TU DORTMUND

DISSERTATION

Measurement of the Branching Ratio of $B_s^0 \to K_S^0 K_S^0$ Decays with the LHCb Experiment

Autor:
Timon SCHMELZER,
geboren am 07.06.1990
in Recklinghausen

Betreuer:
Prof. Dr. Bernhard SPAAN

Dissertation zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften

in der

Fakultät Physik,
Experimentelle Physik V

Dortmund, Juli 2019

1. Gutachter: Prof. Dr. Bernhard Spaan
2. Gutachter: Prof. Dr. Kevin Kröninger
Vorsitzender der Prüfungskommission: Prof. Dr. Markus Betz
Vertreter der wissenschaftlichen Mitarbeiter: Dr. Ulf Berges

Datum des Einreichens der Arbeit: 04.07.2019
Datum der Disputation: 11.09.2019

Abstract

The presented thesis deals with the branching ratio measurement of the decay channel $B_s^0 \to K_S^0 K_S^0$ reconstructed as two charged pion pairs. The data correspond to an integrated luminosity of 5 fb^{-1} collected by the LHCb experiment in proton-proton collisions at the centre-of-mass energies of $\sqrt{s} = 7$ TeV (1 fb^{-1}), 8 TeV (2 fb^{-1}) and 13 TeV (2 fb^{-1}). As normalisation channel the decay $B^0 \to \phi K_S^0$ is used in which the ϕ meson is reconstructed as two charged kaons. The branching ratio is determined to be

$$\begin{aligned}\mathcal{BR}(B_s^0 \to K_S^0 K_S^0) = [&9.5 \pm 1.9(\text{stat}) \pm 0.9(\text{syst}) \\ &\pm 0.9(\mathcal{BR}(B^0 \to \phi K_S^0)) \pm 0.5(f_d/f_s)] \times 10^{-6},\end{aligned}$$

whereby the $B_s^0 \to K_S^0 K_S^0$ component has a significance of $7.24\,\sigma$. The result is consistent with measurements of the Belle collaboration within less than 1 standard deviation and compatible with Standard Model predictions. Furthermore, the analysis represents the first study of b hadrons decaying exclusively into long lived V^0 particles at the LHCb experiment.

Kurzfassung

Dieses Dokument umfasst die Messung des Verzweigungsverhältnisses des Zerfallskanals $B_s^0 \to K_S^0 K_S^0$, rekonstruiert als zwei geladene Pionenpaare. Die zugrundeliegende Datenmenge entspricht einer integrierten Luminosität von 5 fb^{-1}, aufgenommen mit dem LHCb Experiment in Proton-Proton-Kollosionen, bei den Schwerpunktsenergien von $\sqrt{s} = 7$ TeV (1 fb^{-1}), 8 TeV (2 fb^{-1}) und 13 TeV (2 fb^{-1}). Der Zerfallskanal $B^0 \to \phi K_S^0$ wird als Normierungskanal verwendet, wobei das ϕ-Meson als zwei geladene Kaonen rekonstruiert wird. Das Verzweigungsverhältnis wird gemessen als

$$\begin{aligned}\mathcal{BR}(B_s^0 \to K_S^0 K_S^0) = [&9{,}5 \pm 1{,}9(\text{stat}) \pm 0{,}9(\text{syst}) \\ &\pm 0{,}9(\mathcal{BR}(B^0 \to \phi K_S^0)) \pm 0{,}5(f_d/f_s)] \times 10^{-6},\end{aligned}$$

was einer Signifikanz der $B_s^0 \to K_S^0 K_S^0$ Signalkomponente von $7{,}24\,\sigma$ entspricht. Dieses Ergebnis ist konsistent mit Messungen der Belle Kollaboration innerhalb von weniger als einer Standardabweichung und kompatibel mit Standardmodellvorhersagen. Damit stellt diese Analyse die erste Untersuchung von b-Hadronen-Zerfällen in ausschließlich langlebige V^0-Teilchen am LHCb-Experiment dar.

Contents

1 Introduction **1**

2 The Standard Model of Particle Physics **5**
2.1 Particles, Forces and Couplings 6
2.2 CKM mechanism 8
2.3 Mixing and decay of flavoured neutral mesons 12
2.3.1 Types of *CP*-violation 15
2.3.2 *CP*V in the kaon system 17
2.4 Possible Physics beyond the SM 18

3 The LHC and the LHCb Experiment **21**
3.1 Large Hadron Collider 21
3.1.1 Experiments 22
3.2 LHCb Detector 23
3.2.1 Tracking Systems 24
3.2.2 Particle Identification Systems 26
3.2.3 Trigger 28
3.2.4 Software Stack 29
3.3 Running Conditions 32
3.3.1 Upgrade and Prospects 33

4 Statistical Tools and Methods **35**
4.1 Multivariate Analysis 35
4.1.1 Boosted Decision Trees 36
4.2 Maximum Likelihood Estimation 38
4.3 Determination of ${}_s\mathcal{W}$eights 39
4.4 Kolmogorov-Smirnov test 40

5 The Decay Channel $B_s^0 \to K_S^0 K_S^0$ **41**
5.1 Theoretical predictions of the branching ratio 41
5.2 Decay chain properties 44
5.3 Physics observables in $B_s^0 \to K^0 \bar{K}^0$ decays 44
5.4 Previous Measurements 46

6 Branching ratio measurement of $B_s^0 \to K_S^0 K_S^0$ decays **49**
6.1 Datasets and decay topologies . . . 49
6.2 Determination of the branching ratio . . . 51
6.2.1 External inputs . . . 52
6.3 Signal extraction . . . 52
6.3.1 Event reconstruction . . . 53
6.3.2 Preselection . . . 56
6.3.3 Trigger Requirements . . . 58
6.3.4 Background studies . . . 60
6.3.5 Subtraction of the combinatorial background . . . 66
6.4 Fit model . . . 77
6.4.1 Signal component . . . 77
6.4.2 Background component . . . 78
6.4.3 Fit model scaling . . . 79
6.4.4 Non-ϕ K^+K^- background in $m(K^+K^-)$ . . . 79
6.5 Fit results . . . 80
6.5.1 Determination of the mass observable . . . 83
6.5.2 PID corrections . . . 83
6.6 Efficiency determinations . . . 84
6.6.1 Geometrical acceptance . . . 85
6.6.2 Reconstruction and stripping . . . 85
6.6.3 Track categories . . . 85
6.6.4 Trigger efficiencies . . . 85
6.6.5 Further selection efficiencies . . . 86
6.6.6 Efficiency summary . . . 86
6.6.7 Combinations . . . 87
6.6.8 Background subtraction . . . 89
6.7 Combination of the $B_s^0 \to K_S^0 K_S^0$ signal yields . . . 90
6.8 Systematic uncertainties . . . 92
6.8.1 Fit Model Validation . . . 92
6.8.2 Fit Model Choice . . . 93
6.8.3 Multivariate Classification . . . 95
6.8.4 Particle Identification . . . 95
6.8.5 Trigger Systematics . . . 96
6.8.6 Further systematic uncertainties . . . 98
6.8.7 Systematic uncertainty summary . . . 99
6.9 Significance determination . . . 100

7 Conclusion and Outlook **103**

8 Appendix **107**

Bibliography **113**

Chapter 1

Introduction

Even long before the era of modern particle physics the great natural philosopher Democritus (460–370 BC) began to ponder about the smallest components of matter. He postulated that nature is composed of indivisible units, he called them *atoms*, where each individual atom is solid and massive, and there is nothing but empty space between them. Only about 2300 years later Rutherford proved with his famous scattering experiments [1] that atoms are not solid, but consist of a nucleus and a shell. In the following one hundred years new particles were discovered and theories were developed to describe the observed phenomena. The best known and most prominent theory is the Standard Model of particle physics (SM) [2–4], which covers the description of all elementary particles and three of four fundamental interactions. Electromagnetic effects, i.e. the interaction between electrically charged particles by exchanging photons, are described in the theory of quantum electrodynamics (QED). It can be unified with the quantum flavourdynamics (QFD), of which a vivid example is the decay of free neutrons, to the electroweak theory (EWT). Finally, the SM is completed with the quantum chromodynamics (QCD), which through its strong interaction involves the interactions between quarks and gluons and thus enforcing composite quark states such as protons.

Up to now, the history of the SM is very successful. Already in the 70s within the SM the existence of the heavy top quark [5] and the neutrino of the third generation, the tau neutrino [6], were predicted, which were discovered in the years 1995 and 2000, respectively. In 2012, the Higgs boson, an excitation of the Higgs field responsible for the mass generation in the SM, has been discovered as the last particle predicted [7, 8]. Even though the SM has been measured to a great precision and no measurement shows significant deviations from its predictions, it is known that the SM is not an all-embracing theory of fundamental interactions. For example, the SM does not incorporate gravity, which is formulated by general relativity. In fact, only about 5 % of all existing matter and energy is covered by the SM. Observations of the rotation speed of galaxies indicate that the visible mass within the galaxies is not sufficient to explain the high velocity of stars at greater distances from the centre. This invisible mass, predicted to be about six times as common in the universe as visible matter, is called dark matter [9]. Furthermore, the current accelerated expansion of space against the force of gravity is explained by introducing a new type of energy,

the so-called dark energy [10]. However, the SM is inconsistent with these observations from cosmology. Another unsolved phenomenon are the matter-antimatter asymmetries in the current universe. Although the Big Bang should have created matter and antimatter in equal amounts, nowadays, everything in nature consists almost solely of matter. Sakharov [11] postulated three conditions to produce matter and antimatter at different rates, also referred to as baryogenesis. First, the baryon number conservation must be violated. This can be verified by examining e.g. free proton decays, which have not yet been observed [12]. Secondly, interactions out of the thermal equilibrium must exist, for which measurements have already provided evidence [13]. Finally, the discrete symmetries C and CP have to be violated. This has been found in multiple decay channels, most recently in charm decays [14]. Even though all of the aforementioned criteria are represented in the SM, the measured quantities are not sufficient by orders of magnitude to explain the baryogenesis.

Large particle accelerators like the Large Hadron Collider (LHC) [15] at CERN provide an excellent test bed for precise measurements of the SM as well as the search for physics beyond the SM, also called New Physics, NP. Within the LHC, bunches of protons are accelerated to centre-of-mass energies of nominally 14 TeV and brought to collision at four interaction points, where particle detectors are located. Of particular interest for this thesis is the LHCb experiment [16], which specialises in the investigation of b and c hadron decays. While the direct search for new particles is limited to the energy scale of the collider, indirect searches allow tests for NP particles up the 1000 TeV scale [17]. These particles could alter SM predictions in terms of branching ratios or angular distributions. Measurements of such discrepancies are challenging, since NP effects are only expected in higher order processes and their influence is expected to be small. The required precision can only be achieved with the usage of robust statistical techniques combined with a detailed understanding of detector effects.

One promising channel of interest is the decay $B_s^0 \to K_S^0 K_S^0$ [18], which only occurs in loop-level transitions in the SM and therefore offers possibilities for NP contributions. On quark level, this decay describes a $b \to s\bar{d}d$ transition and has a branching ratio of $\mathcal{O}(10^{-5})$ [19–22]. This work discusses the branching ratio measurement of $B_s^0 \to K_S^0 K_S^0$ decays using data collected from 2011 to 2016 by the LHCb detector. Results obtained during this analysis will be published in Physical Review D,

Aaij, R. *et al.* Search for $B_s^0 \to K_S^0 K_S^0$, *Phys. Rev. D*, in preparation.

This is the first measurement and analysis of decays in exclusively long-lived particles in a fully hadronic environment and can be regarded as a benchmark for the LHCb collaboration. To cancel out systematic uncertainties, the branching ratio is determined in relation to the normalisation channel $B^0 \to \phi K_S^0$. Experimentally, the K_S^0 mesons are reconstructed as two oppositely charged pion pairs, while the

ϕ meson is reconstructed as K^+ and K^- mesons. The measurement has been performed in close cooperation with Moritz Demmer, who mainly focused on the Run I part of the analysis. His results are published in the PhD thesis

> M. Demmer, *Analysis of rare hadronic decay modes and simulation studies for the scintillating fibre tracker at the LHCb experiment*, PhD thesis, TU Dortmund, 2018.

The present thesis is structured as follows: Chapter 2 provides an overview of the SM and declare its basic theoretical concepts. In Ch. 3, the LHC and the LHCb experiment is introduced, where detector components which are important for this analysis are discussed in detail. Essential statistical and machine learning tools are explained in Ch. 5. The largest part of this work encompasses the actual measurement, i.e. signal extraction and efficiency determination, which are described in Ch. 6. Finally, this thesis concludes with a summary and outlook given in Ch. 7.

Chapter 2

The Standard Model of Particle Physics

The Standard Model of particle physics (SM) is a quantum field theory describing all known particles and the fundamental forces between them, except gravity. The gauge group of the SM is $SU(3)_c \times SU(2)_L \times U(1)_Y$, where the index c describes the color charge, L the weak isospin and Y the hypercharge. Spontaneous symmetry breaking enters the SM via $SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM}$ causing masses for particles and separating the weak and electromagnetic forces. In this theory, a Lagrangian $\mathcal{L}$ controls the dynamics of the system. Assuming massless neutrinos, 18 arbitrary parameters must be introduced to construct the most general Lagrangian[1]. The values of these parameters cannot be predicted from theory and have to be taken as external input from experiments. The 18 free parameters are:

- 6 quark masses,
- 3 charged lepton masses,
- 3 coupling constants,
- 3 CKM matrix angles,
- 1 CKM matrix phase,
- 1 Higgs vacuum expectation value v, which is connected to the Fermi coupling constant via $\sqrt{2}G_{\mathrm{F}} = 1/v^2$ and
- 1 Higgs boson mass.

Since the discovery of the Higgs boson in 2012 [7, 8], values of all these parameters have been determined. Nevertheless, the parameters are constantly updated in new measurements, performed by various experiments and collaborations, and their uncertainties are reduced.

Further characteristics of the SM are explained and discussed throughout this chapter. At first, the fundamental particles, forces and couplings are introduced

[1] Dependent of the exact formalism, some sources predict a nineteenth parameter the QCD vacuum angle θ_{QCD}, which is consistent with zero.

(Sec. 2.1). The origin of meson mixing and *CP* violation is delineated in Sec. 2.2. Practical examples of *CPV* on the basis of B and K meson decays are dealt with in Sec. 2.3. Finally, limitations of the SM are enumerated and possible physics behind the SM is covered in Sec. 2.4. Unless otherwise stated, the informations provided in this chapter are based on Refs. [23–25].

2.1 Particles, Forces and Couplings

Elementary particles in the SM are fundamentally distinguished between fermions and bosons. The former ones carry half of an integer spin while the latter ones have full integer spin. There are twelve fundamental fermions in the SM, six quarks and six leptons. Additionally, twelve anti-fermions are present, which carry exactly the opposite electric charge as their fermion counter parts. The quarks can be subdivided into three families, each containing an up-type and a down-type quark. The up-type quarks u, c and t carry a positive charge of $+2/3e$, while the down-type quarks d, s and b are negatively charged ($-1/3e$). Generally, only combination of quarks with an integer electric charge, hadrons, can be observed. This fact is known as the confinement of quarks [26]. Bound states of one quark and one anti-quark are called meson, a combination of three quarks baryon. The best known representatives of baryons are protons (uud) and neutrons (udd). Recent measurements from LHCb indicate the existence of another kind of hadrons, tetra- and pentaquarks, represented by $q\bar{q}q\bar{q}$ and $q\bar{q}qqq$, respectively [27–29]. The structure of these hadrons can be interpreted as molecules of mesons and baryons.

Like quarks, leptons are also grouped into three families, which consists of electrons e^-, muons μ^- and taus τ^-. Each of them carries a negative electric charge and has a neutral counterpart, the neutrinos ν_e, ν_μ and ν_τ.

Besides fermions, twelve gauge bosons exist in the SM. The massless photon γ is the force carrier of electromagnetism. Its coupling strength depends on the charge of particles, resulting in a non-existent coupling to neutrinos. Three massive particles, $W^\pm$ and Z^0, mediate the weak interaction and couple to the weak isospin. Both aforementioned couplings are connected to each other via the equation $Q = T_3 + 1/2 Y_W$, where Q is the electric charge, T_3 the weak isospin and Y_W the weak hypercharge. Furthermore, eight gluons, g, mediate the strong interaction. In contrast to mediators of the electromagnetic or weak interaction, gluons themselves carry colour charge, which is the mediator of the strong force. This implies that gluons could not only interact with quarks, but also with each other [30]. Finally, the SM is completed with the massive Higgs boson with spin 0. All fundamental particles are depicted in Fig. 2.1.

The generalisation of Maxwell's equations, which form the foundation of electromagnetism in classical physics, is called quantum electrodynamics (QED). The QED is an abelian gauge theory with a symmetry group of $U(1)$. One can deduce the full Lagrangian $\mathcal{L}_{\text{QED}}$ [31] by starting at the Lagrangian $\mathcal{L}_0$ of a free fermion field ψ,

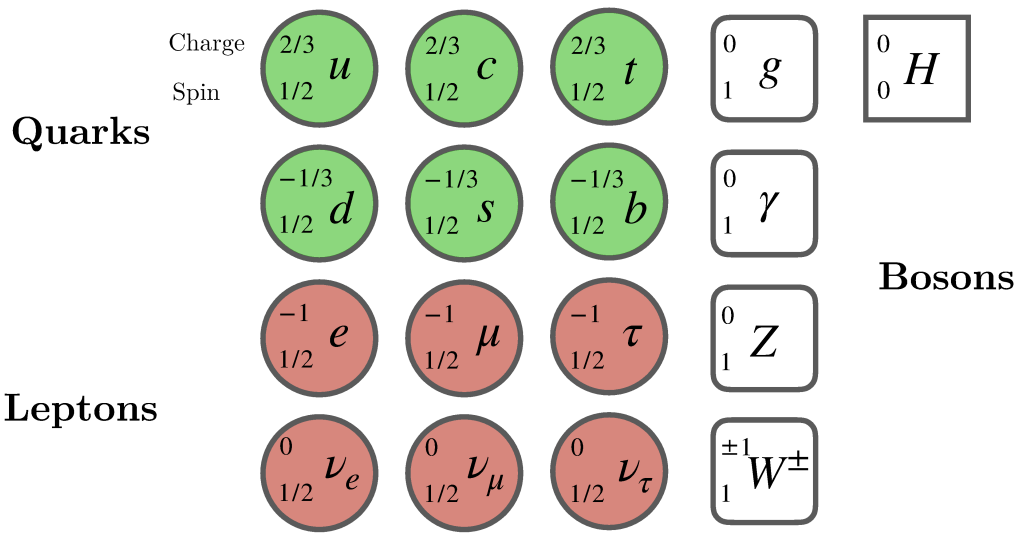


FIGURE 2.1: Summary of all fundamental particles predicted by the SM. The first three particles in the first row show the up-type quarks, the next row the down-type quarks. Charged and uncharged leptons are illustrated below. The fourth column summarize the gauge bosons, which mediate forces. Finally, the Higgs boson is outlined on the far right.

$$\mathcal{L}_0 = \overline{\psi}(i\gamma^\nu \partial_\mu - m)\psi. \tag{2.1}$$

$\mathcal{L}_0$ has a *global* $U(1)$ symmetry, which means that it is invariant under the transformation $\psi(x) \rightarrow \psi^{'}(x) = e^{i\alpha}\psi(x)$. To extend Eq. (2.1) to *local* transformations $\alpha \rightarrow \alpha(x)$, the derivative ∂_μ has to be replaced by the gauge covariant derivative,

$$D_\mu = \partial_\mu + ieA_\mu. \tag{2.2}$$

Here, the vector field A_μ could be interpreted as a photon. Consequently, the Lagrangian is now invariant under the local gauge transformations

$$\psi(x) \rightarrow \psi^{'}(x) = e^{i\alpha(x)}\psi(x) \text{ and} \tag{2.3}$$

$$A_\mu(x) \rightarrow A^{'}_\mu(x) = A_\mu(x) - \frac{1}{e}\partial_\mu \alpha(x). \tag{2.4}$$

Additionally, a kinetic term for A_μ is introduced as $F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ represents the field strength tensor. Finally, a source term causing the photon field J^ν is needed, completing the QED Lagrangian to

$$\mathcal{L}_{\text{QED}} = \overline{\psi}(i\gamma^\nu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu. \tag{2.5}$$

Similar to QED, the theory of the weak interaction is explained in quantum flavourdynamics (QFD). However, QED and QFD can be unified in the SM to the electroweak theory (EWT). One important characteristic of EWT is the CKM mechanism, which will be delineated in Sec. 2.2.

The interaction of quarks and gluons is described by quantum chromodynamics (QCD). Two main properties are embedded in the QCD: Firstly, the colour confinement, what has already been mentioned before. In contrast to this, asymptotic freedom predicts the interaction between particles to become asymptotic weaker with increasing energy scales, causing them to behave like free particles. A complete derivation of the QCD Lagrangian is given in Ref. [32]. Similarly to Eq. (2.5), it can be summarised as

$$\mathcal{L}_{\mathrm{QCD}} = \overline{\psi}_i (i(\gamma^{\mu D_\mu})_{ij} - m\delta_{ij})\psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a, \tag{2.6}$$

where $\psi_i(x)$ is the quark field, δ_{ij} the Kronecker delta, and $G^a_{\mu\nu}$ the gluon field strength tensor, which itself is the abbreviated form of

$$G^a_{\mu\nu} = \partial_\mu \mathcal{A}^a_\nu - \partial_\nu \mathcal{A}^a_\mu + g f^{abc} \mathcal{A}^b_\mu \mathcal{A}^c_\nu. \tag{2.7}$$

Here, $\mathcal{A}^a_\mu$ are the gluon fields, g the quark gluon coupling constant and f^{abc} the structure constant of $SU(3)$. While Greek indices indicate four-vectors, Latin letters represent colours. Due to effects only appearing in QCD, like possible self-interactions between gluons, many QCD calculations are more complex then QED calculations, which can lead to larger uncertainties in its predictions. The complete Lagrangian of the SM can be found in Ref. [33].

2.2 CKM mechanism

In the SM, quark masses are generated via a Yukawa coupling between quark fields and the Higgs field, or more precisely, the vacuum expectation value of the Higgs boson v. Furthermore, these couplings allow transitions between different quark families. This process is called CKM mechanism. In contrast to neutral currents (NC), which couple to all quark types, the charged currents, $W^\pm$, only interact with left-handed quarks. Here, the chirality indicates the property whether the spin of a particle points in the same direction as the impulse (right-handed) or in the opposite direction (left-handed). Additionally, charged currents (CC) are the sole origin of *CP*V in the SM. Formulating a Lagrangian for these currents,

$$\mathcal{L}_{\mathrm{CC}} = -\frac{g}{\sqrt{2}} \left[(\overline{u}, \overline{c}, \overline{t})_L \gamma^\mu W^+_\mu V_{\mathrm{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L + (\overline{d}, \overline{s}, \overline{b})_L \gamma^\mu W^-_\mu V^\dagger_{\mathrm{CKM}} \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L \right], \tag{2.8}$$

introduces the unitary matrix V_{CKM}. It was first postulated by Cabibbo in two dimensions and later extended to three quark generations by Kobayashi and Maskawa [34, 35] — in their honour, this matrix was named after their initials. It transforms the mass eigenstates of left-handed down-type quarks into electroweak eigenstates,

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.9}$$

Here, the usage of down-type quarks is just a convention. Utilizing this matrix, the probability of a down-type quark to transition into an up-type quark is the square of the absolute value of the corresponding matrix element $|V_{\text{ij}}|$. Due to the structure of this matrix, it is possible to reduce the number of free parameters. In general, a complex matrix with a structure of $N \times N$ has $2N^2$ parameters. The condition of unitarity, meaning that the conjugate transpose of a matrix is also its inverse, yields $\sum_{\text{k}} V_{\text{ij}} V^*_{\text{jk}} = \delta_{\text{ij}}$, reduces this number to N^2. Furthermore, $2N - 1$ additional parameters can be absorbed into each of the six quark fields (one overall common phase is unobservable). Finally, the free parameters of V_{CKM} are three $(N(N-1)/2)$ real and one $((N-1)(N-2)/2)$ complex one. In SM, these parameters are parametrised as three Euler angles θ_{12}, θ_{23}, θ_{13} and one phase δ_{13}, which is responsible for the CP-violation. Hence, an exact representation of the CKM matrix is

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\mathrm{e}^{-\mathrm{i}\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta_{13}} & c_{23}c_{13} \end{pmatrix}, \tag{2.10}$$

where s_{ij} (c_{ij}) are abbreviations for the sine (cosine) of the corresponding Euler angles. A priori, the values of these angles could be arbitrary. However, it is experimentally proven that $1 \gg s_{12} \gg s_{23} \gg s_{13}$. To avoid small angles, another parametrisation has been introduced by Wolfenstein called Wolfenstein parametrisation [36], where all parameters are $\mathcal{O}(1)$. The Wolfenstein parametrisation is most commonly used in particle physics. Furthermore, it is an expansion of the CKM matrix in powers of λ, which corresponds to s_{12}. The transition can be accomplished with the definitions of

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \tag{2.11}$$

$$s_{23} = A\lambda^2 = \lambda \left| \frac{V_{cb}}{V_{us}} \right|, \tag{2.12}$$

$$s_{13}\mathrm{e}^{\mathrm{i}\delta} = V^*_{ub} = A\lambda^3(\rho + \mathrm{i}\eta), \tag{2.13}$$

which results into the new four parameters λ, A, ρ and η, where the last one describes the CP violation phase. Thus, the CKM matrix can be represented as

$$V_{\mathrm{CKM}} \approx \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho-\mathrm{i}\eta) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\rho-\mathrm{i}\eta) & -A\lambda^2 & 1 \end{pmatrix}$$

$$+\begin{pmatrix} -\frac{1}{8}\lambda^4 & 0 & 0 \\ \frac{1}{2}A^2\lambda^5[1-2(\rho+\mathrm{i}\eta)] & -\frac{1}{8}\lambda^4(1+4A^2) & 0 \\ \frac{1}{2}A\lambda^5(\rho+\mathrm{i}\eta) & \frac{1}{2}A\lambda^4[1-2(\rho+\mathrm{i}\eta)] & -\frac{1}{2}A^2\lambda^4 \end{pmatrix} \tag{2.14}$$

$$+\mathcal{O}(\lambda^6).$$

To ensure consistency of the SM, one important goal of particle physics is to measure the CKM matrix elements as well as the Wolfenstein parameters in many different decay channels, overconstraining its values. The particular structure of this matrix can also be exploited in this process. Thus, one possible way to visualise inconsistencies between SM parameters is to construct triangles in a complex plane. Due to unitarity, six of these triangles exist. Four of these triangles are "squashed", which means that one side length is significantly large than the others. Nevertheless, all possible triangles share the same area, which is given by half of the Jarlskog invariant [37],

$$J_{\mathrm{CP}} = \pm\mathrm{Im}(V_{\mathrm{ik}}V_{\mathrm{jl}}V_{\mathrm{il}}^*V_{\mathrm{jk}}^*), \text{ where } (\mathrm{i} \neq \mathrm{j},\ \mathrm{l} \neq \mathrm{k}). \tag{2.15}$$

Non-zero values of J_{CP} are the verification for *CP*V in the SM, current determinations yield $J_{\mathrm{CP}} = (3.04^{+0.21}_{-0.20}) \times 10^{-5}$ [38]. One triangle where all side lengths are in the same order is the most common one and called the unitarity triangle (UT). It is defined as

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{2.16}$$

Fig. 2.2 depicts the current world average of the UT. The angles in this triangle, the so called CKM angles, are defined as

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{2.17}$$

Furthermore, by convention, all side lengths in the UT are divided by $V_{cd}V_{cb}^*$, resulting into

$$R_{\mathrm{t}} = \left|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right|, \quad R_{\mathrm{u}} = \left|\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right|, \quad R_{\mathrm{c}} = \left|\frac{V_{cd}V_{cb}^*}{V_{cd}V_{cb}^*}\right| = 1. \tag{2.18}$$

The two parameters $\overline{\rho}$ and $\overline{\eta}$ define the apex position and are defined as

$$\overline{x} = x(1 - \frac{\lambda^2}{2}) \quad \text{for } x \text{ in } [\rho, \eta]. \tag{2.19}$$

In this thesis, another angle β_s is of special interest, which enters due to B_s^0 mixing diagrams (see Ch. 5). In terms of CKM matrix elements, it is defined as

$$\beta_s = \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right). \tag{2.20}$$

Its corresponding triangle is different from the aforementioned UT,

$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0, \tag{2.21}$$

where one side lengths is $\mathcal{O}(\lambda^4)$ and two sides are $\mathcal{O}(\lambda^2)$.

