $B^+ \to Dh^+$, $D \to K\pi\pi\pi$, $1\,\text{fb}^{-1}$ [186],
- $B^+ \to DK^+$, $D \to K^0_S hh$, $3\,\text{fb}^{-1}$ [187],
- $B^+ \to DK^+$, $D \to K^0_S K\pi$, $3\,\text{fb}^{-1}$ [188],
- $B^0 \to DK^{*0}$, $D \to hh$, $3\,\text{fb}^{-1}$ [189], and
- $B^0_s \to D^\mp_s K^\pm$, $1\,\text{fb}^{-1}$, detailed discussion in Chapter 5.

Although facing different experimental challenges, the various LHCb γ measurements follow roughly the same strategy, when considered as an input for the γ combination. A set of CP observables is measured from which the physics parameters can be determined, as described for the $B^0_s \to D^\mp_s K^\pm$ analysis in Section 5.6. The physics parameters, as defined in Section 2.3.1 for the $B^0_s \to D^\mp_s K^\pm$ case, are mainly the amplitude ratios $r^f_{B,D}$ of the interfering decays and the strong phase differences $\delta^f_{B,D}$ for the respective B or D meson decays, and the weak phase γ. However, the parameters differ depending on the final state (marked as superscripts) or the decaying meson, e.g. B or D meson decays (marked in the subscripts). For example, the CP asymmetry $A_{CP}^{DK,KK}$ marks an asymmetry observable defined as the difference of CP conjugated $B \to DK$ meson decays, where the D meson is reconstructed from two-body CP eigenstates K^+K^-. It can experimentally defined and related to the physics parameters as

$$\begin{aligned} A_{CP}^{DK,KK} &\equiv \frac{\Gamma(B^- \to [KK]_D K^-) - \Gamma(B^+ \to [KK]_D K^+)}{\Gamma(B^- \to [KK]_D K^-) + \Gamma(B^+ \to [KK]_D K^+)}\,, \\ &= \frac{2r_B^{DK}\sin\delta_B^{DK}\sin\gamma}{1 + (r_B^{DK})^2 + 2r_B^{DK}\cos\delta_B^{DK}\cos\gamma} + A_{CP}^{\text{dir}}(KK)\,, \end{aligned}$$

where $A_{CP}^{\text{dir}}(KK)$ accounts for possible CP violating effects in the $D \to K^+K^-$ decays. Other examples are the ratio of the partial widths of the interfering suppressed D decay path relative to favoured D decay mode, $R_+^{DK,K\pi}$:

$$\begin{aligned} R_+^{DK,K\pi} &\equiv \frac{\Gamma(B^+ \to [\pi K]_D K^+)}{\Gamma(B^+ \to [K\pi]_D K^+)}\,, \\ &= \frac{(r_B^{DK})^2 + (r_D^{K\pi})^2 + 2r_B^{DK}r_D^{K\pi}\cos(\delta_B^{DK} + \delta_D^{K\pi} + \gamma)}{1 + (r_B^{DK})^2(r_D^{K\pi})^2 + 2r_B^{DK}r_D^{K\pi}\cos(\delta_B^{DK} - \delta_D^{K\pi} + \gamma)}\,, \end{aligned}$$

or the ratio of total $B \to DK$ decays over total $B \to D\pi$ decays in the case, where the D is reconstructed from K^+K^-, $R^{KK}_{K/\pi}$:

$$\begin{aligned} R^{KK}_{K/\pi} &\equiv \frac{\Gamma(B^- \to [KK]_D K^-) + \Gamma(B^+ \to [KK]_D K^+)}{\Gamma(B^- \to [KK]_D \pi^-) + \Gamma(B^+ \to [KK]_D \pi^+)} , \\ &= R_{\text{Cab.}} \frac{1 + (r_B^{DK})^2 + 2 r_B^{DK} \cos\delta_B^{DK} \cos\gamma}{1 + (r_B^{D\pi})^2 + 2 r_B^{D\pi} \cos\delta_B^{D\pi} \cos\gamma} . \end{aligned}$$

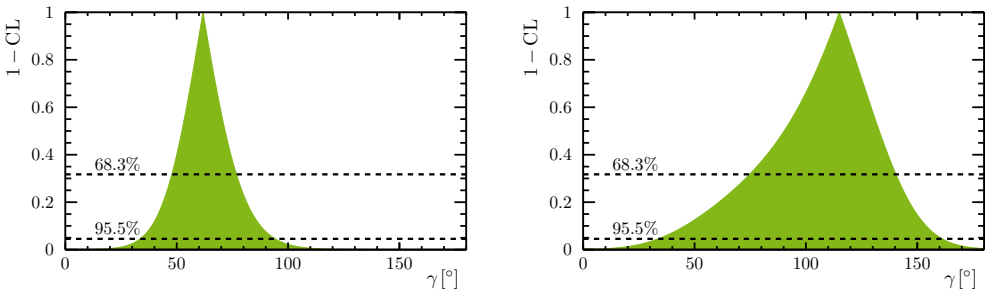
Here, $R_{\text{Cab.}}$ is the ratio of partial widths $R_{\text{Cab.}} \equiv \Gamma(B^- \to D^0 K^-)/\Gamma(B^- \to D^0 \pi^-)$ between the kaon and pion decays. It is evident that a single observable has intrinsic ambiguities with respect to the weak phase γ and depends on several physics parameters. This is also reflected when the single input measurements are interpreted in terms of the CKM angle γ in Fig. 6.1. The results are shown modulo 180°, which means that modes such as $B^+ \to DK^\pm$ with $D \to K^0_S h^+ h^-$ or $B^0_s \to D^\mp_s K^\pm$ have an intrinsic two-fold ambiguity, while the mode $B^0 \to DK^{*0}$ with $D \to hh$ has an eight-fold ambiguity. Hence, the precision on γ arises from the combination of observables and the simultaneous extraction of the CKM angle γ. The full list of observables and the relations to the physics parameters can be found in Refs. [132, 184, 190].

6.2 Combination method and validation

The strategy to extract γ from the set of observables follows the one described in Section 5.6 using multivariate gaussian PDFs, f_i, in a combined product likelihood

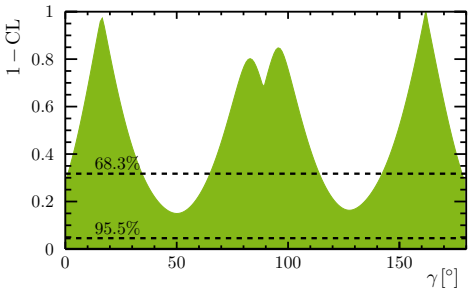
$$\mathcal{L}(\vec{\alpha}) = \prod_i f_i(\vec{A}_{i,\text{obs}} | \vec{A}_i(\vec{\alpha_i})) , \tag{6.1}$$

where the index i marks the input measurements. The precision on the physics parameters of interest (in particular γ) is increased by constraining the D meson system or B mixing parameters to the results of dedicated measurements [65, 69, 191–193]. The constrains are modelled with gaussian PDFs. The non-gaussian constraint on the coherence factor, $\kappa_D^{K3\pi}$, which accounts for possible dilution effects from the interference of intermediate resonances in multi-body D meson decays [186, 194], and the strong phase difference $\delta_D^{K3\pi}$ [192] is provided as a likelihood histogram and is fully taken into account. The flowchart in Fig. 6.2 shows the interplay of three example measurements with the physics parameters of interest (γ, r_B^{DK}, and δ_B^{DK}), the parameters exclusive for a single channel, and the constrained D meson system parameters. While some parameters are only present in one analysis (e.g. $\delta_B^{DK^*}$ or $r_B^{DK^*}$ for the $B^0 \to DK^{*0}$ analysis in Fig. 6.2) other parameters can be constrained from external measurements (e.g. the ratio of interfering D decay amplitudes $r_D^{K\pi}$).

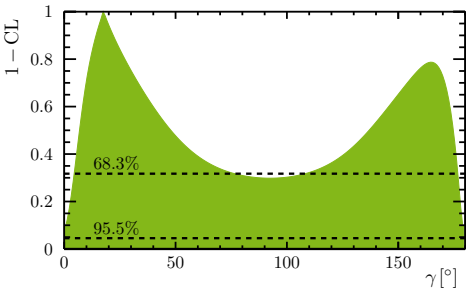


(a) $B^+ \to DK^+$ with $D \to K^0_S hh$

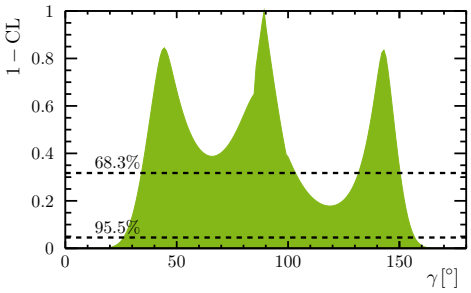
(b) $B^0_s \to D^\mp_s K^\pm$



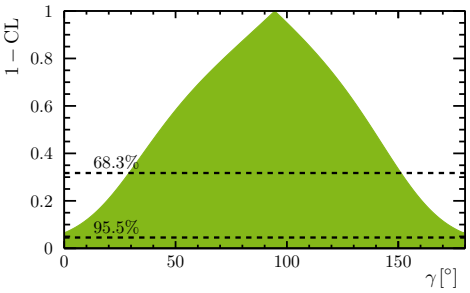
(c) $B^0 \to DK^{*0}$ with $D \to hh$



(d) $B^+ \to DK^+$ with $D \to K^0_S K\pi$



(e) $B^+ \to Dh^+$ with $D \to hh$



(f) $B^+ \to Dh^+$ with $D \to K\pi\pi\pi$

Figure 6.1: $1 - \mathrm{CL}$ curves for the CKM angle γ for all input measurements to the γ combination — the curves are obtained using the profile likelihood ratio scan. The measurements are (a) $B^+ \to DK^+$ with $D \to K^0_S hh$, (b) $B^0_s \to D^\mp_s K^\pm$, (c) $B^0 \to DK^{*0}$ with $D \to hh$, (d) $B^+ \to DK^+$ with $D \to K^0_S K\pi$, (e) $B^+ \to Dh^+$ with $D \to hh$, and (f) $B^+ \to Dh^+$ with $D \to K\pi\pi\pi$.

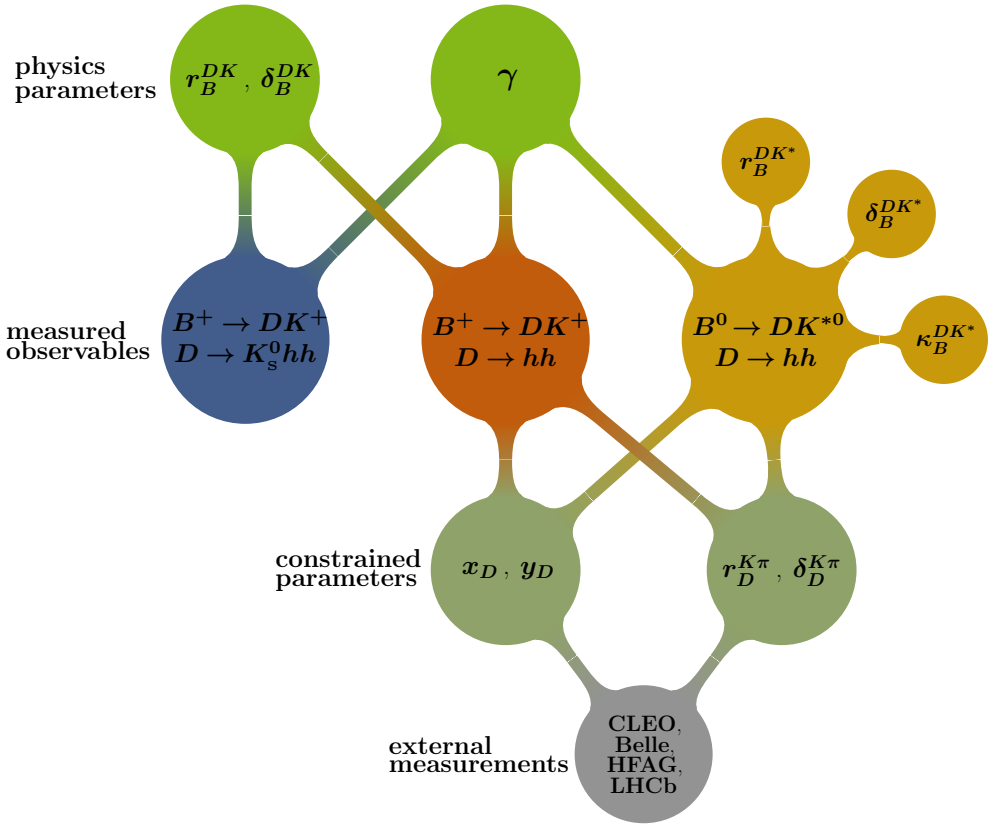


Figure 6.2: Connection of physics parameters (green) in the example combination of two $B \to DK$ and one $B \to DK^*$ decay to measure the physics parameter of interest (γ, r_B^{DK}, and δ_B^{DK}). In the combination there are measurements such as $B^+ \to DK^+$ with $D \to K_S^0 hh$ that are fully described by the parameters of interest and others such as $B^+ \to Dh^+$, $D \to hh$ that introduce D meson system parameter describing the decay ($r_D^{K\pi}$, $\delta_D^{K\pi}$) or the mixing in the D meson system (x_D, y_D). The D system parameters are constrained from external measurements when available.

The precision on the parameters of interest increases by adding disjunct measurements that depend on the parameters of interest.
Two statistical methods are used to extract γ from the set of input measurements. A fast profile likelihood ratio scan and the PLUGIN method (cf. section 3.3.4). As the profile likelihood ratio scan is used to define the values that are used in the generation of the pseudo-experiments, always both results are calculated. Although the PLUGIN method is based on the unified frequentist method, the strict profiling of the nuisance parameters possibly results in an incorrect coverage. The coverage α of both methods is studied based on pseudo-experiments in the following way:

1.) Generate N sets of experiment observables $\vec{A}_{i,\text{toy}}$ from the respective PDFs with all parameters set to the best fit values.

2.) Determine the confidence intervals with the profile likelihood ratio scan and the PLUGIN method using the resulting combined likelihood from Eq. 6.1 and replacing $\vec{A}_{i,\text{obs}}$ by $\vec{A}_{i,\text{toy}}$.

3.) Calculate the fraction of times α' where the true value used in the generation of the pseudo-experiments is covered by the resulting confidence interval.

For an ideal method and $N \to \infty$ the calculated fraction α' is equal to the supposed coverage α. The results of the coverage test is shown in Table 6.1. Although the toy-based PLUGIN method has a better coverage than the profile likelihood ratio scan, still a 6% under-coverage is found.

Table 6.1: Coverage test of the profile likelihood ratio scan (LL scan) and the PLUGIN method full γ combination. The tests are performed for the 1σ, 2σ, and 3σ coverage probabilities of $\alpha = 68.27\%$, $\alpha = 95.45\%$, and $\alpha = 99.73$, respectively. All numbers in the Table are given in percent.

full	$\boldsymbol{\alpha'}$ **(LL scan)**	**residual** $\boldsymbol{(\alpha' - \alpha)}$	$\boldsymbol{\alpha'}$ **(plugin)**	**residual** $\boldsymbol{(\alpha' - \alpha)}$
$\alpha = 68.27$	55.93 ± 0.91	-12.34	61.54 ± 0.89	-6.73
$\alpha = 95.45$	91.60 ± 0.51	-3.85	94.02 ± 0.43	-1.43
$\alpha = 99.73$	99.29 ± 0.15	-0.44	99.29 ± 0.15	-0.44

6.3 Results

The full combination of $B \to Dh$ decays to determine the CKM angle γ results in $\gamma = 78.9^\circ$ with PLUGIN confidence intervals of

$$\gamma \in [71.5^\circ, 84.7^\circ] \text{ @68.3\% CL}\,, \tag{6.2}$$

$$\gamma \in [54.6^\circ, 91.4^\circ] \text{ @95.5\% CL}\,. \tag{6.3}$$

The results are represented modulo 180°. The $1 - \text{CL}$ curve for the combination of CKM angle γ is shown in Fig. 6.3. The impact of a single measurement on the γ combination is studied by removing one input measurement at a time from the combination (only $B \to DK$ decays are used for this study). It can clearly be seen that the $B^+ \to DK^+$, $D \to K^0_S hh$ measurement has the greatest impact on the result. This is fully expected, as the measurement of γ in $B^+ \to DK^+$, $D \to K^0_S hh$ decays provides the most precise values of γ coming from a single analysis.
Compared to the legacy results from Belle [50] $\gamma = (68^{+15}_{-14})^\circ$ and BaBar [49] $\gamma = (69^{+17}_{-16})^\circ$, the result of the LHCb experiment $\gamma = (78.9^{+5.8}_{-7.4})^\circ$ has half the total uncertainty without fully exploiting the complete dataset of the first run of the LHC collider.

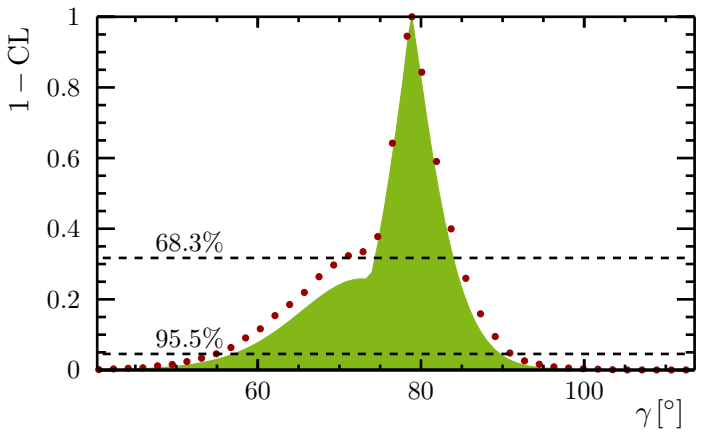


Figure 6.3: 1−CL curves for the CKM angle γ obtained from the profile likelihood ratio scan (light green) and the PLUGIN method (red points).

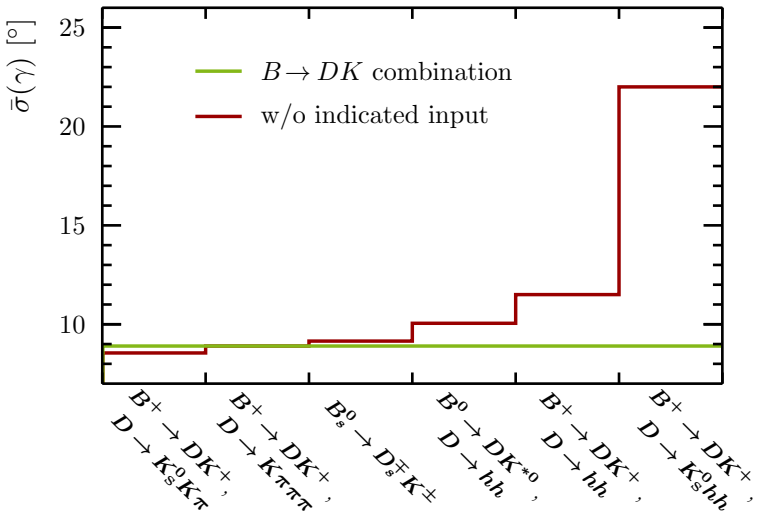


Figure 6.4: Impact of various measurements the γ combination, when only $B \to DK$ are considered. The mean uncertainty $\bar{\sigma}(\gamma)$ is evaluated using the profile likelihood ratio scan and by neglecting each single measurement from the combination but keeping the others.

6.4 Future Prospects

As the highest precision on the CKM angle γ arises from the combination of the single analyses, the expected performance for Run II is studied. The strategy to extrapolate the measurement of the CKM angle γ to higher luminosities is equivalent to the one used in Section 5.7. In order to estimate the combined sensitivity on γ the statistical and systematic uncertainties of all input *CP* observables are scaled with a factor of $1/\sqrt{\mathcal{N}}$ (cf. Eq. 5.10). However, the constraints from external measurements are not modified. The $1-\mathrm{CL}$ curves from profile likelihood scans and the results for γ are shown in Fig. 6.5 for additional integrated luminosities of $L_{13} = 5\,\mathrm{fb}^{-1}$ and $L_{13} = 10\,\mathrm{fb}^{-1}$. It can be seen that the central value shifts for the two extrapolations. This is due to the fact, that the input measurements have different integrated luminosities to start with and are scaled to have the same integrated luminosity. In other words, the relative weight of the input measurements changes with the scaling. The second effect, which alters the central value and the expected uncertainties, is that measurements are combined that have distinct different central values (cf. Fig. 6.1). By scaling the statistical and systematic uncertainties of the input *CP* observables, but keeping the central values as measured in the respective analyses, results in a combination of incompatible measurements. This effectively leads to a plateau in the sensitivity on the CKM angle γ, as it is reflected in the expected uncertainties of $\sigma(\gamma) \approx 1.9^\circ$ and $\sigma(\gamma) \approx 2.2^\circ$ for an additional integrated luminosity of $L_{13} = 5\,\mathrm{fb}^{-1}$ and $L_{13} = 10\,\mathrm{fb}^{-1}$, respectively. In order to increase the precision of the measurement of the CKM angle γ to a level of $< 1^\circ$, more integrated luminosity is needed. This will be provided by the upgrade of the LHCb detector during the second long shutdown of the LHC in 2018 [195].

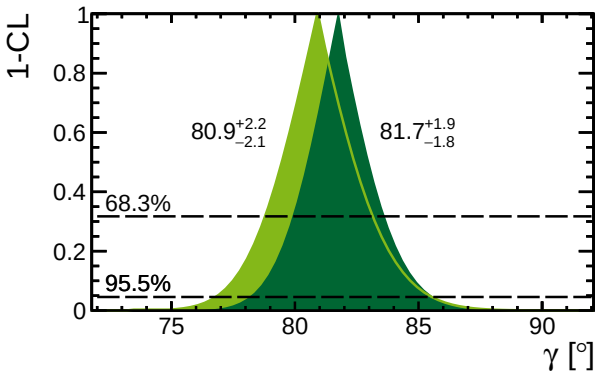


Figure 6.5: $1-\mathrm{CL}$ curve for the full γ combination for assumed additional integrated luminosities of $L_{13} = 5\,\mathrm{fb}^{-1}$ (light green) and $L_{13} = 10\,\mathrm{fb}^{-1}$ (dark green) using the profile likelihood scan method.

7 Measurement of the branching fractions of $B \to \mu^+\mu^-$ decays at the LHCb experiment

This chapter describes the analysis of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays based on the full RunI dataset of the LHCb experiment corresponding to $3\,\text{fb}^{-1}$ of pp interactions at a centre-of-mass energy of $\sqrt{s}=7$ and $8\,\text{TeV}$. The work presented in this thesis is embedded into an analysis group of about 20 scientists and is concentrated on the investigation of the background channels in Section 7.6.

7.1 Analysis strategy

The naive formula to measure the branching fractions for $B \to \mu^+\mu^-$ decays reads

$$\begin{aligned}\mathcal{B}(B_s^0 \to \mu^+\mu^-) &= \frac{N_{B\to\mu^+\mu^-}}{N_B} \\ &= \frac{N^{\text{obs}}_{B\to\mu^+\mu^-}}{\epsilon^{\text{tot}}_{B\to\mu^+\mu^-}} \times \frac{1}{L_{\text{int}} \times \sigma(b\bar{b}) 2 f_{s,d}}\,,\end{aligned}$$

where $N^{\text{obs}}_{B\to\mu^+\mu^-}$ is the observed number of $B \to \mu^+\mu^-$ candidates, $\epsilon^{\text{tot}}_{B\to\mu^+\mu^-}$ the total selection efficiency, L_{int} the integrated luminosity, $\sigma(b\bar{b})$ the production cross-section for a $b\bar{b}$ pair, and f_s (f_d) the hadronisation probability for a b quark to hadronise into a B_s^0 (B^0) meson. However, due to relatively large uncertainties on the $b\bar{b}$ cross-section the total number of B_s^0 mesons is determined from the normalisation channels, $B^+ \to J/\psi\, K^+$ and $B^0 \to K^+\pi^-$, so that the branching fractions for the $B \to \mu^+\mu^-$ decays are measured relatively to two normalisation decays. The analysis studies decays of B^0 and B_s^0 mesons into a muon pair. However, all optimisations are performed with respect to $B_s^0 \to \mu^+\mu^-$ decays, as these have the larger expected branching fraction (cf. Eq. 2.25).
After a loose pre-selection the main separation between background and signal is due to the classifier of a Boosted Decision Tree (BDT) and the invariant mass distribution of the dimuon system. The invariant mass region around the B_s^0 and B^0 peak is excluded from the analysis until the very end to avoid any kind of human bias.

The results for the $B \to \mu^+\mu^-$ branching fractions come from a simultaneous extended maximum likelihood fit to the invariant mass of the dimuon system in eight bins of the BDT classifier. In this way the full information from the selected events can be exploited. The expected and observed significance is evaluated from a fine binning scheme within the signal search windows $\pm 60\,\mathrm{MeV}/c^2$ around the B_s^0 and B^0 mass peaks assuming an SM signal and an extrapolation of the background PDFs into the signal region. In order to determine the correct number of signal candidates from the maximum likelihood fit and to get an accurate extrapolated background estimate in the signal mass region, a list of possible background decays must be carefully studied. Each background which either can bias the combinatorial background estimate or contributes significantly to the signal region must be explicitly modelled in the final fit.

7.2 Selection

The event selection is organized in several steps. The first step is the trigger, where mainly high momentum tracks or good quality muons that form a common vertex with a high associated invariant mass are selected. The most important criteria for the trigger are discussed in Section 4.1.3.
The following step is a pre-selection, which reduces the number of events to a level, where the analysis can be performed on an event-by-event basis. This selection is described in Section 7.2.1. The final classification into signal and background is done by a boosted decision tree (BDT) [159,160] implemented in the TMVA package [196], which is briefly described in Section 7.2.2.

7.2.1 Event selection

After the trigger selection the number events in the dataset, $\mathcal{O}(10^8)$, is still too large to be processed efficiently. Hence, a centrally controlled loose pre-selection, called *stripping*, is applied. These selection steps are unchanged compared to the previous analysis [86] and are therefore only briefly reviewed. Furthermore, the selection criteria for the control channel $B^+ \to J/\psi\, K^+$ are kept as close to the $B \to \mu^+\mu^-$ signal selection as possible by applying an equivalent set of cuts with the exception of particle identification and J/ψ invariant mass cuts.

The main characteristics of the $B \to \mu^+\mu^-$ signals are two well reconstructed muons that form a good common vertex, which can be significantly separated from the primary vertex due to the lifetime of the B mesons. Additionally, the negative reconstructed momentum vector of the dimuon system points straight back to the primary vertex. As a consequence, the main cuts in this selection step are the

separation of the muons from the primary vertex (or impact parameter quality, IPχ^2), vertex quality (χ^2_{vertex}) and flight distance quality (FDχ^2) of the dimuon system, and the impact parameter quality of the B meson. This effectively reduces the dataset to 3×10^5 candidates, in the invariant dimuon range of $m_{\mu\mu} \in [4.9,6.0]\,\text{GeV}/c^2$ while the signal efficiency is $\approx 34\%$. The efficiency combines the stripping and reconstruction efficiencies.
In order to increase the background suppression a first boosted decision tree (BDTS) is trained on simulated signal and background events. As input variables the BDTS uses the impact parameter and the impact parameter quality of the B candidate, the vertex quality, distance of closest approach as well as the minimum impact parameter of the two muons, and the angle between the reconstructed B momentum and the vector defined by the difference of the primary and secondary vertex. The cut on the BDTS classifier (BDTS> 0.05) is unchanged from the previous publication [115]. Additionally, the binary classifier `IsMuon` and a combination of cuts on particle identification variables from Eq. 4.1 of $\text{DLL}_{K\pi} < 10$ and $\text{DLL}_{\mu\pi} > -5.0$ are applied on the muons of the $B \to \mu^+\mu^-$ candidates. The full selection reduces the number of candidates to 5.5×10^4. Hence, the effective background level is reduced by four orders of magnitude, while the signal selection and reconstruction efficiency is $\approx 30\%$.