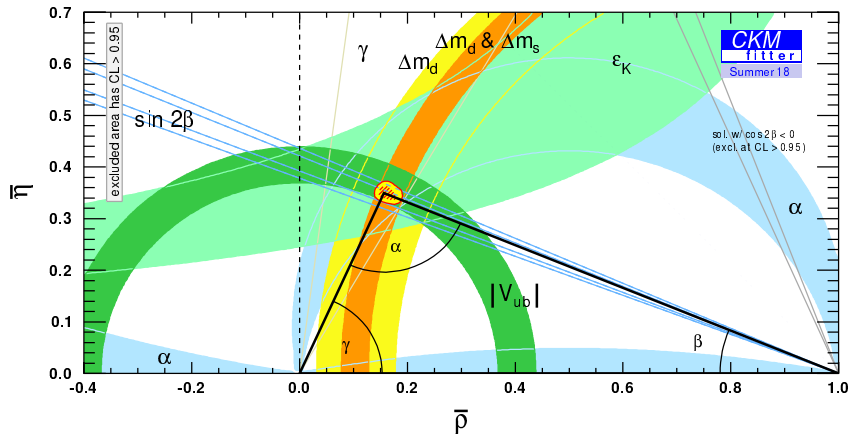


FIGURE 2.2: The current representation of the unitarity triangle in the complex $(\overline{\rho}\overline{\eta})$ plane [39]. The different colours illustrate experimental constraints. The red hashed region of the global fit corresponds to the 68 % confidence interval.

At the beginning of this chapter, 18 free parameters of the SM have been introduced. Now, in 2019, most of them have been measured to a great precision in multiple different measurements [38, 40]. Their current values are given below. Note that the Euler angles can be calculated from the Wolfenstein parametrisation using Eq. (2.11).

$m_u = (2.2^{+0.5}_{-0.4})$ MeV/c^2	$m_c = (1.275^{+0.025}_{-0.035})$ GeV/c^2	$m_t = 173.0(4)$ GeV/c^2
$m_d = (4.7^{+0.5}_{-0.3})$ MeV/c^2	$m_s = (95^{+9}_{-3})$ MeV/c^2	$m_b = (4.18^{+0.04}_{-0.03})$ GeV/c^2
$m_e = 0.510\,998\,946\,1(31)$ MeV/c^2	$m_\mu = 105.658\,374\,5(24)$ MeV/c^2	$m_\tau = 1.776\,86(12)$ GeV/c^2
$g_{\text{em}} = 0.4616(6)$ (at m_{Z^0})	$g_{\text{weak}} = 0.6515(29)$ (at m_{Z^0})	$g_{\text{strong}} = 1.23(7)$ (at m_{Z^0})
$\lambda = 0.224\,53(44)$	$A = 0.836(15)$	$\overline{\rho} = 0.122^{+0.018}_{-0.017}$
$\overline{\eta} = 0.355^{+0.012}_{-0.011}$	$m_{H^0} = 125.18(16)$ GeV/c^2	$v = 246.219\,65(13)$ GeV/c^2

2.3 Mixing and decay of flavoured neutral mesons

To be able to understand the process of meson mixing, it is necessary to introduce symmetries, which are essential operations for the SM. Three *discrete* symmetries exist:

Parity P The system is mirrored along all axes $P(\vec{x}) \Rightarrow -\vec{x}$,

Charge C Particles in the system are exchanged for its anti-particles $C(K^+) \Rightarrow K^-$,

Time T The direction time in this system is reserved $T(\psi(t)) \Rightarrow \psi(-t)$.

The invariance under *CPT* is one of the fundamental properties of the SM and is assumed to be given in the following, unless otherwise stated. In contrast, the Wu experiment proved that the weak force violates P symmetry [41]. Moreover, the C symmetry alone is maximal violated by the weak force as well.

In the SM, there are two groups of neutral mesons. Unflavoured mesons consists of two oppositely charged quarks of the same family. One example is the ϕ meson with the quark content $s\bar{s}$. This type of meson decays dominantly via strong or electromagnetic interactions, leading to small lifetimes, e.g. $\tau(\phi) \approx 10^{-22}$ s. On the other hand, flavoured neutral mesons such as neutral kaons ($d\bar{s}$) only decay via the weak interaction, which is the only source for *CPV* in the SM. Furthermore, both real and virtual transitions, i.e. changing to the *CP* conjugated state of the meson is possible. This phenomenon is called (meson) mixing. In the following section the combination of meson mixing and decay is described theoretically.

Four flavoured neutral mesons exist in the SM: K^0, D^0, B^0 and B_s^0. The theoretical description of mixing and decay for these mesons is the same, hence denoted as P^0 for the particle and $\overline{P^0}$ for the anti particle with opposite flavour content. The decay amplitudes of P^0 to the final state f is then given by

$$A_f = \langle f|\mathcal{H}|P^0\rangle\,, \quad A_{\overline{f}} = \langle \overline{f}|\mathcal{H}|P^0\rangle\,, \tag{2.22}$$

$$\overline{A}_f = \langle f|\mathcal{H}|\overline{P^0}\rangle\,, \quad \overline{A}_{\overline{f}} = \langle \overline{f}|\mathcal{H}|\overline{P^0}\rangle\,, \tag{2.23}$$

where $\mathcal{H}$ is the Hamiltonian of the weak interaction. Applying the *CP* operation on both the initial and final state,

$$CP\,|P^0\rangle = \mathrm{e}^{\mathrm{i}\xi}\,|\overline{P^0}\rangle\,, \quad CP\,|\overline{P^0}\rangle = \mathrm{e}^{-i\xi}\,|P^0\rangle\,, \tag{2.24}$$

$$CP\,|f\rangle = \mathrm{e}^{\mathrm{i}\xi_f}\,|\overline{f}\rangle\,, \quad CP\,|\overline{f}\rangle = \mathrm{e}^{-\mathrm{i}\xi_f}\,|f\rangle\,, \tag{2.25}$$

introduces two arbitrary phases, ξ and ξ_f. Combining the above equations, lead to *CP* invariance conditions for the decay amplitudes

$$\overline{A}_{\overline{f}} = \mathrm{e}^{\mathrm{i}(\xi_f - \xi)} A_f, \quad A_{\overline{f}} = \mathrm{e}^{\mathrm{i}(\xi_f + \xi)} \overline{A}_f, \tag{2.26}$$

or, eliminating the phases by only considering the absolute values,

$$|A_f| = |\overline{A}_{\overline{f}}|, \quad |A_{\overline{f}}| = |\overline{A}_f|. \tag{2.27}$$

In other words, *CP* conservation would lead to the same probabilities for P^0 decaying to f as $\overline{P^0}$ decaying to $\overline{f}$.

Speaking of meson mixing, it is mandatory to introduce a time dependent term. Weisskopf and Wigner [42, 43] showed that this can be accomplished by describing the oscillation and decay of neutral mesons as

$$|\psi(t)\rangle = \psi_1(t)\,|P^0\rangle + \psi_2(t)\,|\overline{P^0}\rangle\,, \tag{2.28}$$

with t as proper time, which should be much larger than the typical timescale of a strong interaction. The wave function evolution is given by the time dependent Schrödinger-like equation

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = \mathcal{H}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = \left(\mathbf{M} - \frac{\mathrm{i}}{2}\mathbf{\Gamma}\right)\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}. \tag{2.29}$$

Here, the complex 2×2 matrices $\mathbf{M}$ and $\mathbf{\Gamma}$ are hermitian, i.e. they are equal to its own conjugate transpose, yielding $\Gamma_{12} = \Gamma_{21}^*$ and $M_{12} = M_{21}^*$. Due to *CPT* invariance, the main diagonal elements of this matrices are the same, $M_{11} = M_{22} = m$ and $\Gamma_{11} = \Gamma_{22} = \Gamma$. This means in particular that P^0 and $\overline{P^0}$ share the same mass m and decay width Γ. Summarized, the Hamiltonian $\mathcal{H}$ becomes

$$\mathcal{H} = \begin{pmatrix} m - \mathrm{i}/2\Gamma & M_{12} - \mathrm{i}/2\Gamma_{12} \\ M_{12}^* - \mathrm{i}/2\Gamma_{12}^* & m - \mathrm{i}/2\Gamma \end{pmatrix}, \tag{2.30}$$

which is non-hermitian. The existence of non-zero off-diagonal elements leads to mixing between the flavour eigenstates P^0 and $\overline{P^0}$. On the other hand, the mass eigenstates can be expressed as a linear combination of the states,

$$|P_L^0\rangle = p\,|P^0\rangle + q\,|\overline{P^0}\rangle \quad \text{and} \quad |P_H^0\rangle = p\,|P^0\rangle - q\,|\overline{P^0}\rangle\,, \tag{2.31}$$

where p and q are complex mixing parameters and the normalisation condition $|p|^2 + |q|^2 = 1$ applies. The indices L and H denote the *light* and *heavy* mass eigenstates, respectively. The masses (decay widths) of such well–defined states can be described as m_L and m_H (Γ_L and Γ_H), their differences are defined as

$$\Delta m = m_H - m_L \quad \text{and} \quad \Delta\Gamma = \Gamma_H - \Gamma_L. \tag{2.32}$$

Furthermore, a meson system can be represented by the average sum and decay width

$$m = \frac{m_H + m_L}{2} \quad \text{and} \quad \Gamma = \frac{\Gamma_H + \Gamma_L}{2}. \tag{2.33}$$

In the B_s^0 system these values have already been experimentally determined to [44]

$$\Delta m_{B_s^0} = (1.1688 \pm 0.0014) \times 10^{-8}\ \text{MeV/c}^2, \qquad \Delta\Gamma_{B_s^0} = (0.088 \pm 0.005) \times 10^{12}\,\hbar\text{s}^{-1}, \tag{2.34}$$

$$m_{B_s^0} = (5366.89 \pm 0.19)\ \text{MeV/c}^2, \qquad \Gamma_{B_s^0} = (0.663 \pm 0.021) \times 10^{12}\,\hbar\text{s}^{-1}. \tag{2.35}$$

However, in terms of matrix elements of $\mathcal{H}$ the mixing parameters p and q can be described as

$$\frac{q}{p} = \sqrt{\frac{2M_{12}^* - \mathrm{i}\Gamma_{12}^*}{2M_{12} - \mathrm{i}\Gamma_{12}}}. \tag{2.36}$$

The time evolution of the mass eigenstates can be expressed as

$$|P_H^0(t)\rangle = \mathrm{e}^{-\mathrm{i}m_H t}\mathrm{e}^{-\frac{\Gamma_H}{2}t}\,|P_H\rangle \tag{2.37}$$

$$|P_L^0(t)\rangle = \mathrm{e}^{-\mathrm{i}m_L t}\mathrm{e}^{-\frac{\Gamma_L}{2}t}\,|P_L\rangle\,. \tag{2.38}$$

In this equations, t is elapsing in the particle's rest frame. Starting at $t = 0$ and assuming a pure flavour state of P^0 resp. $\overline{P^0}$, the above equations can be combined with Eq. (2.31) to

$$|P^0(t)\rangle = g_+(t)\,|P^0\rangle - \frac{q}{p}g_-(t)\,|\overline{P^0}\rangle \tag{2.39}$$

$$|\overline{P^0}(t)\rangle = g_+(t)\,|\overline{P^0}\rangle - \frac{p}{q}g_-(t)\,|P^0\rangle\,, \tag{2.40}$$

with

$$g_\pm(t) = \frac{1}{2}\left(\mathrm{e}^{-\mathrm{i}m_H t}\mathrm{e}^{-\frac{\Gamma_H}{2}t} \pm \mathrm{e}^{-\mathrm{i}m_L t}\mathrm{e}^{-\frac{\Gamma_L}{2}t}\right). \tag{2.41}$$

Experimentally, it is not possible to measure the time evolution directly. Thus, time dependent decay rates are of interest, which describe the rate of an initial P^0 or $\overline{P^0}$ into the final state f or $\overline{f}$ at the time t and are given by the squared matrix element

$$\Gamma(P^0(t) \to f) = |\,\langle f|\,\mathcal{H}\,|P^0(t)\rangle\,|^2. \tag{2.42}$$

Rewriting the complex exponential functions of Eq. (2.41) with trigonometric terms and defining

$$\lambda_f \equiv \frac{1}{\overline{\lambda}_f} = \frac{q}{p}\frac{\overline{A}_f}{A_f} \quad \text{and} \quad \overline{\lambda}_{\overline{f}} \equiv \frac{1}{\lambda_{\overline{f}}} = \frac{p}{q}\frac{A_{\overline{f}}}{\overline{A}_{\overline{f}}}, \tag{2.43}$$

the four decay rates can be described as

$$\begin{aligned}
\frac{\Gamma(P^0(t) \to f)}{\mathrm{e}^{-\Gamma t}} = & \frac{1}{2}|A_f|^2(1+|\lambda_f|^2)\left[\cosh\left(\frac{\Delta\Gamma}{2}t\right) + D_f \sinh\left(\frac{\Delta\Gamma}{2}t\right)\right. \\
& \left. + C_f \cos(\Delta m t) - S_f \sin(\Delta m t)\right], && (2.44)\\
\frac{\Gamma(\overline{P^0}(t) \to f)}{\mathrm{e}^{-\Gamma t}} = & \frac{1}{2}|A_f|^2\left|\frac{p}{q}\right|^2(1+|\lambda_f|^2)\left[\cosh\left(\frac{\Delta\Gamma}{2}t\right) + D_f \sinh\left(\frac{\Delta\Gamma}{2}t\right)\right. \\
& \left. - C_f \cos(\Delta m t) + S_f \sin(\Delta m t)\right], && (2.45)\\
\frac{\Gamma(P^0(t) \to \overline{f})}{\mathrm{e}^{-\Gamma t}} = & \frac{1}{2}|\overline{A}_{\overline{f}}|^2\left|\frac{q}{p}\right|^2(1+|\overline{\lambda}_{\overline{f}}|^2)\left[\cosh\left(\frac{\Delta\Gamma}{2}t\right) + D_{\overline{f}} \sinh\left(\frac{\Delta\Gamma}{2}t\right)\right. \\
& \left. + C_{\overline{f}} \cos(\Delta m t) - S_{\overline{f}} \sin(\Delta m t)\right], && (2.46)\\
\frac{\Gamma(\overline{P^0}(t) \to \overline{f})}{\mathrm{e}^{-\Gamma t}} = & \frac{1}{2}|\overline{A}_{\overline{f}}|^2(1+|\overline{\lambda}_{\overline{f}}|^2)\left[\cosh\left(\frac{\Delta\Gamma}{2}t\right) + D_{\overline{f}} \sinh\left(\frac{\Delta\Gamma}{2}t\right)\right. \\
& \left. - C_{\overline{f}} \cos(\Delta m t) + S_{\overline{f}} \sin(\Delta m t)\right]. && (2.47)
\end{aligned}$$

Typically, the pre-factors of the trigonometric formulas are called *CP* observables. These observables meet the normalisation conditions $D_f^2 + C_f^2 + S_f^2 = 1$ (the same holds for its conjugated values) and are defined as

$$D_f = \frac{2\mathrm{Re}\lambda_f}{1+|\lambda_f|^2}, \qquad C_f = \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}, \qquad S_f = \frac{2\mathrm{Im}\lambda_f}{1+|\lambda_f|^2}, \tag{2.48}$$

$$D_{\overline{f}} = \frac{2\mathrm{Re}\overline{\lambda}_{\overline{f}}}{1+|\overline{\lambda}_{\overline{f}}|^2}, \qquad C_{\overline{f}} = \frac{1-|\overline{\lambda}_{\overline{f}}|^2}{1+|\overline{\lambda}_{\overline{f}}|^2}, \qquad S_{\overline{f}} = \frac{2\mathrm{Im}\overline{\lambda}_{\overline{f}}}{1+|\overline{\lambda}_{\overline{f}}|^2}. \tag{2.49}$$

2.3.1 Types of *CP*-violation

CPV can be expressed as the simple condition $\lambda_f \neq 1$ (see Eq. (2.43)). Nevertheless, different manifestations of *CPV* occur, which are outlined throughout this chapter. Additionally, experimental verifications are listed where possible.

Direct *CP*-violation

Direct *CP*-violation (or *CP*-violation in the decay) can occur when the decay amplitudes are different between the decay and *CP* conjugated decay,

$$\frac{\overline{A}_{\overline{f}}}{A_f} \neq 1. \tag{2.50}$$

This type of *CPV* is the only source of *CPV* for charged mesons. This is achievable due to two different phases in the decay amplitudes, which are the *strong* phase δ_i and the *weak* phase ϕ_i. The weak phase is *CP*-odd, which means that it changes the sign under *CP* transformation. The strong phase on the other hand is *CP*- even, so the phase is the same for both *CP* conjugate states. While the absolute values of these phases are unphysical and convention-dependent, their relative ratio is actually meaningful. Constructing a meson decay consisting of two amplitudes $A_{1,2}$ and their corresponding phases $\delta_{1,2}$ and $\phi_{1,2}$, the complete amplitudes A_f and $\overline{A}_{\overline{f}}$ can be written as

$$A_f = |A_1|\mathrm{e}^{\mathrm{i}(\delta_1+\phi_1)} + |A_2|\mathrm{e}^{\mathrm{i}(\delta_2+\phi_2)} \quad \text{and} \tag{2.51}$$

$$\overline{A}_{\overline{f}} = |A_1|\mathrm{e}^{\mathrm{i}(\delta_1-\phi_1)} + |A_2|\mathrm{e}^{\mathrm{i}(\delta_2-\phi_2)}. \tag{2.52}$$

Assuming that $\delta_{1,2}$ and $\phi_{1,2}$ are non-zero, Eq. (2.50) is fulfilled. Direct *CPV* has already observed in decays like $B^0_{(s)} \to K^+\pi^-$ [45–47], where the most precise measurement was performed by the LHCb experiment [48].

Indirect *CP*-violation

Indirect *CP*-violation (or *CP*-violation in the mixing) can occur if the ratio of q and p is other than one,

$$\left|\frac{q}{p}\right| \neq 1. \tag{2.53}$$

Furthermore, by comparing the above equation with Eq. (2.36), a relative phase difference between M_{12} and Γ_{12}, $|\Gamma_{12}/M_{12}| \neq 0$, emerges. In other words, the mixing rates differ between the particle P^0 and anti-particle $\overline{P^0}$,

$$\Gamma(P^0 \to \overline{P^0}) \neq \Gamma(\overline{P^0} \to P^0). \tag{2.54}$$

Evidence for indirect *CP*-violation in the kaon system was already found in 1964 by Cronin and Fitch [49] (Nobel Prize in Physics in 1980 [50]), being the first hint of *CPV* in general. In the B system, recent measurement of indirect *CPV* yield results compatible SM predictions, see e.g. Ref. [51].

CP-violation in the interference between mixing and decay

It is possible for *CPV* in the interference between mixing and decay to occur if one final state f can be reached from both, the decay of particle P^0 and anti-particle $\overline{P^0}$,

making f to a *CP* eigenstate. However, this does not require the existence of either direct nor indirect *CP*V. By constructing a decay amplitude with only one weak phase ϕ_f and one strong phase δ_f, such as $A_f = |A_f| e^{i(\phi_f + \delta_f)}$, direct *CP*V is forbidden. Additionally, indirect *CP*V can be neglected if $|\Gamma_{12}/M_{12}| \approx 0$, turns the term $q/p = e^{i\phi_f}$ into a pure phase. These assumptions still result in $|\lambda_f = 1|$, while the different phases in the mixing and decay lead to

$$\text{Im}\, \lambda_f \neq 0, \tag{2.55}$$

which is a requirement for *CP*V in the interference between mixing and decay. It is common practice to define a time-dependent asymmetry $A_f(t)$ as

$$A_f(t) = \frac{\Gamma(\overline{P^0}(t) \to f) - \Gamma(P^0(t) \to f)}{\Gamma(\overline{P^0}(t) \to f) + \Gamma(P^0(t) \to f)} \tag{2.56}$$

$$= \frac{S_f \sin(\Delta m t) - C_f \cos(\Delta m t)}{\cosh\left(\frac{\Delta\Gamma t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma t}{2}\right)}, \tag{2.57}$$

where the last equation is expressed in units of *CP* observables defined in Eq. (2.48).

The decays $B^0 \to J/\psi K^0_S$ and $B^0_s \to J/\psi \phi$ are the so-called *golden channels* for *CP*V in the interference between mixing and decay for B^0 and B^0_s mesons. For the former decay, the latest analysis of the *CP* observables S and C was performed by the LHCb experiment [52] yielding values of

$$S_{[c\bar{c}]K^0_S} = 0.760 \pm 0.034 \tag{2.58}$$

$$C_{[c\bar{c}]K^0_S} = -0.017 \pm 0.029, \tag{2.59}$$

where $[c\bar{c}]$ implies that the excited state of the J/ψ, i.e. the $\psi(2S)$ meson, was also included in the measurement.

2.3.2 *CP*V in the kaon system

While the theoretical description of *CP*V in general has already been discussed in the previous chapters, the kaon system has some special characteristics which is further elaborated in the following.

The flavour eigenstates of kaon and anti-kaon are given as $K^0 = \bar{s}d$ and $\bar{K}^0 = s\bar{d}$, respectively. *CP* eigenstates, which are the result of diagonalizing the Hamiltonian in Eq. (2.30), are linear combinations of the flavour eigenstates,

$$K^0_1 = \frac{1}{\sqrt{2}}(K^0 - \bar{K}^0) \quad \text{and} \quad K^0_2 = \frac{1}{\sqrt{2}}(K^0 + \bar{K}^0). \tag{2.60}$$

Applying the *CP* operator to both *CP* eigenstates yields

$$CP(K_1^0) = +K_1^0 \quad \text{and} \quad CP(K_2^0) = -K_2^0. \tag{2.61}$$

Hence, under the assumption of *CP*-preservation, only the decay of K_1^0 into two pions and K_2^0 into three pions should be allowed, since pions possess a *CP*-eigenvalue of -1. Furthermore, the decay $K_2^0 \to 3\pi$ leaves only a small part of the phase-space left, increasing the lifetime of K_2^0 by a factor of about 600 in comparison to the K_1^0. Nevertheless, *CPV* is possible by defining weak eigenstates, which are itself combinations of K_1^0 and K_2^0 with a small contribution ϵ from each other,

$$K_\mathrm{S}^0 = \frac{1}{\sqrt{1+|\epsilon|^2}}(K_1^0 + \epsilon K_2^0) \quad \text{and} \quad K_\mathrm{L}^0 = \frac{1}{\sqrt{1+|\epsilon|^2}}(\epsilon K_1^0 + K_2^0). \tag{2.62}$$

The indices *S short* and *L long* refer to the lifetime of the particle. This is in contrast to other neutral meson systems, where the distinction is based on the particles mass. However, the mass of K_L^0 is greater than the mass of K_S^0. The current value of the *CP* violating parameter ϵ is given as [44]

$$|\epsilon| = (2.228 \pm 0.011) \times 10^{-3}. \tag{2.63}$$

2.4 Possible Physics beyond the SM

Although the SM is the most widespread and best understood theory in particle physics, it is still known to be incomplete. For example, gravity is described by the general theory of relativity, being no component of the SM. Also not explainable is the existence of dark matter and dark energy [53]. The former one is postulated by analysing rotation speeds of stars as a function of distance to the centre of a galaxy. Dark energy on the other hand is needed to explain the acceleration of the expansion of the universe [54, 55]. Furthermore, neutrino masses are assumed massless in the SM. This is not true, since the oscillation of neutrinos require a non-zero mass as was already measured in the Wu experiment [56, 57]. The SM is also not able to explain the matter-antimatter asymmetry in the current universe, since the effect of *CPV* alone is too small by orders of magnitude.

Measurements published by the LHCb collaboration may also hint to physics beyond the SM. One example are lepton flavour universality tests in electroweak-penguin decays. In SM predictions, the fraction of

$$R_X = \frac{b \to se^+e^-}{b \to s\mu^+\mu^-} \tag{2.64}$$

should be compatible to one. However, two analyses where $X \in \{K, K^*\}$ show deviations of this value. The lepton universality test in $B^+ \to K^+l^+l^-$ decays yield a

result being compatible to SM predictions within 2.6 standard deviations [58]. Additionally, the measurement of form-factor-independent observables in $B^0 \to K^* \mu^+ \mu^-$ decays show an anomaly in one specific parameter P_5', whose probability to be a random fluctuation is one in 200 [59]. To validate these anomalies, a lot plenty lepton flavour universality tests are currently studied, e.g. the measurement of $R_{K^+\pi^+\pi^-}$ and an updated analysis of R_K. Therefore, a so-called R_X framework is in development to simplify such kinds of analyses.

Nevertheless, there are many theories trying to explain these phenomena. Supersymmetry (SUSY) [60] introduces a hypothetical symmetry, which allows to convert bosons into fermions. Additionally, this theory postulates a superpartner to each particle existing in the SM with a spin differing by a half-integer. Among other benefits SUSY could solve the hierarchy problem. In the SM, the hierarchy difference between the Planck and electroweak scale is achieved by fine-tuning parameters, while the quantum corrections of the Planck scale cancel out in SUSY between particle and superpartner, explaining the hierarchy in a more natural manner. Besides SUSY, theories exist to unify the strong, weak and elecro-magnetic force, called Grand Unified Theory. This would be a first step to develop a *Theory of Everything* that fully explains all physics in the universe.

Chapter 3

The LHC and the LHCb Experiment

The Large Hadron Collider beauty (LHCb) experiment is part of the European Organization for Nuclear Research (CERN) and is located in Meyrin near Geneva, Switzerland. Founded in 1954, CERN currently has 22 member states, about 3400 employees and 14000 guest scientists, operating the largest particle accelerator in the world, the Large Hadron Collider (LHC), at which the LHCb detector is positioned. This chapter is structured as follows: first, the pre-accelerators as well as the LHC itself are briefly introduced in Sec. 3.1. In addition to the LHCb experiment, other experiments exist at the LHC, of which the largest are ATLAS, CMS and ALICE. These experiments are discussed in more detail in Sec. 3.1.1. The LHCb detector as well as the most important subdetectors relevant for this thesis are outlined in Sec. 3.2. Furthermore, the trigger system, specific software stacks and upgrade plans are discussed here. Finally, upgrades and running conditions of the LHCb experiment are summarised in Sec. 3.3.

3.1 Large Hadron Collider

The LHC [15] is a two-ring hadron accelerator with a circumference of 26.7 km. Inside of it, protons (p) are accelerated in counter-rotating beams and collisioned at four interaction points. It is designed to reach centre-of-mass energies of up to 14 TeV and instantaneous luminosities of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [16]. Furthermore, the protons can be replaced by heavy ions X like lead nuclei, allowing the analysis of pX as well as XX collisions. The LHC supersedes the Large Electron-Positron Collider (LEP) [61, 62] and reuses the same quasi-circular tunnel, 50 m to 175 m below the surface and consisting of eight straight sections and eight arcs. To keep the protons on track, 1232 superconducting dipole magnets, 392 quadrupole magnets and supporting magnets are mounted around the accelerator.

Protons are obtained by ionizing hydrogen and pre-accelerated by the Linear Accelerator 2 (LINAC2), the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) to an energy of 450 GeV before being filled in the LHC, see Fig. 3.1. Afterwards, the proton beam is further accelerated to

an energy of maximal 7 TeV. The beams itself consists of bunches with 1.15×10^{11} particles each and a bunch-spacing of 25 ns, which corresponds to a collision rate of 40 MHz.

To prevent the LHC and detectors for out-of-control beam conditions, the Beam Condition Monitor (BCM) [63] has been developed. Polycrystalline chemical vapour deposited diamonds measure the particles flux closely to the beam pipe at two locations around the interaction point of the LHCb experiment. If the flux is higher than a certain threshold (usually around $200 - 500$ higher than the minimum bias signal), the BCM immediately causes a beam dump. This is also the case when the power supply of the BCM gets interrupted.

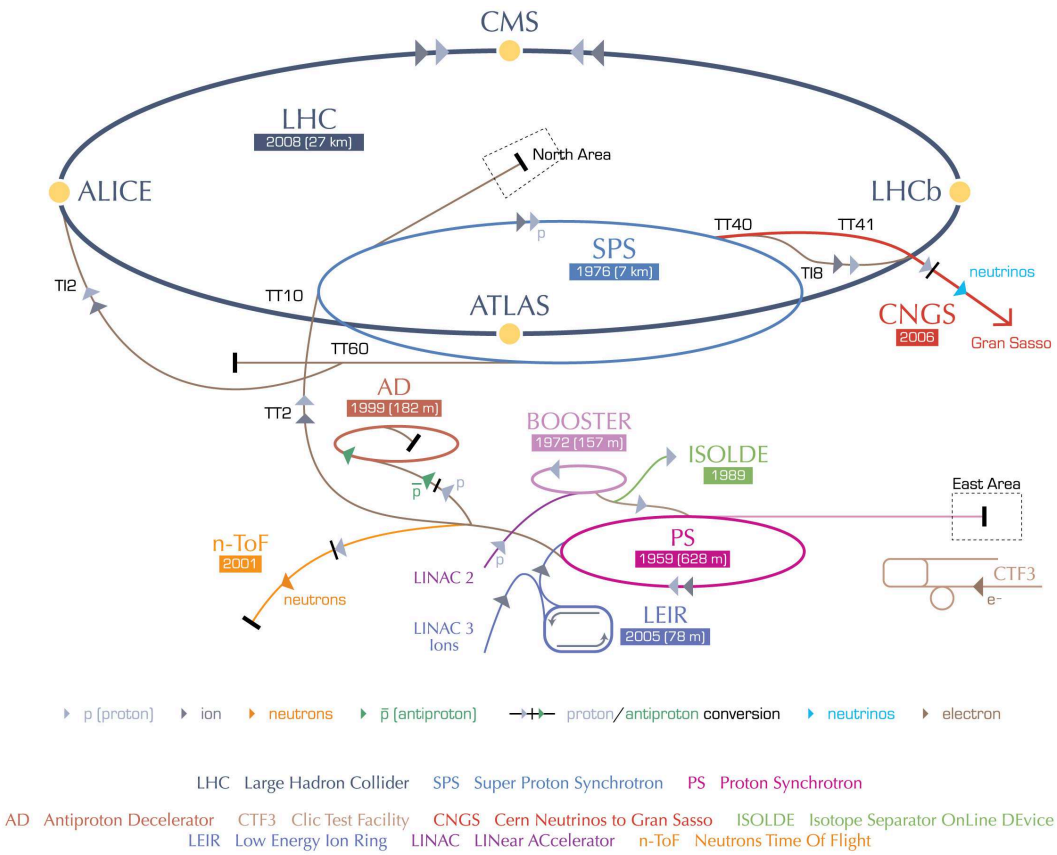


FIGURE 3.1: Schematic overview of the accelerator complex located at CERN [64].

3.1.1 Experiments

In total, seven experiments are installed at the LHC. While the LHCb detector is described in detail in Sec. 3.2, the other experiments will be discussed in more detail in the following.

ATLAS [65] (A Toroidal LHC Apparatus) and CMS [66] (Compact Muon Solenoid) are two General Purpose Detectors (GPD), which cover the entire 4π solid angle. Their physics program includes precision measurements of heavy particles like the

Higgs boson, t quark and $W^{\pm}$ bosons. Furthermore, ATLAS and CMS are searching for New Physic effects, like supersymmetric particles and dark matter candidates. Even though both detectors have similar physic goals, their design and technical solutions are different. For example, the single magnet of CMS is optimised on a maximum magnetic field strength, while the three magnets in ATLAS are focussed on a high-precision momentum measurement of muons.

ALICE [67] (A Large Ion Collider Experiment) is specialised in the analysis lead-ion collisions. It focuses on the analysis of the strong interaction and exotic states. The latter one describes a state akin to a very early state of the universe, where quarks and gluons are asymptotically free due to extremely high temperatures and densities.

The main task of LHCf [68] (Large Hadron Collider forward) is to analyse π^0 decays in the very forward regions of the beam pipe, which could help to understand the origin of high energy cosmic rays. MoEDAL [69] (Monopole and Exotics Detector at the LHC) searches for the existence of magnetic monopoles and dyons, a hypothetical particle postulated in many GUTs. The last experiment at the LHC is TOTEM [70] (TOTal Elastic and diffractive cross section Measurement). Similar to LHCf, the detector components (called Roman Pots) are arranged to cover the forward regions to be able to measure the pp cross-section independently of the luminosity as well as diffractive scattering of the protons.

3.2 LHCb Detector

The LHCb detector [16] is a single arm forward spectrometer and specialised in the analysis of b- and c-hadron decays. The structural difference of the detector design in contrast to GPDs can be clarified in the generation process of this hadrons, which is dominated by gluon–gluon fusion and quark–antiquark annihilation [71]. Since the parton momenta in the pp collision is dominantly asymmetric, the produced quark pair is boosted in the forward/backward direction in respect to the beam pipe. Simulations show that about 25 % of all $b\bar{b}$ pairs are in detector acceptance, see Fig. 3.2. The correlation between the b quarks is exploited in the flavour tagging (FT) [72], a technique to determine the initial flavour of a neutral hadron. FT is an important ingredient for one of the main physics motivations of LHCb, the analysis of CP violation and precise measurement of CP observables, see also Sec. 2.3. Furthermore, CP violation measurements benefit from LHCbs low decay time resolution of about 45 fs [73]. Another key task of LHCb is the search for rare decays like $B^0_{(s)} \to \mu^+\mu^-$. The branching ratio of these decays have been measured by LHCb [74, 75]. Additionally, a combined Run I measurement has been performed together with the CMS collaboration yielding [76]

$$\mathcal{BR}(B_s^0 \to \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9} \text{ and} \tag{3.1}$$

$$\mathcal{BR}(B^0 \to \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10}. \tag{3.2}$$

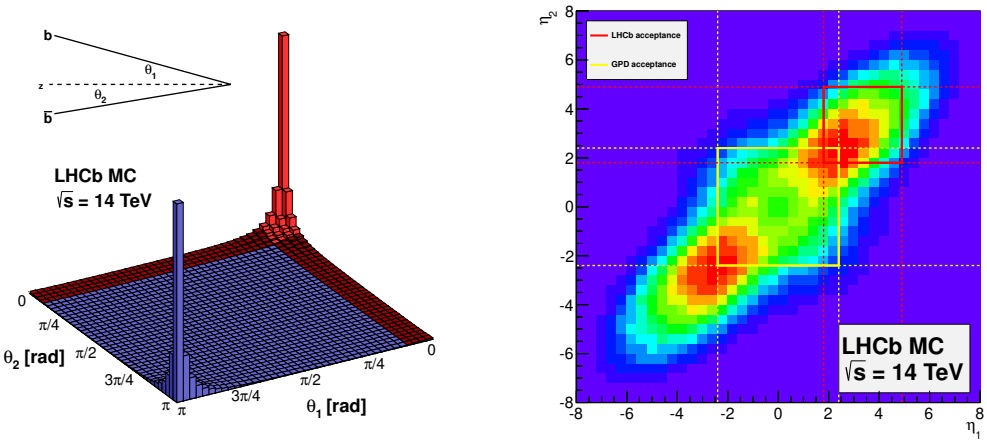


FIGURE 3.2: Production angle distributions of $b\bar{b}$ quark pairs produced in pp collisions, simulated with PYTHIA8 [77]. Angular distributions of both b quarks in radian (left) and pseudorapidity η (right) are shown. The red area (left) and red square (right) highlights the detector acceptance of LHCb, respectively. For comparison, the yellow square in the right plot marks the acceptance of a 4π GPD like ATLAS.

A schematic overview of the LHCb detector is given in Fig. 3.3. The reference system is defined so that the interaction point lies in the coordinate origin, the z-axis points into the detector, the x-axis is horizontal and y-axis is aligned vertical. Furthermore, positive z-axis values are refereed to as *downstream*, while the opposite direction is called *upstream*. The following components are arranged downstream of the detector: A vertex locator, two ring imaging Cherenkov detectors, a dipole magnet, four tracking stations, two calorimeters, one scintillating pad detector and pre-shower as well as five muon chambers.

3.2.1 Tracking Systems

The tracking system of LHCb enables the determination of charged particles track trajectories. It consists of the Vertex Locator (VELO), mounted closely to LHCs beampipe, the Tracker Turicensis (TT) and three tracking stations T1, T2 and T3, which itself consists of an inner (Inner Tracker, IT) and outer (Outer Tracker, IT) part. Between TT and T1, a warm dipole magnet with an integrated magnetic field of 4 T m [16] is assembled. Due to Lorentz force, charged particles get deflected, allowing a momentum measurement. The polarity of the magnet is changeable (*up* and *down*), which can reduce systematic effects. The momentum resolution of the

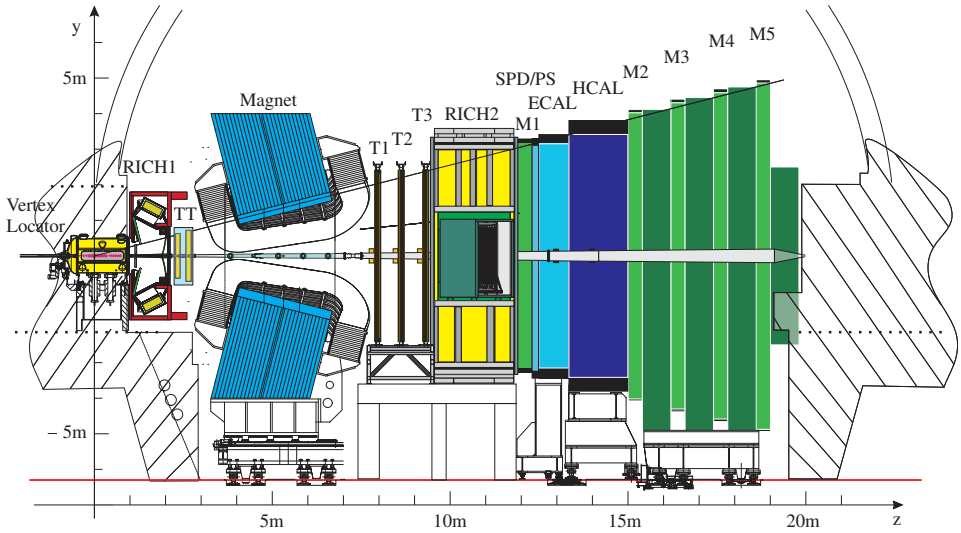


FIGURE 3.3: Schematic overview of the LHCb detector.

hole tracking system is $\Delta p/p = 0.4\,\%$ at 5 GeV/c and 0.6 % at 100 GeV/c, respectively [78]. Furthermore, the impact parameter resolution is 20 μm for tracks with high transversal momenta.

VELO

The VELO [79] has the smallest distance to the beam pipe of all detector components at the LHC. This enables precise track measurements closely to the interaction point, which is crucial to identify displaced vertices of b- and c decays. Such mesons have a mean lifetime of $\mathcal{O}(10^{-12}\,\mathrm{s})$, corresponding to a flight distance of $\mathcal{O}(5\,\mathrm{mm})$ in the detector. The VELO is made up of 42 half-disc-shaped silicon modules, grouped together horizontally in stations, two at a time. 19 stations are located downstream and arranged in a way that at least three modules are hit if the track is in LHCbs acceptance of 10–300 mrad horizontally and 10–250 mrad vertically. The remaining two upstream stations serve as pile-up sensors and estimate the number of pp interactions in each bunch crossing. A schematic overview of the modules arrangement is given in Fig. 3.4. The modules themselves have a radius of 42 mm and measure the radial distance from the beam pipe R and azimuthal angle ϕ of the particles track independently. To not be damaged by unstable beam conditions e.g. at the injection phase, the modules are mounted to a moveable apparatus. Hereby, the VELO can be moved as close as 5 mm during data taking, while being protected in garage position, which is 30 mm away from the beam pipe.

Silicon Trackers and Outer Tracker

The Silicon Trackers (ST) summarise two detectors, the TT [80] and the IT [81]. Both detectors share the same silicon strip technology to obtain a excellent spatial resolution in regions of high track multiplicities. The TT is located upstream of the magnet

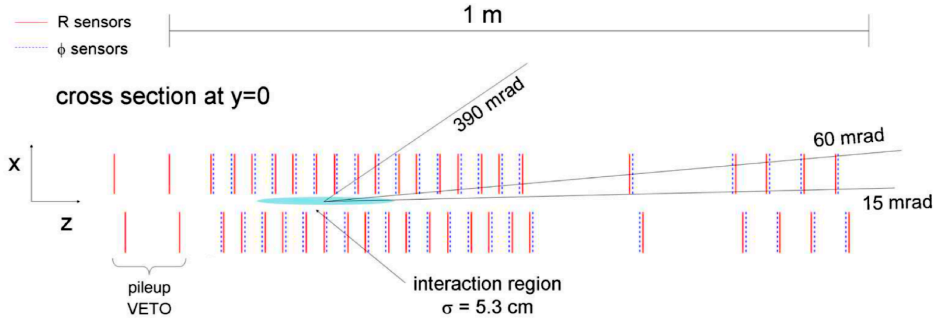


FIGURE 3.4: Scheme of the Vertex Locator [79]. The pp interaction region is marked as a blue ellipse.

and has a sensitive area of $150 \cdot 130\text{cm}^2$, covering the full LHCb acceptance. Downstream of the magnet, the tracking is adopted by three stages IT and OT combinations, where the IT only covers the central part around the beam pipe while being surrounded by the OT. Hence, the detector area of the IT is smaller ($120 \cdot 40\text{cm}^2$). Each station of the ST consists of four detection layers arranged in a x-u-v-x geometry. The first and last layers are vertically aligned with respect to the y-axis, while the second and third layers are rotated around the beam pipe by $+5°$ and $-5°$, respectively.

The OT [82] is a drift time detector and extends the IT to cover the full LHCb acceptance. It consists of about 55000 independent straw-tubes with a diameter of 4.9 mm each. 256 (128) straws are combined to 168 long (96 short) two-ply modules, where the two layers are shifted by one straw radius. Each straw is filled with a gas mixture of 70 % Ar, 28.5 % CO_2 and 1.5 % O_2, allowing drift times below 50 ns.

3.2.2 Particle Identification Systems

The process of determining the flavour of final-state particles, Particle Identification (PID), is a crucial part in most analyses of LHCb. Therefore, multiple subdetectors were constructed to perform a prediction about the particle itself: The RICH1 and RICH2, the hadronic and electromagnetic calorimeters (HCAL and ECAL) and the Muon Chambers. Together, π-K meson misidentification rates of less than 5 % [83] are achieved, which is especially useful in analyses of purely hadronic final states, like $B_s^0 \to K_S^0 K_S^0$.

RICH1 and RICH2

The Ring Imaging Cherenkov detectors RICH1 [84] and RICH2 [85, 86] provide a particle identification in a wide momentum range. Both detectors exploit the Cherenkov radiation, i.e. the emission of electromagnetic radiation by charged particles, which travel to a specific dielectric medium with a speed greater than the phase velocity of light in this medium. This effect can be summarised by the formula

$$\cos\theta_c = \frac{1}{n\beta}, \tag{3.3}$$

where $\beta = v/c$, n is the refractive index of the medium and θ_c the emission angle. The emitted photons are reflected by a combination of a spherical and planar mirror to build cones, which are registered Hybrid Photon Detectors (HPDs). The radii of these cones can then be translated into the angle θ_c. Furthermore, with the additional information about the particles momenta and direction (provided by the tracking stations) and the knowledge about n, it is possible to calculate a probability about a particles identity. RICH1 is located between the VELO and TT, upstream of the magnet. It uses silica aerogel and C_4F_{10} as radiators to be sensitive to particles with low momenta (1 GeV $-$ 60 GeV) in the full LHCb acceptance. In Run II of data taking, the aerogel has been completely removed from RICH1. The RICH2 consists of CF_4 and covers the high-momentum region 15 GeV $-$ 100 GeV in the central angular range 15 mrad $-$ 120 mrad.

Calorimeters

The Hadronic (HCAL) and Electromagnetic (ECAL) Calorimeters [87] are build for two main purposes. First it measures the energy of incoming particles. This information can be used to provide a fast estimate about the particles type, i.e. whether the particle is a hadron, electron or neutral particle such as photon and neutron, what is exploited in the trigger. The complete calorimeter system consists of the Scintillating Pad Detector (SPD), the Preshower Detector (PS), the ECAL and the HCAL, all located downstream of the first muon chamber. In combination, the SPD and PS allow to determine whether a particle is charged and to separate electrons from hadrons. The ECAL is equipped with the so called "shashlik" technology, which means that it consists of alternating scintillating tiles and lead plates. The lead causes particles to shower, an effect where a cascade of secondary particles are produced due to the interaction of a high-energy particle with dense matter. Only if the particle is fully stopped in the calorimeter, the measurement of energy is adequate. The HCAL is build in a similar way as the ECAL, but the lead is replaced with iron plates.

Muon Chambers

The muon system [88–90] is composed of five individual muon chambers. The first station is placed upstream of the calorimeter system and mainly used to obtain transversal momentum measurements, which is an important input for the trigger. The remaining chambers are mounted at the very end (downstream) of the detector, separated by iron absorbers with a thickness of 80 cm. Muons are the only type particles penetrating through all muon chambers, so that their signature is very clean. Hence, a large part of analyses in LHCb contain muons, such as the gold-plated decay channels to measure *CP*-violation $B^0 \to J/\psi K^0_S$ and $B^0_s \to J/\psi \phi$.

3.2.3 Trigger

The main task of LHCbs trigger system [91, 92] is to reduce the nominal collision rate of 40 MHz to a manageable amount while keeping only events compatible with LHCbs physics program. In order to achieve this objective, the trigger is build in a step-wise system and consists of three stages. First, a hardware trigger (L0) reduces the rate by a factor of 40. After passing the requirements of two software High Level Triggers (HLT), the rate is further reduced to around 5 kHz, which is stored for later off-line analyses. This corresponds to a continuous data production of about 0.3 GB per second. Fig. 3.5 summarises the trigger system of LHCb.

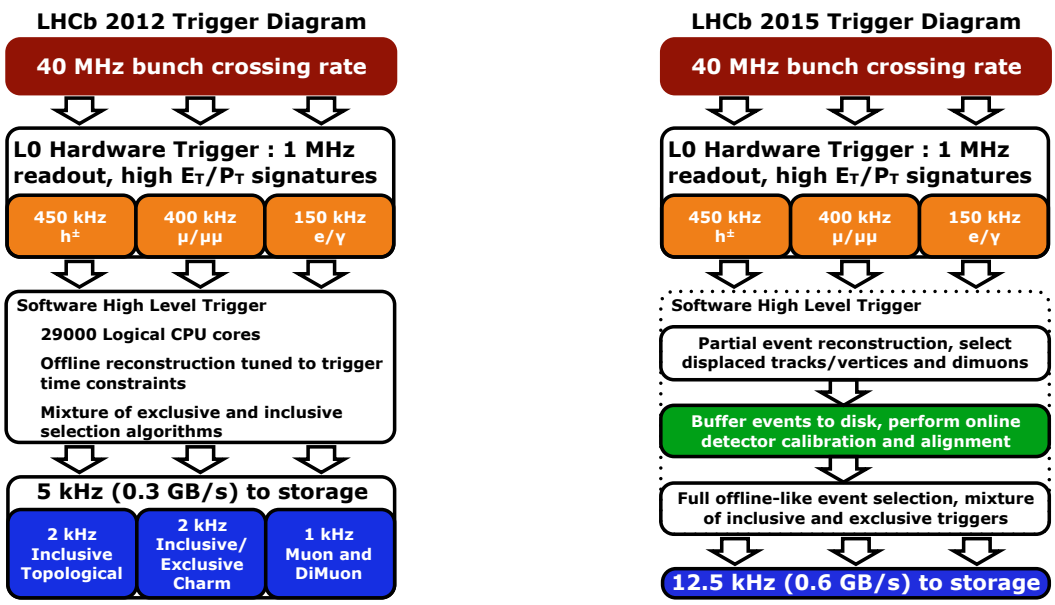


FIGURE 3.5: Schematic overview of the trigger architecture of LHCb. Representative for Run I (Run II), the trigger conditions of in 2012 (2015) are shown on the left (right) diagram [93].