7.2.2 Signal and background classification

The final classification in signal and background events is done by a second BDT, which is trained on simulated samples of $B \to \mu^+\mu^-$ signal and on inclusive $b\bar{b} \to \mu^+\mu^- X$ background events using twelve discriminating variables. These variables fully exploit the characteristics of $B \to \mu^+\mu^-$ decays. Firstly, two muons form a separated secondary vertex and the B meson candidate points back to the primary vertex. Therefore, the decay time and the impact parameter of the B signal candidate, the minimum impact parameter and the distance of closest approach of the two muons are used in the BDT categorisation. Secondly $B \to \mu^+\mu^-$ events are two body decays, whereas random combinations of two muons are associated with additional tracks. Variables exploiting the features of two body decays are the absolute difference of the pseudorapidities ($\Delta\eta$) and the difference of azimuthal angles ($\Delta\varphi$) of the two muon candidates, and two track isolation criteria applied to the B candidate and the muons. The B mesons isolation has been used by CDF before [100] and is defined as

$$I(B) = \frac{p_T(B)}{p_T(B) + \sum_i p_T(i)}\,,$$

where the transverse momentum of the B candidate is normalised to the transverse momentum of the B candidate plus the sum of all tracks that are within a cone around the B meson momentum with a radius of $\sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} < 1$. The muon

isolation is the number of good two-track vertices made of one signal muon and another random track in the event. The other four variables entering the BDT training are the transverse momentum of the B candidate, the angle of the muon momentum in the dimuon rest frame and the vector perpendicular to the beam axis and the momentum of the B candidate, and two angles connecting the signal momenta with the approximated momentum of the second non-signal b hadron produced in the pp interaction. The first angle is between the signal B momentum in the laboratory frame and the other non-signal b hadron. The second angle is between the momentum of the positive signal muon in the rest frame of the B candidate and the momentum of the non-signal b hadron.
The resulting BDT classifier is tested for non-linear correlations with the invariant mass of the dimuon system, which can bias the background estimation and lead to a fake signal peak. Only a small linear correlation of 1% is found between the BDT classifier and the invariant mass of the $b\bar{b} \to \mu^+\mu^- X$ background, which has a negligible effect on the estimation of the background in the signal region. The BDT classifier is constructed to achieve a uniform distribution for the signal, so that the signal over background ratio increases for higher BDT classifier values. The expected distribution of the BDT classifier is shown in Fig. 7.1. The distributions are obtained from $b\bar{b} \to \mu^+\mu^- X$ background simulation and from $B \to h^+h'^-$ decays as a $B \to \mu^+\mu^-$ signal proxy. It can be seen that the distribution of the BDT classifier for the expected signal is uniformly distributed, while the combinatorial background peaks for lower values of the BDT classifier. The binning scheme was optimised for a previous analysis [115] to achieve the maximum expected significance defined as the poissonian likelihood ratio of the signal plus background over the background only hypothesis.

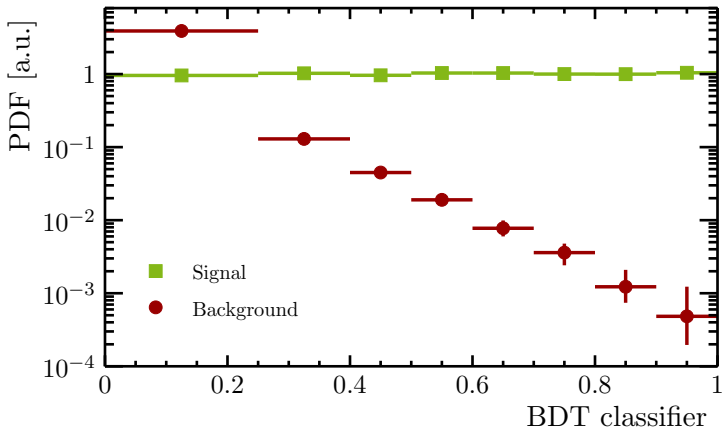


Figure 7.1: Expected distribution of the BDT classifier using $b\bar{b} \to \mu^+\mu^- X$ background simulation (red dots) and $B \to h^+h'^-$ data as a signal proxy for $B \to \mu^+\mu^-$ signal decays (green squares).

The expected BDT distribution shows that the sensitivity for the $B \to \mu^+\mu^-$ signal, approximated by the signal over background ratio, increases for each BDT bin. While the first bin is purely dominated by background, for the BDT bins with BDT>0.7 the combinatorial background is suppressed by more than three orders of magnitude.

7.3 Decay time dependent effects

Due to $B-\overline{B}$ oscillations, the mixture of heavy and light B meson eigenstates depends on the decay time of the mother B particle. This dependence can be parametrised by y_s and $\mathcal{A}_{\mu\mu}^{\Delta\Gamma}$, which are defined as the difference of total decay rates and the difference of $B \to \mu^+\mu^-$ decay rates of the heavy and light B eigenstates normalised to the respective sum:

$$y_s \equiv \frac{\Gamma_L - \Gamma_H}{\Gamma_L + \Gamma_H} \text{ and} \tag{7.1}$$

$$\mathcal{A}_{\mu\mu}^{\Delta\Gamma} \equiv \frac{\Gamma_{B_H^0 \to \mu^+\mu^-} - \Gamma_{B_L^0 \to \mu^+\mu^-}}{\Gamma_{B_H^0 \to \mu^+\mu^-} + \Gamma_{B_L^0 \to \mu^+\mu^-}} \,. \tag{7.2}$$

Here, y_s can be calculated from the averages of the HFAG group [191]. The value for $\mathcal{A}_{\mu\mu}^{\Delta\Gamma}$, however, is unknown and depends on the theoretical model. The SM prediction [197] is $\mathcal{A}_{\mu\mu}^{\Delta\Gamma} = 1$ and therefore the decay time distribution is a single exponential with the slope of Γ_H.
In the samples of simulated $B_s^0 \to \mu^+\mu^-$ events the decay time distribution is a single exponential with a mean lifetime of $\tau = 1.47\,\text{ps}$. Due to the fact that the selection uses properties that are correlated with the B_s^0 meson decay time (flight distance χ^2 or IPχ^2), the decay time distribution is biased after the selection. In order to correct the decay time acceptance bias a factor, δ_ϵ, is calculated that translates efficiencies obtained from simulated events, ϵ^{MC}, into model dependent efficiencies $\epsilon^{\mathcal{A}_{\mu\mu}^{\Delta\Gamma}, y_s}$:

$$\begin{aligned}\delta_\epsilon &= \frac{\epsilon^{\mathcal{A}_{\mu\mu}^{\Delta\Gamma}, y_s}}{\epsilon^{MC}} \\ &= \frac{\int_0^\infty \Gamma(B(t) \to \mu^+\mu^-, \mathcal{A}_{\mu\mu}^{\Delta\Gamma}, y_s)\epsilon(t)\mathrm{d}t}{\int_0^\infty \Gamma(B(t) \to \mu^+\mu^-, \mathcal{A}_{\mu\mu}^{\Delta\Gamma}, y_s)\mathrm{d}t} \times \frac{\int_0^\infty e^{-\Gamma_{MC}t}\mathrm{d}t}{\int_0^\infty e^{-\Gamma_{MC}t}\epsilon(t)\mathrm{d}t} \,.\end{aligned} \tag{7.3}$$

This effect arises in two stages of the analysis. The first stage is the kinematic selection before the final BDT classification of events into signal and background. The time integrated signal efficiencies need to be corrected by the factors

$$\delta_\epsilon(B_s^0) = 1 + (4.57 \pm 0.02)\% \text{ and}$$
$$\delta_\epsilon(B^0) = 1 + (1.50 \pm 0.01)\% \,.$$

The effect is smaller for the $B^0 \to \mu^+\mu^-$ channel, because for B^0 mesons the decay width difference $\Delta\Gamma_d$ is negligible compared to $\Delta\Gamma_s$ in the B_s^0 system.
Due to the fact that the decay time of the B meson is one of the input variables of the final BDT, the resulting decay time distributions are different for each bin of the BDT classifier. In this case, only the B_s^0 signal is affected, because possible correction on $B^0 \to \mu^+\mu^-$ cancel with the B^0 normalisation channels. The correction factors for each bin of the BDT classifier assuming the SM prediction $\mathcal{A}_{\mu\mu}^{\Delta\Gamma} = 1$ are summarised in Table 7.1. The effect of the decay time bias correction is that the efficiencies are increased for most sensitive bins six to eight. The overall sensitivity to $B_s^0 \to \mu^+\mu^-$ decays is effectively increased by about 3%.

Table 7.1: Decay time bias correction factor δ_ϵ per bin of the BDT classifier in the SM ($\mathcal{A}_{\mu\mu}^{\Delta\Gamma} = 1$).

BDT bin	$\boldsymbol{\delta_\epsilon - 1}$ **[%]**
1	-3.11 ± 0.02
2	-1.38 ± 0.03
3	-0.39 ± 0.04
4	$+0.27 \pm 0.04$
5	$+0.72 \pm 0.04$
6	$+1.37 \pm 0.05$
7	$+2.54 \pm 0.05$
8	$+4.74 \pm 0.04$

7.4 Normalisation

The branching fractions of $B \to \mu^+\mu^-$ decays are determined with respect to two well measured normalisation channels in order to cancel systematic uncertainties and eliminate the dependence on the total $b\bar{b}$ cross-section. The normalisation channels are chosen so that one has similar kinematic decay properties, $B^0 \to K^+\pi^-$, and one has similar trigger and muon identification efficiencies, $B^+ \to J/\psi K^+$. The final branching fraction parameters are determined from the number of signal decays measured in the unblinded dataset as

$$\begin{aligned} \mathcal{B}(B \to \mu^+\mu^-) &= \frac{\mathcal{B}_{\text{norm}} \times \epsilon_{\text{norm}}^{\text{tot}}}{N_{\text{norm}}} \times \frac{f_{\text{norm}}}{f_{(s)d}} \times \frac{N_{B\to\mu^+\mu^-}}{\epsilon_{\text{sig}}^{\text{tot}} \times \delta_\epsilon} \\ &= \alpha_{\text{norm}} \times N_{B\to\mu^+\mu^-} \,. \end{aligned} \tag{7.4}$$

Here, $\mathcal{B}_{\text{norm}}$, $\epsilon_{\text{norm}}^{\text{tot}}$ and N_{norm} are the branching fraction, total efficiency and measured number of events for the normalisation channel, respectively. The number of measured $B \to \mu^+\mu^-$ decays is $N_{B\to\mu^+\mu^-}$ and $\epsilon_{\text{sig}}^{\text{tot}} \times \delta_\epsilon$ the model dependent signal efficiency.

The factor $f_{\text{norm}}/f_{(s)d}$ is the hadronisation fraction ratio of the B meson in the normalisation with respect to B meson in the signal channel. By assuming that the hadronisation fractions for B^+ and B^0 mesons are equal, the factor can be obtained from the measurement of $f_{\text{s}}/f_{\text{d}}$ by the LHCb collaboration [198, 199]. The normalisation factor α_{norm} is, hence, the inverse of the total number of B mesons times the model dependent signal efficiency. The total efficiencies are the combination of the acceptance of the LHCb detector, the total trigger efficiency, and the reconstruction and selection efficiency. The acceptance, reconstruction and selection efficiencies are estimated from simulated events and corrected for time dependent effects as described in Section 7.3. Differences seen between the simulated events and the dataset, e.g. the different impact parameter distributions due to wrongly simulated number of total tracks, are assigned as systematic uncertainties. The trigger efficiencies are calculated directly from data using information about which track actually triggered the event [200]. For the $B^+ \to J/\psi K^+$ channel, the trigger efficiency is determined from selected data events, while for the $B^0 \to K^+\pi^-$, due to limited samples sizes, the combined L0 and HLT1 efficiency is estimated from $B^+ \to J/\psi K^+$ and the HLT2 efficiency is evaluated from simulated events. The trigger efficiency for the $B \to \mu^+\mu^-$ signal is taken from $J/\psi \to \mu^+\mu^-$ decays with J/ψ mesons that are separated from the primary vertex. The trigger efficiencies of $J/\psi \to \mu^+\mu^-$ events are calculated in dependence of the maximum transverse momentum and impact parameter of the muons. This trigger efficiency map is folded into the maximum transverse momentum and impact parameter distribution of the muons from $B \to \mu^+\mu^-$ simulated events. The list of input parameter values for the calculation of the normalisation factor α_{norm} is shown in Table 7.2. The final averaged normalisation factors for the selected $3\,\text{fb}^{-1}$ LHCb dataset are

$$\alpha_s = (8.93 \pm 0.62) \times 10^{-11} \text{ and} \tag{7.5}$$

$$\alpha = (2.38 \pm 0.09) \times 10^{-11}\,. \tag{7.6}$$

Assuming the SM branching fractions for the $B \to \mu^+\mu^-$ signal decays (cf. Eq. 2.25) the expected number of total signal candidates are

$$N^{\text{exp}}(B_s^0 \to \mu^+\mu^-) = 39.9 \pm 4.2 \text{ and} \tag{7.7}$$

$$N^{\text{exp}}(B^0 \to \mu^+\mu^-) = 4.5 \pm 0.4\,. \tag{7.8}$$

7.5 Signal shapes

To incorporate signal into the maximum likelihood fit, a parametrisation of the $B \to \mu^+\mu^-$ signal in the invariant dimuon mass is needed. The two $B \to \mu^+\mu^-$ signal components are modelled with a gaussian core function with an power law tail to lower invariant masses to cover final state radiation effects. The mass mean parameters, μ, of the gaussian function are evaluated using $B \to h^+h'^-$ decays that

Table 7.2: Input values for the calculation of the normalisation factor α_{norm}. The inputs that are not studied in the presented analysis are the branching fractions $\mathcal{B}(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+) = (6.025 \pm 0.205) \times 10^{-5}$ and $\mathcal{B}(B^0 \to K^+\pi^-) = (1.94 \pm 0.06) \times 10^{-5}$ [175] as well as the ratio of production fraction $f_s/f_d = 0.259 \pm 0.015$ [199]. The different efficiencies are relative to the corresponding $B_s^0 \to \mu^+\mu^-$ signal efficiency. Given are the ratios of acceptance efficiencies, $r(\epsilon^{\text{acc}})$, selection and reconstruction efficiencies, $r(\epsilon^{\text{sel}})$, and trigger efficiencies, $r(\epsilon^{\text{trig}})$. Values are separately evaluated for the centre-of-mass energies, $\sqrt{s}$, to account for a possible dependence on the different LHCb running conditions.

$\sqrt{s}$	$\delta_\epsilon - 1$ [%]	$r(\epsilon^{\text{acc}})$ [%]	$r(\epsilon^{\text{sel}})$ [%]	$r(\epsilon^{\text{trig}})$ [%]	N_{norm} [10^3]
			$B^+ \to J/\psi K^+$		
7 TeV	4.57 ± 0.02	88.9 ± 0.6	47.8 ± 0.1	95.5 ± 0.2	355.2 ± 1.2
8 TeV			47.3 ± 0.1	93.7 ± 0.3	761.1 ± 2.5
			$B^0 \to K^+\pi^-$		
7 TeV	1.50 ± 0.01	99.9 ± 0.7	84.7 ± 2.4	5.0 ± 0.2	10.8 ± 0.4
8 TeV			84.0 ± 2.6	5.9 ± 0.2	26.7 ± 0.4

pass the requirement of BDTS > 0.1 as a signal proxy. The $B^0 \to \pi^+\pi^-$, $B^0 \to K^+\pi^-$, $B_s^0 \to K^+K^-$, and $B_s^0 \to \pi^+K^-$ decays included in the $B \to h^+h'^-$ background sample are separated by particle identification criteria of $\text{DLL}_{K\pi} > 10$ and $\text{DLL}_{Kp} > 2$ for kaons, and $\text{DLL}_{Kp} < -10$ and $\text{DLL}_{\pi p} > 2$ for pions. The resulting B^0 and B_s^0 invariant mass distributions are fit with a gaussian core function with two power laws to both sides of the gaussian. The mass mean parameters are determined to

$$\mu_{B^0} = (5284.90 \pm 0.22)\,\text{MeV}/c^2 \text{ and}$$
$$\mu_{B_s^0} = (5371.85 \pm 0.25)\,\text{MeV}/c^2\,,$$

where the uncertainty reflects the combined statistical and systematic uncertainty. The mass resolutions or the widths of the gaussian core functions for the signal shapes are studied with two different methods. First, the mass resolution is taken from fits to charmonium and bottomonium decays into two muons. The resolution depends on the mass of the decaying hadron, so the different values for the resolution from the quarkonia dimuon decays are interpolated using a polynomial function. The second method uses the same $B \to h^+h'^-$ decays as for the estimation of the mass mean parameter. The results of both methods are averaged to

$$\sigma_{B^0} = (22.83 \pm 0.43)\,\text{MeV}/c^2 \text{ and}$$
$$\sigma_{B_s^0} = (23.24 \pm 0.45)\,\text{MeV}/c^2\,,$$

where the uncertainty is the combined statistical and systematic uncertainty. The transition points where the total fit function changes from the gaussian core function to the exponential tail are determined from simulated events.

7.6 Estimation of background components

The main background component in the search windows for B_s^0 and B^0 signal decays are random combinations of two muons (combinatorial background). These originate mostly from two different b flavoured hadron decays faking a single B candidate. This background is modelled by a single exponential with a different slope in each BDT bin. The yields are determined by a fit to the invariant mass sidebands defined as $m_{\mu\mu} \in [4.9, 6.0]\,\mathrm{GeV}/c^2$, where the blinded region, $\pm 60\,\mathrm{MeV}/c^2$ around the B^0 and B_s^0 masses are excluded. While most backgrounds from exclusive b hadron decays are negligible in the signal search windows, the lower mass sideband is populated with various background decays. Therefore, fitting the invariant mass sidebands with a single exponential results in a wrong estimation of the background contribution in the signal search windows. Hence, a list of exclusive decays is studied in the Sections 7.6.2, 7.6.3, and 7.6.4. The procedure to estimate the mis-identification probabilities for hadrons reconstructed as muons is described in Section 7.6.1.

7.6.1 Mis-identification procedure

In order to study the importance of certain background components that have at least one wrongly identified hadron, it is necessary to know the rates at which hadrons are mis-identified as muons. This probabilities must be evaluated with the full selection (cf. Section 7.2) applied. The background decays that are studied in the presented analysis involve mis-identification processes for pions, kaons, and protons that are wrongly identified as muons. Hence, the respective mis-identification efficiencies are evaluated. A clean sample of $D^0 \to K^\pm\pi^\mp$ events from $D^{*\pm} \to D^0\pi^\pm$ decays is used for the pion and kaon mis-identification rates, while for the proton rates $\Lambda^0 \to p\pi^-$ decays are used.
In the two body decays one track (the *tag* track) needs to be identified with high precision, while the track for which the mis-identification rate is calculated (the *probe* track) must not be responsible for triggering the event (TIS). The latter is necessary to avoid a possible bias from the first two trigger stages.
The resulting data samples with the full $B \to \mu^+\mu^-$ muon identification criteria applied are fit in dependence of the momentum (p) and transverse momentum (p_T) of the probe track. From these fits the number of wrongly identified signal decays can be measured yielding the mis-identification efficiency. Some bins of the p–p_T

space are empty due to either kinematic restrictions or the lack of events passing the selection. In the following, whenever a track from a potential background source needs to be evaluated in one of these empty bins the value from the next higher filled momentum bin within the same transverse momentum category is used. The efficiencies for identifying a pion, kaon, or proton as a muon under the $B \to \mu^+\mu^-$ muon selection are shown in Figs. 7.2.

For the combined muon requirements of the binary `IsMuon` operator and the particle identification information $\mathrm{DLL}_{\mu\pi} > -5$ and $\mathrm{DLL}_{K\pi} < 10$, the hadron mis-identification rates are found to be less than 1% for all hadrons and almost the complete phase space. The exception are low momentum pions ($\lesssim 10\,\mathrm{GeV}/c$) where the mis-identification probability is between 1–4%.

Possible systematic effects resulting from differences between the evaluation channels and the potential backgrounds are studied in terms of multiplicity (number of tracks in an event) for the $D^0 \to K^\pm\pi^\mp$ sample and flight distance of the heavy hadron for the $\Lambda^0 \to p\pi^-$ decays. Within the statistical uncertainties no dependence is found.

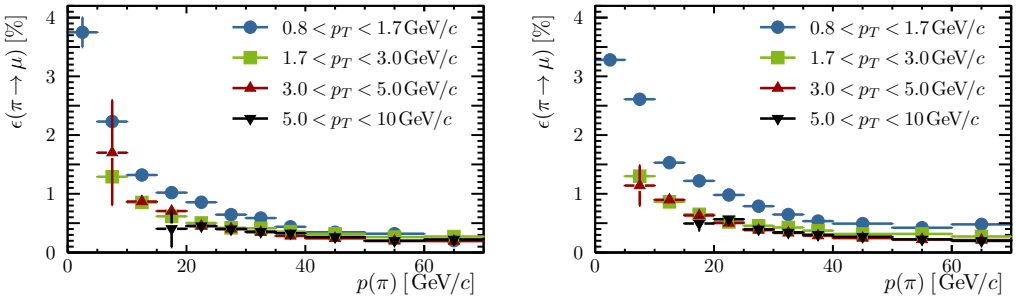
7.6.2 Doubly mis-identified decays

The decays faking a $B \to \mu^+\mu^-$ candidate, with two hadrons wrongly identified as muons, are mainly $B \to h^+h'^-$ decays, where h is either a kaon or a pion. Four exclusive channels ($B^0 \to \pi^+\pi^-$, $B^0 \to K^+\pi^-$, $B_s^0 \to K^+K^-$, and $B_s^0 \to \pi^+K^-$) are studied to evaluate the expected yield of events, the combined invariant mass shape, and the fraction of events in each BDT bin with simulated events.

The mis-identification probabilities from Section 7.6.1 are folded into the momentum and transverse momentum distributions of the $B \to h^+h'^-$ simulation samples. To obtain an average double mis-identification efficiency, $\epsilon_{hh\to\mu\mu}$, the separate efficiencies are weighted according to the branching fraction of the respective decay modes. The branching fractions used for this averaging procedure are summarised in Table 7.3. The average double mis-identification efficiency is cross-checked with a data-driven approach. Here, a $B \to h^+h'^-$ sample triggered independently from the signal tracks (TIS) where one hadron is mis-identified as a muon is used to calculate the single hadron mis-identification probability. This probability is squared to emulate the double mis-identification process. The result is compatible with the efficiencies from simulated events, but has at least a factor of 30 higher uncertainties due to a very small sample size of $B \to h^+h'^-$ TIS events.

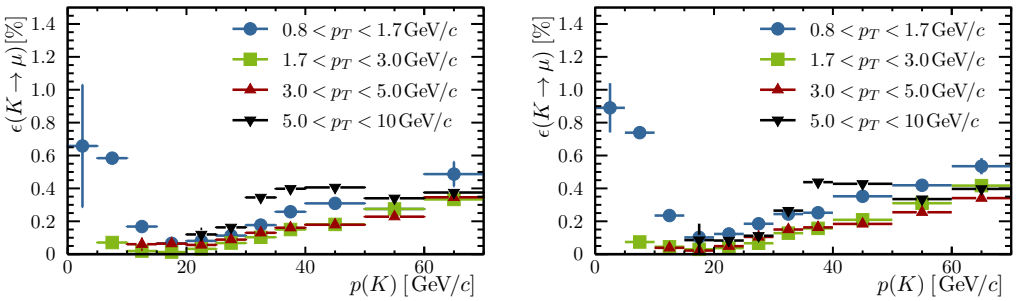
Table 7.3: Branching fractions of the different $B \to h^+h'^-$ decays [175] in units of 10^{-6} and the b hadron fragmentation ratio $f_\mathrm{s}/f_\mathrm{d}$ [199].

$B^0 \to \pi^+\pi^-$	$B^0 \to K^+\pi^-$	$B_s^0 \to K^+K^-$	$B_s^0 \to \pi^+K^-$	$f_\mathrm{s}/f_\mathrm{d}$
5.2 ± 0.2	19.4 ± 0.6	26.4 ± 2.8	5.3 ± 1.0	0.259 ± 0.015



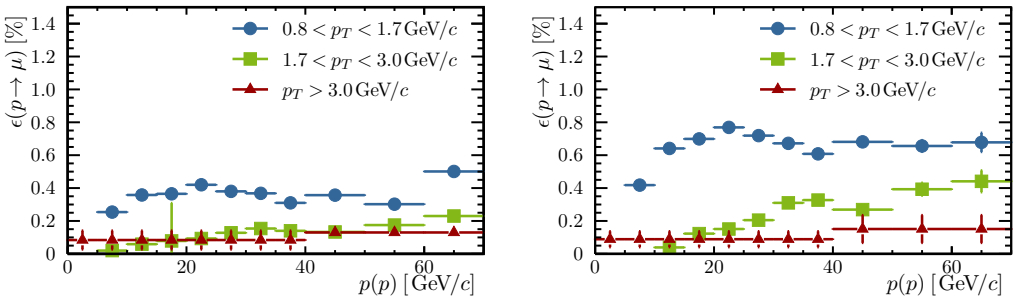
(a) $\pi \to \mu$ at $\sqrt{s} = 7\,\mathrm{TeV}$

(b) $\pi \to \mu$ at $\sqrt{s} = 8\,\mathrm{TeV}$



(c) $K \to \mu$ at $\sqrt{s} = 7\,\mathrm{TeV}$

(d) $K \to \mu$ at $\sqrt{s} = 8\,\mathrm{TeV}$



(e) $p \to \mu$ at $\sqrt{s} = 7\,\mathrm{TeV}$

(f) $p \to \mu$ at $\sqrt{s} = 8\,\mathrm{TeV}$

Figure 7.2: Mis-identification probabilities in percent for pions, kaons, and protons reconstructed as muons for different p_T ranges. (a) and (b) show $\pi \to \mu$, (c) and (d) $K \to \mu$, and (e) and (f) show $p \to \mu$ mis-identification rates for $\sqrt{s} = 7\,\mathrm{TeV}$ data on the left and for $\sqrt{s} = 8\,\mathrm{TeV}$ data on the right. Numbers from Ref. [201].

The average double mis-identification efficiency is used to calculate the number of $B \to h^+h'^-$ decays wrongly reconstructed as $B \to \mu^+\mu^-$ decays according to

$$N_{B\to h^+h'^-\to\mu^+\mu^-} = \epsilon^{\text{TRIG|SEL}}_{B\to\mu^+\mu^-} \times \epsilon_{hh\to\mu\mu} \times \frac{N^{\text{TIS}}_{B\to h^+h'^-}}{\epsilon^{\text{TIS}}\epsilon^{\text{HLT2}}}\,. \tag{7.9}$$

Here, $\epsilon^{\text{TRIG|SEL}}_{B\to\mu^+\mu^-}$ is the signal trigger efficiency after the selection, $N^{\text{TIS}}_{B\to h^+h'^-}$ is the number of $B \to h^+h'^-$ TIS events, and ϵ^{TIS} (ϵ^{HLT2}) is the combined L0 and HLT1 TIS (HLT2 TIS) efficiency for $B \to h^+h'^-$ decays. Using the values given in Table 7.4 the expected number of $B \to h^+h'^-$ background events in the selected LHCb $B \to \mu^+\mu^-$ dataset is $N_{B\to h^+h'^-\to\mu^+\mu^-} = 14.6 \pm 1.3$.

Table 7.4: Values entering the calculation of the expected number of $B \to h^+h'^-$ decays.