The L0 trigger synchronously processes at the full bunch crossing rate of the LHC and has a latency of 4 µs. Its decisions are based on subdetector components with a fast response time such as muon stations and the calorimeter system. The complete 1 MHz output rate is composed of different channels: `L0Hadron` ($\approx$ 50 %), `L0(Di)Muon` ($\approx$ 40 %), `L0Photon` and `L0Electron` (together $\approx$ 10 %). In case of muons, the trigger decision depends on the transversal momentum provided by the muon stations. A minimal transversal energy (E_T) measurement of the calorimeter system is required for all other particles. For example, the `L0Hadron` requirement was $E_T > 3.5$ GeV ($E_T > 3.7$ GeV) for 2011 (2012). Furthermore, the L0 trigger omits events with a high pile-up, i.e. multiple pp interactions per bunch crossing. These events have a few drawbacks, like the track reconstruction is more difficult and time-consuming and the assignment of a secondary b-decay vertex to the PV has a higher possibility to be wrong.

Software triggers

To confirm (or revise) the decision given by the L0 trigger, the software application HLT has been developed. A computing farm called Event Filter Farm (EFF), consisting of more than 25000 logical CPU cores provides the computing power for the HLT. Similar to the L0 channels, the HLT requirements are organised in trigger lines. To keep track of all lines and potential adjustments, Trigger Configuration Keys (TCKs) are introduced. For example, the baseline TCK for 2016 data taking contains ≈ 50 HLT1 and ≈ 400 HLT2 lines.

The HLT1 reduces the rate of the L0 output to 40 kHz − 80 kHz, depending on the run conditions. In contrast to the L0 trigger, information from the VELO and the tracking stations are available at this stage, allowing a partial event reconstruction. Tracks are extrapolated through the detector using the same forward tracking algorithms [94] as done for the later off-line analysis, but in a smaller search window to reduces the computing resources. Of special interest for decays with purely hadronic final states is the trigger line `Hlt1TrackAllL0`. It selects particles with a high p_T and a significant displacement form the primary vertex.

If an event also passes the HLT2 requirements, it is written to permanent storage. The HLT2 has access to the full event information, hence is able to reconstruct a complete decay chain. This allows to build trigger lines for specific decays in an *exclusive* way, like $B_s^0 \to K_S^0 K_S^0$. In the course of this work a dedicated line for such decays has been developed and included in the TCKs starting in 2017 (also see Ch. 7). Another way to trigger hadronic final states are the topological lines. With their *inclusive* nature, their focus is on b hadron decays with at least two charged particles in the final state. Furthermore, they require a high quality track as well as a good vertex quality for to V^0 particles K_S^0 and Λ.

In addition to the binary output of a certain trigger line, the particles involved in this decision are also stored. By comparing this information with the signal candidate decay, three categories can be defined. First, an event is referred to as Triggered On Signal (TOS) if the signal candidate led to the positive trigger decision. Contrary to this, the event is designated as Triggered Independent of Signal (TIS) if it is trigged even after removing the signal candidate and all tracks belonging to it from the event. Triggered On Both (TOB) events are neither TIS nor TOS and only occurs in less than 0.5 % of all events. The latter trigger requirement is only fulfilled by taking into account both, a part of the signal as well as the underlying event.

3.2.4 Software Stack

All LHCb specific software is integrated into the so-called `Gaudi` [95, 96] framework, an experiment-independent open project designed for high energy physics (HEP) tasks. This software is divided into software packages, each specialised on a specific part in the event reconstruction chain. The output data of the LHCb experiment is

forwarded to the Moore [97] software, which determines the HLT decisions as described in Sec. 3.2.3. After a successful trigger decision, the full event reconstruction is performed by Brunel [98]. Here, tracking and particle identification are of particular importance. Therefore this will be dealt with in detail in Sec. 6.3.1. The output of Brunel [98] are proto particles, a summary of all the reconstructed information about a track, stored in Data Summary Tapes (DSTs). The DaVinci [99] package is used to select specific candidates in decay chains and to calculate higher level observables like invariant masses. Finally, the selected events are stored in the so-called nTuples format.

For most HEP analysis, simulated datasets are an indispensable input. Thereby, the aim is to be as close to reality as possible, e.g. in terms of trigger decisions and detector responses. To create those datasets, which are also called Monte Carlo events (MC), the framework Gauss [100, 101] is used. The complete event processing chain is shown in Fig. 3.7.

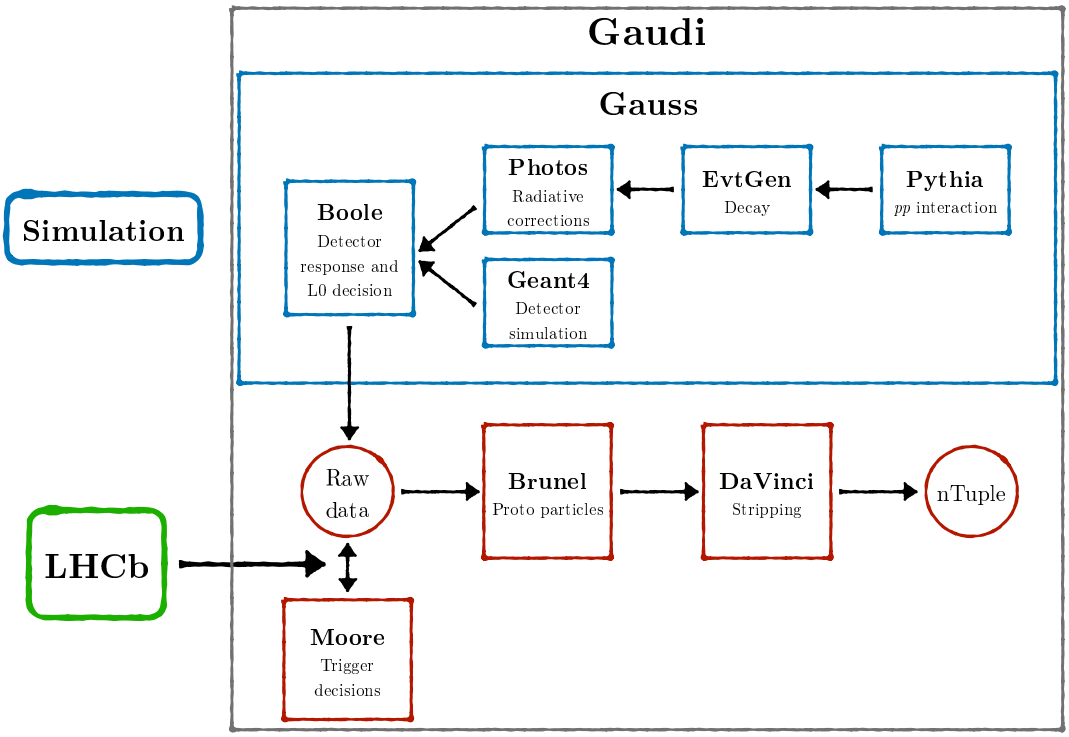


FIGURE 3.6: Overview of LHCbs software stack. In the top row, the software module to generate simulated data samples are shown. Hereby, GAUSS interacts as steering software. After the trigger decisions have been determined in MOORE, both simulated and real detector data are treated in the same way.

Event reconstruction

As mentioned in the previous chapters, three tracking sub-detectors exist at the LHCb detector. The VELO and TT are located upstream of the magnet, while the three T stations T1-T3 are downstream, itself consisting of IT and OT components. There are different algorithms to combine tracking station hits to tracks. Due to the

large distance to the magnet, the magnetic field strength is low. Therefore, a pattern recognition algorithm specialised on straight tracks is sufficient [102]. These tracks are called VELO tracks. The forward tracking algorithm extrapolates these trajectories into the TT to build upstream tracks, taking into account the magnetic field map. Furthermore, if hits of the T stations can be assigned to the extrapolated upstream tracks, these tracks become long tracks, containing the most precise information and best momentum resolution of all track types. The track matching algorithms on the other hand combines VELO tracks with T tracks first, before interpolating the track to the TT stations. If a track is only found in the T stations, this track is called a T track. Another important track category applies especially for V^0 particles. Due to their long lifetime, a decay downstream of the VELO is possible. Thus, such tracks are referred to as downstream tracks. An overview of all track types is also given in Fig. 3.7.

The identification of charged hadrons is dominated by the information provided by the RICH systems. The track trajectory is interpolated through the RICH detector and a global likelihood fit is applied for each track, particle hypothesis and the pixel hits. These likelihoods are combined with likelihoods from the other particle identification systems using similar techniques and then normalised to a pion hypothesis. Afterwards, other particles like kaons or protons can be tested by comparing the likelihood of this particle against the pion, such as $\Delta \log \mathcal{L}(K - \pi)$. The logarithms of these values is taken because of computational reasons. Another way to calculate particle probabilities is based on Artificial Neural Nets. One of the advantages is that the individual output is normalised between 0 and 1, so that its value can be treated as probability.

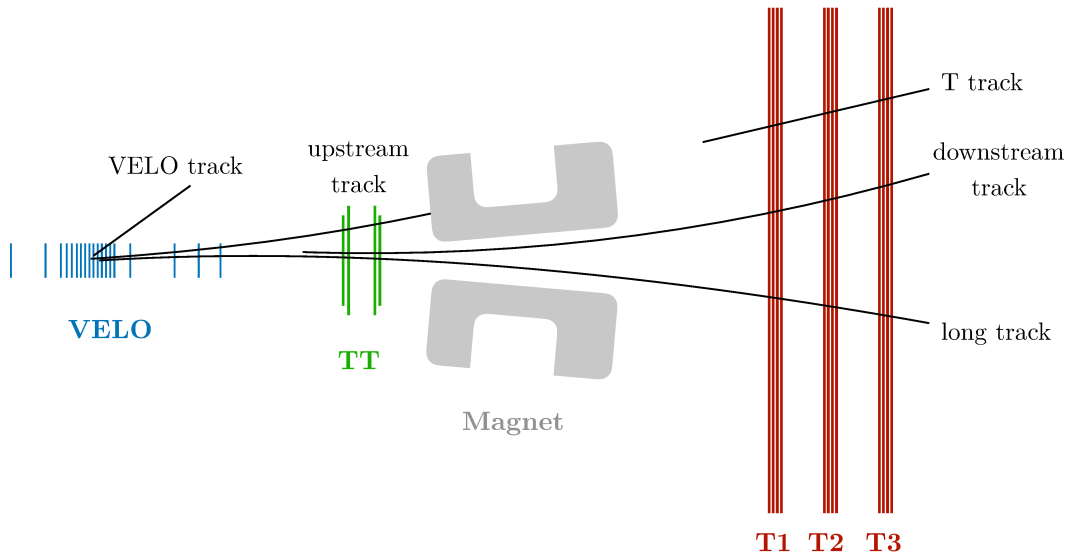


FIGURE 3.7: Overview of all track type categories that can occur in LHCb. For this analysis, the most important track types are long and downstream tracks.

Simulation

Similar to detector data processing chain, the MC production is also performed in multiple steps by tools specialised on specific tasks. The `Gauss` [100, 101] framework forms the central interface for this purpose. It contains important physics properties of LHCb like beam energies and crossing angles, and passes this information to the underlying software. The pp interaction and hadronisation is simulated by `Pythia` [103] in a special configuration adapted to the conditions of LHCb [104]. Depending on the physics analysis, other event generators like `Sherpa` [105] can be used. The decay of heavy hadrons is taken over by `EvtGen` [106], being excellently suitable for simulating meson mixing and *CP* violation. QED corrections like photon radiation is performed by `Photos` [107]. Furthermore, the whole LHCb detector is build in `Geant4` [108, 109], providing simulations of particle interaction with the detector material. To be able to simulate an actual detector response, `Boole` [110] is used, also emulating the hardware trigger decision. Finally, the output of `Boole` can be treated as real detector data and hence being further processed by `Moore`, `Brunel` and `DaVinci`.

Simulated events come with the benefit of truth information, which is stored and can be accessed in the final nTuple. As the event reconstruction does not now about this information, combining signal particles with those of the underlying event may yield a "fake" signal. The process of filtering these events is called truth-matching and is essential step in any analysis. In contrast to signal MC, where a specific decay chain is requested, the production of inclusive MC or even minimum bias samples are possible, which are interesting to identify possible background contributions.

3.3 Running Conditions

At LHC, the data taking periods are divided into runs, separated by Long Shutdown (LS) periods scheduled for accelerator maintenance and detector upgrades. Run I comprises the years $2010 - 2012$ and Run II $2015 - 2018$, whereby most of 2010 data was used for detector calibration. Since the integrated luminosity $\mathcal{L}_{\text{int}} < 0.04\ \text{fb}^{-1}$ is also quite low, this part of data is omitted in most analyses. The LHC is designed for a centre-of-mass energy of $\sqrt{s} = 14$ TeV and a bunch spacing of 25 ns. However, during the first two runs these values have not been reached [78]. Instead, the bunch crossing rate was cut in half in Run I. Furthermore, the centre-of-mass energy increased over the time from $\sqrt{s} = 7$ TeV in 2011, $\sqrt{s} = 8$ TeV in 2012 and $\sqrt{s} = 13$ TeV in the complete Run II period, respectively. The instantaneous luminosity provided by the LHC beam is around $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ [111]. To meet the physics goals of LHCb, a precise determination and correct assignment of decay vertices is crucial. Hence, the crossing angle is adjusted during stable beam conditions to lower the number of visible interactions per bunch-crossing μ to be between 1 and 2, corresponding to an instantaneous luminosity of $4 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$. This process is called

luminosity leveling at LHCb. Furthermore, it brings the advantage that event multiplicities remain constant over a fill, while they decrease continuously for the other experiments. Averaged over all years, the data taking efficiency of LHCb is around 89 %. A summary table over all run statistics is given in Tab. 3.1.

TABLE 3.1: Running conditions summary of the first two LHC runs. From left to right the columns show the run, data taking year, bunch-crossing frequency, center-of-mass energy $\sqrt{s}$, recorded integrated luminosity $\mathcal{L}_{\text{int}}$, average instantaneous luminosity $\mathcal{L}_{\text{inst}}$, average interactions per bunch-crossing μ and data taking efficiency ε, respectively. Here, $\mathcal{L}_{\text{inst}}$ is calculated by dividing $\mathcal{L}_{\text{int}}$ and the time of actual data taking [78, 91, 92, 112].

Run	Year	Rate [MHz]	$\sqrt{s}$ [TeV]	$\mathcal{L}_{\text{int}}$ [fb^{-1}]	$\mathcal{L}_{\text{inst}}$ [$\text{nb}^{-1}\text{s}^{-1}$]	μ	ε
Run I	2010	20	7	0.04	0.02	varies	0.90
	2011	20	7	1.11	0.23	1.4	0.91
	2012	20	8	2.08	0.32	1.7	0.94
Run II	2015	40	13	0.33	0.10	1.1	0.88
	2016	40	13	1.67	0.22	1.1	0.87
	2017	40	13	1.82	0.26	1.1	0.91
	2018	40	13	2.19	0.29	1.1	0.89

3.3.1 Upgrade and Prospects

With Run III starting in 2021, the LHC will for the first time operate on its nominal centre-of-mass energy of $\sqrt{s} = 14$ TeV while doubling the instantaneous luminosity $\mathcal{L}_{\text{int}}$ with regard to Run II [113]. Besides the environment changes, LHCb also introduces the next phase of the experiment, where the instantaneous luminosity is planned to increase by a factor of five to be $2 \times 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$. To be able to handle such rates, LHCb performs an upgrade which applies to both, the hardware and the software side. Furthermore, the detector should be able to withstand radiation damage up to an integrated luminosity of $50\,\text{fb}^{-1}$. In the following lines, the most important changes will be summarised briefly.

Trigger system

To fully profit from the increased luminosity in Run III, the hardware trigger L0 will be removed and replaced by a full software trigger. The latter must make a trigger decision at an average frequency of LHC's nominal collision rate, which corresponds to a speed update needed regarding to the current Run II HLT1 algorithms by more than 10. Additionally, the output bandwidths is expected to be about $5\,\text{GBs}^{-1}$ [114, 115].

Tracker

By removing the L0 trigger, all tracking sub-detectors have to increase their readout rate from 1 MHz to 40 MHz. Moreover, the increased luminosity also comes with a

higher number of interactions per bunch crossing $\mu = 5.2$ as well as a higher occupancy of the tracking stations. To be able to process this higher number of tracks, the tracking stations are replaced in Run III. First, the TT is replaced by the Upstream Tracker (UT), which is based on new high granularity silicon micro-strip planes. Furthermore, the combination of inner and outer tracker is exchanged with a Scintillating Fibre Tracker (SFT). It consists of 2.5 m long modules containing scintillating fibres with a diameter of 250 mm, emitting light when a charged particle passes through them. This light is detected by silicon photomultiplier bundled as read-out boxes, located at the top and bottom of the detector [116].

The higher number of primary vertices is processed by an upgrade of the VELO. Its R and ϕ strips are replaced by pixel sensors consisting of 41 million pixels and the number of modules to is extended to 52. Furthermore, the total data output rate is expected to be 2.8 Tbits^{-1} [117].

Chapter 4

Statistical Tools and Methods

To be able to make a precise and reliable statement in mathematical sciences, the use of statistical methods is indispensable. In this chapter, the main statistical tools and techniques important for this measurement are discussed. A multivariate analysis (MVA) relies on the simultaneous consideration of multiple input observables to provide one or more outcome variables. This technique is applied several times throughout the course of this work, thus it is explained in detail using the examples of a Boosted Decision Tree (BDT) in Sec. 4.1. To estimate parameters like signal yields and decay widths, the Maximum Likelihood Estimation (MLE) method is used and discussed in Sec. 4.2. Furthermore, an introduction to the ${}_s\mathcal{P}$lot technique, a procedure to disentangle different distributions from each other using MLE is given in Sec. 4.3. Finally, this chapter closes with an overview of the Kolmogorov-Smirnov test in Sec. 4.4, which can be used to quantify the compatibility of two distributions.

4.1 Multivariate Analysis

In contrast to univariate analyses, where only one variable is included, MVA use multiple observables, so that possible correlations between the inputs are also taken into account, improving the accuracy of the results. Thus, multivariate methods can be understood as a tool to explain and analyse the joint behaviour of input variables. These methods can be divided into two major processes, structure-discovering and structure-testing. The former is often used to understand and visualise the observables, e.g. in clustering algorithms. No labeled data is required for these structure-discovering processes, which is why it is also referred to as unsupervised learning. The latter method on the other hand computes outcomes that are beneficial in particle physics. Two common structure-testing methods are multivariate regression and multivariate classification. The aim of a regression is to express an observed dependent variable by one or more independent variables. Classification on the other hand is one of the most important tasks in high energy physics, where one example is the separation between signal decays and background. These so-called supervised learning methods rely on labeled data. More information about multivariate analyses especially relevant for particle physics analyses can be found in Ref. [118]. One

application case of a multivariate analysis classification is the Boosted Decision Tree algorithm, which is outlined in the next section.

4.1.1 Boosted Decision Trees

The BDT algorithm relies on either classification or regression trees (CART) [119], whereby in the following only the former will be considered. A classification tree is a binary tree consisting of a root node, internal nodes and at least two leaf nodes. Each connection between a node and sub-node is equivalent to applying a binary decision. The leaf node contains class labels such as *signal* and *background*, paths from root to leaf node are therefore representations of classification rules. A schematic decision tree is illustrated in Fig. 4.1. Since training labels are available, the aim of the decision tree is to maximise the number of correctly classified samples after passing through the decisions chain. One way to measure the quality of a node is to calculate the Gini impurity I_G. Assuming p_i corresponds to the probability of an item to belong to class i, the Gini impurity can be determined as

$$I_G(p) = 1 - \sum_{i=1}^{J} p_i^2, \tag{4.1}$$

with J being the number of classes. In other words, I_G is a measure of how often a randomly chosen item is labeled incorrectly, assuming the item would be labeled randomly according to the overall distribution of labels. Thus, the minimal value of I_G is 0, which can only be reached if a leaf note contains observations of a single class.

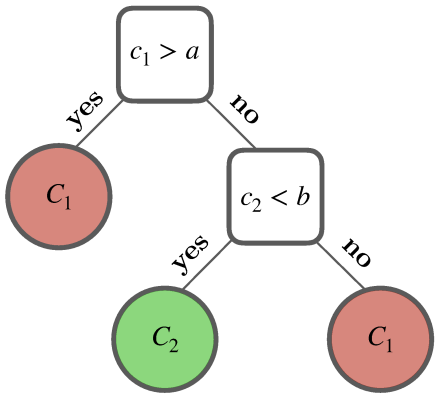


FIGURE 4.1: Representation of a binary decision tree with a root node (top), one internal node and three leaf nodes (circles). Inside the tree, a criterion c_i is applied at each node. The classes are referred to as C_i.

A decision tree with a large enough number of nodes could classify a data sample perfectly. This is not encouraged since the ability to generalise is lost in this case. The behaviour of a classifier performing significantly better on the seen data (training sample) compared to the unseen one (validation sample) is called overfitting. To avoid this problem, multiple trees can be used where the depth of an individual tree is limited. A simple approach would be to take a fixed number of trees, assign each

a randomised sub sample of data and optimise I_G individually. The overall classification of an item to be in a certain class is then the mean decision of all decision trees. This procedure is also known as random forest classification [120]. However, many particle physics analyses have shown that a boosting techniques combined with classification trees as "weak" prediction models show overall better classification results. The most famous boosting methods are adaptive [121] and gradient boosting [122], while only the latter algorithm is relevant for this analysis. The main idea behind gradient boosting is to modify the loss function in order to improve the predicted outcome of observations which are difficult to classify. The following lines provide a short mathematical introduction to gradient boosting [122].

Let $S_{\vec{x},y} = \{(\vec{x}_1, y_1), \cdots, (\vec{x}_n, y_n)\}$ be the training data set containing n observations, a number of variables $\vec{x}$, and corresponding labels y. If the function $F(\vec{x})$ maps $\vec{x}$ to y, the goal is to find an approximation $\hat{F}(\vec{x})$ to minimise the expectation value $\mathbb{E}_{\vec{x},y}$ of a loss function $L(y, F(\vec{x}))$ over $S_{\vec{x},y}$,

$$\hat{F}(\vec{x}) = \underset{F(\vec{x})}{\arg\min}\, \mathbb{E}_{\vec{x},y}[L(y, F(\vec{x}))]. \tag{4.2}$$

The loss function is usually given as quadratic loss, $L(y, F(\vec{x})) = (y - F(\vec{x}))^2$. Gradient boosting allows to replace $\hat{F}(\vec{x})$ by a sum over M decision trees $h(\vec{x})$, each weighted by an integer γ_i, to become

$$\hat{F}(\vec{x}) = \sum_{i=1}^{M} \gamma_i h_i(\vec{x}) + \text{const.} \tag{4.3}$$

Each decision tree, $h_i(\vec{x})$, has a specific number of leave nodes J_m. The input variables can therefore be split in J_m disjoint regions $R_{1m}, \cdots, R_{J_m m}$, whereby a constant value b_{jm} in each region is predicted. Now it is possible to rewrite $h(\vec{x})$ as

$$h_m(\vec{x}) = \sum_{j=1}^{J_m} b_{jm} \mathbf{1}_{R_{jm}}(\vec{x}), \tag{4.4}$$

where $\mathbf{1}_{R_{jm}}(\vec{x})$ is the indicator function,

$$\mathbf{1}_R(x) := \begin{cases} 1 & \text{if } x \in R, \\ 0 & \text{if } x \notin R. \end{cases} \tag{4.5}$$

The actual training process of gradient tree boosting starts initialising $F_0(\vec{x})$ with a constant value and updates every iteration by adjusting γ_m to minimise the loss function,

$$F_0(\vec{x}) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_i, \gamma), \tag{4.6}$$

$$F_m(\vec{x}) = F_{m-1}(\vec{x}) + \gamma_m h_m(\vec{x}) \text{ with} \tag{4.7}$$

$$\gamma_m = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_i, F_{m-1}(\vec{x}_i) + \gamma h_m(\vec{x}_i)). \tag{4.8}$$

A commonly used variation of this procedure is the so called TreeBoost algorithm, which calculates a weight γ_{jm} for each leaf node instead of the whole tree. This makes the prediction of b_{jm} superfluous, changing the above formula to

$$F_m(\vec{x}) = F_{m-1}(\vec{x}) + \sum_{j=1}^{J_m} \gamma_{jm} \mathbf{1}_{R_{jm}}(\vec{x}) \text{ with} \tag{4.9}$$

$$\gamma_m = \underset{\gamma}{\operatorname{argmin}} \sum_{\vec{x}_i \in R_{jm}} L(y_i, F_{m-1}(\vec{x}_i) + \gamma). \tag{4.10}$$

One of the main challenges of supervised learning is to generalise the classifier so that it performs optimal on unknown data samples. A common technique is to split the labeled data in a train and test subset. The classifier will be trained on the first part and validated on the other. To reduce the sensibility on statistical fluctuations of the training subset, most BDT frameworks allow to modify hyperparameters, such as the maximum number of trees or leaves per leaf.

4.2 Maximum Likelihood Estimation

The Maximum Likelihood Estimation (MLE) is commonly used to estimate parameters in particle physics [123]. It is based on a likelihood function $L(\vec{\lambda})$, where $\vec{\lambda}$ is the set of parameters to be estimated. Assuming that the distribution of the data $\vec{x} = \{x_1, \cdots, x_n\}$ corresponds to a properly normalised Probability Density Function (PDF), $P(S_x; \vec{\lambda})$, the likelihood function can be formulated as

$$L(\vec{\lambda}) = \prod_{i=1}^{n} P(x_i; \vec{\lambda}). \tag{4.11}$$

For a specific value of λ_i the result of the likelihood function is the probability to observe x_i. Hence, maximising $L(\vec{\lambda})$ leads to the optimal parameter estimation, if the assumed distribution is equivalent to the underlying distribution. If the number of candidates is also of interest, e.g. when determining the signal yield, an extended form of the MLE is used. A Poisson term is added to the likelihood to become

$$L(\vec{\lambda}) = \nu^n \frac{\mathrm{e}^{-\nu}}{n!} \prod_{i=1}^{n} P(x_i; \vec{\lambda}), \tag{4.12}$$

where $\nu = \nu(\vec{\lambda})$ is the predicted number of candidates. Due to computational reasons, e.g. simplification of the derivation, the likelihood function is often used in its negative and logarithmic form

$$l(\vec{\lambda}) = -\ln\left(L(\vec{\lambda})\right) = \nu - \sum_{i=1}^{n} \ln\left(P(x_i;\vec{\lambda})\nu\right) + \cancel{\ln(n!)}\text{const.} \tag{4.13}$$

This is possible since the monotony of the logarithm keeps the position of the maximum identical. The last part of the equation, $\ln(n!)$, is constant and thus can be ignored in the optimisation process, which has become a minimisation problem of the negative log-likelihood $l(\vec{\lambda})$. In the course of this analysis, the Minuit minimisation framework is used [124]. Besides the parameter estimation, Minuit also provides an uncertainty estimation based on the parabolic minimum of $l(\vec{\lambda})$.

4.3 Determination of ${}_s\mathcal{W}$eights

Even after applying the full selection chain the data samples contain parts of signal and combinatorial background components. To disentangle these distributions from each other the ${}_s\mathcal{P}$lot technique can be used [125]. Prerequisite for this method is a *discriminating* variable x and separate PDFs describing the distributions of each category in x. In addition, x must be uncorrelated to the *control* variables, which can follow any distribution. Examples for control variables in particle physics are kinematic distributions like momenta and pseudo-rapidities. Since the invariant mass is usually not directly correlated to this control variables, it is often considered as discriminating variable. The separation is accomplished by determining per-event weights, or ${}_s\mathcal{W}$eights. Assuming n_c is the number of categories and $P_i(x)$ the PDF for category i, the weights are obtained as

$${}_s w(x_i) = \frac{\sum_{j=1}^{n_c} \boldsymbol{V}_{nj} P_j(x_i)}{\sum_{k=1}^{n_c} N_k P_k(x_i)}. \tag{4.14}$$

The indices j and k correspond to the categories, where N_k represent its yield and $\boldsymbol{V}_{nj}$ the covariance matrix between the yields. The latter one is obtained in a two step procedure. First, an extended maximum likelihood fit is applied to x, resulting in best estimates of the parameters $\vec{\lambda}$. Next, all parameters are fixed except the corresponding yields of each category and a further fit is performed. The ${}_s\mathcal{W}$eights thus obtained are normalised so that the sum of all weights represent the yield in each category. To visualise control variables in histograms of separate categories, the bin uncertainties can be calculated as

$$\sigma = \sqrt{\sum_{i}^{n_i} w_i^2}, \tag{4.15}$$

where n_i is the number of entries per bin.

4.4 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (KS) test statistics [126] can be used to quantify the equality of two distributions. Two versions of this test exists. First, the one-sample KS test compares the empirical distribution function of a sample with a given cumulative probability distribution. For example, this test can be exploited to determine the quality of a fit. The two-sample KS method tests whether two data samples originate from the same distribution. The latter algorithm is especially useful when comparing e.g. MC with data distributions, which becomes necessary in the further analysis, see Sec. 6.3.5. Hence, the formalism of the two-sample KS test is discussed shortly in this section.

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$ be two samples of ordered one-dimensional independent and identically distributed random variables, their empirical distribution function is defined as

$$F_q(p) = \frac{1}{q} \sum_{i=1}^{q} \mathbf{1}_{\{p_i \leq p\}}, \tag{4.16}$$

for $p \in \{X, Y\}$, $q \in \{n, m\}$ and $\mathbf{1}_{\{p_i \leq p\}}$ being the indicator function (Eq. (4.5)). If X and Y share the same probability distribution, the following hypothesis is true:

$$H_0 : F_X(x) = F_Y(y). \tag{4.17}$$

Alternatively, this hypothesis must be rejected and must be replaced by

$$H_1 : F_X(x) \neq F_Y(y). \tag{4.18}$$

The two-sample KS test statistics is given by the expression

$$d_{n,m} = ||F_X(x) - F_Y(y)|| = \underset{\text{sup}}{x, y} |F_X(x) - F_Y(y)|, \tag{4.19}$$

where sup is the supremum. For a given confidence level α and with n and m being the number observations of the samples X and Y, a critical value for $d_{n,m}$ can be calculated, at which the null hypothesis must be rejected. For a small number of observations, a corresponding table of critical values, d_{crit}, can be found in Ref. [127]. However, for larger sample sizes the value of d_{crit} can be approximated as

$$d_{\text{crit}} \approx \sqrt{\frac{1}{2} \ln(\alpha)} \sqrt{\frac{n+m}{nm}}. \tag{4.20}$$

Furthermore, Eq. (4.20) can be resolved to α to directly determine the significance.

Chapter 5

The Decay Channel $B_s^0 \to K_S^0 K_S^0$

The physics motivation to consider $B_s^0 \to K_S^0 K_S^0$ decays is versatile, as described in the following. This chapter discusses the main characteristics and properties of the decay channel. At first, the particles involved in the decay chain are introduced. Afterwards, the decay amplitude calculation of $B_s^0 \to K_S^0 K_S^0$ is motivated using an effective theory. Based on this and other theoretical calculations, important key observables of this decay are enumerated and explained, e.g. the branching ratio ($\mathcal{BR}$) and *CP* violation parameters. A measurement of the $\mathcal{BR}$ has already been performed by the Belle experiment [128] and is discussed in Sec. 5.4.

5.1 Theoretical predictions of the branching ratio

The decay $B_s^0 \to K^0 \overline{K}^0$ can be expressed as a $\overline{b} \to \overline{s} d \overline{d}$ transition at the quark level. Since the $\overline{b}$ and $\overline{s}$ have the same charge, a transition between them via exchange of a $W^{\pm}$ boson is not possible. Hence, the process $\overline{b} \to \overline{s}$ can only occur via flavour changing neutral currents (FCNC). FCNCs only happen through loop diagrams in the SM, i.e. a process with at least four vertices, which are forbidden at tree level. The dominating Feynman diagram is shown in Fig. 5.1, whereby other loop contributions are at a level of below 5 % [129]. In Tab. 5.1 the branching ratio calculations from different theoretical models are shown. The different predictions for $B_s^0 \to K^0 \overline{K}^0$ decays agree within errors and yield a value of about 20×10^{-6}. However, the branching ratio prediction for $B_s^0 \to K_S^0 K_S^0$ decays differs by a factor of two, where the factor is explained in the following.

By converting the flavour eigenstates $K^0(d\overline{s})$ and $\overline{K}^0(\overline{d}s)$ into the weak eigenstates K_S^0 and K_L^0, four different physical kaon mode combinations are plausible: $K_S^0 K_S^0$, $K_S^0 K_L^0$, $K_L^0 K_S^0$ and $K_L^0 K_L^0$. Applying the *CP* transformation to the flavour eigenstates yield a positive (or even) *CP* eigenvalue,

$$CP(K^0 \overline{K}^0) = (-\overline{K}^0) \cdot (-K^0) \cdot (-1)^L = +K^0 \overline{K}^0, \tag{5.1}$$

since the angular momentum L between K^0 and $\overline{K}^0$ is 0. The same *CP* eigenvalue must be achieved for the weak eigenstates of the kaons. For each combination, the *CP* eigenvalue can be determined to be

$$CP(K_S^0 K_S^0) \approx CP(K_1^0 K_1^0) = +K_1^0 K_1^0, \tag{5.2}$$

$$CP(K_S^0 K_L^0) = CP(K_L^0 K_S^0) \approx CP(K_1^0 K_2^0) = -K_1^0 K_2^0 \quad \text{and} \tag{5.3}$$

$$CP(K_L^0 K_L^0) \approx CP(K_2^0 K_2^0) = +K_2^0 K_2^0, \tag{5.4}$$

where $K^0_{S(L)} \stackrel{!}{=} K^0_{1(2)}$ when assuming CP conservation (see also Sec. 2.3.2). Thus, only the former and latter decay mode are physically allowed and equally likely, so that the branching ratio of $B_s^0 \to K_S^0 K_S^0$ is exactly half of $B_s^0 \to K^0 \overline{K}^0$, i.e. about 10×10^{-6}.

A common method to determine the branching ratio theoretically is based on an effective field theory (EFT). An in depth discussion of this formalism would exceed the scope of this work, hence only the idea of these calculations is given below. The interested reader shall be referred to Refs. [22, 130, 131].

TABLE 5.1: Theoretical predictions of the branching ratio in $B_s^0 \to K^0 \overline{K}^0$. The calculation of the $\mathcal{BR}$ was performed with different models: Soft Collinear Effective Theory (SCET), QCD factorisation (QCDF) and pertubative QCD (pQCD). The latter one is further divided into leading order (LO) and next-to-leading order (NLO) calculations.

Model	$\mathcal{BR}$ prediction (in 10^{-6})	Source
SCET	17.7 ± 6.6	[19]
QCDF	$24.7 \pm {}^{29.3}_{14.0}$	[20]
pQCD (L0)	$15.6 \pm {}^{9.7}_{6.0}$	[21]
pQCD (NL0)	$20.2 \pm {}^{6.9}_{5.4}$	[22]

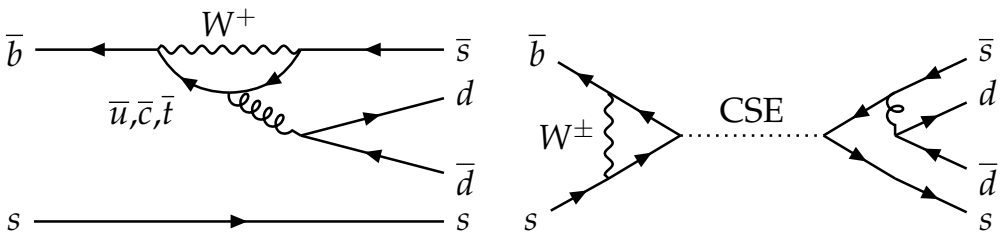


FIGURE 5.1: Possible Feynman diagrams for the $B_s^0 \to K^0 \overline{K}^0$ decay channel. The left figure shows the dominant contribution, a $b \to s$ penguin diagram (P). On the right side, a penguin annihilation diagram (PA) is depicted. The abbreviation CSE means colour singlet exchange and can be represented by e.g. a Z^0 boson or at least two gluons, since both, the initial and final state are colour neutral.

An EFT combined with an operator product expansion (OPE) approach can be used to determine the decay amplitude $\mathcal{A}_{B_s^0 \to K^0 \overline{K}^0}$. First, an effective Hamiltonian for $b \to s$ transitions is constructed

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} \left\{ \lambda_u \left[C_1(\mu)\mathcal{O}_1^u(\mu) + C_2(\mu)\mathcal{O}_2^u(\mu) \right] - \lambda_t \left[\sum_{i=3}^{10} C_i(\mu)\mathcal{O}_{i(\mu)} \right] \right\} + \text{h.c.} \quad (5.5)$$

In this formula, G_F is the Fermi constant, C_i are the Wilson coefficients and $\mathcal{O}_i$ is the four-fermion operator, both evaluated at the renormalisation scale μ. The use of a renormalisation scale allows to separate short distance (operators) from long distance (coefficient) effects, where large values correspond to high masses which in turn could be interpreted as a short distance. Hence, hadronic effects and short-distance QCD/QED corrections are described by the former, electro-weak interactions and possible New Physics by the latter. The Wilson coefficients could also be seen as effective coupling strengths for the corresponding operator. Additionally, λ_q describes the CKM factors. The operators can be classified in three classes and are defined as:

Current–current/tree operators

$$\mathcal{O}_1^u = (\overline{u}_\alpha b_\beta)_{V-A}(\overline{q}_\beta u_\alpha)_{V-A}$$

$$\mathcal{O}_2^u = (\overline{u}_\alpha b_\alpha)_{V-A}(\overline{q}_\beta u_\beta)_{V-A}$$

QCD penguin operators

$$\mathcal{O}_3 = (\overline{q}_\alpha b_\alpha)_{V-A} \sum_{q' \neq t} (\overline{q}'_\beta q'_\beta)_{V-A}$$

$$\mathcal{O}_4 = (\overline{q}_\beta b_\alpha)_{V-A} \sum_{q' \neq t} (\overline{q}'_\alpha q'_\beta)_{V-A}$$

$$\mathcal{O}_5 = (\overline{q}_\alpha b_\alpha)_{V-A} \sum_{q' \neq t} (\overline{q}'_\alpha q'_\alpha)_{V+A}$$

$$\mathcal{O}_6 = (\overline{q}_\beta b_\alpha)_{V-A} \sum_{q' \neq t} (\overline{q}'_\alpha q'_\beta)_{V+A}$$

Electroweak penguin operators

$$\mathcal{O}_7 = \frac{3}{2}(\overline{q}_\alpha b_\alpha)_{V-A} \sum_{q' \neq t} e_{q'} (\overline{q}'_\beta q'_\beta)_{V+A}$$

$$\mathcal{O}_8 = \frac{3}{2}(\overline{q}_\beta b_\alpha)_{V-A} \sum_{q' \neq t} e_{q'} (\overline{q}'_\alpha q'_\beta)_{V+A}$$

$$\mathcal{O}_9 = \frac{3}{2}(\overline{q}_\alpha b_\alpha)_{V-A} \sum_{q' \neq t} e_{q'} (\overline{q}'_\beta q'_\beta)_{V-A}$$

$$\mathcal{O}_{10} = \frac{3}{2}(\overline{q}_\beta b_\alpha)_{V-A} \sum_{q' \neq t} e_{q'} (\overline{q}'_\alpha q'_\beta)_{V-A}.$$

Here, α and β are the colour indices, q' cover all quark types except t-quark and $V(A)$ describes the vector (axial-vector) current. For $B_s^0 \to K_S^0 K_S^0$ decays, QCD penguin operators are dominating. A common way to take advantage of the similar structure of operators is to define electro-weak dynamical variables a_i, which only depend on constants and the Wilson coefficients. The decay amplitude $\mathcal{A}(B_s^0 \to K^0\overline{K}^0)$ can be formulated as

$$\mathcal{A}(B_s^0 \to K^0\overline{K}^0) = \left\langle K^0\overline{K}^0 \,\middle|\, \mathcal{H}_{\text{eff}} \,\middle|\, B_s^0 \right\rangle, \quad (5.6)$$

where the exact formulation of $\mathcal{A}(B_s^0 \to K^0\bar{K}^0)$ can be expressed as a sum over factorisable form factors and non-factorizable matrix elements, itself depending on the dynamical variables [130]. Finally, the branching fraction can be calculated using the equations

$$\Gamma(B_s^0 \to K^0\bar{K}^0) = \frac{G_F^2 m_{B_s^0}^3}{32\pi}|\mathcal{A}(B_s^0 \to K^0\bar{K}^0)|^2 \quad \text{and} \tag{5.7}$$
$$\mathcal{BR}(B_s^0 \to K^0\bar{K}^0) = \frac{\Gamma(B_s^0 \to K^0\bar{K}^0)}{\Gamma(B_s^0)},$$

with $m_{B_s^0}$ being the mass of the B_s^0 meson. Summarising, a precise measurement of $\mathcal{BR}(B_s^0 \to K_S^0 K_S^0)$ could constraint the values of dynamical variables and thus its corresponding Wilson coefficients.

5.2 Decay chain properties

The mother particle of the decay chain $B_s^0 \to K^0\bar{K}^0$ is the B_s^0 meson, consisting of a $\bar{b}s$ quark pair. Its mass is around 5360 GeV[1], which is strongly dominated by the b-quark, i.e. $m_b = 4.18^{+0.04}_{-0.03}$ GeV [38]. The K_S^0 meson on the other hand is a superposition of K^0 and $\overline{K^0}$, hence has the valence quarks $\frac{d\bar{s}-\bar{d}s}{\sqrt{2}}$ and a mass of about 490 MeV. Both aforementioned mesons carry a spin of 0 and have odd parity ($J^P = 0^-$), expressing them by definition as pseudo-scalar particles. The lifetime of K_S^0 mesons is about 60 times longer than that of B_s^0 hadrons, whereby a special treatment in the reconstruction of K_S^0 mesons is required in the further analysis, see Sec. 6.1. These mesons were reconstructed as a pair of two oppositely charged pions, which have a significantly higher lifetime due to the favoured weak decay. All particle properties are summarized in Tab. 5.2.

TABLE 5.2: Information about the particles appearing in the decay chain $B_s^0 \to K_S^0 K_S^0$. All values are taken from [38].

Particle	Mass m [MeV]	Lifetime τ [s]	Quark content	$I(J^P)$
B_s^0	5366.77 ± 0.24	$1.519 \pm 0.005 \times 10^{-15}$	$s\bar{b}$	$0(0^-)$
K_S^0	497.614 ± 0.024	$89.54 \pm 0.04 \times 10^{-12}$	$\approx \frac{d\bar{s}-\bar{d}s}{\sqrt{2}}$	$1/2(0^-)$
$\pi^\pm$	$139.570\,18 \pm 0.000\,35$	$2.6 \pm 0.1 \times 10^{-9}$	$u\bar{d}/d\bar{u}$	$1(0^-)$

5.3 Physics observables in $B_s^0 \to K^0\bar{K}^0$ decays

Besides the branching ratio of $B_s^0 \to K^0\bar{K}^0$, the measurement of other physics observables might be interesting. This is especially the case for future LHC upgrades and

[1] In the following chapters natural units are used, i.e. $c = \hbar = 1$. This leads to impulses and masses having the same unit eV.

other B factories, entailing higher statistics and better understanding of systematic effects. Refs. [18, 132] suggest to measure the $B_s^0 - \bar{B}_s^0$ mixing angle β_s^{eff}, which is discussed in the following.

In SM calculations, β_s is predicted to be roughly zero, since the imaginary phase entering in V_{ts} is strongly suppressed, see Eq. (2.20) and Eq. (2.14). Hence a measurement of a non-zero value would indicate NP effects. At the time of writing this thesis, a combination of different measurements provided by the PDG yield

$$\beta_{s,\,\text{PDG}} = (1.1 \pm 1.6) \times 10^{-2}\,\text{rad}, \tag{5.8}$$

which is dominated by LHCb measurements. Evaluating the Feynman diagrams given in Fig. 5.1, it is possible to deduce β_s in the $B_s^0 \to K^0\bar{K}^0$ decays. As the penguin annihilation diagram PA is suppressed by more than one order of magnitude (and to simplify the calculation), only the penguin diagram, P, is taken into account. In the diagram, three penguin contributions have to be considered,

$$P = \lambda_u^s P_u' + \lambda_c^s P_c' + \lambda_t^s P_t', \tag{5.9}$$

where P_q expresses the penguin loop (the single quote signals the $\Delta s = 1$ transition) and λ_q^s the CKM factors, $\lambda_q^{q'} = V_{qb}^* V_{qq'}$. The unitarity properties of the CKM matrix defined in Sec. 2.2 allow to omit one summand because of

$$\lambda_u^s + \lambda_c^s + \lambda_t^s = 0 \quad \leftrightarrow \quad \lambda_c^s = -\lambda_u^s - \lambda_t^s. \tag{5.10}$$

This leads to an alternative expression of Eq. (5.9),

$$P = \lambda_u^s (P_u' - P_c') + \lambda_t^s (P_t' - P_c'). \tag{5.11}$$

The decay amplitude $\mathcal{A}(B_s^0 \to K^0\bar{K}^0)$ now only depends on P,

$$\mathcal{A}(B_s^0 \to K^0\bar{K}^0) = \exp(i\gamma)|\lambda_u^s|P_{uc}' - \exp(i\beta_s)|\lambda_t^s|P_{tc}'. \tag{5.12}$$

In the formula above, the definition of $P_{q^1q^2} \equiv P_{q^1} - P_{q^2}$ is used the CKM angle dependences are explicitly stated, see also Eq. (2.14). The CP conjugate decay amplitude $\overline{\mathcal{A}}(B_s^0 \to K^0\bar{K}^0)$ can be expressed the same way by changing the signs in front of the exponents. Given these amplitudes it is now possible to phrase the measurements (M) of the branching ratio, the direct CP violation and the CP violation in mixing as

$$M_{\mathcal{BR},\,B_s^0\to K^0\bar{K}^0} \propto \frac{1}{2}\left(|\mathcal{A}(B_s^0 \to K^0\bar{K}^0)|^2 + |\overline{\mathcal{A}}(B_s^0 \to K^0\bar{K}^0)|^2\right), \tag{5.13}$$

$$M_{\text{Direct } CPV,\,B_s^0\to K^0\bar{K}^0} \propto \frac{1}{2}\left(|\mathcal{A}(B_s^0 \to K^0\bar{K}^0)|^2 - |\overline{\mathcal{A}}(B_s^0 \to K^0\bar{K}^0)|^2\right), \tag{5.14}$$

$$M_{\text{Mixing } CPV,\,B_s^0\to K^0\bar{K}^0} \propto \text{Im}\left(\exp(-2i\beta_s)\mathcal{A}(B_s^0 \to K^0\bar{K}^0)\overline{\mathcal{A}}(B_s^0 \to K^0\bar{K}^0)\right). \tag{5.15}$$

These equations yield seven observables, $|\lambda_u^s|$, $|\lambda_t^s|$, γ, P'_{uc}, P'_{tc}, β_s and δ', which is the strong phase between the magnitudes of P'_{qc}. The values of the first three observables are already well known from other independent measurements.

Comparing the absolute sizes of the two components entering Eq. (5.12), the influence of the term $|\lambda_u^s|P'_{uc}$ is expected to be small. First, the factor $|\lambda_u^s|$ yield a high CKM suppression of λ^4. Secondly, the GIM mechanism [133] suppresses the contribution of P'_{uc}, which is proportional to the so-called Inami-Lim functions $S(x_q) \approx \mathcal{O}(x_q)$ [134]. The parameter x_q is the squared fraction of an up-type quark and the $W^\pm$ boson mass $x_q = (m_q/m_{W^\pm})^2$. Since $m_u^2, m_c^2 \ll m_t^2$ and $P'_{uc} = P'_u - P'_c$, the penguin contributions cancel out to a certain degree for P'_{uc}, while the difference is larger in case of P'_{tc} due to the heavier t quark mass. Hence, assuming that $|\lambda_u^s|P'_{uc}$ is negligible would yield a total of two measurable observables, P'_{tc} and β_s. These values can be extracted from Eq. (5.13) and Eq. (5.15), while Eq. (5.14) vanishes. However, this simplified approach leads to a high theoretical uncertainty on the angle β_s and could overshadow NP effects in an experimental measurement.

One possible way to constraint P'_{uc} is to measure the same observables in the U-spin [2] partner decay $B^0 \to K^0\overline{K}^0$, a $\overline{b} \to \overline{d}s\overline{s}$ penguin transition. The decay amplitude of this decay is given as

$$\mathcal{A}(B^0 \to K^0\overline{K}^0) = \exp(i\gamma)|\lambda_u^d|P_{uc} - \exp(-i\beta)|\lambda_t^d|P_{tc}. \tag{5.16}$$

Under the SU(3) symmetry, the values of P_{tc} and P'_{tc} are equal, except the SU(3) breaking factor, which also increases the theoretical uncertainty of β_s.

The decay $B^0 \to K^0\overline{K}^0$ has not been measured to a great accuracy yet. The theoretical predictions described in this section hold also, besides minor corrections, for the excited states of the kaons, i.e. $B_{s,d}^0 \to K^0\overline{K}^{*0}$ and $B_{s,d}^0 \to K^{*0}\overline{K}^{*0}$.

5.4 Previous Measurements

When determining a branching ratio for a rare decay it is common to specify a significance σ expressed as number of Gaussian standard deviations, which represents the p-value of the background only hypothesis. The first evidence of the decay $B_s^0 \to K^0\overline{K}^0$ was found in an analysis performed by the Belle collaboration in 2010 [135], which searched for B_s^0 decays to two hadron decays. Due to the low statistics the significance of this decay was only 1.2 standard deviations. Nevertheless, an upper limit of $\mathcal{BR}(B_s^0 \to K^0\overline{K}^0) < 6.6 \times 10^{-5}$ could be measured. In 2016, Belle published another analysis focussing on the decay $B_s^0 \to K_S^0 K_S^0$ [136]. The measurement yields $29^{+8.5}_{-7.6}$ signal candidates at an significance of 5.1 standard deviations. The branching ratio could be determined to

[2] The U-spin technique describes the exchange of all s and d quarks in a given decay mode.

$$\mathcal{BR}(B_s^0 \to K^0\bar{K}^0) = (19.6^{+5.8}_{-5.1} \pm 1.0 \pm 2.0) \times 10^{-6}, \tag{5.17}$$

where the first uncertainty is statistical (dominated by the final mass fit), the second is systematic and the third is due to the total number of $B_s^0\bar{B}_s^0$ pairs.

The U-spin partner decay $B^0 \to K^0\bar{K}^0$ has also been measured at e^+e^- collider experiments. The most precise measurements are coming from the BaBar [137] and Belle [45] collaborations. A combined result of the branching fractions yield

$$\mathcal{BR}(B^0 \to K^0\bar{K}^0) = (1.21 \pm 0.16) \times 10^{-6}, \tag{5.18}$$

which is more than an order of magnitude smaller than in the case of the B_s^0 decay. In the same decay mode, a time-dependent CP-violation measurement has been performed [138].