Parameter	$\sqrt{s} = 8\,\text{TeV}$ **dataset**	$\sqrt{s} = 7\,\text{TeV}$ **dataset**
$N^{\text{TIS}}_{B\to h^+h'^-}$	49653 ± 507	20143 ± 572
$\epsilon_{hh\to\mu\mu}$	$(1.2 \pm 0.1) \times 10^{-5}$	$(1.1 \pm 0.1) \times 10^{-5}$
ϵ^{HLT2} [%]	91.6 ± 0.2	91.5 ± 0.3
ϵ^{TIS} [%]	$5.92 \pm 0.04_{\text{stat}} \pm 0.4_{\text{syst}}$	$5.05 \pm 0.04_{\text{stat}} \pm 0.4_{\text{syst}}$
$\epsilon^{\text{TRIG|SEL}}_{B\to\mu^+\mu^-}$ [%]	$92.4 \pm 0.3_{\text{stat}} \pm 1.9_{\text{syst}}$	$92.1 \pm 0.5_{\text{stat}} \pm 1.6_{\text{syst}}$

Besides the total number of expected background events, it is important to estimate how these events are distributed in the bins of the discriminating BDT classifier. Therefore, an invariant dimuon mass shape is extracted from the combined sample of simulated $B \to h^+h'^-$ events. Due to the low mis-identification rates, the invariant mass shape is influenced by decay in flight processes, $\pi^+/K^+ \to \mu^+\nu_\mu$ [201]. The muon is measured in the muon stations and the neutrino escapes the detector without being measured and, hence, alters the invariant dimuon mass distributions. Due to the limited number of simulated events, it is not possible to request simultaneous decays in flight for each final state hadron. Therefore, one decay in flight is requested from which a distribution of relative momentum change, $(p_{\text{rec}} - p_{\text{true}})/p_{\text{true}}$, is measured. This distribution is used to smear the momentum distribution of the second hadron to emulate double decay in flight events. The resulting non-parametric invariant mass template [174] together with the two signal distributions (cf. Section 7.5) normalised to the number of expected events in the full $3\,\text{fb}^{-1}$ dataset using the SM predictions is shown in Fig. 7.3. It is found that $(8.8^{+3.0}_{-2.1})\%$ and $(48.0^{+20}_{-8})\%$ of the $B \to h^+h'^-$ background events are situated in the $\pm 60\,\text{MeV}/c^2$ signal windows around the B^0_s and B^0 mass, respectively. Hence, the $B \to h^+h'^-$ background component needs to be modelled in the final fit and is the main background for the $B^0 \to \mu^+\mu^-$ signal. It can be seen that in order to gain precision on the $B^0 \to \mu^+\mu^-$ signal the PID requirements need to be re-optimized.

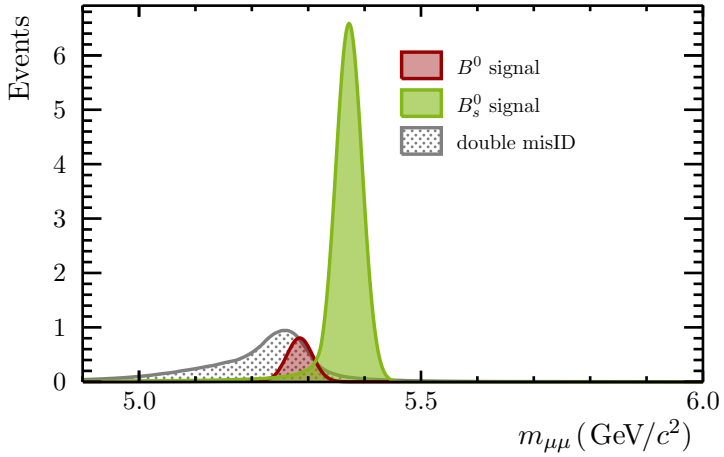


Figure 7.3: Mass shapes of the double mis-identified $B \to h^+h'^-$ decays in comparison of the B_s^0 and B^0 signal mass shapes. All shapes are normalized to the expected number of events in the full $3\,\mathrm{fb}^{-1}$ LHCb dataset (14.6 double misID, 39.9 $B_s^0 \to \mu^+\mu^-$, and 4.5 $B^0 \to \mu^+\mu^-$ events) assuming SM branching fraction.

The invariant mass template for the double mis-identification background is assumend to be the same for each bin of the BDT classifier. However, the normalisation for the double mis-identification invariant mass template is calculated for each bin of the BDT classifier separately.

7.6.3 Background from semi-leptonic decays

Besides $B \to h^+h'^-$ decays, a group of semi-leptonic B decays where one hadron is mis-identified as a muon can contribute to the invariant mass distribution of selected $B \to \mu^+\mu^-$ events. The influence of these decays in the signal mass windows need to be studied as well as the shape of the respective invariant mass distributions. Whenever there is a significant expected number of events in the signal windows or the shape differs from an exponential, the background needs to be taken into account.
The method of evaluation is similar for each of the backgrounds. Dedicated samples of fully simulated events are used to evaluate the trigger and total selection efficiencies. From these efficiencies the expected yield of events is calculated with respect to the $B^+ \to J/\psi\, K^+$ normalisation channel as follows

$$\begin{aligned} N_x^{\mathrm{exp}} &= N(B^+ \to J/\psi\, K^+)\frac{f_x}{f_u}\frac{\mathcal{B}_x}{\mathcal{B}(B^+ \to J/\psi\, K^+)}\frac{\epsilon_x^{\mathrm{tot}}}{\epsilon^{\mathrm{tot}}(B^+ \to J/\psi\, K^+)} \\ &= \beta_x \times \epsilon_x^{\mathrm{tot}} \times \mathcal{B}_x\,. \end{aligned} \tag{7.10}$$

Here, the number of $B^+ \to J/\psi K^+$ events, $N(B^+ \to J/\psi K^+)$, the total efficiency, $\epsilon^{\text{tot}}(B^+ \to J/\psi K^+)$, and the branching fraction, $\mathcal{B}(B^+ \to J/\psi K^+)$, of the normalisation channel, as well as the hadronisation fraction ratio, f_x/f_u are combined in the normalisation factors β_x. Assuming that the probability that a b quark hadronises into a B^+ meson is equal to the one where the b quark hadronises into a B^0 meson and taking the input values from Section 7.4, the normalisation constants for B^0 and B_s^0 background processes are

$$\beta = (8.10 \pm 0.33) \times 10^{11} \quad \text{and} \tag{7.11}$$

$$\beta_s = (2.12 \pm 0.16) \times 10^{11} \, . \tag{7.12}$$

The respective total efficiencies, ϵ_x^{tot}, of the background decays are defined as the combination of the acceptance, trigger, and reconstruction and selection efficiencies, which are evaluated from samples of fully simulated events. The different branching fractions are taken from the latest measurements or from theory when not yet measured.
To investigate the invariant dimuon mass shape of the relevant backgrounds the selection from Section 7.2 is applied to the simulated events with exception of the muon identification criteria for a mis-identified hadron. The muon, in turn, is required to pass the muon identification. The selected events are weighted according the the mis-identification probabilities from Section 7.6.1. The resulting invariant mass distribution is split into bins of the BDT classifier. Then, each bin is fit with the same parametric model of a combination of exponential and gaussian functions. It is assumed that the overall invariant mass shape is the same for each BDT bin, while the parameters are allowed to vary bin-by-bin. In the final fit, the shape is fixed for each bin, while the total number of events and the fractions of events per BDT bin for each background component is left floating with a gaussian constraint around the expected values.

$B^0 \to \pi^-\mu^+\nu_\mu$ and $B_s^0 \to K^-\mu^+\nu_\mu$ decays

The semileptonic decays $B^0 \to \pi^-\mu^+\nu_\mu$ and $B_s^0 \to K^-\mu^+\nu_\mu$ are backgrounds to $B \to \mu^+\mu^-$ when the hadron ($h^- = K^-, \pi^-$) is mis-identified as a muon. The final state neutrino leaves the detector without being measured. As a consequence, the distribution of the invariant mass of the hadron-muon system, $m(h^-\mu^+)$, depends on the neutrino momentum. The simulated events for both background channels are generated with a $q^2 \equiv m^2(\mu^+\nu_\mu)$ dependent decay rate. The q^2 dependence is modelled by the *ISGW2* model [202–204], which predicts the decay rate over the whole q^2 range. To increase the efficiency of the event generation, only events are kept where the $h^-\mu^+$-pair is in the LHCb acceptance and the invariant hadron-muon mass is above $m^2(h^-\mu^+) > (4.5\,\text{GeV}/c^2)^2$. In this way, the shapes of the background components can be estimated with greater precision, as one event passing these

requirements corresponds to about 30 events in the full $m^2(h^-\mu^+)$ region. This invariant mass requirement is reflected in the acceptance efficiency.
The branching fraction of $B^0 \to \pi^-\mu^+\nu_\mu$ decays is taken from Ref. [175], while for $B_s^0 \to K^-\mu^+\nu_\mu$ decays the branching fraction is not yet measured. The theory prediction of Ref. [205] is used to estimate the contribution of the $B_s^0 \to K^-\mu^+\nu_\mu$ background. All input numbers for the calculation of the expected number of events are summarised in Table 7.5.

Table 7.5: Summary of $B^0 \to \pi^-\mu^+\nu_\mu$ and $B_s^0 \to K^-\mu^+\nu_\mu$ results for the normalisation constant, β_x, total efficiency, ϵ_x^{tot}, branching fraction values, $\mathcal{B}_x$, and total expected events, N_x^{exp}.

		$B^0 \to \pi^-\mu^+\nu_\mu$	$B_s^0 \to K^-\mu^+\nu_\mu$
β_x	$[10^{+11}]$	8.10 ± 0.33	2.12 ± 0.16
ϵ_x^{tot}	$[10^{-7}]$	9.69 ± 0.14	3.80 ± 0.10
$\mathcal{B}_x$	$[10^{-4}]$	1.44 ± 0.05	1.27 ± 0.49
N_x^{exp}		114 ± 6	10.2 ± 4.0

In the selected LHCb dataset a total number of 114 ± 6 $B^0 \to \pi^-\mu^+\nu_\mu$ events and a total number of 10.2 ± 4.0 $B_s^0 \to K^-\mu^+\nu_\mu$ events are expected. The difference is due to lower $K^+ \to \mu^+$ mis-identification rates and the lower number of produced B_s^0 meson.
In order to investigate the contributions of the two backgrounds to the various BDT bins, the background samples are split into the eight bins of the BDT classifier. The fractions of events per BDT bin and the corresponding number of expected events are shown in Table 7.6.

Table 7.6: Fraction and number of expected events in each BDT bin for the $B^0 \to \pi^-\mu^+\nu_\mu$ and $B_s^0 \to K^-\mu^+\nu_\mu$ background channels.

BDT bin	$B^0 \to \pi^-\mu^+\nu_\mu$		$B_s^0 \to K^-\mu^+\nu_\mu$	
	fraction [%]	N^{exp}	fraction [%]	N^{exp}
1	37.58 ± 0.20	43.0 ± 2.4	31.72 ± 0.17	3.2 ± 1.3
2	18.87 ± 0.14	21.6 ± 1.2	16.88 ± 0.13	1.7 ± 0.7
3	10.80 ± 0.11	12.6 ± 0.7	10.47 ± 0.10	1.1 ± 0.4
4	9.38 ± 0.10	10.7 ± 0.6	10.24 ± 0.10	1.0 ± 0.4
5	8.26 ± 0.09	9.5 ± 0.5	10.08 ± 0.10	1.0 ± 0.4
6	7.16 ± 0.08	8.2 ± 0.5	9.54 ± 0.09	1.0 ± 0.4
7	5.27 ± 0.06	6.0 ± 0.3	7.29 ± 0.08	0.7 ± 0.3
8	2.68 ± 0.04	3.1 ± 0.2	3.78 ± 0.06	0.4 ± 0.2

It can be seen that the expected number of $B_s^0 \to K^-\mu^+\nu_\mu$ events is small against the $B^0 \to \pi^-\mu^+\nu_\mu$ contributions. Hence, the $B_s^0 \to K^-\mu^+\nu_\mu$ is only studied as a systematic effect in the combinatorial background estimation.
The invariant $B^0 \to \pi^-\mu^+\nu_\mu$ mass distribution is fit in bins of the BDT classifier to find the invariant mass shape for each BDT bin. The fits to the various bins of the BDT classifier are shown in Fig. 7.4. Indeed, the same effective function can be used to fit all BDT bins. In the final fit, however, all shape parameters are fixed and only to total number of events is floating within a gaussian constraint.

$\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ decays

The $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ decay can contribute to the $B_s^0 \to \mu^+\mu^-$ background sources, if the proton is wrongly identified as a muon. However, the branching fraction of $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ decays is not measured, so a theoretical prediction [206]

$$\mathcal{B}\left(\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu\right)_{\text{th}} = 3.3^{+1.5}_{-1.2} \times 10^{-4} \times \left(\frac{V_{ub}}{3.5 \times 10^{-3}}\right)^2 \tag{7.13}$$

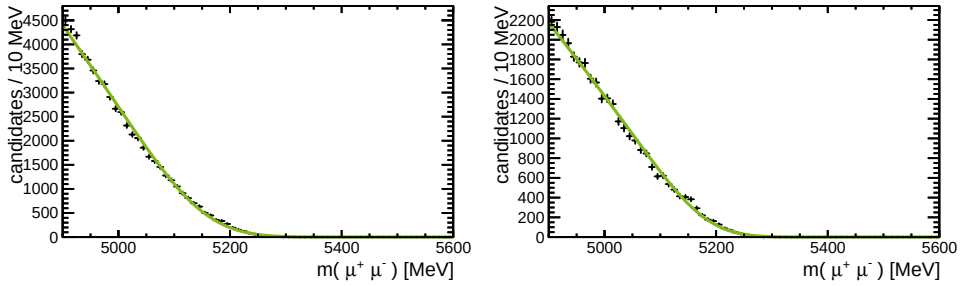
is used to estimate the branching fraction. This result was calculated using the Λ_b^0 lifetime of $\tau_{\Lambda_b^0,o} = 1.391^{+0.038}_{-0.037}\,\text{ps}$ [207], which is updated to $\tau_{\Lambda_b^0,n} = 1.425 \pm 0.032\,\text{ps}$ [175] resulting in an additional factor of $\tau_{\Lambda_b^0,n}/\tau_{\Lambda_b^0,o}$ in Eq. 7.13. With the CKM element $V_{ub} = (4.15 \pm 0.49) \times 10^{-3}$ [175] the Λ_b^0 branching fraction is determined to

$$\mathcal{B}\left(\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu\right) = (4.75 \pm 2.11) \times 10^{-4}\,. \tag{7.14}$$

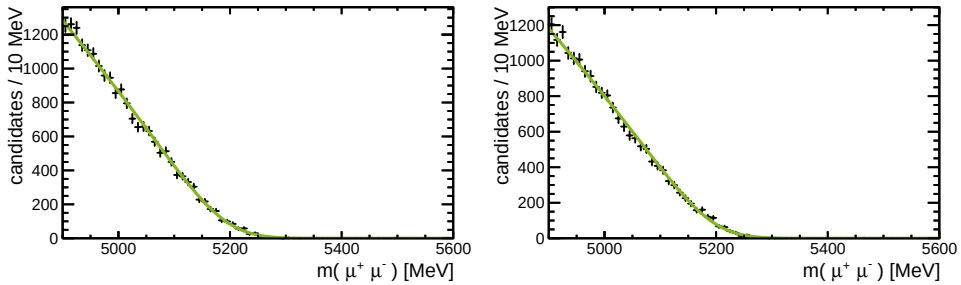
To calculate the number of expected events with Eq. 7.10 the production fraction of Λ_b^0 baryons relative to B^+ mesons is needed. The p_T dependent production fraction

$$r_{\Lambda_b^0} \equiv \frac{f_{\Lambda_b^0}}{f_u + f_d} = (0.404 \pm 0.110) \times [1 - (0.031 \pm 0.005) \times p_{\text{T}}(\,\text{GeV}/c)] \tag{7.15}$$

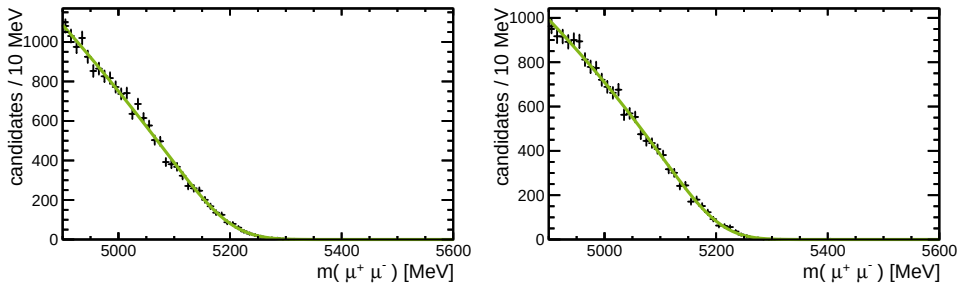
is measured by the LHCb experiment [208]. Assuming $f_u = f_d$, each event in the simulated sample gets a weight according to the transverse momentum of the p-μ system, $\tilde{p}_T$, and the momentum and transverse momentum of the proton. This weight is the product of the $p \to \mu$ mis-identification probability (cf. Section 7.6.1) and the production fraction ratio, $2r_{\Lambda_b^0}(\tilde{p}_T)$. The factor of two accounts for the sum of $f_u + f_d$ in the denominator of $r_{\Lambda_b^0}$ in Eq. 7.15.
The total efficiency, which includes the event weights, and the contributions to the BDT bins depend on the missed neutrino momentum and, hence, on $q^2 = m^2_{\mu\nu_\mu}$. The sample of simulated events used to estimate the efficiency is generated with an equal probability model for the decay products to end up in any region of the phase space. No spin correlations are taken into account. Therefore, the phase space is



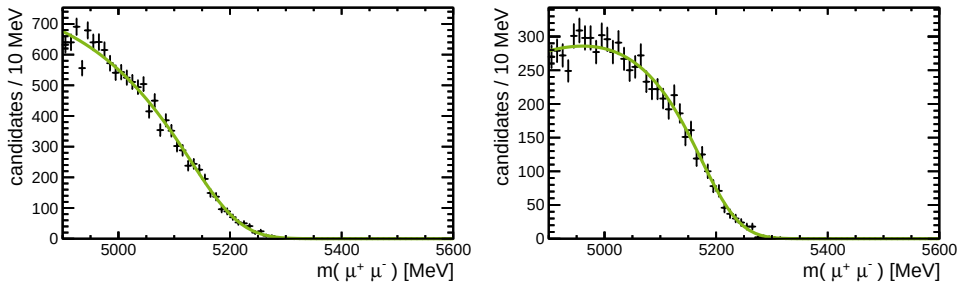
(a) BDT ≤ 0.25 (b) $0.25 <$ BDT ≤ 0.4



(c) $0.4 <$ BDT ≤ 0.5 (d) $0.5 <$ BDT ≤ 0.6



(e) $0.6 <$ BDT ≤ 0.7 (f) $0.7 <$ BDT ≤ 0.8



(g) $0.8 <$ BDT ≤ 0.9 (h) $0.9 <$ BDT ≤ 1

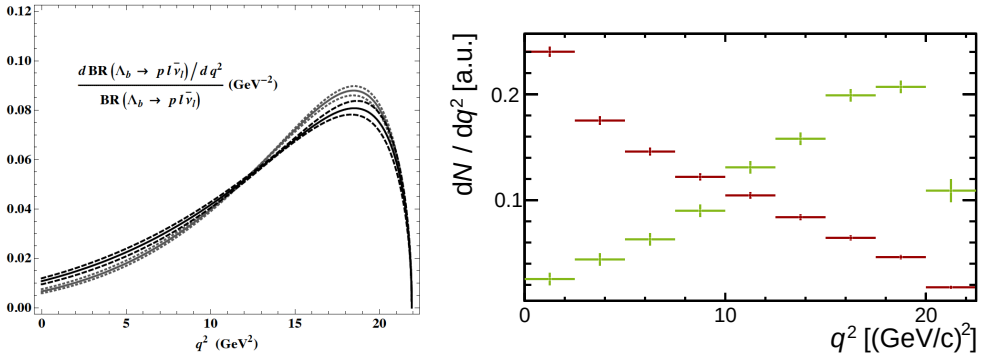
Figure 7.4: Invariant mass distributions and fits to the $B^0 \to \pi^- \mu^+ \nu_\mu$ channel in the eight bins of the BDT classifier. The simulated events are weighted according to the misidentification efficiencies from Section 7.6.1.

weighted to match the theory prediction of Ref. [206], by modelling the theoretical q^2 distribution in nine bins of q^2. Then, the quotient of the theory predictions and the flat generated phase space model for the q^2 bin is taken as a weight for the simulated events. The theoretical prediction for the q^2 dependence of the branching fraction and the comparison of q^2 distributions for the generated phase space model in the sample of simulated events with the binned theory prediction are shown in Fig. 7.5(a) and Fig.7.5(b), respectively. After the complete reweighting and the complete $B_s^0 \to \mu^+\mu^-$ selection, the total number of expected $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ events is $N^{\text{exp}}_{\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu} = 67.6 \pm 30.3$. The total efficiency, including the events weights for the q^2 dependence and the production fraction ratio $2r_{\Lambda_b^0}(\tilde{p}_T)$, the branching fraction prediction, and normalisation are summarised in Table 7.7.

Table 7.7: Summary of $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ results for the normalisation constant, β, total efficiency, ϵ^{tot}, branching fraction, $\mathcal{B}$, and total expected events, $N^{\text{exp}}_{\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu}$.

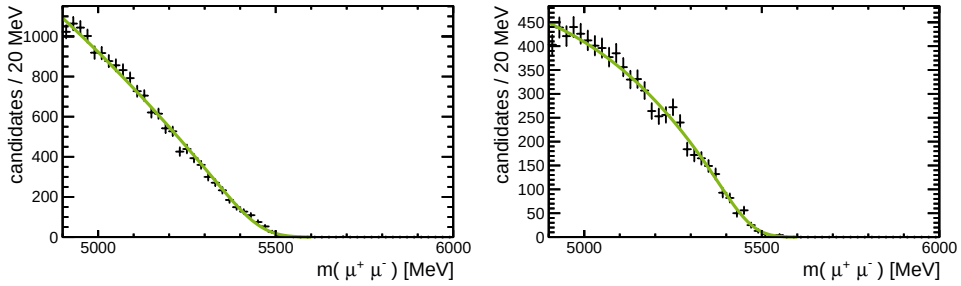
$\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$		
β	$[10^{+11}]$	8.10 ± 0.33
ϵ^{tot}	$[10^{-7}]$	1.74 ± 0.07
$\mathcal{B}$	$[10^{-4}]$	4.75 ± 2.11
$N^{\text{exp}}_{\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu}$		67.6 ± 30.3

The large uncertainty on the expected number of events is driven by the poorly known $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ branching fraction. The fits to the bins of the BDT classifier are shown in Fig. 7.6 and the expected number of events per BDT bin is listed in Table 7.8.

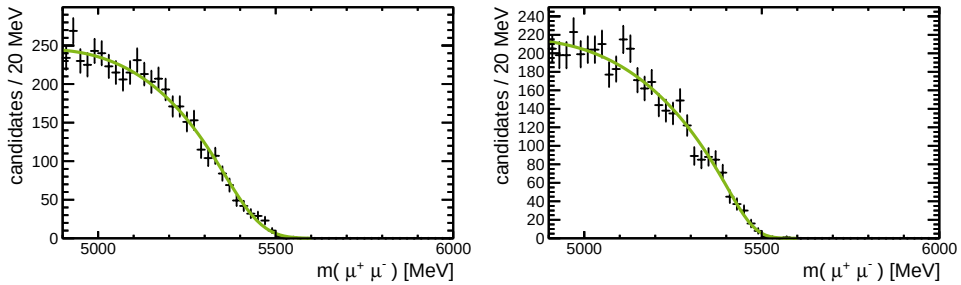


(a) (b)

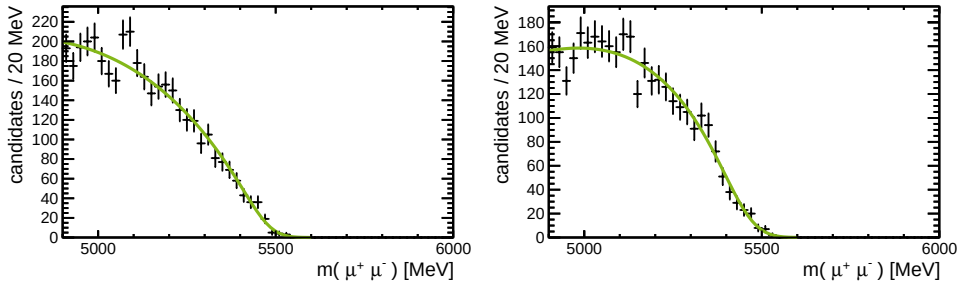
Figure 7.5: (a) Theoretical prediction for the q^2 dependence of the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ branching fraction [206]. (b) The distribution of q^2 from $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ decays generated with the flat "phase space" model (red) and the binned theory prediction (green).



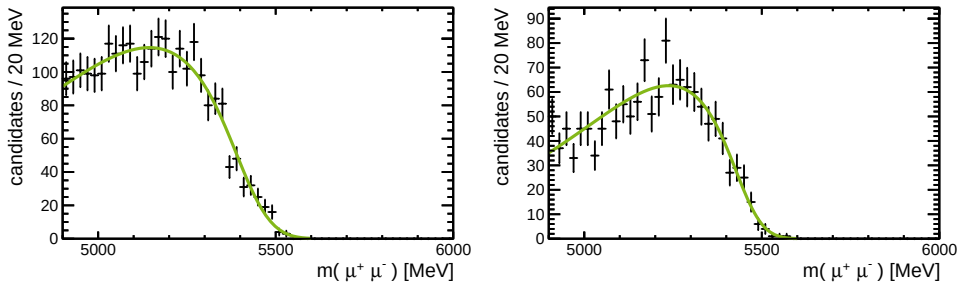
(a) BDT ≤ 0.25 (b) $0.25 <$ BDT ≤ 0.4



(c) $0.4 <$ BDT ≤ 0.5 (d) $0.5 <$ BDT ≤ 0.6



(e) $0.6 <$ BDT ≤ 0.7 (f) $0.7 <$ BDT ≤ 0.8



(g) $0.8 <$ BDT ≤ 0.9 (h) $0.9 <$ BDT ≤ 1

Figure 7.6: Invariant mass distributions and fits to the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ channel for each BDT bin.

Table 7.8: Fraction and number of expected events in each BDT bin for the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ channel.

BDT bin	$\boldsymbol{\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu}$	
	fraction [%]	N^{exp}
1	45.75 ± 0.51	30.95 ± 13.87
2	18.74 ± 0.30	12.68 ± 5.68
3	10.04 ± 0.21	6.79 ± 3.04
4	8.05 ± 0.17	5.45 ± 2.44
5	6.69 ± 0.15	4.52 ± 2.07
6	5.40 ± 0.12	3.65 ± 1.64
7	3.59 ± 0.09	2.43 ± 1.09
8	1.73 ± 0.06	1.17 ± 0.53

Although the number of expected events in the most sensitive bins is not as high as for the $B^0 \to \pi^-\mu^+\nu_\mu$ background component, the invariant mass distribution extends into the signal mass windows. For the two most sensitive bins of the BDT classifier the invariant mass shape of $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ decays develops a peaking structure right in the signal region. This effect needs to be studied in the final fit, because the assumption that the combinatorial background can be extrapolated into the signal region and solely accounts for the total background is violated with a significant $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ contribution in the signal region.

7.6.4 Partially reconstructed decays

Partially reconstructed events have two *real* muons in the final state, but one or more particles are missed in the event reconstruction. Hence, no weight according to mis-identification probabilities needs to be applied and the muon identification is required for both muons. Else, the procedure is equivalent to classes of backgrounds discussed previously.