The similar $P \to PV$ decay $B_s^0 \to K^0\bar{K}^{*0}$ has been observed in a LHCb measurement [139]. One challenge to determine the branching ratio is to disentangle the P-wave, i.e. the K^{*0} meson, from the non-resonant $K\pi$ meson pair S-wave contributions. The analysis determined a branching ratio of

$$\mathcal{BR}(B_s^0 \to \bar{K}^0K^{*0}) + \mathcal{BR}(B_s^0 \to K^0\bar{K}^{*0}) = (16.4 \pm 3.4 \pm 2.3) \times 10^{-6}, \tag{5.19}$$

where the first uncertainty is statistical and the second is systematic.

Chapter 6

Branching ratio measurement of $B_s^0 \to K_S^0 K_S^0$ decays

This chapter describes the search for $B_s^0 \to K_S^0 K_S^0$ decays as well as the determination of its branching ratio. The corresponding data was recorded in the years 2011, 2012, 2015 and 2016 and correspond to 5 fb^{-1} of integrated luminosity. Besides the signal channel $B_s^0 \to K_S^0 K_S^0$, the decay channel $B^0 \to \phi K_S^0$ has been chosen for normalisation. To reduce a possible bias of the analyst, the measurement is performed blinded. This means, that the signal region of the invariant B_s^0 mass is cut out throughout the analysis until the selection and systematics effects are fully understood.

The datasets relevant for this analysis are explained in Sec. 6.1. Next, the normalisation procedure as well as external inputs are described in Sec. 6.2. A huge amount of work has been spend on the extraction of signal candidates. Hence, the signal selection, elaborated in Sec. 6.3, is divided into several parts. To extract the signal yield, a special fit model is developed, which is further described in Sec. 6.4 and whose results are given in Sec. 6.5. A similar signal extraction procedure is applied for the decay channel $B^0 \to \phi K_S^0$, also explained in these sections. The signal efficiencies are discussed in Sec. 6.6. This chapter closes with the summary of systematic uncertainties as well as the significance determinations in Sec. 6.8 and Sec. 6.9, respectively.

6.1 Datasets and decay topologies

The whole dataset used in this analysis is split into several smaller subsets, whereby the analysis is performed separately in each part. To ensure that the distributions inside a data subset are as homogeneous as possible, the division is made by centre-of-mass energies and track types. As discussed in Sec. 3.3, the data collecting conditions differ between the years. While the difference between $\sqrt{s} = 7$ and 8 TeV are small so that the samples between 2011 and 2012 can be unified to a dataset named Run I, the higher energies in Run II of $\sqrt{s} = 13$ TeV and changes in the stripping procedure yield different distributions. The different track type possibilities exist due to the long lifetime of the K_S^0 mesons. The fraction of K_S^0 candidates decaying inside or rather outside the VELO is roughly $1/3$ to $2/3$. Throughout this analysis, only K_S^0 candidates reconstructed from two oppositely charged pions of the same

track type, i.e. two long track or two downstream pion candidates, were considered. Therefore, the K_S^0 candidates are divided into long track or downstream K_S^0 mesons, abbreviated as L and D, respectively. Since the decay $B_s^0 \to K_S^0 K_S^0$ consists of two K_S^0 candidates, the reconstructed B_s^0 candidate has the possibility to consist of LL, LD or DD K_S^0 mesons. Downstream tracks have a lower momentum resolution which propagates to the invariant mass

$$m_0 = \sqrt{E^2 - |\vec{p}|^2}, \tag{6.1}$$

resulting into a broader mass peak. The invariant mass distributions of the B_s^0 candidate as well as K_S^0 candidate is shown in Fig. 6.1. The ϕ candidate in the normalisation channel $B^0 \to \phi K_S^0$ has a relatively short lifetime of about 10^{-22} s, hence decaying into long track kaons. Thus, this decay mode is only reconstructed as LL and LD.

Although the $B_s^0 \to K_S^0 K_S^0$ DD dataset contains the highest number of signal candidates, it is omitted in this analysis. Up to the year 2016 no high level trigger line existed performing a decision purely based on downstream tracks. As all events recorded must have pass a certain trigger, a possible trigger selection would be based on the underlying event. Thus, a combination of all physical trigger lines which have triggered independent of the signal candidate (TIS) can be used. However, this information suffers from a bad description in MC and therefore leads to high uncertainties when determining the trigger efficiencies. Secondly, no suitable normalisation channel exists for such kind of decays, further increasing the expected uncertainty of the $\mathcal{BR}$. Finally, the increased number of combinatorial background events and poor resolution of physics observables like the B_s^0 decay vertex would reduce the performance of a MVA needed to separate signal from background events, hence increasing the statistical uncertainty in the later mass fit. Concluding, the total dataset is separated into four parts: Run I, LL, Run I, LD, Run II, LL and Run II, LD, respectively.

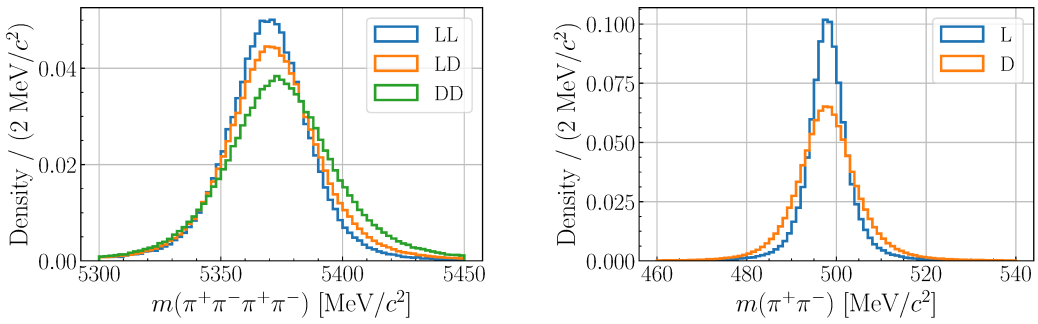


FIGURE 6.1: Comparison of invariant mass distributions calculated from different track types. On the left (right), the B_s^0 (K_S^0) mass taken from $B_s^0 \to K_S^0 K_S^0$ signal MC is shown. The mass peaks broadens, the more downstream tracks are involved in the particle reconstruction.

Besides detector data, this analysis relies on the usage of MC samples for the signal, normalisation and possible background contribution channels, which is a topic of a later section. Fully simulated MC includes all, the particle generation, decay and detector response, whereby the detailed production procedure is given in Sec. 3.2.4. It comes with the downside of being computing intensive and consuming a lot of storage space. To ensure the reliability of these productions, requests must pass through several validation steps and are produced centrally by the collaboration. Besides fully simulated, two other types of MC are used throughout the analysis. For studies where only the mass shape of a decay's distribution is needed, it is sufficient to only simulate the generation and decay of the particles. To accomplish this, the framework RapidSim [140] is used. It applies similar algorithms as in the nominal MC production chain, but omits the interaction with the detector, speeding up the production time. Toy Monte Carlo (ToyMC) datasets, i.e. extremely simplified pseudo-experiments, can be used to simulate a specific aspect of the analysis which is often used extensively in the determination of systematic uncertainties.

6.2 Determination of the branching ratio

The branching ratio ($\mathcal{BR}$) describes the fraction of a particle decaying into an specific decay mode with respect to all possible decay modes. A naive formulation of the $\mathcal{BR}$ in the decay channel $B_s^0 \to K_S^0 K_S^0$ can be written as

$$\mathcal{BR}(B_s^0 \to K_S^0 K_S^0) = \frac{N(B_s^0 \to K_S^0 K_S^0)}{N_{B_s^0}} = \frac{N(B_s^0 \to K_S^0 K_S^0)}{2\mathcal{L}_{\text{int}} \sigma_{b\bar{b}} f_s \varepsilon_{K_S^0 K_S^0}}, \tag{6.2}$$

where $N(B_s^0 \to K_S^0 K_S^0)$ is the signal yield, $N_{B_s^0}$ is the total number of B_s^0, $\mathcal{L}_{\text{int}}$ the integrated luminosity recorded, $\sigma_{b\bar{b}}$ the $b\bar{b}$ production cross section, f_s the probability for a b quark to hadronise with a s quark and $\varepsilon_{K_S^0 K_S^0}$ the total selection efficiencies for the decay channel. The factor of 2 raises from the generation process of b quarks, which are produced in pairs (b and $\bar{b}$). Furthermore, this analysis does not differentiate between B_s^0 and $\bar{B}_s^0$ quark decays. Since the uncertainties of $\sigma_{b\bar{b}}$ are relatively large, a precise determination of the $\mathcal{BR}$ is not possible with this approach. Thus, the normalisation channel is used to estimate the total number of B decays and the measurement is done relatively to the normalisation decay,

$$\mathcal{BR}(B_s^0 \to K_S^0 K_S^0) = \left(\cancel{\frac{2\mathcal{L}_{\text{int}}\sigma_{b\bar{b}}}{2\mathcal{L}_{\text{int}}\sigma_{b\bar{b}}}}\right) \frac{\varepsilon_{\phi K_S^0}}{\varepsilon_{K_S^0 K_S^0}} \frac{f_d}{f_s} \frac{\mathcal{BR}(B^0 \to \phi K_S^0)}{N(B^0 \to \phi K_S^0)} N(B_s^0 \to K_S^0 K_S^0). \tag{6.3}$$

In addition to the advantage of being independent of $\sigma_{b\bar{b}}$, this method also reduces systematic uncertainties in the efficiency determination since they cancel out in the ratio. The final state particles of the measurement are four pions (two kaons and two pions) for the signal (normalisation) channel. Hence, the $\mathcal{BR}$ of the decays

$K_S^0 \to \pi^+\pi^-$ and $\phi \to K^+K^-$ must also be taken into account, modifying the aforementioned formula to

$$\mathcal{BR}(B_s^0 \to K_S^0 K_S^0) = \underbrace{\frac{\varepsilon_{\phi K_S^0}}{\varepsilon_{K_S^0 K_S^0}} \frac{f_d}{f_s} \frac{\mathcal{BR}(\phi \to K^+K^-)}{\mathcal{BR}(K_S^0 \to \pi^+\pi^-)} \frac{\mathcal{BR}(B^0 \to \phi K_S^0)}{N(B^0 \to [K^+K^-]_\phi [\pi^+\pi^-]_{K_S^0})}}_{=\alpha} \cdot N(B_s^0 \to [\pi^+\pi^-]_{K_S^0}[\pi^+\pi^-]_{K_S^0}), \tag{6.4}$$

where a major part of this equation is summarised in the so-called normalisation constant α.

6.2.1 External inputs

Not all parameters of Eq. (6.4) are determined in the present analysis. Thus, external input from previous measurements is needed. The mother particle of the signal and normalisation channel consist of different b mesons, which have individual fragmentation fractions. The ratio of f_s/f_d has been measured by the LHCb collaboration [141–143] depending on different centre-of-mass energies, resulting in values of

$$\frac{f_s}{f_d}(\sqrt{s} = 7\,\text{TeV}) = 0.259 \pm 0.015 \quad \text{and} \tag{6.5}$$

$$\frac{f_s}{f_d}(\sqrt{s} = 13\,\text{TeV}) = 0.244 \pm 0.012. \tag{6.6}$$

For the latter measurement, it is assumed that the hadronisation fractions of d and u (f_d and f_u) quarks are the same. In this analysis, the $\sqrt{s} = 7\,\text{TeV}$ measurement is used as input for the Run I data samples, while the result of the $\sqrt{s} = 13\,\text{TeV}$ analysis is used for Run II data.

Furthermore, the $\mathcal{BR}$'s of the intermediate states are taken from the summary of the particle data group (PDG) [44], which provides combined values of different measurements of the same decay channel. The following values are used:

$$\mathcal{BR}(B^0 \to \phi K_S^0)_{\text{PDG}} = (7.3 \pm 0.7) \times 10^{-6}, \tag{6.7}$$

$$\mathcal{BR}(\phi \to K^+K^-)_{\text{PDG}} = 0.492 \pm 0.005, \tag{6.8}$$

$$\mathcal{BR}(K_S^0 \to \pi^+\pi^-)_{\text{PDG}} = 0.6920 \pm 0.0005. \tag{6.9}$$

6.3 Signal extraction

To minimize the statistical uncertainty of a branching ratio measurement, a high suppression of background decays is mandatory and therefore a major part of this

analysis. The signal candidates are reconstructed in the stripping, further elaborated in Sec. 6.3.1. Next, preselection cuts are applied and described in Sec. 6.3.2, introduced to remove distinct background candidates while being very efficient on the signal. Furthermore, the trigger lines chosen for this analysis are discussed in Sec. 6.3.3. The possible influences of various background decays is examined in Sec. 6.3.4. After applying this selection chain, only combinatorial background is left, which is suppressed by applying a multivariate analysis (MVA) as described in Sec. 6.3.5.

6.3.1 Event reconstruction

In the stripping, the tracks reconstructed in the detector are combined to build the decay of interest. Furthermore, decay chains (or decay trees), i.e. decays where intermediate resonances are present, are constructed. A full topology of the decay $B_s^0 \to K_S^0 K_S^0$ is given in Fig. 6.2. The default strategy to build decay chains at LHCb is by using a bottom-up method. First of all, two charged pion candidates are combined to a K_S^0 candidate using fits, determining kinematic properties and a common decay vertex. The two intermediate K_S^0 are then unified in the same manner to build a B_s^0 candidate. Having access to the fit results allows to omit candidates at a very early stage in the event reconstruction. Thus, to reduce the data sample size to a manageable level, loose cuts were applied directly in the stripping. Each pion is required to have a momentum greater than 2 GeV, while the transverse momentum of long track pions has additionally to be greater than 250 MeV. The minimum χ^2 distance of the pions trajectory to the primary vertex (PV) must be > 9 for long track pions and > 4 for downstream pions, respectively. Due to the higher event multiplicity in Run II, the pion tracks have to pass two further requirements. The track χ^2 per degree of freedom has to be > 4 and the probability of the track to be a ghost, i.e. artificial noise in the detector, should be below 50 %. While combining the two long track (downstream) pions to a L (D) K_S^0 meson, the invariant mass has to be in a 35 MeV (64 MeV) mass window around the nominal K_S^0 mass. The end vertex quality and the χ^2 of the distance of closest approach (DOCA) must both be less than 25, independent of the track type. In Run I, the absolute distance between the two pion tracks also has to be $< 1\,\mathrm{mm}(4\,\mathrm{mm})$ for L (D) K_S^0 candidate. Furthermore, the χ^2 distance from the related PV must be greater than 5 in Run II stripping. The invariant mass of the combination of two K_S^0 mesons has to be in a wide mass window around the nominal B_s^0 mass. This includes mass ranges with only combinatorial background contributions, which is needed as input sample for the MVA, described in Sec. 6.3.5. Like the K_S^0 candidate, the distance of closest approach must be below $< 1\,\mathrm{mm}(4\,\mathrm{mm})$ for the combination of LL (LD) kaons. One difference between the Run I and Run II stripping for the $B_s^0 \to K_S^0 K_S^0$ decays is the usage of an MVA classifier in the Run I stripping, consisting of a random forest classifier with ten trees. As input observables, the impact parameter significance of the best PV, the χ^2 separation from related PV, the child p_T's and the child's track type were used.

The B_s^0 candidates are then required to possess a classifier output of > 0.5, which corresponds to excellence signal efficiency while suppressing a fair amount of background. The stripping line cuts are also summarised in Tab. 6.1.

Decay Tree Fitter

The aforementioned bottom-up approach, also known as leaf-by-leaf fitting, comes with the disadvantage that constraints upstream of the decay vertex do not add information to the parameters of the vertex. Therefore, another method has been developed, which enables a simultaneous fit of the whole decay tree. This method is called decay tree fitter (DTF) [144]. In a global fit, the parameters vertex position, momenta and decay times of all intermediate and final state particles are determined. In addition to internal constraints such as momentum conservation at each vertex, external constraints can be defined. For example, it is possible to constraint the mass of an intermediate particle, reducing track resolution effects of the detector. To reduce computational resources, the χ^2 minimization method is performed by a Kalman filter [145]. In the scope of this work, the DTF is particular used to determine the invariant mass observables. A study on the mass resolution as function of different constraints is given in Sec. 6.5.1.

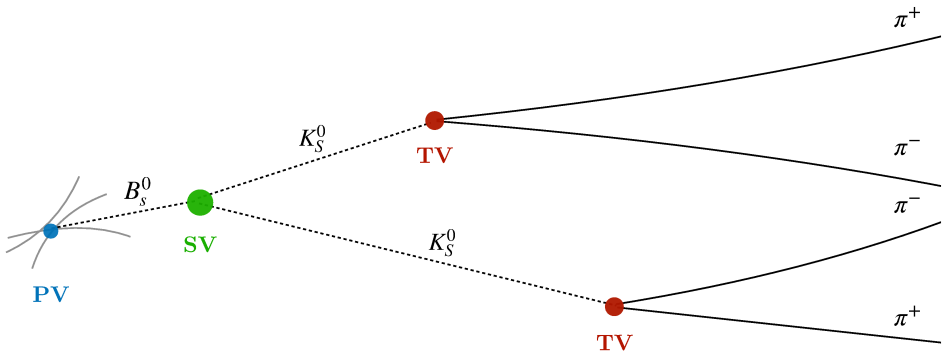


FIGURE 6.2: Schematic overview of the decay $B_s^0 \to K_S^0 K_S^0$. Only the final state particles (pions) leave hits in the tracking stations and are reconstructed in the detector, represented as solid lines. Trajectories of neutral resonances like the B_s^0 and K_S^0 mesons are presented as dashed lines.

TABLE 6.1: Stripping line requirements for $B_s^0 \to K_S^0 K_S^0$ decays. The observables are the total momentum, p, the transverse momentum, p_T, the significance of the impact parameter(IP), χ^2_{IP}, the track quality divided by the number of degrees of freedom, $\chi^2_{track}/\text{ndf}$, the probability of a track to be a ghost particle, ProbNN Ghost, the invariant mass of the π^+ and π^- (two K_S^0) candidates, $m_{\pi^+\pi^-}$ ($m_{K_S^0 K_S^0}$), the endvertex quality, χ^2_{vtx}, the distance of closest approach (DOCA) and its significance, χ^2_{DOCA}, and finally an output of a boosted decision tree classifier. The column "Data Samples" comprises the data sample on which the requirement is applied to.

Decay	Particle	Observable	Requirements	Data Samples
$K_S^0 \to \pi^+\pi^-$ (long track)	$\pi^\pm$	p	> 2 GeV	
		p_T	> 250 MeV	
		$\min_{\chi^2_{IP}}(PV)$	> 9	
		$\chi^2_{track}/\text{ndf}$	< 4	Run II only
		ProbNN Ghost	< 0.5	Run II only
	K_S^0	$m_{\pi^+\pi^-}$	$\lvert m_{\pi^+\pi^-} - m_{K_S^0,\,PDG}\rvert < 35$ MeV	
		χ^2_{vtx}	< 25	
		χ^2_{DOCA}	< 25	
		$\max_{DOCA}$	< 1 mm	Run I only
		χ^2_{IP}	> 5	Run II only
$K_S^0 \to \pi^+\pi^-$ (downstream)	$\pi^\pm$	p	> 2 GeV	
		χ^2_{IP}	> 4	
		$\min_{\chi^2_{IP}}(PV)$	> 4	Run II only
		$\chi^2_{track}/\text{ndf}$	< 4	Run II only
	K_S^0	$m_{\pi^+\pi^-}$	$\lvert m_{\pi^+\pi^-} - m_{K_S^0,\,PDG}\rvert < 64$ MeV	
		χ^2_{vtx}	< 25	
		χ^2_{DOCA}	< 25	
		$\max_{DOCA}$	< 4 mm	Run I only
		χ^2_{IP}	> 5	Run II only
$B_s^0 \to K_S^0 K_S^0$	B_s^0	$m_{K_S^0 K_S^0}$	$\lvert m_{K_S^0 K_S^0} - m_{B_s^0,\,PDG}\rvert < 600$ MeV	Run I
		$m_{K_S^0 K_S^0}$	4000 MeV $\leq m_{K_S^0 K_S^0} \leq 6500$ MeV	Run II
		$\max_{DOCA}$	< 1 mm	LL
		$\max_{DOCA}$	< 4 mm	LD
		clf. output	> 0.5	Run I only

6.3.2 Preselection

The stripping reduces the number of possible signal candidates to be $\mathcal{O}(10^6)$. However, this number is still many orders of magnitude higher than expected. The preselection aims at removing distinct background events while being very efficient on signal decays. Therefore, loose requirements on certain observables are applied to both, the signal and normalisation channel.

The average number of interactions per bunch crossing follow a Poisson distribution with a mean of about 1.7 for Run I and about 1.1 for Run II, respectively (see also Tab. 3.1). Candidates can be reconstructed to originate from different PVs, where all possible decay candidate and PV combinations are stored after the stripping. Since one candidate only stems from one PV, the superfluous PV are removed from the data samples. Two procedures can be used for this purpose. A random PV selection determines an arbitrary PV and assigns it to the candidate. A best PV selection on the other hand chooses the PV with the lowest B candidate fit uncertainty according to the DTF. As this could also increase the separation power between real and fake B candidates, the latter approach is applied. This cut does not reduce the overall number of events, since the stripping implies at least one PV per candidate. Furthermore, the DTF algorithm fails to converge in about 0.5 % of all events. These events are removed.

While the separation between hadrons is challenging and relies on the PID system, muons can be identified more easily because of their ability to penetrate the whole detector. Therefore, the LHCb framework provides an informational flag for each final state particle called `isMuon` [146]. This variable is true, if hits in the muon stations with a certain minimum momentum are assigned to the tracks trajectory. Thus, the requirement have to be false for each pion and kaon candidate.

To separate K_S^0 candidates from non-resonant pion pairs a loose requirement on the decaying position of the K_S^0 is sufficient due to the high lifetime of K_S^0 mesons. Hence, a new observable is introduced, which describes the distance between the $B_{(s)}^0$ and K_S^0 decay vertex position along the beam axis,

$$\Delta Z = \text{Endvertex}_z(K_S^0) - \text{Endvertex}_z(B_s^0). \tag{6.10}$$

A comparison between pion tracks stemming from K_S^0 and non-resonant pions is shown in Fig. 6.3.

To validate a cut point ΔZ, the signal efficiency taken from signal MC is compared to the background reduction determined on the upper mass side band (UMSB) of the invariant mass $m_{\pi^+\pi^-\pi^+\pi^-}$ ($m_{K^+K^-\pi^+\pi^-}$). A requirement of $\Delta Z > 15\,\text{mm}$ is chosen for this analysis. For Run I, the signal efficiency is 95 %, while reducing the background by 10 %. In Run II however, the background reduction is more than 60 % for $B_s^0 \to K_S^0 K_S^0$ decays while the signal efficiency remains compatible. This is expected because of the weaker stripping selection in this period. All aforementioned selection steps are also summarised in Tab. 6.2.

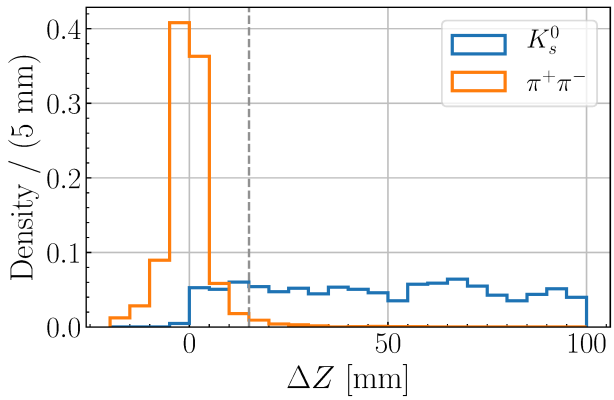


FIGURE 6.3: Comparison between a non-resonant pion pair (orange) and two pions of a K_S^0 decay (blue). The dashed grey line marks the chosen ΔZ cut position. The distributions are based on signal MC data of $B_s^0 \to K_S^0\pi^\pm\pi^\mp$ decays.

TABLE 6.2: Preselection requirements for $B_s^0 \to K_S^0 K_S^0$ and $B^0 \to \phi K_S^0$ decays.

Observable	Cut
# PV from DTF	1, Best PV selection
DTF status	Converged
`isMuon`	False for all π (and K)
ΔZ	$> 15\,\text{mm}$

Multiple candidates

One event, i.e. one bunch crossing, can contain multiple reconstructed particles of interest or (signal) candidates. Depending on the analysis and selection strategy, the range of events containing multiple candidates is typically between 0.1 % and 20 % [147]. There are several reasons for its occurrence: first and most common, overlaps between candidate components can yield multiple candidates. For example, one K_S^0 candidate could be combined with two other K_S^0 candidates meeting all criteria to build a B_s^0 meson. Another type are reflections, where all reconstructed objects are shared between to candidates except particle identification. In the present analysis, however, the share of these events is negligible. The third possibility are genuine multiples, which use completely independent event information to build multiple candidates. Since the branching ratio of the decays of interest are in the order of $\mathcal{O}(10^{-5})$, the fraction of these candidates is also neglectable.

After stripping and preselection, the number of events containing multiple candidates is about 0.2 % for the complete sample of $B^0 \to \phi K_S^0$ decays and Run I of $B_s^0 \to K_S^0 K_S^0$ decays. Due to the looser stripping requirements of $B_s^0 \to K_S^0 K_S^0$ decays in Run II, this fraction raises to 8 %, where the possibility of events containing three or more candidates is less than 0.05 %. To avoid biases due to different selections in the normalisation and signal channel, one candidate is chosen randomly after applying the complete selection chain.

6.3.3 Trigger Requirements

As outlined in Sec. 3.2.3, an event has to pass multiple trigger stages in order to be recorded. An overview of the trigger lines used to select $B_s^0 \to K_S^0 K_S^0$ and $B^0 \to \phi K_S^0$ decays divided by trigger level is given in this section. Additionally, a summary of all relevant trigger lines is shown in Tab. 6.3. For the decays of ϕ mesons a specialised trigger line exist, which takes advantage of its unique decay structure (`IncPhi` lines).

TABLE 6.3: Trigger line decisions required in the present analysis. For $B^0 \to \phi K_S^0$ decays, additional lines are accepted and denoted in round brackets. Between the stages, a logical `AND` is required. At least one line has to be passed by the candidate (`OR`) within a stage.

Stage	Run I	Run II
L0	`HadronDecision_TOS OR Global_TIS`	
HLT1	`TrackAllL0Decision_TOS`	`TrackMVADecision_TOS` `OR TwoTrackMVADecision_TOS`

HLT2	`Topo2BodyBBDTDecision_TOS`
	`OR Topo3BodyBBDTDecision_TOS`
	`OR Topo4BodyBBDTDecision_TOS`
	`(OR IncPhiDecision_TOS)`
		`Topo2BodyDecision_TOS`
		`OR Topo3BodyDecision_TOS`
		`OR Topo4BodyDecision_TOS`
		`(OR {Phi}IncPhiDecision_TOS)`

L0

A pure hardware trigger represents the first stage, relying on the information by the calorimeter and muon systems, the latter being of no relevance to this analysis. The `L0_Hadron_TOS` trigger requirement is utilised in this analysis, which is based on the transverse energy

$$E_T = E \sin\theta. \tag{6.11}$$

For hadron candidates originating from b decays, the transverse energy is expected to be higher than for particles directly created in the pp collision, since the transverse momentum rises with the mass of the decaying particle. The energy E_T is determined as the projection of the energy in calorimeter clusters of (2×2) cells given the polar angle θ with respect to the beam axis [148]. It corresponds to the energy deposited in the HCAL cluster or the sum of ECAL and HCAL clusters, if the candidate can be associated to both clusters. The requirement in 2011 is $E_T > 3.5$ GeV and is adjusted throughout the years, so that the output rate is constantly at 400 kHz. Furthermore, events with a high number of track multiplicities are omitted to prevent the HLT processing farms from saturation. The occupancy reduction is realized by a cut on the SPD multiplicity of $n_{\mathrm{SPD}} < 600$. It has to be noted that this trigger is not optimal for symmetrical decays like $B_s^0 \to K_S^0 K_S^0$ or $B^0 \to \phi K_S^0$ as the decision

only depends on one final state track. Therefore, besides the hadron trigger requirement the event is also kept if the rest of the event triggers one of the physical lines `L0_Phys_TIS`.

Hlt 1

Similar to the L0 decision, the `Hlt1_TrackAllL0` decision depends on one track. These tracks are retrieved in a successive approach. First, VELO tracks and the position of PV(s) are reconstructed. The number of VELO hits must be greater than 10, which excludes the possibility of downstream tracks to be triggered by `Hlt1_TrackAllL0`. Using the forward tracking algorithm (see also Sec. 6.3.1), these track trajectories are propagated to the T stations (see Sec. 3.2.1) to build long tracks. In contrast to the full event reconstruction, where all VELO tracks were extended simultaneously allowing them to compete for VELO hits, the trigger forward tracking is accomplished independently per VELO track [149]. The search windows in the T stations are also narrowed because the tracks must meet (transverse) momentum criteria of $p > 10$ GeV ($p_T > 1.7$ GeV). Further prerequisites for the track are a minimal number of T station hits $n_{\mathrm{T}} > 16$, a minimal impact parameter significance of $\chi^2_{\mathrm{IP}} > 16$ and a maximal track fit uncertainty of $\chi^2_{\mathrm{track}}/\mathrm{ndof} < 2.5$ [150].

Hlt 2

While many trigger lines exist for muonic decay modes, purely hadronic decays are selected most efficiently by the topological triggers like the `Topo[2-4]BodyBBDTLine`, also known as inclusive B triggers [151]. To speed up the decision-making process, the mass hypothesis of a kaon is assigned to all final state particles. This can be done since all occurring particles have small masses comparing the $B^0_{(s)}$ meson mass. To save computing resources, the input particles for these lines have to satisfy the (transverse) momentum requirements of $p > 5$ GeV ($p_T > 0.5$ GeV). Furthermore, they should not stem from the PV ($\chi^2_{\mathrm{IP}} > 4$) and fulfil a lose track quality criterion, $\chi^2_{\mathrm{Track}}/\mathrm{ndof} < 3$. In addition to single tracks, V^0 mesons like K^0_{S} can be used as input, too. Their decay products have to satisfy the same requirements as before, but a harder IP significance of $\chi^2_{\mathrm{IP}} > 16$. The V^0 mesons should have a sufficient vertex resolution of $\chi^2_{\mathrm{vtx}} < 10$, a high vertex separation significance of $\chi^2_{\mathrm{VS}} > 1000$, match in a lose mass window around the nominal V^0 mass and an angle pointing downstream of the detector. These input particles are used to build proto-particles in an iterative way. First, two particles are combined to two-body proto-particles by applying a selection, whose outcome can be used for the two body line decision or be an input for the three body line. This newly added input particle must again fulfil certain conditions in combination two-body proto-particles to build three-body proto-particles. This procedure is repeated for the four body line. An overview of the requirements in this process is given in Tab. 6.4.

TABLE 6.4: Requirements to determine trigger decision of the topological inclusive B trigger lines [150].

	Observable	2 body	3 body	4 body	Unit
Input proto-particle	$m(\sum \text{Particles})$	-	< 6	< 6	GeV
	χ^2_{vtx}	-	< 10	-	
Proto-particle	$m(\sum \text{Particles})$		< 7		GeV
	max(DOCA)		< 0.2		mm
	Direction angle		> 0		
	χ^2_{VS}		> 100		
Decision	$\sum p_T$	> 3000	> 4000	> 4000	MeV
	Classifier decision	> 0.4	> 0.4	> 0.3	

The final trigger line decisions are given by the output of a Bonsai BDT [152] classifier, which is trained separately for each line. The input variables are the sum over all p_T, the minimal p_T, the invariant mass, the DOCA between the particles, the candidates χ^2_{IP} and χ^2_{FD} and the corrected mass being calculated as

$$m_{\text{corr}} = \sqrt{m^2_{\text{PP}} + |p^*_{T\text{miss}}|^2} + |p^*_{T\text{miss}}|. \tag{6.12}$$

Here, m_{PP} is the invariant mass of the n-body proto-particle and $p^*_{T\text{miss}}$ the missing momentum transverse to the flight direction of the candidate, determined by the combination between PV and SV.

For the normalisation channel $B^0 \to \phi K_S^0$, also a dedicated trigger line for the ϕ mesons is used which focusses on the resonant kaon pair. The trigger decision is determined in two steps. Similar to the topological lines, first geometric cuts are applied on a per track basis. Afterwards, the probability for the track to be a kaon should be greater than to be a pion. This information is provided by the particle identification systems, see Sec. 3.2.2. Finally, the reconstructed ϕ mesons must be in a mass window of 20 MeV around its nominal mass. By adding the `(Phi)IncPhi` line, the number of $B^0 \to \phi K_S^0$ candidates can be increased by about 30 %, compared to using the topological lines exclusively.

6.3.4 Background studies

The determination of specific contributions which overlay the physically relevant data and hence are referred to as backgrounds is a crucial part of every particle physics analysis. Wrong assumptions can lead to incorrect descriptions of PDF components, which can falsify the signal yield and thus the branching ratio. A way to handle background distributions is either to exclude them by applying selections to the data set or consider them in the final fit model. Different types of backgrounds exist. First, there are fully reconstructed backgrounds which share the same final state as the signal mode, like $B_s^0 \to K_S^0 \pi^+ \pi^-$ decays. Except the missing K_S^0 resonance, this decay mode exactly mimes the signal and therefore is undistinguishable

in the B_s^0 meson mass. Partly reconstructed B hadron decays yield the same final state particles, but originate from different sources, e.g. assuming a $B \to K_S^0 X$ decay combined with a K_S^0 candidate from the PV by omitting X. When reconstructing such backgrounds in the invariant $m(K_S^0 K_S^0)$ mass, these distribution smear out to lower masses. In the case of misidentified backgrounds, a false assumption was made regarding a particles mass hypothesis. An example of this background is the decay $B_s^0 \to K_S^0 \pi^+ K^-$, where the K^- candidate has been reconstructed as π^- candidate, which yields a shifted invariant mass while rebuilding the B_s^0 meson. Finally combinatorial background exists, a random combination of particles originating from different sources. In this analysis, combinatorial background contributions dominate all aforementioned ones and are of flat or exponential-shape in the invariant mass $m(K_S^0 K_S^0)$ ($m(K_S^0 \phi)$). A discussion of important background distributions is given in this section.

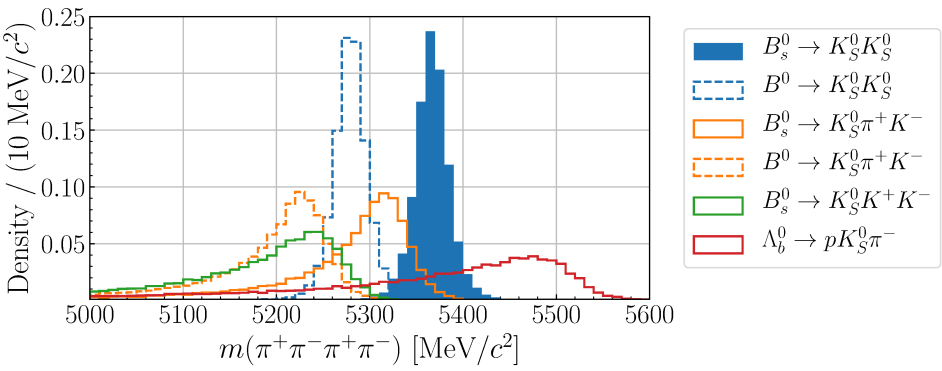


FIGURE 6.4: Possible background contributions for the decay channel $B_s^0 \to K_S^0 K_S^0$. The invariant masses have been determined using a four pion mass hypothesis.

$B_s^0 \to K_S^0 K_S^0$

To get a quick overview of the shapes of certain background contributions, the package RapidSim [140] is used. It generates MC samples according to a phase space model of the given decay mode. Although only the decay and not the interaction with the detector is simulated, all final state particles are forced to lie in the acceptance of LHCb. A summary of possible background decays is shown in Fig. 6.4. As can be seen, $B_s^0 \to K_S^0 \pi^+ K^-$ and $\Lambda_b^0 \to p K_S^0 \pi^-$ decays share the same mass region as the $B_s^0 \to K_S^0 K_S^0$ signal. In the former (latter) decay, the kaon (proton) is misidentified as pion. These decay modes share a similar branching fraction as $B_s^0 \to K_S^0 K_S^0$, see Tab. 6.5.

However, by requiring that the invariant mass of the two non-resonant particles should be in a mass window around the nominal K_S^0 meson mass, these modes are suppressed by a factor of 100. The mass distributions and the corresponding K_S^0 mass window is given in Fig. 6.5. In the further course of this analysis an additional PID requirement is applied (see Sec. 6.3.5), which also suppresses the aforementioned

TABLE 6.5: Possible background decay modes in the same mass region as $B_s^0 \to K_S^0 K_S^0$ [44].

Decay	Branching Ratio
$B_s^0 \to K_S^0 \pi^+ K^-$	$(4.2 \pm 0.5) \times 10^{-5}$
$\Lambda_b^0 \to p K_S^0 \pi^-$	$(0.7 \pm 0.2) \times 10^{-5}$

decay channels. Thus, the expected yields of $B_s^0 \to K_S^0 \pi^+ K^-$ and $\Lambda_b^0 \to p K_S^0 \pi^-$ decays are found to be negligible.

Two other possible backgrounds are the decays $B_s^0 \to K_S^0 \bar{K}^{*0}(\to K^- \pi^+)$ and $\Lambda_b^0 \to K_S^0 N^*(\to p\pi^-)$, where N^* is an abbreviation for the excited neutron states $N(1440 \ldots 2700)$. Similar to non-resonant $B_s^0 \to K_S^0 \pi^+ K^-$ decays, the kaon has been misidentified as pion in the former channel, while the measured branching ratio $\mathcal{BR}(B_s^0 \to K_S^0 \bar{K}^{*0}) = (1.6 \pm 0.4) \times 10^{-5}$ [44] is compatible to the signal channel $B_s^0 \to K_S^0 K_S^0$. The $\Lambda_b^0 \to K_S^0 N^*$ decays on the other hand, where a confusion between proton and pion hypothesis exists, has not been measured before. Thus, no branching ration prediction exists for this channel. The background rejection due to the K_S^0 mass window cut is 85 % (91 %) for the decay $\bar{K}^{*0} \to K^+ \pi^- (N^* \to p\pi^-)$, while each decay is reconstructed in the two pion mass hypothesis. Both resonances decay via strong interaction, which lead to a short lifetime and as a consequence of further suppression due to the ΔZ cut. Hence, these decay channels will also not be considered any further in the analysis.

The fully reconstructed decay $B_s^0 \to K_S^0 \pi^+ \pi^-$ is mainly reduced by the ΔZ cut. To determine the impact of this cut, $B_s^0 \to K_S^0 \pi^+ \pi^-$ MC has been generated and reconstructed as $B_s^0 \to K_S^0 K_S^0$. The ΔZ cut has then been applied to the non-resonant pion pair, yielding a background rejection of 98 %. Finally, before extracting the signal yields in the unblinded mass fit, the K_S^0 mass window is made more tight to further reduce the possible phase space of non-resonant hadron pairs. Hence, also this decay mode shows no significant contributions after full selection.

The decay $B^0 \to K_S^0 K_S^0$ is very similar to the signal mode, except the mother particle. No action was taken to suppress this channel. Therefore, a $B^0 \to K_S^0 K_S^0$ component is added in the final mass fit.

$B^0 \to \phi K_S^0$

The advantage of two long-lived mesons and the resulting suppression of physical backgrounds by the ΔZ cut is not given in the normalisation channel. Thus, more background contributions must be considered. The main physical background is expected to be B_s^0 decays into a K_S^0 meson and hadron pair, consisting of kaons and pions. In general, the spectrum of the hadron pair is complex and contains plenty of resonances, which are for example studied in Ref. [153]. However, as only the ϕ resonance and thus a narrow peak in the K^+K^- spectrum is of interest, only non-resonant decay modes were considered as background. A summary of branching

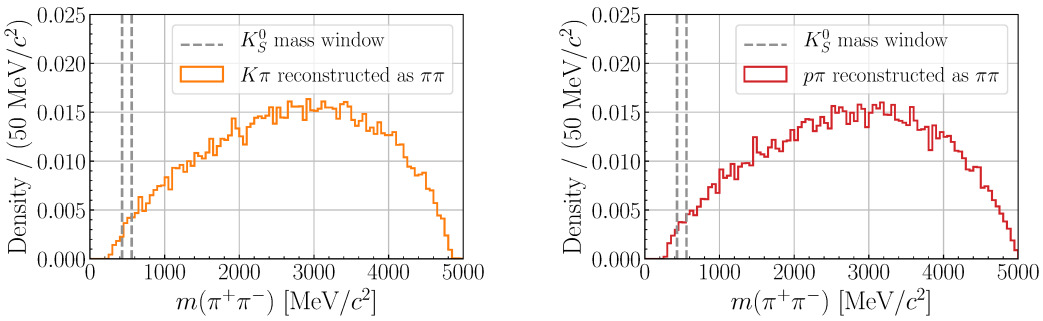


FIGURE 6.5: Invariant $K\pi$ (left) and $p\pi$ (right) mass distributions under the $\pi\pi$ hypothesis of the decays $B^0_s \to K^0_S\pi^+K^-$ and $\Lambda^0_b \to pK^0_S\pi^-$, respectively. Furthermore, the K^0_S mass window required in the stripping selection is given in the plots (grey line).

ratios of these decays as well as a proportion with respect to the branching ratio of $B^0 \to \phi K^0_S$ decays is shown in Tab. 6.6.

TABLE 6.6: Possible background contributions of the decay channel $B^0 \to \phi K^0_S$. The branching ratios of the decays given in the table corresponds to the non-resonant decay modes only [44].

Decay	$\mathcal{BR}$ (Decay)	$\frac{\mathcal{BR}(\text{Decay})}{\mathcal{BR}(B^0\to\phi K^0_S)}$
$B^0 \to K^0_S\pi^+\pi^-$	$7.4 \pm 2.0 \times 10^{-6}$	2.0 ± 0.6
$B^0 \to K^0_S K^-\pi^+$	$3.10 \pm 0.35 \times 10^{-6}$	0.85 ± 0.13
$B^0 \to K^0_S K^+K^-$	$1.6 \pm 0.5 \times 10^{-5}$	4.5 ± 1.4
$B^0_s \to K^0_S\pi^+\pi^-$	$4.7 \pm 1.1 \times 10^{-6}$	1.29 ± 0.31
$B^0_s \to K^0_S K^+\pi^-$	$4.2 \pm 0.4 \times 10^{-5}$	11.5 ± 1.7
$B^0_s \to K^0_S K^+K^-$	$6.5 \pm 3.0 \times 10^{-7}$	0.18 ± 0.08

While the preselection and topological trigger efficiencies are similar between all channels, the stripping efficiency differs up to three orders of magnitude between normalisation and possible background. The main reason for this is that invariant mass of the hadron pair reconstructed as two kaons must be in a mass window around ± 50 MeV of the nominal ϕ mass. To estimate the background fraction, the background candidates are reconstructed identical as $B^0 \to \phi K^0_S$ decays. Tab. 6.7 summarises the efficiencies of the selection chain up to trigger selection. To further reduce the probability of a non-resonant hadron pair that matches the selection criterion, the mass ϕ mass window is tightened to ± 12 MeV, which is applied after the MVA selection. As can be seen in the last column of Tab. 6.7, the largest contribution is due to the decay channel $B^0_s \to K^0_S K^+\pi^-$. This channel is further suppressed by the PID selection by more then 95 %, which is applied in combination with the MVA selection, reducing the contribution of $B^0_s \to K^0_S K^+\pi^-$ decays to a negligible level.

Other background contribution are the decay channels $B^0 \to \phi K^{*0}$, $B^0_s \to \phi K^{*0}$ and $B^+ \to \phi K^{*+}$, where $K^{*0} \to K^0\pi^0$ and $K^{*+} \to K^0\pi^+$, respectively. Omitting

the π^0 (π^+) of the excited kaon decay, these decay exactly mimic the normalisation channel and hence should be considered as partly reconstructed backgrounds. Fig. 6.6 shows distributions of the reconstructed mass, while one graph represents the distribution of all particles reconstructed and one the partly reconstructed decay. As can be seen the partly reconstructed mass is shifted to lower masses. Both, the B^0 and B^+ decays have a kinematic limit at around 5150 MeV, while the B_s^0 decay reaches masses up to 5240 MeV.

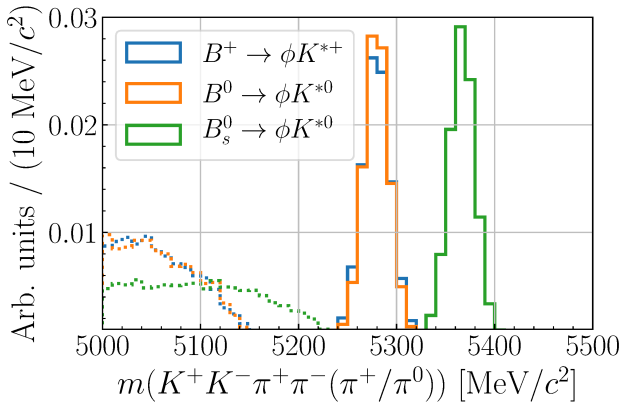


FIGURE 6.6: Invariant mass distributions of the decays $B^+ \to \phi K^{*+}$ (blue), $B^0 \to \phi K^{*0}$ (orange) and $B_s^0 \to \phi K^{*0}$ (green). The solid lines represent the fully reconstructed mass, while the dotted ones depict the decays reconstructed as the normalisation channel $B^0 \to \phi K_S^0$.

To estimate a contribution of these channels, the expected fraction of candidates regarding to $B^0 \to \phi K_S^0$ is calculated in a simplified approach, assuming that the selection efficiencies are the same for all channels, in the full mass range of 5000 to 5600 MeV. The ratios obtained are 0.041 ± 0.009 for $B^+ \to \phi K^{*+}$, 0.0204 ± 0.0022 for $B^0 \to \phi K^{*0}$ and $0.001\,03 \pm 0.000\,29$ for $B_s^0 \to \phi K^{*0}$. While the latter channel can be ignored due to its low efficiency, the contribution of the other two decays are at a few percent level and could alter the mass shape in the final fit, biassing the determined yields. For this reason, the final mass region for the normalisation channel is shrunk to be 5150 MeV $> m(K^+K^-K_S^0) >$ 5600 MeV. After this modification, all aforementioned partly reconstructed decay modes were reduced to a per mill level and thus neglected.

Hence, besides combinatorial the only expected background component are non-ϕ $B^0 \to K_S^0 K^+ K^-$ decays, which is considered in a fit to the invariant mass of the hypothetical ϕ meson, see also Sec. 6.5.

Combinatorial background

While combinatorial background consists in principle of arbitrarily combined tracks, the vertex separation due to the ΔZ cut requires that the decaying particles must feature a significant flight distance. Besides K_S^0 candidates, the Λ candidates have

TABLE 6.7: Efficiencies for possible background contributions of the decay channel $B^0 \to \phi K^0_S$. The efficiency ε_{sel} summarises the efficiency of stripping, preselection cuts and trigger selection. Afterwards, the invariant mass of both reconstructed kaons must lie in a mass window of ± 12 MeV around the nominal ϕ mass (second columns). Finally, $\varepsilon_{\text{comb},\,\mathcal{BR}}$ denotes the combined efficiency of $\varepsilon_{\text{tight}\,\phi}$ divided by the branching ratio difference between background decay and normalisation channel. Thus, this number of events is expected per 1000 events of $B^0 \to \phi K^0_S$ decays.

Decay	$\varepsilon_{\text{sel}}\,[10^{-5}]$	$\varepsilon_{\text{tight}\,\phi}$	$\varepsilon_{\text{comb},\,\mathcal{BR}}\,[10^{-3}]$
$B^0 \to K^0_S \pi^+\pi^-$	1.34 ± 0.18	0.33 ± 0.06	1.9 ± 0.7
$B^0 \to K^0_S K^-\pi^+$	1.62 ± 0.20	0.22 ± 0.05	0.62 ± 0.19
$B^0 \to K^0_S K^+K^-$	2.08 ± 0.23	0.32 ± 0.05	6.3 ± 2.3
$B^0_s \to K^0_S \pi^+\pi^-$	1.43 ± 0.19	0.16 ± 0.05	0.60 ± 0.25
$B^0_s \to K^0_S K^+\pi^-$	1.61 ± 0.20	0.35 ± 0.06	13.8 ± 3.5
$B^0_s \to K^0_S K^+K^-$	2.07 ± 0.23	0.28 ± 0.05	0.22 ± 0.11

a similar lifetime of $\mathcal{O}(10^{-10}\,\text{s})$, thus also decaying downstream of the PV. These V^0 particles can be visualised by the so-called Armenteros-Podolanski plot [154]. It compares the asymmetries of the z-component of two final state particles (P_1, P_2) momentum,

$$p_{z,\,\text{asym}} = \frac{p_z(P_1) - p_z(P_2)}{p_z(P_1) + p_z(P_2)}, \tag{6.13}$$

with the sum of both particles transversal momentum in a 2D histogram. If a decay is symmetrical regarding to its final state particles, like $K^0_S \to \pi^+\pi^-$ decays, it is distributed centrally in the plot. In contrast, asymmetric decays like $\Lambda \to p\pi^-$ decays are shifted to the sides, because the proton candidates on average carry a larger part of the momentum due to their higher mass. Fig. 6.7 shows a Armenteros-Podolanski plot for both, upper mass side bands of $B^0_s \to K^0_S K^0_S$ stripped data as well as $B^0_s \to K^0_S K^0_S$ signal MC.

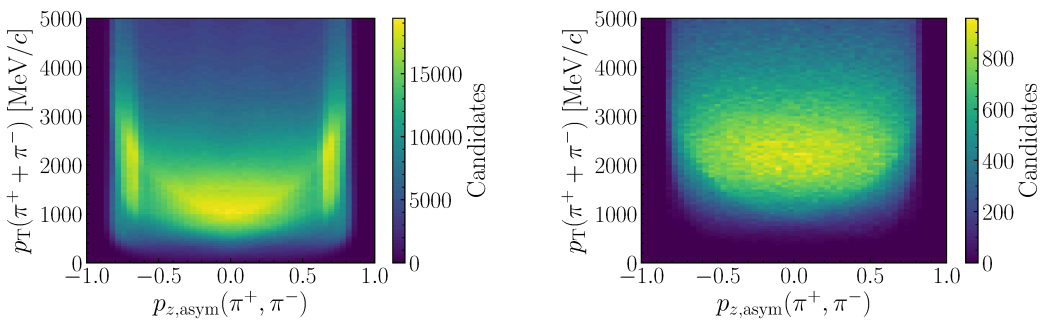


FIGURE 6.7: The Armenteros-Podolanski plot, determined on upper mass side band data (left) and signal MC (right) of $B^0_s \to K^0_S K^0_S$ decays. On the left plot, two further components can be seen at $|p_{z,\,\text{asym}}| > 0.6$, which can be identified as $\Lambda \to p\pi^-$ decays.