$B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ decays

The FCNC decays $B \to \pi\ \mu^+\mu^-$, which denotes $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ decays, can fake the $B \to \mu^+\mu^-$ signal, because the two final state muons form a good common vertex. However, completely missing a particle in the reconstruction results in a significantly lower invariant dimuon mass estimate. The $B \to \pi\ \mu^+\mu^-$ decays can bias the estimation of the combinatorial background by polluting the lower mass sideband with non-exponential invariant mass distributions.

The $B^+ \to \pi^+\mu^+\mu^-$ branching fraction is measured by the LHCb collaboration [209] to

$$\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-) = (2.3 \pm 0.6 \pm 0.1) \times 10^{-8}\,,$$

where the first uncertainty is the statistical and the second one the systematic uncertainty.
The $B^0 \to \pi^0\mu^+\mu^-$ branching fraction on the other hand has not been measured so far, but the decay rate is expected to be similar. From the theoretical predictions in [205]

$$\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-) = 1.95^{+0.61}_{-0.48} \times 10^{-8} \text{ and}$$
$$\mathcal{B}(B^0 \to \pi^0\mu^+\mu^-) = 0.91^{+0.29}_{-0.23} \times 10^{-8}\,.$$

the ratio of branching fractions is calculated

$$\frac{\mathcal{B}(B^0 \to \pi^0\mu^+\mu^-)}{\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-)} = 0.47^{+0.21}_{-0.17}\,. \tag{7.16}$$

The ratio is used to estimate the $B^0 \to \pi^0\mu^+\mu^-$ contribution, while the absolute scale is defined by the experimentally measured $B^+ \to \pi^+\mu^+\mu^-$ branching fraction. A sample of simulated events is used to measure the total efficiency, which together with the input numbers in Table 7.9 results in a total number of expected $B \to \pi\,\mu^+\mu^-$ events of $N^{\text{exp}}_{B\to\pi\mu^+\mu^-} = 28.01^{+8.51}_{-8.16}$.

Table 7.9: Summary of $B \to \pi\,\mu^+\mu^-$ results for the normalisation constant, total efficiency, branching fraction, and total expected events. To get the total number of expected $B \to \pi\,\mu^+\mu^-$ decays, the branching fraction needs to be multiplied by a factor $1.47^{+0.21}_{-0.17}$ following Eq. 7.16.

$\boldsymbol{B \to \pi\mu^+\mu^-}$		
β	$[10^{+11}]$	8.10 ± 0.33
ϵ^{tot}	$[10^{-3}]$	1.01 ± 0.02
$\mathcal{B}_x$	$[10^{-8}]$	2.3 ± 0.6
$N^{\text{exp}}_{B\to\pi\mu^+\mu^-}$		$28.01^{+8.51}_{-8.16}$

Although the total number of expected events is not that large, the distribution within the bins of the BDT classifier is studied. The $B \to \pi\,\mu^+\mu^-$ final state particles after the reconstruction and the kinematic properties are indistinguishable and therefore the two decays are treated as a common component. The fraction and number of events per BDT bin is given in Table 7.10 and the fits to the invariant dimuon mass distribution are shown in Fig. 7.7. For the estimation of the shape parameters, the invariant dimuon mass range is widened from $[4.9, 6.0]\,\text{GeV}/c^2$ to $[4.2, 6.0]\,\text{GeV}/c^2$ in order to increase the number of events in the simulation sample and therefore get

more reliable parameter estimates. It is found that the distributions do not leak into the signal regions. However, all distributions show a steep decrease of events right at the lowest end of the nominal fit range of $[4.9, 6.0]\,\mathrm{GeV}/c^2$. As a consequence, $B \to \pi\ \mu^+\mu^-$ can alter the slope of the combinatorial exponential function and, therefore, biases the background estimate in the signal search windows. Hence, $B \to \pi\ \mu^+\mu^-$ is added to the final fit model.

Table 7.10: Fraction and number of expected events in each BDT bin for the $B \to \pi\ \mu^+\mu^-$ channel.

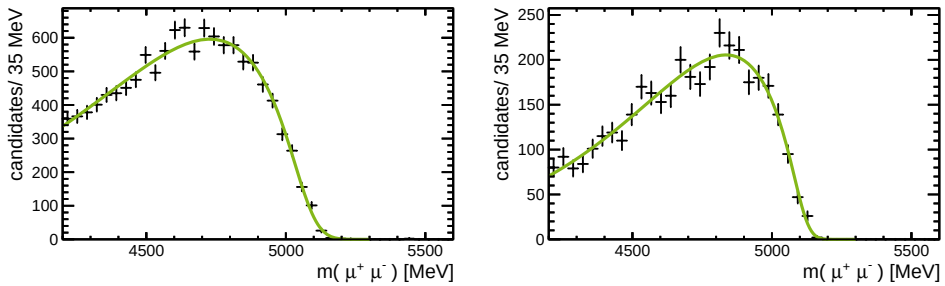
BDT bin	$\boldsymbol{B \to \pi\mu^+\mu^-}$	
	fraction [%]	N^{exp}
1	38.03 ± 0.96	$10.65^{+3.25}_{-3.12}$
2	18.46 ± 0.67	$5.17^{+1.58}_{-1.52}$
3	10.42 ± 0.50	$2.92^{+0.90}_{-0.86}$
4	10.40 ± 0.50	$2.91^{+0.90}_{-0.86}$
5	7.55 ± 0.43	$2.11^{+0.65}_{-0.63}$
6	7.43 ± 0.43	$2.08^{+0.64}_{-0.62}$
7	5.02 ± 0.35	$1.41^{+0.44}_{-0.42}$
8	2.70 ± 0.26	$0.76^{+0.24}_{-0.23}$

$B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decays

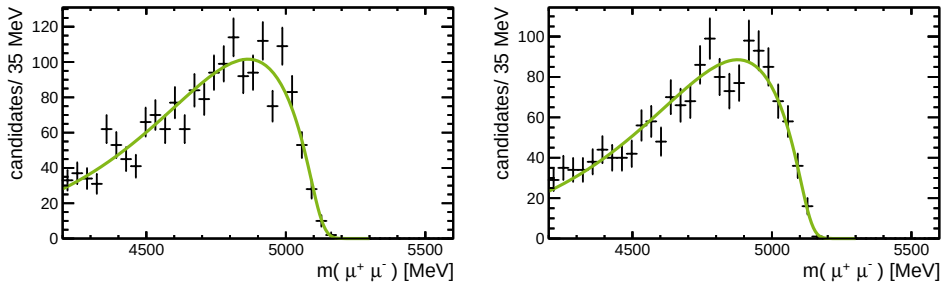
Due to the larger mass of the B_c^+ meson, partially reconstructed $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ with $J/\psi \to \mu^+\mu^-$ events can potentially contribute to the $B \to \mu^+\mu^-$ signal classification. If one of the muons from the J/ψ is combined with the muon from the $W^+ \to \mu^+\nu_\mu$ decay, the resulting invariant mass can be shifted into the mass window of the analysis.
The expected number of $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ events reconstructed as $B \to \mu^+\mu^-$ candidates is calculated with a slightly modified version of Eq. 7.10. The production and absolute branching fraction of B_c^+ mesons is largely unknown. The CDF collaboration measured the production times branching fraction relative to the $B^+ \to J/\psi\,K^+$ channel [210, 211] to be

$$\begin{aligned}\mathcal{R} &= \frac{\sigma(B_c^+)\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu_\mu)}{\sigma(B^+)\mathcal{B}(B^+ \to J/\psi\,K^+)} \\ &= \frac{1}{2} \times 0.132^{+0.041}_{-0.037}(\mathrm{stat}) \pm 0.031(\mathrm{sys})^{+0.032}_{-0.020}(\text{decay time})\,,\end{aligned}$$

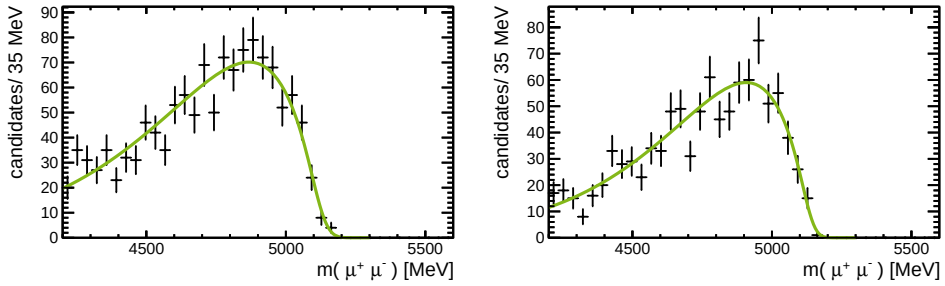
where the first uncertainty is statistical, the second one systematic and the third uncertainty is assigned due to a decay time dependent selection. The division by a factor of two is due to the fact that the CDF result is valid for the sum of



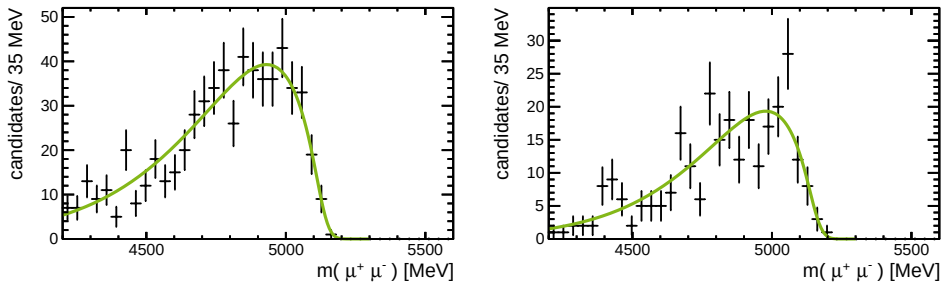
(a) BDT ≤ 0.25 (b) $0.25 <$ BDT ≤ 0.4



(c) $0.4 <$ BDT ≤ 0.5 (d) $0.5 <$ BDT ≤ 0.6



(e) $0.6 <$ BDT ≤ 0.7 (f) $0.7 <$ BDT ≤ 0.8



(g) $0.8 <$ BDT ≤ 0.9 (h) $0.9 <$ BDT ≤ 1

Figure 7.7: Invariant mass distributions and fits to the $B \to \pi\ \mu^+\mu^-$ channel for each BDT bin.

$B_c^+ \to J/\psi\, e^+\nu_e$ and $B_c^+ \to J/\psi\, \mu^+\nu_\mu$[1]. Because all the exclusively studied backgrounds are normalised to $B^+ \to J/\psi\, K^+$, the CDF measurement can be used to calculate the expected number of $B_c^+ \to J/\psi\, \mu^+\nu_\mu$ events as follows

$$N^{\text{exp}} = \mathcal{R} \times \beta \times \mathcal{B}(B^+ \to J/\psi\, K^+) \times \epsilon^{\text{tot}}_{B_c^+} \,. \tag{7.17}$$

Here, β is the normalisation factor for B^+ mesons from Eq. 7.11, $\mathcal{B}(B^+ \to J/\psi\, K^+)$ the product of the branching fractions of $B^+ \to J/\psi\, K^+$ and $J/\psi \to \mu^+\mu^-$ [175], and $\epsilon^{\text{tot}}_{B_c^+}$ the total efficiency evaluated using simulated events. The numerical values used to calculate the expected number of $B_c^+ \to J/\psi\, \mu^+\nu_\mu$ events are summarised in Table 7.11.

Table 7.11: Summary of $B_c^+ \to J/\psi\, \mu^+\nu_\mu$ results for the normalisation constant, total efficiency, relative branching fraction times production cross section, and total expected events. The uncertainties for $2 \times \mathcal{R}$ are square roots of the sum of squared uncertainties.

$\boldsymbol{B_c^+ \to J/\psi\, \mu^+\nu_\mu}$		
β	$[10^{+11}]$	8.10 ± 0.33
$\epsilon^{\text{tot}}_{B_c^+}$	$[10^{-5}]$	3.33 ± 0.05
$2 \times \mathcal{R}$		$0.132^{+0.061}_{-0.052}$
$N^{\text{exp}}_{B_c^+ \to J/\psi\, \mu^+\nu_\mu}$		$108.3^{+42.3}_{-43.1}$

The number of expected $B_c^+ \to J/\psi\, \mu^+\nu_\mu$ events that survive the reconstruction and $B \to \mu^+\mu^-$ selection is

$$N^{\text{exp}}_{B_c^+ \to J/\psi\, \mu^+\nu_\mu} = 108.3^{+42.3}_{-43.1} \,.$$

The expected number of events per BDT bin and the corresponding fractions of the total number of expected events are listed in Table 7.12. It is found that the contribution of $B_c^+ \to J/\psi\, \mu^+\nu_\mu$ decays in the most sensitive BDT bins seven and eight is negligible in comparison with the other exclusive backgrounds (cf. Table 7.13 in Section 7.6.6). As a result, $B_c^+ \to J/\psi\, \mu^+\nu_\mu$ events are not explicitly modelled in the final fit.

7.6.5 Other potential background channels

Two commonly discussed background sources are $B_s^0 \to \mu^+\mu^-\gamma$ and $B_s^0 \to \tau^+\tau^-$ decays. These have been studied for earlier publications [86]. Both decays were found not to contribute significantly. The arguments are reviewed for the analysis of the full $3\,\text{fb}^{-1}$ dataset of LHCb.

[1] The factor of 0.5 is conservative. The actual fitted electron fraction in the CDF sample is 0.59 ± 0.04.

Table 7.12: Fraction and number of expected events in each BDT bin for the $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ channel.

BDT bin	$B_c^+ \to J/\psi\,\mu^+\nu_\mu$	
	fraction [%]	N^{exp}
1	73.37 ± 0.79	$79.5^{+31.0}_{-31.6}$
2	13.41 ± 0.34	$14.53^{+5.68}_{-5.79}$
3	5.50 ± 0.22	$5.96^{+2.34}_{-2.38}$
4	3.72 ± 0.18	$4.03^{+1.56}_{-1.62}$
5	2.15 ± 0.13	$2.33^{+0.92}_{-0.94}$
6	1.34 ± 0.11	$1.45^{+0.58}_{-0.59}$
7	0.42 ± 0.06	$0.46^{+0.19}_{-0.19}$
8	0.08 ± 0.03	$0.08^{+0.04}_{-0.04}$

$B_s^0 \to \mu^+\mu^-\gamma$ decays

The categorisation of $B_s^0 \to \mu^+\mu^-\gamma$ events as background for $B \to \mu^+\mu^-$ decays depends on the origin of the outgoing photon. If the photon is radiated from the initial state quarks the process is called *initial state radiation* (ISR) and if the photon comes from the final state particles it is called *final state radiation* (FSR). As these two production sources are indistinguishable from the experimental point of view the interference of both is also present. The FSR component is part of the signal and is accounted for in the simulation using the PHOTOS package [212].
The relevance of the $B_s^0 \to \mu^+\mu^-\gamma$ background depends on the branching fraction of $B_s^0 \to \mu^+\mu^-\gamma$ decays and on the momentum distribution of the photon which is not reconstructed. The branching fraction of $B_s^0 \to \mu^+\mu^-\gamma$ is not known, so a theoretical prediction [213] of $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) \sim 19 \times 10^{-9}$ is assumed[2]. In order to estimate the number of $B_s^0 \to \mu^+\mu^-\gamma$ events that possibly contribute to the background of $B \to \mu^+\mu^-$ a sample of simulated $B_s^0 \to \mu^+\mu^-\gamma$ events is used. The simulation accounts for all three photon radiations (ISR,FSR and the interference). As the FSR component is part of the signal the following statements can be interpreted as upper limits on the $B_s^0 \to \mu^+\mu^-\gamma$ events.
From the samples of simulated events the total efficiency ratio including an additional selection cut on the BDT classifier of BDT>0.8 is determined to be

$$\frac{\epsilon\left(B_s^0 \to \mu^+\mu^-\gamma\right)}{\epsilon\left(B_s^0 \to \mu^+\mu^-\right)} = 0.0029 \pm 0.0002\,.$$

The efficiency ratio together with the expected number of $B_s^0 \to \mu^+\mu^-$ events can be used to estimate the $B_s^0 \to \mu^+\mu^-\gamma$ contribution to the background. With 7.6

[2]This prediction is larger than older theory predictions [214, 215] and therefore more conservative in terms of estimating a possible background yield of $B_s^0 \to \mu^+\mu^-\gamma$ events.

expected $B_s^0 \to \mu^+\mu^-$ events in the most sensitive region of BDT > 0.8 the number of expected $B_s^0 \to \mu^+\mu^-\gamma$ events is

$$N^{\text{exp}}_{B_s^0\to\mu^+\mu^-\gamma} < \frac{\epsilon^{\text{BDT}}_{B_s^0\to\mu^+\mu^-\gamma}}{\epsilon^{\text{tot}}_{B_s^0\to\mu^+\mu^-}} \times \frac{\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)} \times N_{B_s^0\to\mu^+\mu^-}(\text{BDT} > 0.8)$$
$$N^{\text{exp}}_{B_s^0\to\mu^+\mu^-\gamma} < 0.13\,.$$

A precise uncertainty estimation is difficult, because the theory predictions vary by a factor of ten and the efficiency of $B_s^0 \to \mu^+\mu^-\gamma$ depends on the theoretical description for the photon momentum distribution. However, even if a factor of ten is multiplied to the number of expected $B_s^0 \to \mu^+\mu^-\gamma$ events, the resulting yield is just in the order of the uncertainty of the estimate of the combinatorial background component. That gives confidence that the $B_s^0 \to \mu^+\mu^-\gamma$ background contribution remains negligible.

$B_s^0 \to \tau^+\tau^-$ decays

Another potential background source can result from $B_s^0 \to \tau^+\tau^-$ decays when the two τ leptons decay into muons. Compared to the branching fraction of $B_s^0 \to \mu^+\mu^-$ the SM prediction for the $B_s^0 \to \tau^+\tau^-$ branching fraction is higher by a factor of $(m_\tau/m_\mu)^2 \approx 300$. This is due to the breaking of the helicity suppression in $B_s^0 \to \mu^+\mu^-$ decays. However, the probability of τ leptons to decay into muons is $\mathcal{B}(\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau) = 17.41\%$ so that actually only a factor of nine more $B_s^0 \to \tau^+\tau^-$ with two $\tau^\mp \to \mu^\mp\nu_\mu\nu_\tau$ decays are expected in the data sample compared to $B_s^0 \to \mu^+\mu^-$. The $B_s^0 \to \tau^+\tau^-$ final state which results in a background for $B_s^0 \to \mu^+\mu^-$ events contains six particles from which four are not detectable. This shifts the reconstructed invariant mass significantly to lower values. An exemplary invariant dimuon mass distribution for $B_s^0 \to \tau^+\tau^- \to \mu^+\mu^-\nu_\mu\overline{\nu}_\mu\nu_\tau\overline{\nu}_\tau$ is shown in Fig. 7.8. This distribution is obtained from a simple phase space simulation, where B_s^0 mesons with momenta from simulated $B_s^0 \to \mu^+\mu^-$ candidates decay into two τ leptons, which in turn decay into $\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau$. It does not include any detector simulation or selection efficiencies. Although the invariant mass distribution of selected $B_s^0 \to \tau^+\tau^-$ events does not necessarily need to follow the shape in Fig. 7.8, $B_s^0 \to \tau^+\tau^-$ is clearly separated from the $B_s^0 \to \mu^+\mu^-$ signal. Additionally, the mean lifetime of a τ lepton is $\tau_\tau = (290.3 \pm 0.5) \times 10^{-15}\,\text{s}$ which leads to a slight separation of the two reconstructed muons. Due to the requirement of a good vertex quality for the signal muons, $B_s^0 \to \tau^+\tau^-$ decay are mainly categorised as background (or result in lower values of the BDT classifier). As a consequence $B_s^0 \to \tau^+\tau^-$ is not expected to contribute to the relevant invariant dimuon mass region.

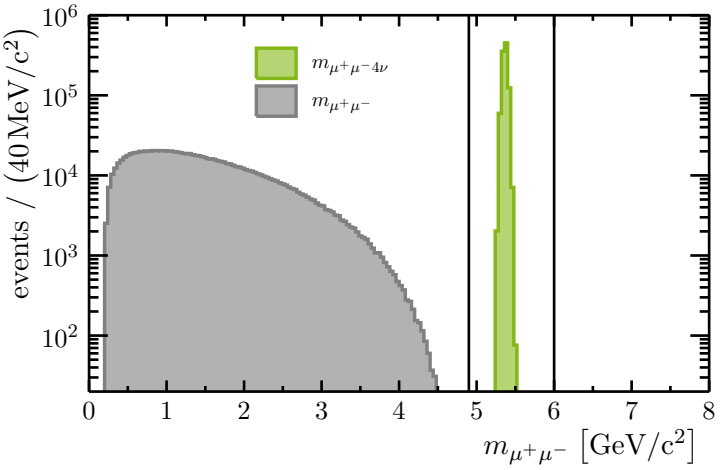


Figure 7.8: Invariant mass distribution of $B_s^0 \to \tau^+\tau^- \to \mu^+\mu^-\nu_\mu\overline{\nu}_\mu\nu_\tau\overline{\nu}_\tau$ phase space decays. The invariant mass of the two muons, $m_{\mu^+\mu^-}$, is shown in grey and the full $m_{2\mu4\nu}$ invariant mass is displayed in green. A mass resolution of $30\,\mathrm{MeV}/c^2$ is assumed for this study. The vertical black lines indicate the $B \to \mu^+\mu^-$ invariant mass window.

7.6.6 Summary of the exclusive background studies

A list of exclusive background channels are studied. The radiative $B_s^0 \to \mu^+\mu^-\gamma$ and the $B_s^0 \to \tau^+\tau^-$ are found to be negligible. The expected contributions from $B^0 \to \pi^-\mu^+\nu_\mu$, $B_s^0 \to K^-\mu^+\nu_\mu$, $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$, $B_c^+ \to J/\psi\,\mu^+\nu_\mu$, $B \to \pi\mu^+\mu^-$, and $B \to h^+h'^-$ decays for the most sensitive bins of BDT > 0.8. Compared to the ≈ 8 expected $B_s^0 \to \mu^+\mu^-$ events in the highest sensitivity bins, the importance to model the distributions of the exclusive background channels becomes evident. The exception is $B_c^+ \to J/\psi\,\mu^+\nu_\mu$, where the contribution can be absorbed by the fluctuations of the other backgrounds.

Table 7.13: Summary of the exclusive backgrounds in the most sensitive BDT bins seven and eight (BDT > 0.8). The expected yield of events is calculated for the full $3\,\mathrm{fb}^{-1}$ LHCb dataset.

Background	$\boldsymbol{N}^{\mathrm{exp}}$
$B^0 \to \pi^-\mu^+\nu_\mu$	9.1 ± 0.5
$B_s^0 \to K^-\mu^+\nu_\mu$	1.1 ± 0.4
$\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$	3.6 ± 1.6
$B_c^+ \to J/\psi\,\mu^+\nu_\mu$	0.5 ± 0.2
$B \to \pi\ \mu^+\mu^-$	2.2 ± 0.6
$B \to h^+h'^-$ misID	2.6 ± 0.3

7.6.7 Estimation of the combinatorial background

One of the main backgrounds for the search for $B \to \mu^+\mu^-$ decays are random combinations of two muons that form a B candidate (combinatorial background). These muons originate mostly from two separate b hadron decays. In order to estimate the number of combinatorial background events the sidebands of the invariant dimuon mass distribution are used. The sidebands are defined as $m_{\mu\mu} \in [4.90, 5.22]\,\mathrm{GeV}/c^2$ and $m_{\mu\mu} \in [5.43, 6.00]\,\mathrm{GeV}/c^2$. The gap in invariant mass is defined by the blinded region expanding from $m_{B^0} - 60\,\mathrm{MeV}/c^2$ to $m_{B_s^0} + 60\,\mathrm{MeV}/c^2$ around the measured B masses[3].
The combinatorial background component is evaluated for each BDT bin separately. The shape is assumed to follow a single exponential function. In order to determine the slope and the number of expected combinatorial background events per BDT bin the sidebands are fit with different model configurations. The baseline model in addition to the exponential function for the combinatorial background contains the shapes of the $B \to h^+h'^-$, $B^0 \to \pi^-\mu^+\nu_\mu$, and $B \to \pi\ \mu^+\mu^-$ channels, while the influence of $B_s^0 \to K^-\mu^+\nu_\mu$ is found to be negligible. The parametrisation of the exclusive background shapes are fixed, while the respective normalisation and fraction of events per BDT bin is allowed to vary within a gaussian constraint centred around the corresponding expected number of events. The width of the gaussian constraint equals the total uncertainty (statistical and systematic) of the respective normalisation and BDT fraction. The slope and the number of events of the exponential component are freely floating.
The second configuration adds the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ component to the fit model. Compared to the baseline configuration the addition of the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ component does not change the result for the combinatorial background estimate significantly. In fact, the result of the combinatorial background estimate without accounting for the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ channel explicitly is statistically equivalent with the sum of the combinatorial background and the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ component. The extrapolated expected number of combinatorial and $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ background events in the signal region ($[5.22, 5.43]\,\mathrm{GeV}/c^2$) for the two fit configurations in dependence of the BDT bin are summarised in Table 7.14. The systematic uncertainty which is introduced by fixing the shape parameters is evaluated by re-fitting the sidebands and varying the shape parameters within their respective uncertainties. The effect on the combinatorial background estimate is negligible.

[3] The measured mass differs by $\approx 0.1\%$ with respect to the PDG averages. The mass scale is calibrated on data with $B \to h^+h'^-$ decays (cf. Section 7.5).

Table 7.14: Extrapolated expected number of combinatorial and $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ background events in $[5.22, 5.43]\,\mathrm{GeV}/c^2$ for each BDT bin. The first configuration is the baseline model including the $B \to h^+h'^-$, $B^0 \to \pi^-\mu^+\nu_\mu$, and $B \to \pi\ \mu^+\mu^-$ components and the second configuration, addΛ_b^0, adds the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ background explicitly.

BDT bin	**baseline**	**add Λ_b^0**	
	$N^{\mathrm{exp}}_{\mathrm{comb}}$	$N^{\mathrm{exp}}_{\mathrm{comb}}$	$N^{\mathrm{exp}}_{\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu}$
1	10915 ± 53	10910 ± 53	4.0
2	214.0 ± 7.7	211.9 ± 7.8	1.9
3	49.4 ± 3.8	48.3 ± 3.9	1.1
4	21.2 ± 2.6	20.4 ± 2.6	0.9
5	8.4 ± 2.0	7.6 ± 2.2	0.8
6	3.8 ± 1.5	3.5 ± 1.5	0.7
7	1.4 ± 0.8	1.2 ± 0.8	0.5
8	0.4 ± 0.4	0.2 ± 0.4	0.3

7.7 Expected significance

The expected significance is only evaluated for $B_s^0 \to \mu^+\mu^-$ decays, as these have the larger expected branching fraction and the analysis updates the results, which claimed the first evidence for $B_s^0 \to \mu^+\mu^-$ decays [86]. Hence, the expected significance for $B_s^0 \to \mu^+\mu^-$ decays is defined as the benchmark measure for the expected sensitivity. The signal search window ($m_{B_s^0} \pm 60\,\mathrm{MeV}/c^2$) is binned in nine bins of the invariant mass and eight bins of the BDT classifier. The considered background components are the combinatorial and the peaking $B \to h^+h'^-$ background component. The test statistic used to evaluate the p-values is the logarithm of the product of the likelihood ratios of signal-plus-background (S+B) over the background-only hypothesis per mass-BDT-bin. Each likelihood is a poisson distribution for the combinatorial and peaking background, and the signal in case of the S+B hypothesis. The mean parameters of the poisson distributions are set to the expected number of events in each of the 72 bins.
In the first step the distribution of the test statistc under the background-only hypothesis is sampled. Therefore, a set of random numbers are generated according to the sum of poissonian PDFs of the combinatorial and peaking background. Here, the systematic uncertainties are folded in by varying the poissonian mean parameters according to the total uncertainty on the expected number of events in each background component.
The second step is to generate random numbers for the expected S+B hypothesis. This distribution depends on the assumed branching fraction of the $B_s^0 \to \mu^+\mu^-$ decay. The best estimate for the branching fraction is the SM value in Eq. 2.25.

Therefore, the expectation is defined as the SM signal. In order to be able to test the significance level necessary for a 5σ observations, 14 million sets of random numbers for the S+B hypothesis are generated. For each set a p-value is calculated from the integral of the tail of the test statistic under the background-only hypothesis. The integrals start at the test statistic values observed under the S+B hypothesis. The expected sensitivity is defined as the median of the distribution of the significances and is evaluated to be $5.0 \pm 0.1\sigma$, which indicates a probability of 50% to observe the $B_s^0 \to \mu^+\mu^-$ signal with more than 5σ significance.

7.8 Results

After fixing the nominal invariant mass description the complete invariant mass range is included and fit simultaneously in eight bins of the BDT classifier. The freely floating parameters are the signal branching fractions, the combinatorial background slopes and event yields. The shape parameters and fractions of events per BDT bin of the $B \to \mu^+\mu^-$ signals are constrained to the values obtained from $B \to h^+h'^-$ decays and allowed to vary within their total uncertainties. The other constrained parameters are the normalisation parameters from $B^+ \to J/\psi\, K^+$, the theoretical and experimental input values for the branching fractions of the background channels, the production fraction ratio $f_\mathrm{s}/f_\mathrm{d}$, and the fractions of events per BDT bin of the background channels. All constraints are gaussian functions with the width set to the respective total uncertainties. The influence of adding the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ background component as well as fixing the parameters of the background shapes is evaluated separately and taken into account as systematic uncertainties. Therefore, the fit is repeated in different configurations, where the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ component is added and the values for the background shapes are varied within the respective uncertainties. The total fit yields

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \left(2.9^{+1.1}_{-1.0}(\mathrm{stat})^{+0.3}_{-0.1}(\mathrm{syst})\right) \times 10^{-9} \text{ and} \tag{7.18}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = \left(3.7^{+2.4}_{-2.1}(\mathrm{stat})^{+0.6}_{-0.4}(\mathrm{syst})\right) \times 10^{-10}\,, \tag{7.19}$$

for the branching fractions of B_s^0 and B^0 mesons into a muon pair, respectively. The projection of the fit results in all eight BDT bins is shown in Fig. 7.9. The sensitivity increases with the BDT bin. While the first BDT bins in Fig. 7.9 is dominated by the combinatorial background, the most sensitive bins for high BDT values contain dominantly signal and exclusive background events. The combined invariant mass distribution of the three most sensitive bins (BDT > 0.7) is shown in Fig. 7.10. A clear excess of events in the B mass region is visible.
In order to quantify the significances of the two signals the p-value with respect to the background-only hypothesis is calculated analogously to section 7.7. However, for the observed significance the cross-feed from the respective second signal decay is taken into account. The probability to observe an at least as extreme B_s^0 signal given

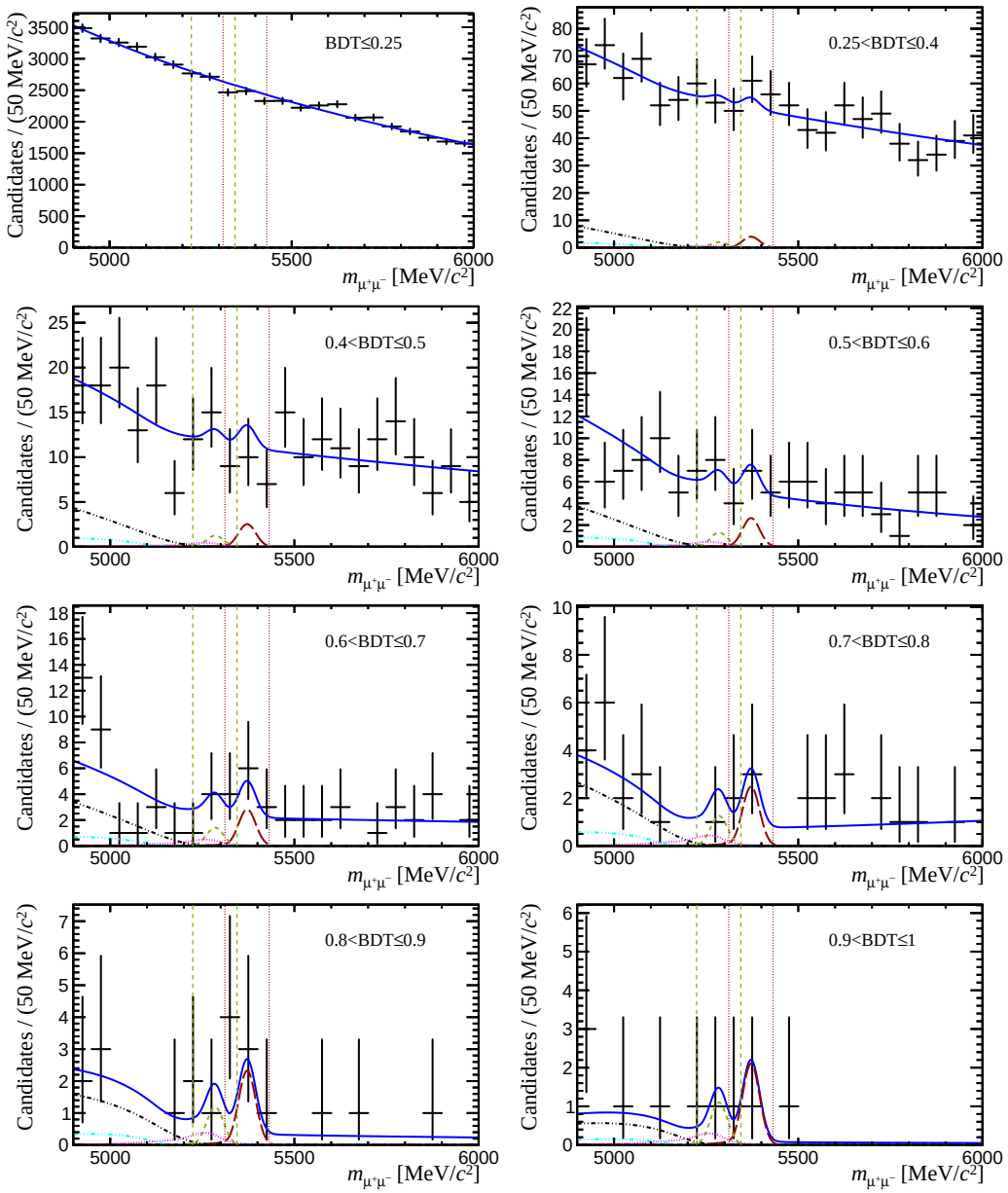


Figure 7.9: Projections of the fit result to the invariant mass distribution of the selected $B \to \mu^+\mu^-$ candidates (black dots) in all eight bins of the BDT classifier. The total fit result is displayed in solid blue, while the dashed red and green lines mark the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signals, respectively. The different background components are in black the $B^0 \to \pi^-\mu^+\nu_\mu$, in cyan the $B \to \pi\mu^+\mu^-$, and in violet the $B \to h^+h'^-$ background decays [116].

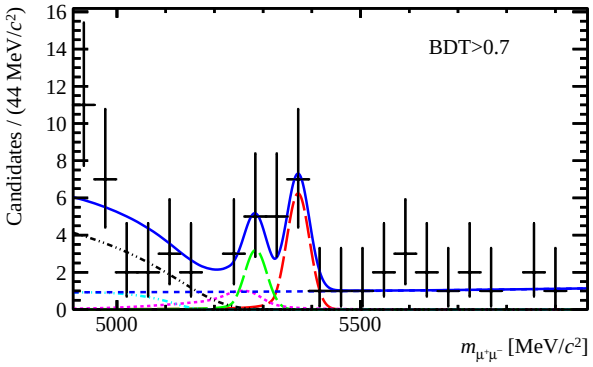


Figure 7.10: Invariant dimuon mass distribution of the three most sensitive BDT bins. The data is given by the black points, the total fit is shown as a solid blue line, and the signal decays are highlighted as dashed red and dashed green lines for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ channels, respectively. The background components are the combinatorial in dotted blue, $B^0 \to \pi^-\mu^+\nu_\mu$ in dashed-dotted black, $B \to \pi\mu^+\mu^-$ in dashed-dotted cyan, and the $B \to h^+h'^-$ in dotted violet [116].

the background-only hypothesis is $p = 1.4 \times 10^{-4}$ corresponding to a significance of 3.8σ. The $B^0 \to \mu^+\mu^-$ signal is not statistically significant, therefore an upper limit is set on the branching fraction using the CL_s method [216]. The expected and observed limit in dependence of the $B^0 \to \mu^+\mu^-$ branching fraction hypothesis is shown in Fig. 7.11. The upper limits on the branching fraction of $B^0 \to \mu^+\mu^-$ are

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 6.3 \times 10^{-10} \text{ with } 90\% \text{ CL and} \tag{7.20}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 7.4 \times 10^{-10} \text{ with } 95\% \text{ CL.} \tag{7.21}$$

The fact that the observed upper limit is constantly above the expected upper limit for all $B^0 \to \mu^+\mu^-$ branching fraction hypotheses indicates a slight excess of events. Whether the excess indicates a signal of $B^0 \to \mu^+\mu^-$ decays or is a statistical fluctuation cannot be determined from the analysed LHCb dataset alone.

7.9 Future prospects

Future prospects for the analysis of $B \to \mu^+\mu^-$ decays at the LHCb experiment are studied for the same two scenarios in terms of integrated luminosity as in Section 5.7 and Section 6.5. Two aspects are studied: the expected signal significances and the evolution of the uncertainties for the $B \to \mu^+\mu^-$ signals. The significances depends

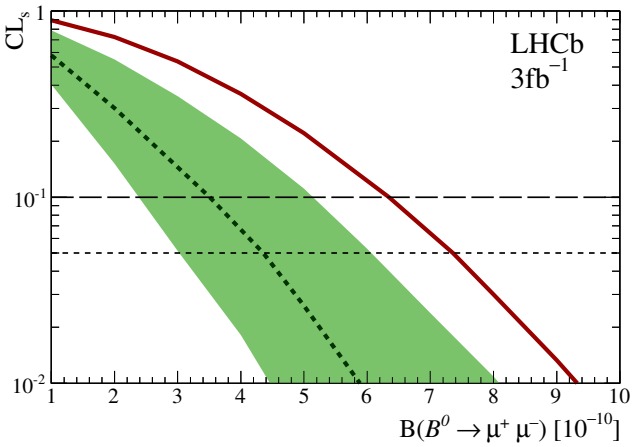
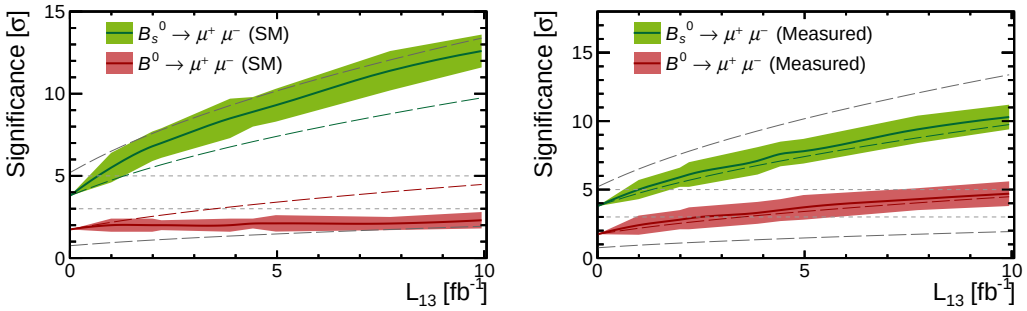


Figure 7.11: Result of the calculation of the expected and observed upper limit in dependence of the $B^0 \to \mu^+\mu^-$ branching fraction. The horizontal lines mark the 90% and 95% confidence level used to calculate the upper limits. The dashed black line shows the upper limit under the background-only hypothesis with the 1σ uncertainty band in green. The solid red line highlights the observed limit. The measured limits on the $B^0 \to \mu^+\mu^-$ branching fraction are the intersection of the horizontal lines with the solid red line [116].

on the branching fraction hypotheses. Hence, the extrapolations are performed for the SM predictions from Eq. 2.25 and from the measured results in Eq. 7.18 and Eq. 7.19. The expected significances are calculated from sets of pseudo-experiments where the number of expected background events is scaled to the total integrated luminosity and the number of signal events is calculated according to the SM or the *measured* hypothesis.
The significances are calculated from pseudo-datasets using Wilks' theorem. The influence of the statistical approximation is negligible compared to the level of approximations made on the future integrated luminosities and efficiencies[4]. The expected significances for the $B \to \mu^+\mu^-$ signals in dependence of the additional integrated luminosity are shown in Fig 7.12. A simple extrapolation of the expected and measured significances of the two $B \to \mu^+\mu^-$ signals using a scale factor of $1/\sqrt{\mathcal{N}}$ (cf. Eq. 5.10). It can be seen that for both branching fraction hypotheses a discovery of $B_s^0 \to \mu^+\mu^-$ decays at the LHCb experiment has a probability of more than 50% with an additional integrated luminosity of $0.7{-}1.3\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$, which might be achieved at the end of the year 2015.
The situation for $B^0 \to \mu^+\mu^-$ is different as the SM prediction and the measured result differ by a factor of four. Starting with the measured dataset, adding pseudo-

[4]For the combined LHCb and CMS analysis the level of approximation of Wilks' theorem in terms of significance is found to be less than 10%, cf. Section 8.4.



(a) (b)

Figure 7.12: Expected significances with 1σ uncertainty bands in dependence of the integrated luminosity normalised to $\sqrt{s} = 13\,\text{TeV}$ centre-of-mass energies. The dotted lines show the extrapolations using a simple $\sqrt{L_{\text{tot}}/L_{\text{RunI}}}$ approach. The dotted grey lines show the extrapolation from the expected RunII significance under the SM prediction hypothesis. (a) Scenario with SM predictions for the $B \to \mu^+\mu^-$ branching fractions. (b) Scenario with the measured central values from the $B \to \mu^+\mu^-$ combination in Chapter 8.

data under the SM hypothesis effectively decreases the signal to background ratio. This is illustrated by the fact that the extrapolated B^0 significance approaches the extrapolation of the pure SM expectation in Fig. 7.12(a) with increasing luminosities. Even with the full RunII dataset it is not expected to observe an excess of $B^0 \to \mu^+\mu^-$ events with a significance of greater 3σ ($\mathcal{P}(z \geq 3\sigma) < 1{-}5\%$). However, this picture changes for the measured hypothesis. In this case, LHCb will be able to claim an evidence for $B^0 \to \mu^+\mu^-$ events with an additional integrated luminosity of $2.5\,\text{fb}^{-1}$ ($\mathcal{P}(z \geq 3\sigma) \approx 50\%$). In the optimistic scenario with $10\,\text{fb}^{-1}$ collected in RunII there even is a chance to observe $B^0 \to \mu^+\mu^-$ decays with a significance of 5σ or more with $\mathcal{P}(z \geq 5\sigma) \approx 30{-}40\%$.

The second interesting figure to extrapolate is the expected uncertainty on the branching fraction measurements. It can be seen from Fig. 7.12 that the scaling with the factor $1/\sqrt{\mathcal{N}}$ works reasonably well. Additionally, the expected uncertainty does, at first order, not depend on the measured central values. Therefore, the uncertainties are extrapolated without the use of dedicated pseudo-datasets. This implicitly assumes that the systematic uncertainties scale according to the statistical uncertainties. However, the measured uncertainty is dominated by the statistical uncertainty so that this assumption is valid up to the point where the systematic uncertainties are in the same order of the statistical uncertainty. This will be the case at the end of RunII for $B_s^0 \to \mu^+\mu^-$, while for $B^0 \to \mu^+\mu^-$ the measurement is statistically limited throughout RunII. The scaled uncertainties on the branching fraction measurements for $B \to \mu^+\mu^-$ decays are shown in Fig. 7.13 for the results of the LHCb experiment. The theoretical uncertainties (cf. Eq. 2.25) are given

as a scale to quantify the size of the experimental uncertainties. After RunII of the LHC, the experimental uncertainties from the results of the LHCb experiment will still be a factor of 2–9 larger than the corresponding theory uncertainties.

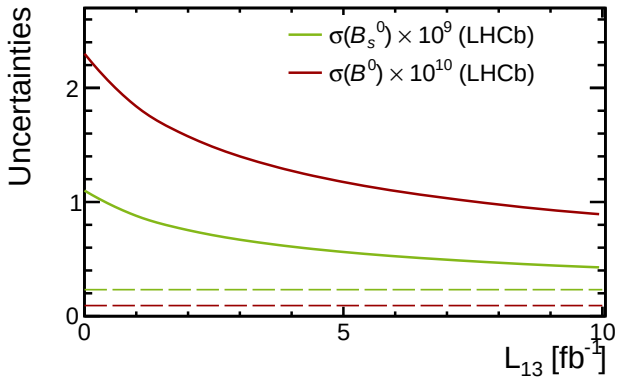


Figure 7.13: Extrapolations of the measured uncertainties on the branching fractions of $B_s^0 \to \mu^+\mu^-$ (green) and $B^0 \to \mu^+\mu^-$ (red) decays. As a comparison the respective theoretical uncertainties are given as dashed lines.

8 Combination of $B \to \mu^+\mu^-$ results from the LHCb and CMS experiments

This chapter describes the combination of the two $B \to \mu^+\mu^-$ analyses performed by LHCb [116] and CMS [217] based on the full Run I datasets corresponding to 3 and $25\,\text{fb}^{-1}$ for the LHCb and CMS experiments, respectively. The LHCb measurement is in detail described in Chapter 7. The CMS analysis is briefly summarised in Section 8.1. The focus of this summary are the parts which are relevant for the combined analysis with the LHCb dataset. More details on the CMS analysis can be found in Ref. [217]. The full likelihood combination is described in Section 8.2. The expected and observed significances, and the determination of confidence intervals are discussed in the Sections 8.3, 8.4, and 8.5, respectively. The necessary statistical tools and methods for the combination of the LHCb and CMS $B \to \mu^+\mu^-$ analyses are implemented in the GammaCombo framework [182] within this thesis. The description of the combination focuses on the LHCb contributions to the analysis, however for each of the shown results a cross-check implementation from the CMS analysis team is present and the results are consistent throughout.

8.1 Summary of the CMS analysis

The overall strategy used by the CMS collaboration is similar to the one established by the LHCb group. After the full selection steps the branching fractions are determined relative to a normalisation channel ($B^+ \to J/\psi\, K^+$) using a simultaneous fit to the invariant dimuon mass, $m_{\mu\mu}$, distribution in different categories of the BDT outputs. However, the detector geometry and running conditions of the CMS detector imply subtle differences in the analysis. Foremost, the pileup in the CMS detector is large compared to LHCb. While recording $5\,\text{fb}^{-1}$ integrated luminosity of proton-proton interactions at a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ in 2011 and $20\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$ in 2012, the average pileup increased from 9 to 21. Another issue is that for the barrel and endcap region different sub-detector layouts lead to significantly different trigger and detection efficiencies and mass resolutions. As a consequence, the $B \to \mu^+\mu^-$ candidates that pass the trigger and the muon selections are split into four categories in terms of centre-of-mass energy ($\sqrt{s} = 7\,\text{TeV}$ or $\sqrt{s} = 8\,\text{TeV}$) and detector region (barrel or endcap). An event is categorised

as *barrel* only if both reconstructed muons have a pseudorapidity of $|\eta| < 1.4$. If one of the two muons has a $|\eta| > 1.4$ the reconstructed particle is categorised as an *endcap* candidate. Due to the different conditions a single BDT is trained for each category to separate signal from background. As training samples simulated events where used as signal and data sidebands dominated by random combination of muons as background. The resulting BDT output is binned in two (four) bins for the $\sqrt{s} = 7\,\text{TeV}$ ($\sqrt{s} = 8\,\text{TeV}$) dataset. The bin boundaries are chosen such that each bin contains the same number of expected signal events. In total, twelve categories of invariant dimuon mass distributions are fit simultaneously using the unbinned extended maximum likelihood method (cf. Eq. 3.12). The shapes of the signal, peaking and "semileptonic" background components are taken from simulated samples. Here, peaking backgrounds are $B \to h^+h'^-$ decays and "semileptonic" labels the combination of $B \to h^-\mu^+\nu_\mu$, $B \to h\mu^+\mu^-$, and $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ decays. The combinatorial component is a third-order polynomial. The invariant dimuon mass resolution is determined on an event-by-event basis and varies significantly between in different categories. Hence, a PDF describing the per-event mass resolution is added as a conditional product, such that the signal PDFs reflect the correct widths according to the per-event mass resolution. The simultaneous fit to the twelve categories yields

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \left(3.0^{+1.0}_{-0.9}\right) \times 10^{-9}\,, \tag{8.1}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = \left(3.5^{+2.1}_{-1.8}\right) \times 10^{-10}\,. \tag{8.2}$$

The statistical significance of the B_s^0 and B^0 signals are 4.3σ and 2.0σ, respectively. Similar to the $B_s^0 \to \mu^+\mu^-$ analysis performed at the LHCb experiment (cf. in particular Fig. 7.1 and Fig 7.9) the sensitivity per bin increases with the output of the BDT classifier. Thus, in the representation of the full fit result in Fig. 8.1 three bins are not shown. These are the lowest BDT bins for the barrel region for $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ data and the lowest BDT bin for the endcap region for $\sqrt{s} = 8\,\text{TeV}$ data. A more illustrative version of the invariant dimuon mass distribution is shown in Fig. 8.2. Here all events are considered but are weighted according to $\text{S}/(\text{S}+\text{B})$, where S is the expected signal and B the total background both evaluated under the B_s^0 peak.

Comparing Fig. 8.2 to the version published by the LHCb collaboration in Fig. 7.10 the differences of the experiments become evident. While the LHCb experiment is able to resolve the B^0–B_s^0 mass difference and thus can distinguish $B^0 \to \mu^+\mu^-$ from $B_s^0 \to \mu^+\mu^-$ events, this is more complicated with the CMS detector due to the worse and varying mass resolution. However, the CMS experiment has in total a larger event yield, which of course also results in a higher background level. The results agree within a standard deviation with the results published by the LHCb collaboration [116]. Both experiments observed nearly the same significances.
As the analysis of $B_{(s)}^0 \to \mu^+\mu^-$ decays is highly recognised throughout the high energy physics community, the goal is to publish a combined LHC RunI result based

on the available result in summer 2013. However, as the experimental results are correlated through the theoretical inputs, simple methods of combining results do not yield satisfying results. Especially, there is no way to consistently combine p-values to access the combined significance of a signal channel.

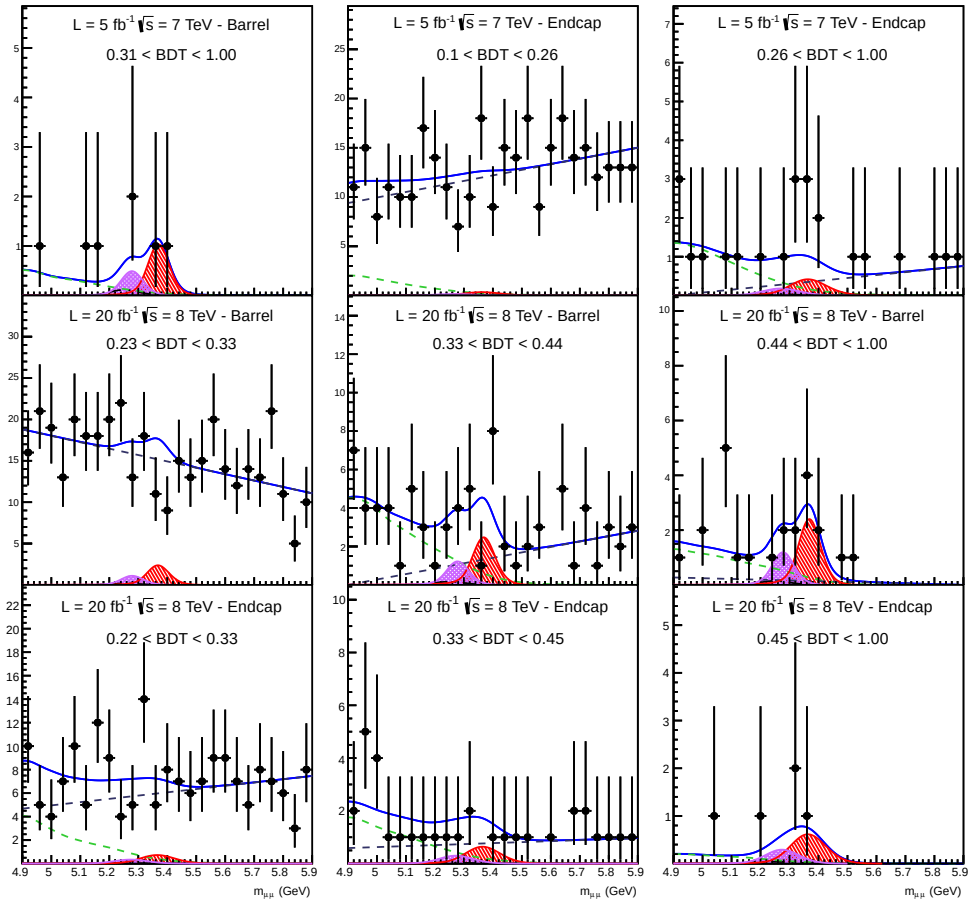


Figure 8.1: Results of the simultaneous fit to the invariant dimuon mass distribution in categories of the centre-of-mass energy, CMS detector region, and BDT classifiers. Shown are the fit result in solid blue, the signal is shaded in red and violet for the B_s^0 and B^0 components, respectively. The various background components are the combinatorial as dashed dark blue line, the semileptonic in dashed green and the peaking backgrounds in dashed-dotted black. Figure modified from [217].

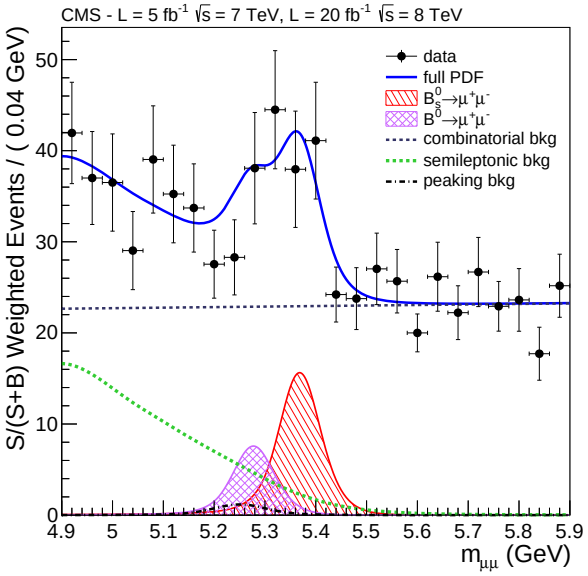


Figure 8.2: Invariant dimuon mass distribution weighted by S/(S + B) over all twelve categories measured by the CMS experiment, where S is the expected signal and B the total background under the B_s^0 peak. This plot is used for illustrative purposes only. Figure taken from [217].

8.2 Combined maximum likelihood fit

A method to combine correlated measurements is to perform an analysis based on combined datasets. In this way, correlations can directly be accounted for in a common description of the data. A simultaneous maximum likelihood fit to the datasets then yields the correct combined result.
As both experiments provide independent sets of proton-proton interactions with different systematic uncertainties evaluated separately for each detector, the datasets remain unchanged compared to the single experiment publications. As a consequence, the simultaneous fit to the invariant dimuon mass distribution is performed in twenty categories split by BDT output for the LHCb dataset and centre-of-mass energy, detector region, and BDT classifiers for the CMS data.
The invariant mass models are modified with respect to the single experiment probability density functions to achieve a common background treatment. Therefore, the $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ background component is added to the list of explicitly modelled background contributions in the LHCb dimuon mass description. For the CMS dimuon mass model the branching fraction prediction for $\Lambda_b^0 \to p^+\mu^-\overline{\nu}_\mu$ is updated and the kinematic properties of the simulated events are adjusted to result in more realistic distributions of the final state particles (cf. Section 7.6.3). In addition, the decay time dependent bias arising from the use of the average B_s^0 lifetime in the Monte

Carlo simulation is accounted for in each of the CMS categories (cf. Section 7.3). The parameters which are assumed to be fully correlated and therefore are used as common variables in all the categories of the simultaneous fit are the B_s^0 and B^0 branching fractions and the dominant theoretical inputs: the branching fraction of the normalisation channel $B^+ \to J/\psi K^+$ and the ratio of hadronisation fractions $f_\mathrm{s}/f_\mathrm{d}$.
The statistical and systematic uncertainties on the various nuisance parameters are reflected in the fit model by allowing the nuisance parameters to float within gaussian constraints with the width of the respective total uncertainty.
The total simultaneous fit to twenty invariant dimuon mass distributions yields

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9} \text{ and} \tag{8.3}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10}, \tag{8.4}$$

where the uncertainty combines the statistical and systematic uncertainty. The latter accounts for 35% and 18% of the total uncertainty for the B_s^0 and B^0 branching fraction, respectively. The resulting invariant dimuon mass distribution together with the fit projection for the six categories with highest $\mathrm{S}/\sqrt{\mathrm{S}+\mathrm{B}}$, as defined in Section 8.1, is shown in Fig. 8.3. A clear excess of events is visible in the region around the B^0 and B_s^0 masses. Two different parametrisations are used in order to extract the strength parameters $\mathcal{S}(B^0_{(s)})$ defined as the measured signal branching fraction over the respective SM prediction and the ratio of the two signal branching fractions $\mathcal{R}$ (Eq. 2.29). For the strength parameters $\mathcal{S}(B^0_{(s)})$ the SM predictions are incorporated into the nominal fit, so that the theoretical uncertainties get propagated into the fit result. The simultaneous fit yields

$$\mathcal{S}(B_s^0) = 0.76^{+0.20}_{-0.18} \quad \text{and} \quad \mathcal{S}(B^0) = 3.7^{+1.6}_{-1.4}. \tag{8.5}$$

The parametrisation for extracting $\mathcal{R}$ cancels the relative uncertainties from the measurements of the $B^+ \to J/\psi K^+$ normalisation channels, so that the parameter depends only on the two measured yields of signal events, the respective efficiencies and the ratio of hadronisation fractions $f_\mathrm{s}/f_\mathrm{d}$ (cf. Eq. 7.4). The simultaneous fit results in

$$\mathcal{R} = \frac{\mathcal{B}(B^0 \to \mu^+\mu^-)}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)} = 0.14^{+0.08}_{-0.06}. \tag{8.6}$$

The compatibility of the results with the respective SM predictions is discussed in Section 8.5 in detail.

8.3 Expected signal significances

The expected significance is an important information for searches of rare particle decays, as it quantifies the sensitivity of the experiment. In order to access the

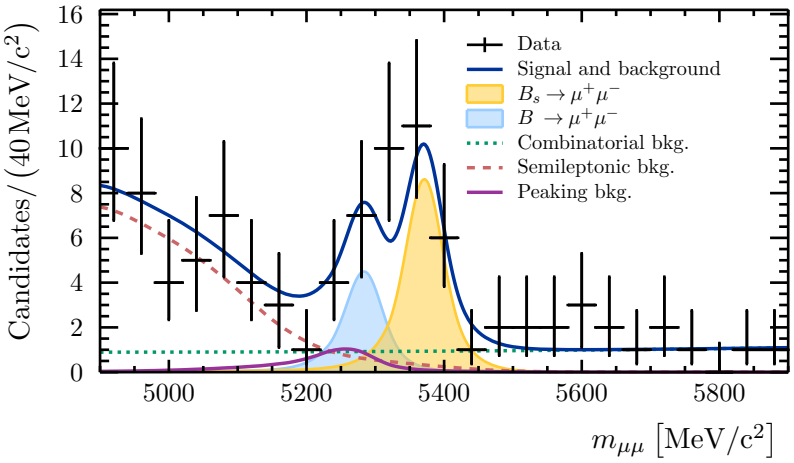


Figure 8.3: Illustrative invariant mass distribution in the six bins with highest $S/\sqrt{S+B}$, where S is the number of expected B_s^0 signal events and B the number of background events under the B_s^0 signal peak. The combined LHCb and CMS data is represented by black crosses and the total signal and background is shown as a solid blue line. The signals are shown in light yellow and blue for the B_s^0 and B^0 modes, respectively. The various background components are the combinatorial (dotted green), the semileptonic (dotted red), and the peaking background (dashed violet).

expected significance of the combined analysis, pseudo experiments are generated using the SM predictions for the branching fractions of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. The generation of the pseudo-datasets follows the PLUGIN method described in Section 3.3.4. In this case, a pseudo-dataset is a set of twenty invariant mass distributions following the PDFs of the nominal combined fit model. The nuisance parameters are fixed to the results of a fit to the measured data with the signal branching fractions fixed to the SM predictions. Given the SM predictions, about 95 B_s^0 and 11 B^0 events are expected in the combined LHCb and CMS dataset.
The significance for the signal decays in each pseudo-dataset is determined by calculating the test statistic $\Delta\chi^2$ for the best estimate of the respective branching fraction, $\chi^2(\hat{\mathcal{B}})$, and the hypotheses for background only, $\chi^2(\mathcal{B}=0)$, where $\mathcal{B}$ denotes either the branching fraction for $B_s^0 \to \mu^+\mu^-$ or $B^0 \to \mu^+\mu^-$ decays. Therefore, each pseudo-experiment is fit three times:

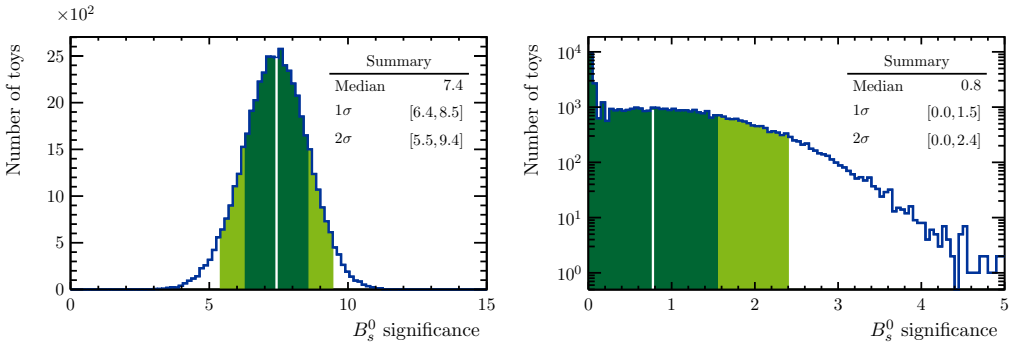
1.) To get the best estimate $\chi^2(\hat{\mathcal{B}})$, the signal branching fractions are left freely floating.

2.) A fit where the branching fraction $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is fixed to zero, while $\mathcal{B}(B^0 \to \mu^+\mu^-)$ is freely floating determining $\chi^2(\mathcal{B}_s = 0)$.

3.) Like 2.) with reversed treatment of the signal parameters.

From these results two $\Delta\chi^2(\mathcal{B}_{s,d}) = \chi^2(\mathcal{B}_{s,d} = 0) - \chi^2(\hat{\mathcal{B}})$ values are calculated (see Eq. 3.13) from which the significance z is calculated using $z = \sqrt{\Delta\chi^2}$ (cf. Section 3.4). The resulting distributions for the expected significances for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signal are shown in Fig. 8.4. The dark and light green areas highlight the 1σ and 2σ uncertainty regions, which are constructed by ensuring that the fraction of pseudo-experiments is as close to the desired coverage of 68.27% and 95.45% as possible and do not undercover.

The expected significance is defined as the median of these distributions and is found to be 7.4σ for $B_s^0 \to \mu^+\mu^-$ and 0.8σ for $B^0 \to \mu^+\mu^-$ decays. It is worth noting that the peak at expected significances $z = 0$ for the B^0 decay is due to the best estimator for $\mathcal{B}(B^0 \to \mu^+\mu^-)$ being equal or close to zero. As the parametrisation for the signals strictly constrains the branching fractions to positive values, all estimations resulting in negative values are accumulated at $\mathcal{B}(B^0 \to \mu^+\mu^-) = 0$. The calculation of the p-value gets therefore biased, as the test statistic $\Delta\chi^2_{B^0}$ equals zero in a significant amount of times. This is due to the best estimate in the physically allowed range is equal to the estimate of the null hypothesis, $\chi^2(\mathcal{B}_d = 0) = \chi^2(\hat{\mathcal{B}})$. Using Wilks' theorem in this case results in a p-value of $p = 1$ and according to Eq. 3.16 in a significance of $z = -\infty$. For the distributions of the expected significances in Fig. 8.4 all negative significances are counted as $z = 0$.



(a) $B_s^0 \to \mu^+\mu^-$ significance (SM)

(b) $B^0 \to \mu^+\mu^-$ significance (SM)

Figure 8.4: Significance distributions for the SM (a) $B_s^0 \to \mu^+\mu^-$ and (b) $B^0 \to \mu^+\mu^-$ signals over the background-only hypothesis. The dark and green areas mark the corresponding one and two sigma regions. The solid white line marks the median of the distributions.

This shows that the physical boundary of positive branching fractions significantly affects the calculation of the significance and the confidence levels for the B^0 signal. Hence, it is crucial to apply a robust statistical method which does not rely on assumptions and can deal with boundaries. The application of such a method to the B^0 signal is described in Chapter 8.4 and 8.5.2.
Assuming the SM predictions, the joint analysis of the combined CMS and LHCb datasets expects to observe the $B_s^0 \to \mu^+\mu^-$ decay. The probability to measure a significance greater than five is $P(z \geq 5) = 97.88\%$. For the $B^0 \to \mu^+\mu^-$ decay the expected sensitivity is much smaller, but the distribution has a long tail towards higher significances. The probability to observe a significance of at least 3σ, given the SM assumption, is $P(z \geq 3) = 1.66\%$.

8.4 Observed signal significances

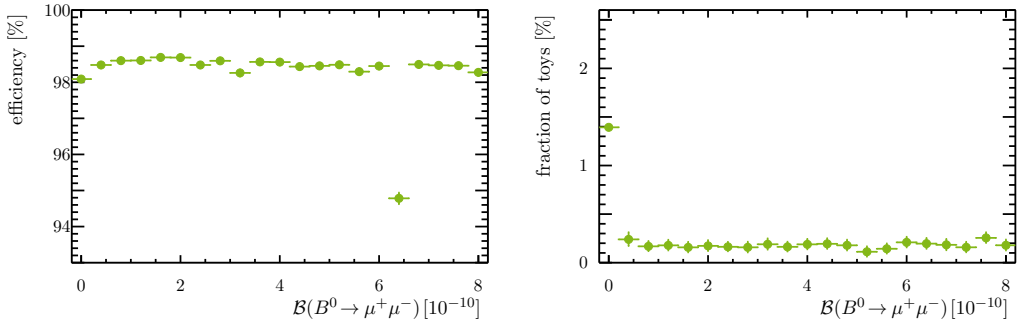
The statistical significance is related to the rejection of the null hypothesis. To test how significant a signal is the background-only hypothesis is chosen as the null hypothesis. As described in Section 3.4 there are a couple of methods to evaluate the significance. The approximative significances calculated by the $z = \sqrt{\Delta\chi^2}$ value (and Wilks' theorem) yield $z = 6.15\sigma$ ($z = 6.04\sigma$) for the B_s^0 signal and $z = 3.17\sigma$ ($z = 2.96\sigma$) for the B^0 signal. Both approximations rely on an asymptotic gaussian behaviour. However, as described in Section 3.3.1 and Section 8.3 the presence of a strict physical boundary breaks the gaussian approximation. Hence, the approximate signal significance yield wrong results.

8.4.1 Significance calculation using pseudo experiments

As no gaussian approximation of the significance can be used, the significance is calculated based on pseudo-experiments. It is necessary to construct the $\Delta\chi^2$ distribution under the background-only hypothesis. Therefore, the values for all parameters are determined by a fit to the measured data where $\mathcal{B}(B^0 \to \mu^+\mu^-)$ is fixed to zero. To ensure a sufficient statistical precision in the lower populated tails of the $\Delta\chi^2$ distribution, $\mathcal{O}(10^5)$ pseudo-datasets are generated. Analogously to Section 8.3, each toy dataset is fit twice to obtain the likelihoods for the best estimate ($\chi^2(\hat{\mathcal{B}})$) and the null hypothesis ($\chi^2(\mathcal{B} = 0)$). However, the minimisations of the negative logarithms of the likelihood is not successful for all pseudo-datasets or can result in unphysical (negative) $\Delta\chi^2$ values. This can result in problematic interpretations, when the fraction of unsuccessful toys is in the order of the supposed p-value.

Convergence criteria

For the calculation of the test statistic in Eq. 3.13 two maximisations of the experimental likelihood are necessary. These are carried out using extended maximum likelihood fits using the common minimisation package MINUIT [130,131] to minimise the negative logarithm of the likelihood. The combined fit model used in Section 8.2 has in total 41 free and additional 111 constrained parameters, which makes the numerical minimisation challenging. For each fit MINUIT needs to converge without issues (`Minuit::Status` equal to zero) and the value of the minimised negative logarithm of the likelihood must be a valid number, i.e. not $-2\ln\mathcal{L} = \pm\infty$ or `nan` (not-a-number flag). The efficiency is evaluated on the complete considered $\mathcal{B}(B^0 \to \mu^+\mu^-)$ parameter space and is shown in Fig. 8.5(a). Despite a single less efficient point at $\mathcal{B}(B^0 \to \mu^+\mu^-) = 6.4 \times 10^{-10}$, the selection criteria remove about 2% of the fits. This particular point shows no distinct behaviour in the evolution of the nuisance parameter over the scan range. Hence, there is no particular reason for the efficiency loss, but it is argued that this does not affect any conclusions inferred from the pseudo-experiments. Although the convergence criteria ensure that the likelihood values are finite and are obtained from successful minimisations, a considerable fraction of toy datasets result in negative test statistic values (cf. Fig. 8.5(b)) which are mathematically not allowed.
The point for $\mathcal{B}(B^0 \to \mu^+\mu^-) = 0$ causes more instabilities. This is due to the fact that the branching fraction is constrained to positive values. To save as much computing time as possible per toy experiment this boundary is implemented as a strict limitation of the range of the branching fraction parameter.



(a) Efficiency of convergence criteria

(b) Fractions of negative test statistic toys

Figure 8.5: (a) Efficiency of convergence criteria. The less efficient point at $\mathcal{B}(B^0 \to \mu^+\mu^-) = 6.4 \times 10^{-10}$ is studied in detail. No distinct feature in the distributions of the nuisance parameters is found. (b) Fraction of pseudo-experiments that result in a negative test statistic $\Delta\chi^2$ in dependence of the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ hypotheses.

In this way the likelihood is directly maximised under the right conditions in one single fit, but parameters converging closer to their range limit increase the instability. Furthermore, due to the boundary, fits to a pseudo-dataset that is generated with $\mathcal{B}(B^0 \to \mu^+\mu^-) = 0$ result in B^0 branching fractions equal or close to zero in over 50% of the times. As a consequence, an increased rate of non-converging fits and negative test statistic values are expected for $\mathcal{B}(B^0 \to \mu^+\mu^-)$ hypotheses close to zero. A possible effect on the statistical evaluation is studied in the following section.

Treatment of unsuccessful toys

The point $\mathcal{B}(B^0 \to \mu^+\mu^-) = 0$ has two special characteristics. Firstly it introduces the highest instabilities (cf. Fig. 8.5), but on the other hand it is used as the null hypothesis for calculating the B^0 signal significance. As it is necessary to study significances up to three standard deviations, which corresponds to a p-value of 0.135%, it needs to be ensured that the fraction of rejected toy datasets (1.9%) does not bias the significance calculation. In cases where the likelihood fit results in either an error or in a non-finite number, the likelihood ratio cannot be constructed. Hence, it is necessary to find another, alternative, test statistic which can be used to calculate the p-value. Therefore, the value of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ during the last successful minimisation step before raising the convergence error is re-interpreted as an alternative test statistic. Although the meaning of this value is somehow limited if the fit did not converge, the hope is that the value is nevertheless close to the actual minimum. The distributions of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ results where either the fit with fixed $\mathcal{B}(B^0 \to \mu^+\mu^-)$ or the fit with freely floating B^0 branching fraction failed is shown in Fig. 8.6. The $\mathcal{B}(B^0 \to \mu^+\mu^-)$ distribution from the nominal toy experiments, i.e. the distribution of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ results from successfully converging fits is shown for comparison. In order to estimate the effect of the disregarded toy experiment the p-Value with respect to the background-only hypothesis is calculated using the distribution of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ values where either of the two fits failed. The result is consistent with the p-Value calculated from the same $\mathcal{B}(B^0 \to \mu^+\mu^-)$ test statistic but with the nominal convergence criteria:

$$\text{p(failed fits)} = 1.29 \times 10^{-3} \;\text{; p} \in [0.78,2.06] \times 10^{-3} \text{ with 68\% CL and} \tag{8.7}$$

$$\text{p(converged)} = 1.26 \times 10^{-3} \;\text{; p} \in [1.20,1.36] \times 10^{-3} \text{ with 68\% CL .} \tag{8.8}$$

From this it is concluded that the calculation of the significance is not biased by the rejection of non-converging fits. The 68% confidence intervals in Eq. 8.7 represent binomial Clopper-Pearson intervals [126] which prevent under coverage.
After applying the convergence criteria the test statistic still has a negative or unphysical contribution (1.4%). As can be seen from Fig. 8.7(a), most of the negative values are close to zero. This can be explained by the numerical minimisation.

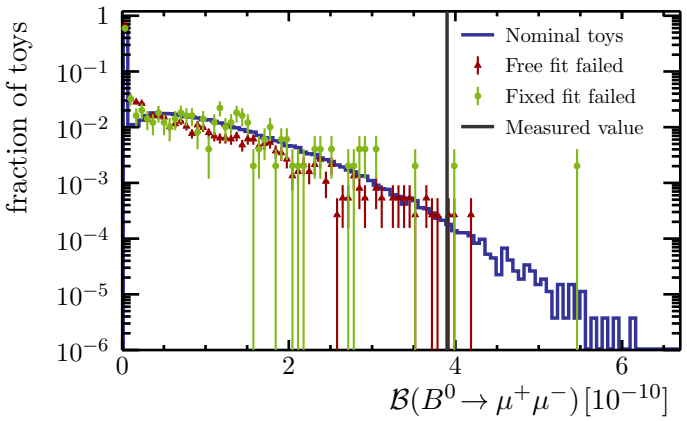
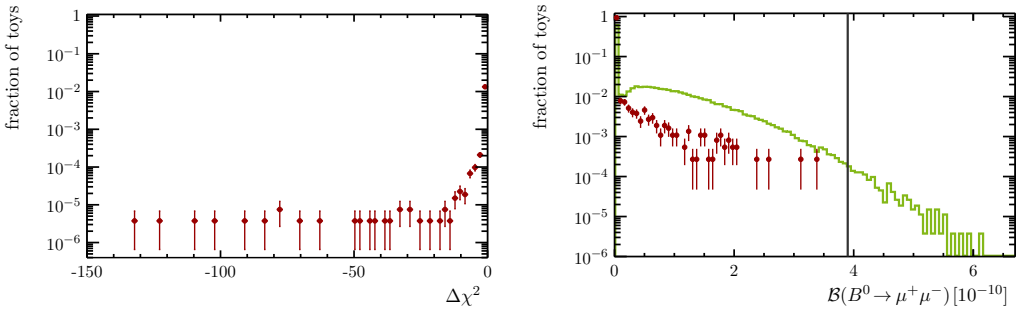


Figure 8.6: Distribution of floating $\mathcal{B}(B^0 \to \mu^+\mu^-)$ results for the nominal toys in green and the fraction of fits where either the fit with fixed $\mathcal{B}(B^0 \to \mu^+\mu^-)$ failed (blue) or the fit with freely floating $\mathcal{B}(B^0 \to \mu^+\mu^-)$ failed (red).



(a) Fraction of toys with $\Delta\chi^2 < 0$

(b) $\mathcal{B}(B^0 \to \mu^+\mu^-)$ for toys with $\Delta\chi^2 < 0$

Figure 8.7: Fraction of all pseudo-datasets with negative $\Delta\chi^2$ values (a). Distribution of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ (b) for the nominal toys (green) and these resulting in a negative test statistic (blue). The measured value for $\mathcal{B}(B^0 \to \mu^+\mu^-)$ is shown as a vertical black line.

Larger negative values can occur when e.g. the fit with freely floating $\mathcal{B}(B^0 \to \mu^+\mu^-)$ converges in a local instead of the global minimum. Again, the distribution of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ values is used to interpret the effect negative $\Delta\chi^2$ values. The distribution of the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ values for the fit with freely floating branching fraction that result in a negative $\Delta\chi^2$ value is shown in Fig. 8.7(b) compared to the nominal $\mathcal{B}(B^0 \to \mu^+\mu^-)$ distribution. It can be seen that indeed most (94%) of the freely floating fits that produce a negative test statistic converge at $\mathcal{B}(B^0 \to \mu^+\mu^-) \approx 0$. In this case the negative $\Delta\chi^2$ is numerically equal to zero. In contrast, when the $\Delta\chi^2$ value is not close to zero, it is a priori not clear how to proceed. Any value up

to which the $\Delta\chi^2$ is be treated as zero, would be arbitrary. However, Fig. 8.7(b) suggests that none of the pseudo-experiments enters the numerator of the p-value calculation, as all branching fraction values are smaller than the measured value. As a consequence, the negative $\Delta\chi^2$ values only change the number of total pseudo-experiments, i.e. the normalisation of the p-value calculation. To avoid any bias towards higher significances for the $B^0 \to \mu^+\mu^-$ signal, the pseudo-datasets that results in a negative $\Delta\chi^2$ value are discarded from the nominal calculation of the p-value and significance.

8.4.2 Results

The resulting distribution of the nominal likelihood ratio test statistic from Eq. 3.13 is shown in Fig. 8.8. The uncertainties represent a 68% CL interval of a poissonian distribution with a mean equal to the bin content. The p-value is equal to the fraction of pseudo-experiments that have a larger or equal test statistic value as the measured dataset

$$\mathrm{p} = \frac{N\left(\Delta\chi^2 \geq \Delta\chi^2_{\mathrm{data}}\right)}{N_{\mathrm{tot}}} \quad . \tag{8.9}$$

The p-value of the data under the background-only hypothesis is $p = (1.35 \pm 0.01) \times 10^{-3}$. Using the definition of gaussian significance in Eq. 3.16 this p-value results in significance of $z = (3.00 \pm 0.02)\sigma$. This result is reproduced by the independent implementation from the CMS group yielding $p = 1.34 \pm 0.01$ ($z = 3.00 \pm 0.01$). The comparison of the nominal results with the alternative approaches using the distribution $\mathcal{B}(B^0 \to \mu^+\mu^-)$ values as a test statistic and the influence of the negative $\Delta\chi^2$ values are summarised in Tab. 8.1. All performed tests indicate a significance of the $B^0 \to \mu^+\mu^-$ signal of at least three gaussian standard deviations. Hence, the first evidence for the decay $B^0 \to \mu^+\mu^-$ can be claimed.

Table 8.1: Summary of p-value and significances calculated using different test statistics and normalisations. No discrepancies are found between the results.

Method	$\mathbf{p}\,[10^{-3}]$	**68% CL**	$\mathbf{z}\,[\sigma]$	**68% CL**
nominal $\Delta\chi^2$	1.33	[1.26, 1.41]	3.00	[2.99, 3.02]
$\mathcal{B}(B^0 \to \mu^+\mu^-)$	1.26	[1.20, 1.36]	3.02	[3.00, 3.04]
failed $\mathcal{B}(B^0 \to \mu^+\mu^-)$	1.29	[0.78, 2.06]	3.01	[2.87, 3.16]
incl. $\Delta\chi^2 < 0$	1.33	[1.26, 1.41]	3.00	[2.99, 3.02]
negl. $\Delta\chi^2 < 0$	1.35	[1.28, 1.42]	3.00	[2.98, 3.02]

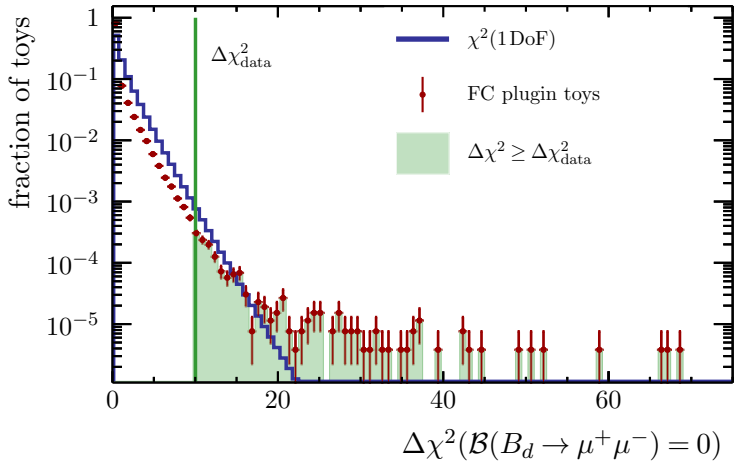


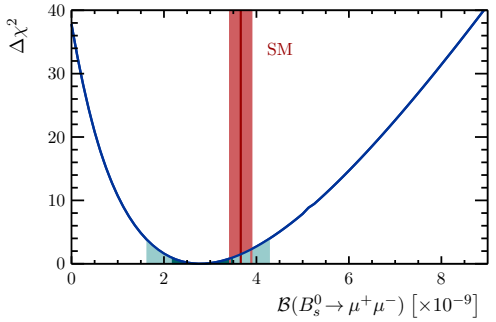
Figure 8.8: Comparison of the outcome of the pseudo-experiments (red points) and the χ^2 approximation (solid blue). The $\Delta\chi^2$ of the measured LHCb and CMS dataset is shown as vertical green line and the p-Value is the area of the light green region.

8.5 Confidence intervals

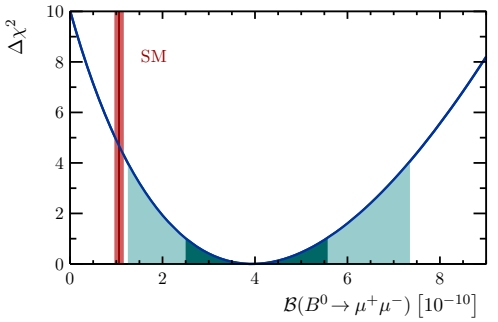
A crucial point in the measurement of $B \to \mu^+\mu^-$ decays is a correct estimate of the uncertainties. Therefore, it is necessary to quantify the amount of disagreement between the approximative methods usually used to calculate the uncertainties and a method, which provides better coverage probabilities. Hence, the method proposed by Feldman and Cousins (cf. Section 3.3) is used to construct one and two sigma confidence levels for the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ estimation in Section 8.5.2. These intervals are compared to the results of the conventional intervals using Wilks' theorem presented in Section 8.5.1.

8.5.1 One dimensional intervals

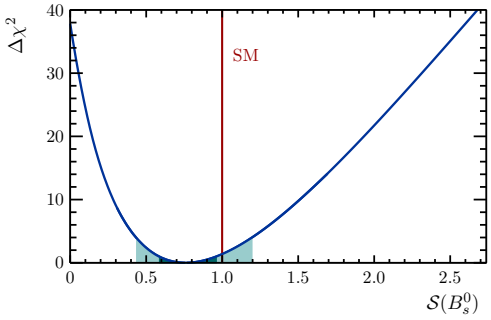
For each parametrisation of the observables ($\mathcal{B}(B \to \mu^+\mu^-)$, $\mathcal{S}(B^0_{(s)})$, and $\mathcal{R}$, cf. Section 8.2) a one-dimensional profile likelihood scan is performed in order to evaluate the respective confidence intervals. Therefore, the measured dataset is fit with different fixed values of the parameter of interest while all other parameters are left freely floating or within their respective constrains. These fits are compared to the nominal result by calculating the likelihood ratio $\Delta\chi^2$. The distributions of the $\Delta\chi^2$ values in dependence of the branching fractions, signal strengths and ratio hypotheses are shown in Fig. 8.9.



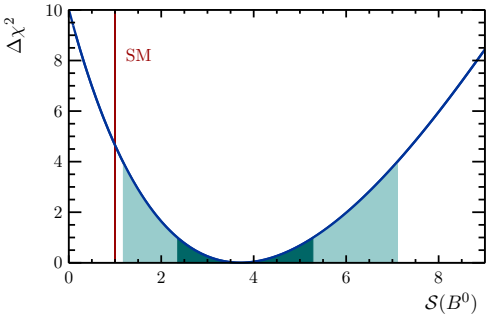
(a) Scan of the B_s^0 branching fracion



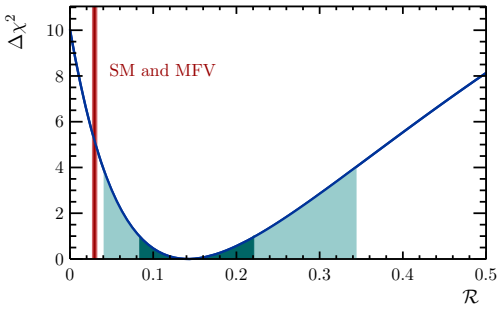
(b) Scan of the B^0 branching fracion



(c) Scan of the B_s^0 signal strength



(d) Scan of the B^0 signal strength



(e) Scan of the ratio of branching fractions $\mathcal{B}(B^0 \to \mu^+\mu^-)/\mathcal{B}(B_s^0 \to \mu^+\mu^-)$

Figure 8.9: Test statistic in dependence of the parameter hypotheses evaluated on the measured LHCb and CMS dataset. The 1σ and 2σ confidence intervalls are highlighted in dark and light cyan, respectively. The different parameters are (a) the branching fraction for the B_s^0 signal, (b) the branching fraction for B^0 signal, (c) the signal strength for the B_s^0 signal, (d) the signal strengh of the B^0 signal, and (e) the ratio of branching fractions.

According to Table A.1, the one and two sigma confidence intervals are the unions of all hypotheses that result in a $\Delta\chi^2$ values less than one and four. In addition, the SM predictions and uncertainties are shown. By construction the SM prediction for the signal strength parameters $\mathcal{S}(B^0_{(s)})$ is equal to unity and has no uncertainty (cf. Section 8.2). Due to the fact that the statistical experimental uncertainty is the dominating source of uncertainty, the single branching fraction measurement and the parametrisation using the strength parameters result in similar likelihood scans. The numerical values for the confidence intervals and the distance of the SM predictions to the best fit central value in units of gaussian standard deviations are summarised in Table 8.2 in Section 8.5.2.

8.5.2 Feldman-Cousins confidence interval

The 68.27% and 95.45% confidence intervals for the $B^0 \to \mu^+\mu^-$ signal are calculated using the FC method. The PLUGIN procedure (cf. Section 3.3.4) is adopted for the nuisance parameters. The $\mathcal{B}(B^0 \to \mu^+\mu^-)$ parameter is scanned between $[0,8] \times 10^{-10}$ using 21 scan points. For each point 2×10^4 pseudo-datasets are generated. An exception is the point $\mathcal{B}(B^0 \to \mu^+\mu^-) = 0$, where the pseudo-experiments from the significance calculation (cf. Section 8.4.1) are used. For each fixed scan point a p-value with respect to the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ hypothesis is calculated. Due to the discrete nature of the scan the resulting $1 - \mathrm{CL}$ values are interpolated using segments of third order polynomials. The results are shown in Fig. 8.10 together with approximative curve using Wilks' theorem. In addition the 3σ significance or 3σ one-sided confidence interval is displayed. The 1σ and 2σ confidence intervals are found to be

$$\mathcal{B}(B^0 \to \mu^+\mu^-) \in [2.48, 5.58] \times 10^{-10} \quad \text{at } 68.27\% \text{ CL}\,,$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) \in [1.36, 7.42] \times 10^{-10} \quad \text{at } 95.45\% \text{ CL}\,.$$

The differences between the χ^2 approximation and the Feldman-Cousins approach, which are evident for small and large $\mathcal{B}(B^0 \to \mu^+\mu^-)$ hypotheses in Fig. 8.10, arise from different sources. Firstly, the influence of the physical boundary at $\mathcal{B}(B^0 \to \mu^+\mu^-) = 0$ and secondly, the a small bias in the estimation of the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ values. As seen in Section 3.3.2 the confidence intervals for biased estimators are widened to remain the correct coverage. Hence, the slightly wider FC intervals for large $\mathcal{B}(B^0 \to \mu^+\mu^-)$ hypotheses can be explained.
A very good agreement between the approximative and the accurate Feldman-Cousins calculation of the confidence intervals up to at least 2σ is found. This indicates that the Wilks' theorem can be used to evaluate confidence intervals and contours with decent precision. The resulting confidence intervals for the different observables together with the deviations from the respective SM predictions are summarised in Table 8.2.
Although the central value for the B^0 signal branching fraction is almost a factor of four higher than the SM prediction, due to the large statistical uncertainties, all measurements are consistent with the SM.

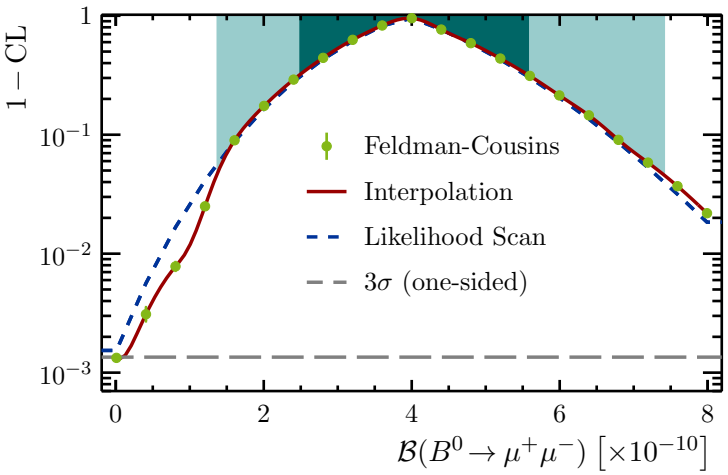


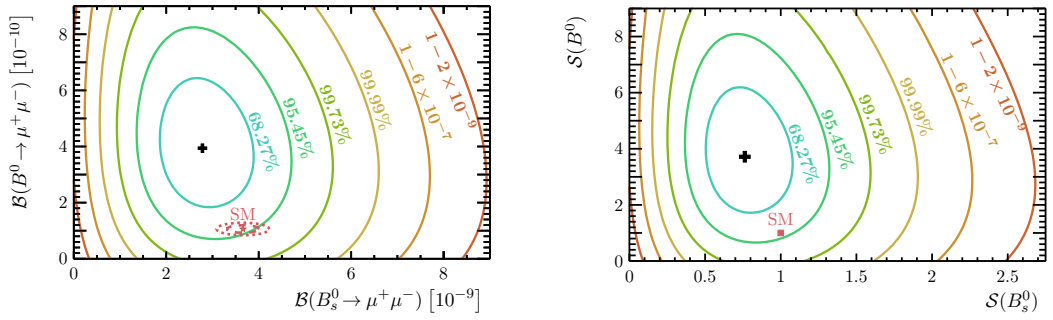
Figure 8.10: Confidence intervals from the unified frequentist method, based on the combined LHCb and CMS dataset, are shown in light (1σ) and dark (2σ) cyan. The green points show the $1-\mathrm{CL}$ value for each fixed $\mathcal{B}(B^0 \to \mu^+\mu^-)$ hypothesis, while the light green line is the interpolation to these points. For comparison, the χ^2 approximation from Wilks' theorem is shown as a dashed blue line.

8.5.3 Two-dimensional confidence regions

In order to calculate two-dimensional confidence contours, the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ vs $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, or equivalently the $\mathcal{S}(B^0)$ vs $\mathcal{S}(B_s^0)$ plane is divided into a 81×81 point grid in the ranges of $\mathcal{B}(B^0 \to \mu^+\mu^-) \in [0,10] \times 10^{10}$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-) \in [0,10] \times 10^{-9}$. For each scan point two maximum likelihood fits are performed where either the two signal branching fractions are fixed to their respective scan point values or the branching fractions are left freely floating. From these two fits the $\Delta\chi^2$ values are calculated and translated into confidence contours by including all points to the confidence region where $\Delta\chi^2 \leq \Delta\chi^2_{\mathrm{n}\sigma}$. Here, $\Delta\chi^2_{\mathrm{n}\sigma}$ is the value where the χ^2 distribution with two degrees of freedom has the same coverage as a one-dimensional gaussian distribution within n standard deviations (cf. Table A.1). The results for the branching fraction and signal strength observables are shown in Fig. 8.11 together with the SM predictions.
Due to the dominating statistical uncertainty in the experiment, the contour plots for the branching fractions and the signal strengths are nearly identical. The SM point is included into the 2σ ellipse in both parametrisations and also the one-dimensional scans show agreement with the SM. The measured branching fractions for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are consistent with the SM within 1.2σ and 2.2σ, respectively. The ratio of branching fractions is consistent with the SM prediction within 2.3σ.

Table 8.2: Summary of the confidence intervals ($CL_{n\sigma}$) and the distance to the SM predictions (ΔSM) for the observables $\mathcal{B}(B\to\mu^+\mu^-)$, $\mathcal{S}(B^0_{(s)})$, and $\mathcal{R}$. The $\mathcal{B}(B^0\to\mu^+\mu^-)$ confidence intervals from the Feldman-Cousins construction are shown for comparison. The values for $\mathcal{B}(B\to\mu^+\mu^-)$ are in units of 10^{-9}.

	$\mathbf{CL_{1\sigma}}$	$\mathbf{CL_{2\sigma}}$	$\mathbf{\Delta SM}$
$\mathcal{B}(B^0\to\mu^+\mu^-)$ (FC)	$[0.25, 0.56]$	$[0.14, 0.74]$	
$\mathcal{B}(B^0\to\mu^+\mu^-)$	$[0.25, 0.56]$	$[0.13, 0.74]$	2.2σ
$\mathcal{B}(B^0_s\to\mu^+\mu^-)$	$[2.17, 3.48]$	$[1.62, 4.28]$	1.2σ
$\mathcal{S}(B^0)$	$[2.35, 5.28]$	$[1.17, 7.11]$	2.2σ
$\mathcal{S}(B^0_s)$	$[0.58, 0.96]$	$[0.43, 1.20]$	1.2σ
$\mathcal{R}$	$[0.08, 0.22]$	$[0.04, 0.34]$	2.3σ



(a) Profile likelihood scan: $\mathcal{B}(B^0\to\mu^+\mu^-)$ vs $\mathcal{B}(B^0_s\to\mu^+\mu^-)$

(b) Profile likelihood scan: $\mathcal{S}(B^0)$ vs $\mathcal{S}(B^0_s)$

Figure 8.11: Two-dimensional profile likelihood scans, where the coverages of the respective regions are given next to the contours. The best fit point is shown as a black cross. (a) Confidence regions in the $\mathcal{B}(B^0\to\mu^+\mu^-)$ versus $\mathcal{B}(B^0_s\to\mu^+\mu^-)$ plane while the SM prediction is marked by a red star. The dashed (dotted) red ellipses show the 68.27% (95.45%) confidence regions for the SM prediction assuming no theoretical correlations. (b) Confidence regions in the $\mathcal{S}(B^0)$ versus $\mathcal{S}(B^0_s)$ plane. The SM uncertainties are folded into the confidence regions. Given the dominating statistical uncertainties of the experiment, the difference in the plots is negligible.

8.6 Future prospects

For the prospects of a possible future combination, the estimations from Section 7.9 and Ref. [218] are combined. An additional luminosity of $L_{13} = 5\,\text{fb}^{-1}$ is assumed for the LHCb and $100\,\text{fb}^{-1}$ for the CMS experiment. A simple error estimate is calculated as $\sigma \approx \sqrt{\sigma(\text{LHCb})^2 + \sigma(\text{CMS})^2}/2$. This results in a sensitivity of $\sigma(B_s^0) \approx 0.3 \times 10^{-9}$ and $\sigma(B^0) \approx 0.8 \times 10^{-10}$. This is still a factor of 1.5 and 9 larger than the SM theory uncertainties of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions, respectively. Still, if the central value of the $B^0 \to \mu^+\mu^-$ signal is as measured by the LHCb and CMS experiments, the precision is sufficient to observe $B^0 \to \mu^+\mu^-$ decays in a combined analysis during Run II and to possibly separate the experimental value from the SM prediction by more than 3σ. In order to test $B \to \mu^+\mu^-$ decays to a precision of $< 10\%$ an increased luminosity is needed, which will be provided by the upgrade phases of the LHCb [195] and CMS [219] experiments.

9 Conclusions and prospects

In today's particle physics experiments, very large datasets are recorded and analysed on a statistical basis. The methods used are crucial for the interpretation of the measured datasets. Within this thesis statistical methods are implemented in a framework and applied to the measurement of key observables in the SM flavour sector.

This thesis presents the first time-dependent measurement of the CKM angle γ from $B_s^0 \to D_s^{\mp} K^{\pm}$ decays. With a dataset of $1\,\text{fb}^{-1}$ of pp interactions recorded with the LHCb detector at a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ it is possible to measure the CKM angle γ as

$$\gamma = (115^{+27}_{-43})^\circ \,,$$

where the uncertainties are evaluated using the frequentist PLUGIN method. The same statistical method is then applied to the LHCb combination of a set of measured CP observables from tree-level $B \to Dh$ decays, which results in the most precise measurement of the CKM angle γ:

$$\gamma = (78.9^{+5.8}_{-7.4})^\circ \,.$$

Compared to the legacy results of the BaBar and Belle experiments, the LHCb result has less than half of the uncertainties. This marks an important step in the measurement of the CKM angle γ from tree-level decays, as a benchmark test for the SM.

In the second part of this thesis rare decays, which are highly sensitive to physics beyond the SM, are studied. Two of the most important processes are the decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. The branching fractions of these decays can cleanly be calculated within the SM and, therefore, provide a precision test of the SM flavour sector. This thesis presents the analysis of $B \to \mu^+\mu^-$ decays based on the full $3\,\text{fb}^{-1}$ LHCb RunI dataset. With precise estimates of the background components, the analysis yields

$$\begin{aligned}
&\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \left(2.9^{+1.1}_{-1.0}(\text{stat})^{+0.3}_{-0.1}(\text{syst})\right) \times 10^{-9} \text{ and} \\
&\mathcal{B}(B^0 \to \mu^+\mu^-) = \left(3.7^{+2.4}_{-2.1}(\text{stat})^{+0.6}_{-0.4}(\text{syst})\right) \times 10^{-10} \,,
\end{aligned}$$

for the branching fractions of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, respectively. To achieve the most precise determination of the branching fractions of $B \to \mu^+\mu^-$ decays from the Run I LHC data, the available $B \to \mu^+\mu^-$ datasets are combined in the last part of this thesis. This is the first time, where LHC experiments combine datasets and perform a joint analysis. The necessary statistical methods and tools are developed within this thesis, which yields the first observation of the decay $B_s^0 \to \mu^+\mu^-$ and the first evidence for $B^0 \to \mu^+\mu^-$ decays with statistical significances of 6.2σ and 3.0σ, respectively. Additionally, for the first time the ratio of branching fractions of $B \to \mu^+\mu^-$ decays is measured as

$$\mathcal{R} = 0.14^{+0.08}_{-0.06},$$

which is compatible with the SM prediction within 2.3σ. Although the measurement of the branching fraction of $B_s^0 \to \mu^+\mu^-$ decays at a rate consistent with the SM prediction can be considered a great success for the SM and the CKM mechanism [220], there is still room for physics beyond the SM to enter the rate of the $B_s^0 \to \mu^+\mu^-$ decay. An excess of events is found in the $B^0 \to \mu^+\mu^-$ channel resulting in a measured branching fraction that is a factor of four larger than the SM prediction. However, in order to determine if this is a statistical fluctuation or a hint for a possible non-MFV structure of physics beyond the SM, more data is needed.

The second data-taking period of the LHC has started in June 2015 with a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and will last until the end of 2018. During this period it will be possible to measure the CKM angle γ to a precision of $\approx 2^\circ$ from combined $B \to Dh$ decays.
For the analysis of $B \to \mu^+\mu^-$ decays the focus during Run II of the LHC will be on the $B^0 \to \mu^+\mu^-$ decay channel. The expectations for this channel, however, depend on the actual branching fraction. The decay channel could be observed at the end of Run II, if the presented measurements reflect an increased branching fraction of $B^0 \to \mu^+\mu^-$ decays and not a statistical fluctuation. Hence, the precise determination of the $B \to \mu^+\mu^-$ branching fractions and their ratio remain key observables throughout Run II and beyond.

Bibliography

[1] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[2] A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. **13** (1964) 168.

[3] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[4] D. Hanneke, S. Fogwell, and G. Gabrielse, *New Measurement of the Electron Magnetic Moment and the Fine Structure Constant*, Phys. Rev. Lett. **100** (2008) 120801.

[5] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[6] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[7] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

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

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

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

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

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

[13] ATLAS collaboration, G. Aad *et al.*, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. **B726** (2013) 120.

[14] CMS collaboration, V. Khachatryan *et al.*, *Observation of the diphoton decay of the Higgs boson and measurement of its properties*, Eur. Phys. J. **C74** (2014) 3076.

[15] CMS collaboration, V. Khachatryan *et al.*, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, Eur. Phys. J. **C75** (2015) 212.

[16] ATLAS and CMS collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV*, 2015. ATLAS-CONF-2015-044.

[17] Planck collaboration, J. Tauber *et al.*, *The Scientific programme of Planck*, `arXiv:astro-ph/0604069`.

[18] M. Markevitch *et al.*, *Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56*, Astrophys. J. **606** (2004) 819.

[19] G. Steigman, *When Clusters Collide: Constraints On Antimatter On The Largest Scales*, JCAP **0810** (2008) 001.

[20] L. Canetti, M. Drewes, and M. Shaposhnikov, *Matter and Antimatter in the Universe*, New J. Phys. **14** (2012) 095012.

[21] A. D. Sakharov, *Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32, [Usp. Fiz. Nauk161, 61 (1991)].

[22] A. J. Buras, D. Buttazzo, J. Girrbach-Noe, and R. Knegjens, *Can we reach the Zeptouniverse with rare K and $B_{s,d}$ decays?*, JHEP **11** (2014) 121.

[23] LHCb collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115** (2015) 072001.

[24] LHCb collaboration, R. Aaij *et al.*, *Observation of the resonant character of the $Z(4430)^-$ state*, Phys. Rev. Lett. **112** (2014) 222002.

[25] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Perseus Books Group, Reading, USA, 1995.

[26] D. H. Perkins, *Introduction to High Energy Physics*, Cambridge University Press, Cambridge, UK, 4th ed., 2000.

[27] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, John Wiley & Sons, New York, USA, 1984.

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

[29] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

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

[31] BaBar collaboration, B. Aubert *et al.*, *Observation of CP Violation in the* B^0 *Meson System*, Phys. Rev. Lett. **87** (2001) 091801.

[32] Belle collaboration, K. Abe *et al.*, *Observation of Large CP Violation in the Neutral B Meson System*, Phys. Rev. Lett. **87** (2001) 091802.

[33] LHCb collaboration, R. Aaij *et al.*, *Observation of CP violation in* $B^{\pm} \to DK^{\pm}$ *decays*, Phys. Lett. **B712** (2012) 203.

[34] LHCb collaboration, R. Aaij *et al.*, *First observation of CP violation in the decays of* B_s^0 *mesons*, Phys. Rev. Lett. **110** (2013) 221601.

[35] G. C. Branco, R. G. Felipe, and F. R. Joaquim, *Leptonic CP Violation*, Rev. Mod. Phys. **84** (2012) 515.

[36] T. Mannel, *Theory and Phenomenology of CP Violation*, Nucl. Phy. B, Proc. Supp. **167** (2007) 170 , Proceedings of the 7th International Conference on Hyperons, Charm and Beauty Hadrons.

[37] C. A. Baker *et al.*, *An Improved experimental limit on the electric dipole moment of the neutron*, Phys. Rev. Lett. **97** (2006) 131801.

[38] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. D **2** (1970) 1285.

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

[40] Particle Data Group, K. A. Olive *et al.*, *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.

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

[42] J. Charles *et al.*, *Current status of the Standard Model CKM fit and constraints on $\Delta F = 2$ New Physics*, Phys. Rev. **D91** (2015) 073007.

[43] K. G. Wilson and W. Zimmermann, *Operator product expansions and composite field operators in the general framework of quantum field theory*, Comm. Math. Phys. **24** (1972) 87.

[44] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, Rev. Mod. Phys. **68** (1996) 1125.

[45] A. J. Buras, *Flavor dynamics: CP violation and rare decays*, Subnucl. Ser. **38** (2002) 200.

[46] K. G. Wilson, *Non-lagrangian models of current algebra*, Phys. Rev. **179** (1969) 1499.

[47] Belle and BaBar collaborations, A. J. Bevan *et al.*, *The Physics of the B Factories*, Eur. Phys. J. **C74** (2014) 3026.

[48] LHCb collaboration, B. Adeva *et al.*, *Roadmap for selected key measurements of LHCb*, `arXiv:0912.4179`.

[49] BaBar collaboration, J. P. Lees *et al.*, *Observation of direct CP violation in the measurement of the Cabibbo-Kobayashi-Maskawa angle γ with $B^\pm \to D^{(*)}K^{(*)\pm}$ decays*, Phys. Rev. **D87** (2013) 052015.

[50] Belle collaboration, K. Trabelsi, *Study of direct CP in charmed B decays and measurement of the CKM angle γ at Belle*, in *7th Workshop on the CKM Unitarity Triangle (CKM 2012) Cincinnati, Ohio, USA, September 28th–October 2nd, 2012*, 2013. `arXiv:1301.2033`.

[51] M. Gronau and D. London, *How to determine all the angles of the unitarity triangle from $B^0 \to DK^0_S$ and $B^0_s \to D^0\phi$*, Phys. Lett. **B253** (1991) 483.

[52] M. Gronau and D. Wyler, *On determining a weak phase from CP asymmetries in charged B decays*, Phys. Lett. **B265** (1991) 172.

[53] D. Atwood, I. Dunietz, and A. Soni, *Enhanced CP violation with $B \to KD^0$ ($\overline{D}^0$) modes and extraction of the CKM angle γ*, Phys. Rev. Lett. **78** (1997) 3257.

[54] A. Giri, Y. Grossman, A. Soffer, and J. Zupan, *Determining γ using $B^{\pm} \to DK^{\pm}$ with multibody D decays*, Phys. Rev. **D68** (2003) 054018.

[55] Y. Grossman, Z. Ligeti, and A. Soffer, *Measuring γ in $B^{\pm} \to K^{\pm}(KK^{*0})_D$ decays*, Phys. Rev. **D67** (2003) 071301.

[56] Du, Dongsheng and Dunietz, Isard and Wu, Dan, *Systematic study of large CP violations in decays of neutral b-flavored mesons*, Phys. Rev. D **34** (1986) 3414.

[57] I. Dunietz and J. L. Rosner, *Time-dependent CP-violation effects in $B^0 - \overline{B}^0$ systems*, Phys. Rev. D **34** (1986) 1404.

[58] R. Aleksan, I. Dunietz, and B. Kayser, *Determining the CP violating phase γ*, Z. Phys. **C54** (1992) 653.

[59] R. Fleischer, *New strategies to obtain insights into CP violation through $B_s^0 \to D_s^{*\pm}K^{\mp}$, $D_s^{*\pm}K^{\mp}$, ... and $B^0 \to D^{\pm}\pi^{\mp}$, $D^{*\pm}\pi^{\mp}$, ... decays*, Nucl. Phys. **B671** (2003) 459.

[60] BaBar collaboration, B. Aubert *et al.*, *Measurement of time-dependent CP-violating asymmetries and constraints on* $\sin(2\beta+\gamma)$ *with partial reconstruction of $B \to D^{*\mp}\pi^{\pm}$ decays*, Phys. Rev. **D71** (2005) 112003.

[61] BaBar collaboration, B. Aubert *et al.*, *Measurement of time-dependent CP asymmetries in $B^0 \to D^{(*)\pm}\,\pi^{\mp}$ and $B^0 \to D^{\pm}\rho^{\mp}$ decays*, Phys. Rev. **D73** (2006) 111101.

[62] Belle collaboration, F. J. Ronga *et al.*, *Measurements of CP violation in $B^0 \to D^{*-}\pi^+$ and $B^0 \to D^-\pi^+$ decays*, Phys. Rev. **D73** (2006) 092003.

[63] Belle collaboration, S. Bahinipati *et al.*, *Measurements of time-dependent CP asymmetries in $B \to D^{*\mp}\pi^{\pm}$ decays using a partial reconstruction technique*, Phys. Rev. **D84** (2011) 021101.

[64] G. C. Branco, L. Lavoura, and J. P. Silva, *CP Violation*, Oxford University Press, New York, USA, 1999.

[65] CLEO collaboration, J. Insler *et al.*, *Studies of the decays* $D^0 \to K^0_S K^- \pi^+$ *and* $D^0 \to K^0_S K^+ \pi^-$, Phys. Rev. **D85** (2012) 092016.

[66] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP asymmetry in* $B^0_s \to D^\mp_s K^\pm$ *decays*, JHEP **11** (2014) 060.

[67] M. A. Baak, *Measurement of CKM angle* γ *with charmed* B^0 *meson decays*, PhD thesis, Vrije U. Amsterdam, Feb, 2007, CERN-THESIS-2007-165.

[68] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of CP violation in* $B^0_s \to J/\psi K^+ K^-$ *decays*, Phys. Rev. Lett. **114** (2015) 041801.

[69] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation and the* B^0_s *meson decay width difference with* $B^0_s \to J/\psi\, K^+ K^-$ *and* $B^0_s \to J/\psi\, \pi^+ \pi^-$ *decays*, Phys. Rev. **D87** (2013) 112010.

[70] A. J. Buras, *Operator product expansion, renormalization group and weak decays*, Lect. Notes Phys. **558** (2000) 65.

[71] T. Inami and C. S. Lim, *Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes* $K^0_{\rm L} \to \mu^+ \mu^-$, $K^+ \to \pi^+ \nu\overline{\nu}$ *and* $K^0 \leftrightarrow \overline{K}^0$, Prog. Theor. Phys. **65** (1981) 297, [Erratum: Prog. Theor. Phys.65 , 1772 (1981)].

[72] G. Buchalla and A. J. Buras, *QCD corrections to rare K and B decays for arbitrary top quark mass*, Nucl. Phys. **B400** (1993) 225.

[73] C. Bobeth, T. Ewerth, F. Kruger, and J. Urban, *Analysis of neutral Higgs boson contributions to the decays* $B^0_s \to \ell^+ \ell^-$ *and* $\overline{B} \to K \ell^+ \ell^-$, Phys. Rev. **D64** (2001) 074014.

[74] Fermilab Lattice and MILC collaboration, A. Bazavov *et al.*, *B- and D-meson decay constants from three-flavor lattice QCD*, Phys. Rev. **D85** (2012) 114506.

[75] H. Na *et al.*, *The* B^0 *and* B^0_s *Meson Decay Constants from Lattice QCD*, Phys. Rev. **D86** (2012) 034506.

[76] O. Witzel, *B-meson decay constants with domain-wall light quarks and non-perturbatively tuned relativistic b-quarks*, PoS **LATTICE2013** (2014) 377.

[77] ATLAS, CDF, CMS and D0 collaborations, *First combination of Tevatron and LHC measurements of the top-quark mass*, 2014. ATLAS-CONF-2014-008.

[78] C. Bobeth *et al.*, $B_{s,d} \to \ell^+ \ell^-$ *in the Standard Model with Reduced Theoretical Uncertainty*, Phys. Rev. Lett. **112** (2014) 101801.

[79] W. Altmannshofer, P. Paradisi, and D. M. Straub, *Model-Independent Constraints on New Physics in $b \to s$ Transitions*, JHEP **04** (2012) 008.

[80] C.-S. Huang, L. Wei, Q.-S. Yan, and S.-H. Zhu, *$B_s \to l^+l^-$ in a type-II two-Higgs-doublet model and the minimal supersymmetric standard model*, Phys. Rev. D **63** (2001) 114021.

[81] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, *Minimal flavor violation: An Effective field theory approach*, Nucl. Phys. **B645** (2002) 155.

[82] G. Isidori, Y. Nir, and G. Perez, *Flavor Physics Constraints for Physics Beyond the Standard Model*, Ann. Rev. Nucl. Part. Sci. **60** (2010) 355.

[83] A. J. Buras, *Relations between $\Delta M_{s,d}$ and $B_{s,d} \to \mu\bar{\mu}$ in models with minimal flavor violation*, Phys. Lett. **B566** (2003) 115.

[84] CMS and LHCb collaborations, V. Khachatryan *et al.*, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature 522, 68–72 (2015) .

[85] CLEO collaboration, R. Giles *et al.*, *Two-body decays of B mesons*, Phys. Rev. D **30** (1984) 2279.

[86] LHCb collaboration, R. Aaij *et al.*, *First Evidence for the Decay $B_s^0 \to \mu^+\mu^-$*, Phys. Rev. Lett. **110** (2013) 021801.

[87] CLEO collaboration, P. Avery *et al.*, *Limits on rare exclusive decays of B mesons*, Phys. Lett. B **183** (1987) 429.

[88] CLEO collaboration, P. Avery *et al.*, *A search for exclusive penguin decays of B mesons*, Phys. Lett. B **223** (1989) 470.

[89] CLEO collaboration, R. Ammar *et al.*, *Search for B^0 decays to two charged leptons*, Phys. Rev. D **49** (1994) 5701.

[90] CLEO collaboration, T. Bergfeld *et al.*, *Search for decays of B^0 mesons into pairs of leptons: $B^0 \to e^+e^-$, $B^0 \to \mu^+\mu^-$ and $B^0 \to e^\pm\mu^\mp$*, Phys. Rev. D **62** (2000) 091102.

[91] ARGUS collaboration, H. Albrecht *et al.*, *B meson decays into charmonium states*, Phys. Lett. B **199** (1987) 451.

[92] UA1 collaboration, C. Albajar *et al.*, *Low mass dimuon production at the CERN proton-antiproton collider*, Phys. Lett. B **209** (1988) 397.

[93] UA1 collaboration, C. Albajar *et al.*, *A search for rare B meson decays at the CERN SppS collider*, Phys. Lett. B **262** (1991) 163.

[94] L3 collaboration, M. Acciarri *et al.*, *Search for neutral B meson decays to two charged leptons*, Phys. Lett. B **391** (1997) 474.

[95] CDF collaboration, F. Abe *et al.*, *Search for flavor-changing neutral current B meson decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\,\text{TeV}$*, Phys. Rev. Lett. **76** (1996) 4675.

[96] CDF collaboration, F. Abe *et al.*, *Search for the decays $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\,\text{TeV}$*, Phys. Rev. D **57** (1998) 3811.

[97] CDF collaboration, D. Acousta *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,\text{TeV}$*, Phys. Rev. Lett. **93** (2004) 032001.

[98] CDF collaboration, A. Abulencia *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays in $p\bar{p}$ collisions with CDF II*, Phys. Rev. Lett. **95** (2005) 221805.

[99] CDF collaboration, T. Aaltonen *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with CDF II*, Phys. Rev. Lett. **107** (2011) 191801.

[100] CDF collaboration, T. Aaltonen *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays with the full CDF Run II data set*, Phys. Rev. **D87** (2013) 072003.

[101] DØ collaboration, B. Abbott *et al.*, *Search for the decay $B \to X_s\mu^+\mu^-$*, Phys. Lett. B **423** (1998) 419.

[102] DØ collaboration, V. Abazov *et al.*, *A search for the flavor-changing neutral current decay $B_s^0 \to \mu^+\mu^-$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,\text{TeV}$ with the DØ detector*, Phys. Rev. Lett. **94** (2005) 071802.

[103] DØ collaboration, V. Abazov *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ at DØ*, Phys. Rev. D **76** (2007) 092001.

[104] DØ collaboration, V. M. Abazov *et al.*, *Search for the rare decay $B_s^0 \to \mu^+\mu^-$*, Phys. Lett. B **693** (2010) 539.

[105] D0 collaboration, V. M. Abazov *et al.*, *Search for the rare decay $B_s^0 \to \mu^+\mu^-$*, Phys. Rev. **D87** (2013) 072006.

[106] ATLAS collaboration, G. Aad *et al.*, *Search for the decay $B_s^0 \to \mu^+\mu^-$ with the ATLAS detector*, Phys. Lett. **B713** (2012) 387.

[107] ATLAS collaboration, *Limit on $B_s^0 \to \mu + \mu^-$ branching fraction based on 4.9 fb^{-1} of integrated luminosity*, 2013. ATLAS-CONF-2013-076.

[108] CMS collaboration, S. Chatrchyan *et al.*, *Search for B_s^0 and B^0 to dimuon decays in pp collisions at 7 TeV*, Phys. Rev. Lett. **107** (2011) 191802.

[109] CMS collaboration, S. Chatrchyan *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays*, JHEP **04** (2012) 033.

[110] Belle collaboration, M. Chang *et al.*, *Search for $B^0 \to \ell^+\ell^-$ at Belle*, Phys. Rev. D **68** (2003) 111101.

[111] BaBar collaboration, B. Aubert *et al.*, *Search for decays of B^0 mesons into pairs of charged leptons: $B^0 \to e^+e^-$, $B^0 \to \mu^+\mu^-$, $B^0 \to e^\pm\mu^\mp$*, Phys. Rev. Lett. **94** (2005) 221803.

[112] BaBar collaboration, B. Aubert *et al.*, *Search for decays of B^0 mesons into e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$ final states*, Phys. Rev. D **77** (2008) 032007.

[113] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$*, Phys. Lett. **B699** (2011) 330.

[114] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$*, Phys. Lett. **B708** (2012) 55.

[115] LHCb collaboration, R. Aaij *et al.*, *Strong constraints on the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$*, Phys. Rev. Lett. **108** (2012) 231801.

[116] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment*, Phys. Rev. Lett. **111** (2013) 101805.

[117] G. Cowan, *Statistical data analysis*, Oxford University Press, New York, USA, 1998.

[118] R. J. Barlow, *Statistics, A guide to the Use of Statistical Methods in the Physical Sciences*, John Wiley & Sons, Chichester, UK, 1989.

[119] L. Lyons, *Statistics for nuclear and particle physisists*, Cambridge University Press, Cambridge, UK, 1986.

[120] F. James, *Statistical methods in experimental physics*, World Scientific, Singapore, 2nd ed., 2006.

[121] G. J. Feldman and R. D. Cousins, *A Unified approach to the classical statistical analysis of small signals*, Phys. Rev. **D57** (1998) 3873.

[122] J. Neyman, *Outline of a theory of statistical estimation based on the classical theory of probability*, Philosophical Transactions of the Royal Society of London **236** (1937), no. 767 333.

[123] M. Kendall and A. Stuart, *The advanced theory of statistics, Volume 2: Inference and Relationship*, Charles Griffin & Company Ltd, London, UK, 3rd ed., 1973.

[124] G. J. Feldman, *Multiple measurements and parameters in the unified approach*, 2000. Talk at the workshop on confidence limits, https://conferences.fnal.gov/cl2k/copies/feldman2.pdf (October 2015).

[125] T. M. Karbach, *Feldman-Cousins Confidence Levels - Toy MC Method*, `arXiv:1109.0714`.

[126] C. J. Clopper and E. S. Pearson, *The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial*, Biometrika **26** (1934), no. 4 404.

[127] K. Cranmer, *Practical Statistics for the LHC*, in *Proceedings, 2011 European School of High-Energy Physics (ESHEP 2011)*, pp. 267–308.

[128] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007.

[129] I. Antcheva *et al.*, *ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization*, Computer Physics Communications **182** (2011), no. 6 1384 .

[130] F. James and M. Roos, *Minuit - a system for function minimization and analysis of the parameter errors and correlations*, Computer Physics Communications **10** (1975) 343 .

[131] F. James, *MINUIT - Function Minimization and Error Analysis - Reference Manual*, CERN Program Library Long Writeup D506, 1998.

[132] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle γ from a combination of $B^{\pm} \to Dh^{\pm}$ analyses*, Phys. Lett. **B726** (2013) 151.

[133] R. L. Berger and D. D. Boos, *P values maximized over a confidence set for the nuisance parameter*, Journal of the American Statistical Association **89** (1994), no. 427 pp. 1012.

[134] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Annals Math. Statist. **9** (1938) 60.

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

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

[137] CMS collaboration, S. Chatrchyan *et al.*, *The performance of the CMS muon detector in proton-proton collisions at sqrts = 7 TeV at the LHC*, JINST **8** (2013) P11002.

[138] CMS collaboration, V. Khachatryan *et al.*, *CMS Tracking Performance Results from early LHC Operation*, Eur. Phys. J. **C70** (2010) 1165.

[139] LHCb Speakers' Bureau, *$b\bar{b}$ production angle plots* , http://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html (September 2015).

[140] R. Alemany-Fernandez, F. Follin, and R. Jacobsson, *The LHCB Online Luminosity Control and Monitoring*, , CERN-ACC-2013-0028.

[141] LHCb collaboration, R. Jacobsson, *Performance of the LHCb detector during the LHC proton runs 2010–2012*, in *Proceedings, 2012 IEEE NSS/MIC*, pp. 1479–1486, 2012.

[142] R. Lindner, *LHCb layout 2*, LHCb-PHO-GENE-2008-002, Feb, 2008.

[143] LHCb collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011, CERN, Geneva, 2001.

[144] LHCb collaboration, R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) 09007.

[145] LHCb collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029, CERN, Geneva, 2002.

[146] LHCb collaboration, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30** (2014) 1530022. 82 p.

[147] P. A. Cherenkov, *Visible luminescence of pure fluids induced by gamma rays*, Dokl. Akad. Nauk Ser. Fiz. **2** (1934) 451.

[148] P. A. Cherenkov, *Visible radiation produced by electrons moving in a medium with velocities exceeding that of light*, Phys. Rev. **52** (1937) 378.

[149] LHCb RICH group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2012) 2431. 25 p.

[150] LHCb collaboration, S. Amato *et al.*, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036, CERN, Geneva, 2000.

[151] LHCb collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010, CERN, Geneva, 2001.

[152] LHCb collaboration, *LHCb muon system: second addendum to the Technical Design Report*, CERN-LHCC-2005-012, CERN, Geneva, 2005.

[153] F. Sauli, *GEM: A new concept for electron amplification in gas detectors*, Nucl. Instrum. Meth. **A386** (1997) 531.

[154] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020.

[155] LHCb collaboration, R. Antunes-Nobrega *et al.*, *LHCb trigger system: Technical Design Report*, CERN-LHCC-2003-031, CERN, Geneva, 2003.

[156] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022.

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

[158] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and Performance Studies*, Tech. Rep. LHCb-2007-015, CERN, Geneva, May, 2007.

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

[160] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577.

[161] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013.

[162] V. V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, Tech. Rep. LHCb-PUB-2011-016, CERN, Geneva, Sep, 2011.

[163] R. Aaij and J. Albrecht, *Muon triggers in the High Level Trigger of LHCb*, Tech. Rep. LHCb-PUB-2011-017, CERN, Geneva, Sep, 2011.

[164] M. Grabalosa, *Flavour Tagging developments within the LHCb experiment*, PhD thesis, Barcelona University, Mar, 2012, CERN-THESIS-2012-075.

[165] LHCb collaboration, *Optimization and calibration of the same-side kaon tagging algorithm using hadronic B_s^0 decays in 2011 data*, , LHCb-CONF-2012-033.

[166] LHCb collaboration, R. Aaij *et al.*, *Opposite-side flavour tagging of B mesons at the LHCb experiment*, Eur. Phys. J. **C72** (2012) 2022.

[167] J. T. Wishahi, *Measurement of CP Violation in $B^0 \to J/\psi K_S^0$ Decays with the LHCb Experiment*, PhD thesis, TU Dortmund, Nov, 2013, CERN-THESIS-2013-297.

[168] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation in $B^0 \to J/\psi K_S^0$ decays*, Phys. Rev. Lett. **115** (2015) 031601.

[169] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, Psychological Review **65(6)** (1958) 386.

[170] LHCb collaboration, R. Aaij *et al.*, *Measurements of the branching fractions of the decays $B_s^0 \to D_s^{\mp} K^{\pm}$ and $B_s^0 \to D_s^- \pi^+$*, JHEP **06** (2012) 115.

[171] T. M. Karbach and M. Schlupp, *Constraints on Yield Parameters in Extended Maximum Likelihood Fits*, `arXiv:1210.7141`.

[172] M. Pivk and F. R. Le Diberder, *sPlot: a statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356.

[173] Y. Xie, *sFit: a method for background subtraction in maximum likelihood fit*, `arXiv:0905.0724`.

[174] K. S. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136** (2001) 198.

[175] Particle Data Group, J. Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86** (2012) 010001.

[176] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431.

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

[178] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the ratio of the Λ_b^0 to $\overline{B}^0$ lifetimes*, Phys. Lett. **B734** (2014) 122.

[179] LHCb collaboration, R. Aaij *et al.*, *A model-independent Dalitz plot analysis of $B^\pm \to DK^\pm$ with $D \to K^0_S h^+ h^-$ ($h = \pi, K$) decays and constraints on the CKM angle γ*, Phys. Lett. **B718** (2013) 43.

[180] LHCb, R. Aaij *et al.*, *Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region*, Phys. Lett. **B694** (2013) 209.

[181] LHCb, R. Aaij *et al.*, *Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13\,\text{TeV}$*, `arXiv:1509.0077`.

[182] T. M. Karbach, GammaCombo, HepForge Project, http://gammacombo.hepforge.org/web/HTML/index.html (October 2015).

[183] LHCb collaboration, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, Nature Phys. **11** (2015) .

[184] LHCb collaboration, R. Aaij *et al.*, *Improved constraints on γ: CKM2014 update*, 2014. LHCB-CONF-2014-004.

[185] LHCb collaboration, R. Aaij *et al.*, *Observation of CP violation in $B^+ \to DK^+$ decays*, Phys. Lett. **B712** (2012) 203, [Erratum: Phys. Lett.B713, 351 (2012)].

[186] LHCb collaboration, R. Aaij *et al.*, *Observation of the suppressed ADS modes $B^\pm \to [\pi^\pm K^\mp \pi^+ \pi^-]_D K^\pm$ and $B^\pm \to [\pi^\pm K^\mp \pi^+ \pi^-]_D \pi^\pm$*, Phys. Lett. **B723** (2013) 44.

[187] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation and constraints on the CKM angle γ in $B^\pm \to DK^\pm$ with $D \to K^0_S \pi^+ \pi^-$ decays*, Nucl. Phys. **B888** (2014) 169.

[188] LHCb collaboration, R. Aaij *et al.*, *A study of CP violation in $B^\pm \to DK^\pm$ and $B^\pm \to D\pi^\pm$ decays with $D \to K^0_S K^\pm \pi^\mp$ final states*, Phys. Lett. **B733** (2014) 36.

[189] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation parameters in $B^0 \to DK^{*0}$ decays*, Phys. Rev. **D90** (2014) 112002.

[190] LHCb collaboration, *Improved constraints on γ from $B^\pm \to DK^\pm$ decays including first results on 2012 data*, May, 2013. LHCb-CONF-2013-006.

[191] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of early 2012*, `arXiv:1207.1158`, updated results and plots available at: `http://www.slac.stanford.edu/xorg/hfag/osc/PDG_2013/`.

[192] J. Libby *et al.*, *New determination of the $D^0 \to K^-\pi^+\pi^0$ and $D^0 \to K^-\pi^+\pi^+\pi^-$ coherence factors and average strong-phase differences*, Phys. Lett. **B731** (2014) 197.

[193] CLEO collaboration, G. Bonvicini *et al.*, *Updated measurements of absolute D^+ and D^0 hadronic branching fractions and $\sigma(e^+e^- \to D\overline{D})$ at $E_{\rm cm} = 3774\,\text{MeV}$*, Phys. Rev. **D89** (2014) 072002, [Erratum: Phys. Rev.D91 (2015) 019903].

[194] D. Atwood and A. Soni, *Role of charm factory in extracting CKM phase information via $B \to DK$*, Phys. Rev. **D68** (2003) 033003.

[195] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007, Geneva, Apr, 2012.

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

[197] K. De Bruyn *et al.*, *Probing new physics via the $B^0_s \to \mu^+\mu^-$ effective lifetime*, Phys. Rev. Lett. **109** (2012) 041801.

[198] LHCb collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP **04** (2013) 1.

[199] LHCb collaboration, *Updated average f_s/f_d b-hadron production fraction ratio for 7 TeV pp collisions*, Jul, 2013. LHCb-CONF-2013-011.

[200] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, May, 2014. LHCb-PUB-2014-039.

[201] C. Adrover *et al.*, *Search for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with 3 fb^{-1} at LHCb*, Mar, 2013. LHCb-ANA-2013-032 (not public).

[202] BaBar collaboration, P. del Amo Sanchez *et al.*, *Study of $B \to \pi\ell\nu$ and $B \to \rho\ell\nu$ Decays and Determination of $|V_{ub}|$*, Phys. Rev. **D83** (2011) 032007.

[203] N. Isgur, D. Scora, B. Grinstein, and M. B. Wise, *Semileptonic B and D Decays in the Quark Model*, Phys. Rev. **D39** (1989) 799.

[204] D. Scora and N. Isgur, *Semileptonic meson decays in the quark model: An update*, Phys. Rev. **D52** (1995) 2783.

[205] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays $B/B_s \to (\pi,K)(\ell^+\ell^-,\ell\nu,\nu\bar{\nu})$ in the perturbative QCD approach beyond the leading-order*, Phys. Rev. **D86** (2012) 114025.

[206] A. Khodjamirian, C. Klein, T. Mannel, and Y.-M. Wang, *Form Factors and Strong Couplings of Heavy Baryons from QCD Light-Cone Sum Rules*, JHEP **1109** (2011) 106.

[207] Particle Data Group, K. Nakamura *et al.*, *Review of particle physics*, J. Phys. **G37** (2010) 075021.

[208] LHCb collaboration, R. Aaij *et al.*, *Measurement of b hadron production fractions in* 7 TeV *pp collisions*, Phys. Rev. **D85** (2012) 032008.

[209] LHCb collaboration, R. Aaij *et al.*, *First observation of the decay $B^+ \to \pi^+\mu^+\mu^-$*, JHEP **1212** (2012) 125.

[210] CDF collaboration, F. Abe *et al.*, *Observation of the B_c meson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. **81** (1998) 2432.

[211] CDF collaboration, F. Abe *et al.*, *Observation of B_c mesons in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. **D58** (1998) 112004.

[212] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97.

[213] D. Melikhov and N. Nikitin, *Rare radiative leptonic decays $B_{d,s} \to l^+l^-\gamma$*, Phys. Rev. **D70** (2004) 114028.

[214] Y. Dincer and L. M. Sehgal, *Charge asymmetry and photon energy spectrum in the decay $B_s \to l^+l^-\gamma$*, Phys. Lett. **B521** (2001) 7.

[215] C. Geng, C. Lih, and W.-M. Zhang, *Study of $B_{s,d} \to l^+l^-\gamma$ decays*, Phys. Rev. **D62** (2000) 074017.

[216] A. L. Read, *Presentation of search results: The CL_s technique*, J. Phys. **G28** (2002) 2693.

[217] CMS collaboration, S. Chatrchyan *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ with the CMS Experiment*, Phys. Rev. Lett. **111** (2013) 101804.

[218] CMS collaboration, C. Collaboration, *B Physics analyses for the Phase-II Upgrade Technical Proposal*, 2015. CMS-PAS-FTR-14-015.

[219] CMS collaboration, *Technical proposal for the upgrade of the CMS detector through 2020*, CERN-LHCC-2011-006, Geneva, Jun, 2011.

[220] J. Ellis, *Summary and Outlook: 2015 Lepton-Photon Symposium*, 2015. `arXiv:1509.0733`.

A Appendix

A.1 Central limit theorem

The central limit theorem (in formulation following [118]) is one of the most important theorems in applied statistics and physics. It states that the distribution of the sum $S = \sum_i x_i$ of random variables x_i with $i = 1, 2, \ldots, N$, each from a distribution with the mean μ_i and a finite variance of $V_i = \sigma_i^2 < \infty$ has the following attributes:

a) The expectation value is

$$\langle S \rangle = \sum_i \mu_i \, .$$

b) The variance is

$$V = \sum_i V_i = \sum_i \sigma_i^2 \, .$$

c) For $N \to \infty$ the S follows a gaussian distribution

$$g(S) = \frac{1}{\sqrt{2\pi V}} e^{-\frac{1}{2}\frac{(S-\langle S \rangle)^2}{V}} \, .$$

For finite N the approximation to a gaussian function works better in the core region close the expected mean $\langle S \rangle$ compared to the tail region of the resulting distribution.

A.2 Coverage probabilities for a χ^2 distribution

In order to construct confidence intervals from a χ^2 distributed variable it is necessary to know the $\Delta\chi^2$ values at which the certain confidence level α is reached. These values differ depending on the number of degrees of freedom, r. For a χ^2 distribution with one degree of freedom the square root of the $\Delta\chi^2$ value provides the two-sided confidence boundary corresponding to a standard gaussian $n\sigma$ coverage. Table A.1 lists the gaussian $n\sigma$ coverage probabilities or confidence level α and the corresponding $\Delta\chi^2$ values in one and two degrees of freedom. In practice, the α values

Table A.1: Values of $\Delta\chi^2$ for r degrees of freedom corresponding to a confidence level α equivalent to an interval given by n gaussian standard deviations.

$\mathbf{n\sigma}$	$\boldsymbol{\alpha}$ [%]	$\mathbf{r=1}$	$\mathbf{r=2}$
1σ	68.27	1.0	2.3
2σ	95.45	4.0	6.2
3σ	99.73	9.0	11.8
4σ	$1-6.3\times10^{-5}$	16.0	19.3
5σ	$1-5.7\times10^{-7}$	25.0	28.7
6σ	$1-2.0\times10^{-9}$	36.0	40.1

corresponding to a given $\Delta\chi^2$ value can be calculated with $1-$`TMath::Prob`($\Delta\chi^2$,ndf) or `ROOT::Math::chisquared_cdf(`$\Delta\chi^2$`,`ndf`)`. The translation from the gaussian coverage to the $\Delta\chi^2$ value is given by the `ROOT::Math::chisquared_quantile(`α`,`ndf`)`. The gaussian standard deviations are given by the square root of the $\Delta\chi^2$ values with one degree of freedom.

A.3 Correlation tables for the $B_s^0 \to D_s^{\mp} K^{\pm}$ CP observables

The correlation matrices of the cFit from the $B_s^0 \to D_s^{\mp} K^{\pm}$ analysis (cf. Chapter 5) are given in Table A.2 and Table A.3 for the statistical and systematic uncertainties, respectively. The tables are taken from Ref. [66] and serve as an input to the measurement of the CKM angle γ from $B_s^0 \to D_s^{\mp} K^{\pm}$ decay (cf. Chapter 5.6) and to the γ combination in Chapter 6.

Table A.2: Statistical correlation matrix of the $B_s^0 \to D_s^{\mp} K^{\pm}$ cFit *CP* observables. Other fit parameters have negligible correlations with the *CP* observables and are omitted for brevity [66].

Parameter		C_f	$\mathcal{A}_f^{\Delta\Gamma}$	$\mathcal{A}_{\bar{f}}^{\Delta\Gamma}$	S_f	$S_{\bar{f}}$
cFit	C_f	1.000	0.084	0.103	−0.008	−0.045
	$\mathcal{A}_f^{\Delta\Gamma}$		1.000	0.544	−0.117	−0.022
	$\mathcal{A}_{\bar{f}}^{\Delta\Gamma}$			1.000	−0.067	−0.032
	S_f				1.000	0.002
	$S_{\bar{f}}$					1.000

Table A.3: Systematic uncertainty correlations of the *CP* observables for the cFit [66].

Parameter		C_f	$\mathcal{A}_f^{\Delta\Gamma}$	$\mathcal{A}_{\bar{f}}^{\Delta\Gamma}$	S_f	$S_{\bar{f}}$
cFit	C_f	1.00	0.22	0.22	−0.04	−0.03
	$\mathcal{A}_f^{\Delta\Gamma}$		1.00	0.96	−0.17	−0.14
	$\mathcal{A}_{\bar{f}}^{\Delta\Gamma}$			1.00	−0.17	−0.14
	S_f				1.00	0.09
	$S_{\bar{f}}$					1.00

Acknowledgements

First of all, I have to thank Prof. Dr. Bernhard Spaan, who was willing to support and supervise me since I decided to move on from studies in theoretical physics. I am grateful that I got the possibility to work in a truly international environment with highly motivated collaborators on difficult yet interesting problems. During the time of my PhD thesis, I met a lot of interesting people at CERN, conferences, or outreach activities, who are responsible for great memories in one or the other way. This resulted in a once-in-a-lifetime experience—thank you all, colleagues, friends, and strangers!

What makes a PhD thesis a lot more fun is an excellent work environment. That's a given at the Experimentelle Physik 5 group of the TU Dortmund university, so thank you everybody. As many new students joined the group over the last few years, I can only mention some long term fellows and friends with whom I shared offices or valuable memories. Thank you (in no particular order) Christophe, Ramon, Michael, Stefan, Robert, Moritz, Laura, Alex, Nina, Florian, Julian, Tobi, Mirco, Uli, and Frank. Special thanks to whomsoever for letting me survive some of the car trips with members of the prior list (you all know!).

Finally, there are two guys I explicitly need to thank, as their extreme knowledge and expertise is undeniably present in my whole thesis. However, I am especially grateful for their support that goes (way) beyond the supervision of my work.
Thank you, Johannes! You think that it is your job to give advice and help your students through tough times. This is very true, but the occasions when we shared barbecues, beer, dinners, cakes, paddling pools, holiday, renovation, or relocation are as—if not even more—valuable to me!
Thank you, Moritz! You've been one of the reasons for me to join experimental particle physics and LHCb in particular. Besides close collaboration in research throughout the years, some of the best memories of my PhD time are because of you. This includes table soccer at the *Missin' Link*, *Kraftstoff*, or at Neckarzimmern, mounting parts of the LHCb detector, hiking on snowy mountains or sharing beer at lake Geneva. You'll never be forgotten!