To remove background components stemming from Λ decays, one pion mass hypothesis is changed to the proton mass and the invariant mass is calculated for this combination. As shown in Fig. 6.8, a clear mass peak is visible in both, signal and normalisation channel. It is required, that the PID observable ProbNNp of the proton candidate must be less than 0.1 in a mass window around 8 MeV of the nominal Λ baryon mass, where a comparison between the invariant proton-pion mass against the probability of the pion candidate to be an actual proton is given in Fig. 8.1. This cut is optimised so that the distribution after the cut should be compatible to the remaining distribution. The same procedure was tested in the area of $p_{z,\,\mathrm{asym}} > 0.6$ ($p_{z,\,\mathrm{asym}} < -0.6$), yielding a worse signal efficiency and was therefore discarded. The signal efficiency on $B_s^0 \to K_S^0 K_S^0$ ($B^0 \to \phi K_S^0$) decays is around 98 % (99 %), while reducing the combinatorial background in the upper mass side band by 15 % (9 %). More signal is removed in the $B_s^0 \to K_S^0 K_S^0$ sample, because the Λ veto is applied on both K_S^0 mesons.

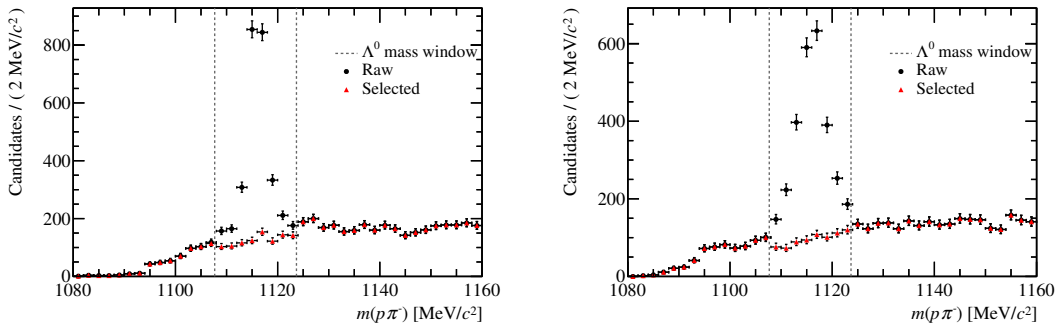


FIGURE 6.8: Reconstructed pion tracks under the $p\pi$ hypothesis for one hypothetical K_S^0 candidate of the decay channel $B_s^0 \to K_S^0 K_S^0$ (left) and $B^0 \to \phi K_S^0$ (right). A clear Λ peak is visible in both channels. The dashed grey line marks the Λ mass window, where the ProbNNp cut is applied. The distribution before (after) the cut is given in black circles (red triangles).

To get an overview of the the combinatorial background component and if respective amount after selection as described in this chapter, the upper mass side band is plotted in a 2D histogram, see Fig. 6.9. For $B_s^0 \to K_S^0 K_S^0$ decays, the two invariant masses of the hypothetical K_S^0 mesons are shown, indicating that the combinatorial background is dominated by two randomly combined K_S^0 mesons. In $B^0 \to \phi K_S^0$ decays on the other hand, one K_S^0 candidate is combined with two arbitrary kaons, where most of them are stemming from the ϕ resonance.

6.3.5 Subtraction of the combinatorial background

Many analyses have shown that a multivariate classification is the ideal tool for separating signal from background in High Energy Physics [155], see Sec. 4.1. For such

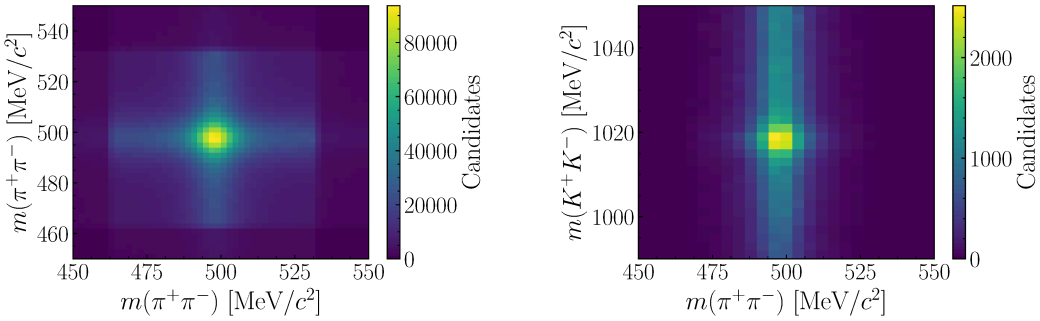


FIGURE 6.9: Two dimensional histogram of the invariant masses of the hypothetical K_S^0 and K_S^0 (K_S^0 and ϕ) mesons. The data corresponds to the upper mass side bands of the decay channel $B_s^0 \to K_S^0 K_S^0$ (left) and $B^0 \to \phi K_S^0$ (right), respectively.

data operations, a gradient boosted decision tree algorithm is used, which is examined in more detail in Sec. 4.1.1. To validate the classifier output and minimize the probability of overfitting, a specialised training strategy is developed. This strategy is applied to each data sample separately, so that a total of eight classifiers is obtained. As this is a binary classification problem, two clean samples (or proxies) of each class are needed. For the signal proxy selected simulated samples of $B_s^0 \to K_S^0 K_S^0$ ($B^0 \to \phi K_S^0$) decays are used. Perfect candidates for combinatorial background are the upper mass side bands, since no resonant particle decays are expected there. Thus, the combinatorial background dataset is made up of candidates with invariant masses of $m(K_S^0 K_S^0) > 5600$ MeV ($m(K_S^0 K^+ K^-) > 5600$ MeV). The total number of candidates per class is shown in Fig. 6.10.

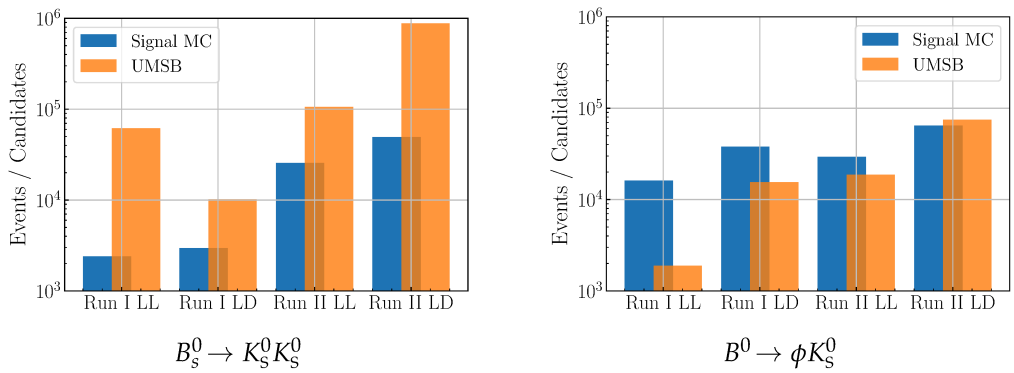


FIGURE 6.10: Number of candidates used as input for the classifier training, for the decay channel $B_s^0 \to K_S^0 K_S^0$ on the left and $B^0 \to \phi K_S^0$ on the right, respectively. The blue (orange) bars indicate the statistics for the corresponding signal MC (upper mass side band) samples. Due to the low selection efficiency of the main channel, the number of signal proxy candidates is very limited, especially for Run I.

MVA strategy

In the nomenclature of machine learning, input observables (or variables) are called features and candidates are referred to as observations. The following lines will stick to this terminology. Furthermore, a few common concepts of machine learning are introduced.

To be able to validate the training process, the complete dataset is split into k parts, where $k-1$ parts are used as training and the remaining as validation data. This process is known a $k$folding. If the complete data is unbalanced, e.g. the number of background candidates is much higher than the number of signal candidates, which is the case for $B_s^0 \to K_S^0 K_S^0$ decays, $k$folding could yield splits with almost no candidates of a specific class, biasing the validation process. To prevent this, a stratified version of $k$folding has been used, where the fraction of classes remains equal in each split.

By defining a metric, the performance of a training process can be validated via $k$folded cross-validation. The training process is repeated k times, with each of the k subsamples used exactly once as validation data. As a result, the signal estimation of a random classifier is chosen for each candidate separately. A typical metric in binary classification is the area under curve (AUC) of the receiver operating characteristic (ROC) [156]. The ROC curve is illustrated by plotting the true positive rate, i.e. the number of signal observations predicted as signal and background as background, against the false positive rate. The ROC AUC value itself is a measure of separability. Generally, the ROC AUC can take values between 0 and 1, where 1 describes perfect classification and 0.5 random guessing (values between 0 and 0.5 can be omitted by a reversal of the prediction). The ROC AUC can also be used to prevent overfitting by applying the early stopping algorithm. This is done by calculating the ROC AUC score for both, the training and validation dataset, after adding each decision tree. If the performance on the validation data did not improve over the last n trees added, early stopping causes the training process to stop at iteration m. The optimised classifier is then chosen to be the one trained of $m-n$ trees. A visualisation of this process is called learning curve. For the nominal classifier for Run II data, the learning curves are given in Fig. 8.2.

The training strategy is defined as follows:

1. A set of beneficial features is selected. These consist of topological ones and fit qualities of the B meson and its daughter particles. Signal and background proxies are compared and features with a potentially high separation power are selected. In addition, it is ensured that the features are not directly correlated to the invariant $B_{(s)}^0$ meson mass. Three important features are the transverse momentum, decay time and impact parameter significance of the B candidate, which are visualised in Fig. 6.11. A description of the input features is also given in Sec. 6.3.5.

2. Determination of the feature importances and feature ranking. Generally, there are multiple metrics which can be used to accomplish this. First, the relative number of times a particular feature occurs in each tree can be calculated. This does not account for the number of observations affected by the decision. Therefore, the *cover* metric is introduced, which corresponds to the relative number of observations connected to a feature. Finally, *gain* represents the accuracy improvement of features with respect to its branches, related to the Gini impurity, see Eq. (4.1). Although the different metrics provide similar results, the latter approach has been chosen. A 3-folding is applied to the complete dataset, causing each feature to be ranked three times. To receive a robust decision, this procedure is repeated ten times with randomised splits of the data. The final rank of a feature is the average of all ranks and summarised in Tab. 8.2 for the decay channel $B_s^0 \to K_S^0 K_S^0$.

3. Determination of the optimal number of features. For a first training, all features are used to apply a 3-fold cross validation. The ROC AUC score is determined for each test dataset and the mean value and its uncertainty are calculated. Afterwards, the least important feature is removed and the procedure is repeated, until only two features are left. The optimal number of features is found at the starting point of the plateau in the ROC AUC score plot, shown in Fig. 6.13.

4. Optimisation of hyperparameters with this subset of features. Again, a 3-fold cross validation with early stopping is deployed. It turns out, that the two tuning parameters are of special interest. The parameter γ represents the minimum loss reduction which is required to create another branch in a decision tree. It can be understood as pseudo-regularisation parameter. Secondly, the maximal depth `max_depth` of each decision tree is varied. If it is too deep, the BDT starts to classify single events in the training dataset by "memorising" them, yielding a bad performance on the validation dataset. Too shallow trees are not able to exploit the correlation between input features, which can lead to an overall worse performance. These hyperparameters are optimised by comparing the classifier output on the training and validation dataset, where a similar performance for both samples reduce the possibility of overfitting. Due to the different number of observations per class in the different samples, γ and `max_depth` are optimised individually for each data subset. In Run I and for the decay channel $B_s^0 \to K_S^0 K_S^0$, a γ of 0.1 and `max_depth` of 3 are found to be optimal, where the other samples uses values of $\gamma = 1$ and `max_depth` $= 2$.

5. Writing classifier decisions to the datasets. To ensure that predictions of the classifier are not biased, the classifier only determines decisions on the test data. Nevertheless, the background proxy is not needed any more throughout the analysis and is therefore omitted. In the case of unseen data, i.e. the invariant B mass regions below 5600 MeV on data, for each candidate one of the

three classifiers obtained is selected randomly to predict the classifier decision. The predictions of the classifiers are compared in Fig. 6.12. Furthermore, the ROC curve is illustrated in Fig. 8.2.

To avoid the classifier to be trained on data and MC differences instead of distinguishing between signal and background, the training is also performed using weights for the simulated samples. The procedure is further outlined in below.

Description of input features

A full set of input features can be found in Tab. 8.2. To increase the classifier performance, features which show an exponential-like behaviour have been modified by a linear transformation, i.e. the logarithm of base 10 ($\log_{10}$). In the case of the LD samples of $B_s^0 \to K_S^0 K_S^0$ decays, the input K_S^0 candidate is expressed in square brackets, e.g. K_S^0[D] refers to a downstream reconstruction. For the $B_s^0 \to K_S^0 K_S^0$ LL data sample, the K_S^0 candidates are sorted by minimum and maximum of the corresponding feature value. The input features are:

Particle τ: Lifetime of the **Particle**.

Particle FD: Flight distance of the **Particle**.

Particle IP: Impact parameter with regard to the primary vertex of the **Particle**.

Particle η: Pseudorapidity of the **Particle**.

Particle $p_{(T)}$: (Transverse) momentum of the **Particle**.

Particle Vertex χ^2: Significance (or quality) of the **Particle**'s vertex.

ΔZ: Distance between the B_s^0 and K_S^0 meson decay vertex with respect to the z-axis , see Eq. (6.10).

Feature reweighting

The MVA heavily relies on the accurate description of the simulated events in all input variables. Hence, small data and MC differences can have a big impact on the determination of signal efficiencies. To estimate the effect, the signal MC has been reweighted using a data-driven approach, which is discussed in detail in this section.

After the nominal MVA and a loose cut on the classifier predictions, it is possible to separate the signal component of the normalisation channel $B^0 \to \phi K_S^0$ from underlying background. A maximum likelihood fit to the invariant mass $m(K_S^0 K^+ K^-)$ is applied using individually shaped models for the expected signal and background components, further described in Sec. 6.4. The ${}_s\mathcal{P}$lot technique (Sec. 4.3) is then used

to determine per-event ${}_s\mathcal{W}$eights, indicating the likelihood for each event to be signal. Those weights are then applied to variables which are not correlated to the mass yield distributions of the signal component only. Hence, these distributions can be compared between signal MC and (${}_s\mathcal{W}$eighted) data. Besides visual comparison of the same variable from data or MC in a binned histogram, the two-sided KS test, which is explained in Sec. 4.4, provides a measure for the equality of two distributions. In this analysis, the variables (transverse) momentum $p_{(T)}$, pseudorapidity η and track multiplicity `nTracks` are chosen as indicators for data and simulation agreement studies. The latter one is not included in the MVA training, but is known to be imperfectly described in MC, since the full hadronisation process is difficult to simulate. An overview of the four variables is shown in Fig. 8.3. As expected, `nTracks` exhibits the largest deviations ($\alpha \approx 0.15$), while the other variables are reasonably well described ($\alpha \approx 0.05$). Nevertheless, the simulated candidates are reweighted to better match the ${}_s\mathcal{W}$eighted data distributions. One approach would be to simply bin the variable which differs the most (`nTracks`), divide the data sample by the simulation and apply weights to the other variables based on the result. This yields a better result for this specific variable, but lacks of performance by applying the weights to other variables. Due to the curse of dimensionality [157], i.e. the population of bins becoming sparse for multiple dimensions, this approach is only applicable on one variable. The Gradient Boosting reweighting algorithm (GB reweighting, [158]) solves this problem. Similar to the Boosted Decison Tree algorithm (see Sec. 4.1.1) it uses binary decision trees to differentiate between simulation and ${}_s\mathcal{W}$eighted data. Then, per-candidate weights are calculated for the MC samples to match the data distributions. Afterwards, a new iteration of the classifier tries to distinguish between the weighted simulation and data. This procedure is repeated until the algorithm is not able to differentiate between the two distributions any more. The weights calculated in this way are applied to simulated datasets and again compared to the ${}_s\mathcal{W}$eighted data distributions, also shown in Fig. 8.3. Furthermore, results of the KS test are given in the figure. It can be seen that applying the GB reweighting algorithm reduces data and MC disagreements.

Nonetheless, this approach can lead to additional uncertainties. The low statistics available for $B^0 \to \phi K^0_S$ decays leads to noticeable uncertainties on the ${}_s\mathcal{W}$eighted distributions, which further leads to uncertainties in the kinematic reweighting by the GB reweighter. Therefore, a systematic uncertainty is assigned, see Sec. 6.8.3. To evaluate the impact of using (non-)reweighted simulation for the MVA training, the nominal MVA training is performed two times, once with "raw" simulation and once with reweighted MC samples. The signal efficiency ε_{sig} of both classifiers are then compared at a constant background reduction level of 95 %. For the LL data samples, the ε_{sig} difference is below 1 %. However, for the LD samples the difference is quite large, especially in the Run II case. A graphical presentation of this procedure on the Run II samples is given in Fig. 6.15. Thus, the reweighing is applied to the $B^0_s \to K^0_S K^0_S$ decays as well as the $B^0 \to \phi K^0_S$ decay channel.

Cut point optimisation

After the MVA, for each candidate a classifier decision is calculated, what is a measure for the probability for the candidate to be signal or background. To optimise a cut point on the classifier decisions, a Figure of Merit (FoM) is used. It takes into account the fraction of the signal and background contributions, returning a metric which needs to be maximised. For both decay channels, $B_s^0 \to K_S^0 K_S^0$ and $B^0 \to \phi K_S^0$, different FoMs are used. The signal channel is blinded and has not been measured by the LHCb collaboration yet, thus a FoM specialised on the observation of new signals is used, namely the Punzi FoM [159],

$$\mathrm{FoM}_{\mathrm{Punzi}}(\alpha) = \frac{\varepsilon_{\mathrm{sig}}}{\frac{\alpha}{2} + \sqrt{N_{\mathrm{bkg}}}}. \tag{6.14}$$

The significance level α has been chosen to be 3 in this optimisation, which is the expected significance in each data subset. Furthermore, the signal efficiency $\varepsilon_{\mathrm{sig}}$ is determined by a cut on the simulated samples of the signal channel. To estimate the number of background candidates N_{bkg}, an exponential function is fitted to the upper and lower mass side bands ($5000 - 5230\,\mathrm{MeV}$ and $5420 - 5600\,\mathrm{MeV}$) of the blinded $m(K_S^0 K_S^0)$ mass, with loose cuts on the classifier decision. The shape of this curve is then used to describe the data with tightened requirements, i.e. the region where the optimised cut values are expected, normalised to its number of entries. The reason of not fitting the exponential function after each specific cut applied is that the slope fluctuates heavily, especially when fitted to a small amount of data in the mass side bands. Since the number of background events is then interpolated in the expected signal mass window of 5230 to 5420 MeV, the FoM value depends heavily on these fluctuations, which makes it difficult to optimise its value.

The optimisation of $\mathrm{FoM}_{\mathrm{Punzi}}(\alpha)$ is done in two dimensions simultaneously, classifier output and the minimum probability of all long track pion candidates to be actual pions, $\min(\pi\mathrm{ProbNN}\pi)$. The latter observable is expected to be large if all candidate are real pions and small if one or more are misidentified particles, such as kaons. Therefore, high values of $\min(\pi\mathrm{ProbNN}\pi)$ are to be preferred. No PID requirement is applied to downstream pions, since the only two particles producing downstream tracks are Λ^0 and K_S^0 candidates, of which the first one has already been excluded, see Sec. 6.3.4. The scan range of the classifier decision ($\min(\pi\mathrm{ProbNN}\pi)$) is 0.9 to 0.99 (0 to 0.5) in steps of 0.01 (0.02), the resulting $\mathrm{FoM}_{\mathrm{Punzi}}(\alpha)$ is shown in Fig. 6.16 for the Run II samples. A red dot marks the optimised cut point position, also listed in Tab. 6.8. Both, the classifier and $\min(\pi\mathrm{ProbNN}\pi)$ cut points differ for Run I and Run II, indicating that the fraction between signal and background changed. This can be explained by the usage of different stripping requirements between the Runs (see also Fig. 6.10). Furthermore, the sample containing downstream decaying K_S^0 candidates require the application of a harder PID cut in both runs.

Similar to the signal channel, the cut point for the $B^0 \to \phi K_S^0$ decay is optimised

TABLE 6.8: Optimized cut point values for the decay channel $B_s^0 \to K_S^0 K_S^0$ (top) and $B^0 \to \phi K_S^0$ (bottom), split by run and track type.

Decay	Data sample	Classifier output	min(πProbNNπ)	max(KProbNNπ)
$B_s^0 \to K_S^0 K_S^0$	Run I, LL	> 0.95	> 0.02	
	Run I, LD	> 0.92	> 0.20	
	Run II, LL	> 0.98	> 0.34	
	Run II, LD	> 0.98	> 0.42	
$B^0 \to \phi K_S^0$	Run I, LL	> 0.78		< 0.96
	Run I, LD	> 0.88		< 0.92
	Run II, LL	> 0.78		< 0.76
	Run II, LD	> 0.80		< 0.76

in two dimensions, the classifier output and the maximum probability of a kaon candidate to be a pion, max(KProbNNπ). The latter variable should be small for real kaons, so the cut is "less than". Since the $B^0 \to \phi K_S^0$ decay has already been observed, a FoM has been chosen that is more related to the relative uncertainty of the (signal) yield, known as significance FoM,

$$\mathrm{FoM}_{\mathrm{Significance}} = \frac{N_{\mathrm{sig}}}{\sqrt{N_{\mathrm{sig}} + N_{\mathrm{bkg}}}}. \tag{6.15}$$

Because these samples are not blinded, it is possible to determine the number of signal candidates, N_{sig}, directly on data by performing a fit of the signal and background PDF after each cut. The optimised cut values are also given in Tab. 6.8. As expected by choosing Eq. (6.15) as FoM, the optimised cut points are looser, which means that a higher fraction of combinatorial backgrounds is allowed.

Tight selection

To further reduce the combinatorial background components, the mass windows around the intermediate particles are tightened by applying mass window cuts of ± 12 MeV(18 MeV) for L (D) K_S^0 meson and ± 12 MeV for the ϕ meson of their nominal masses. Additionally, the events containing multiple candidates are removed by choosing a random candidate per event.

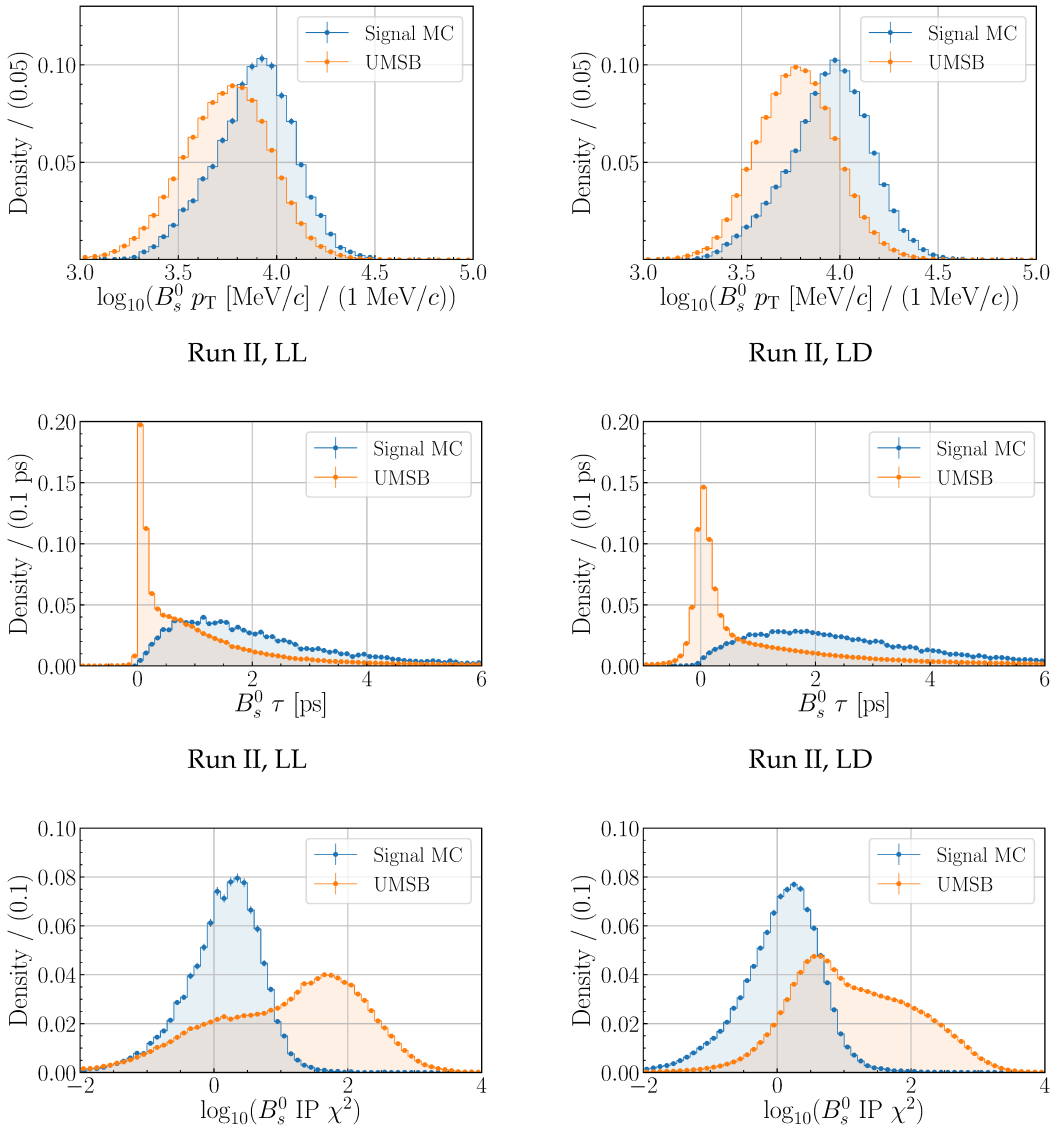
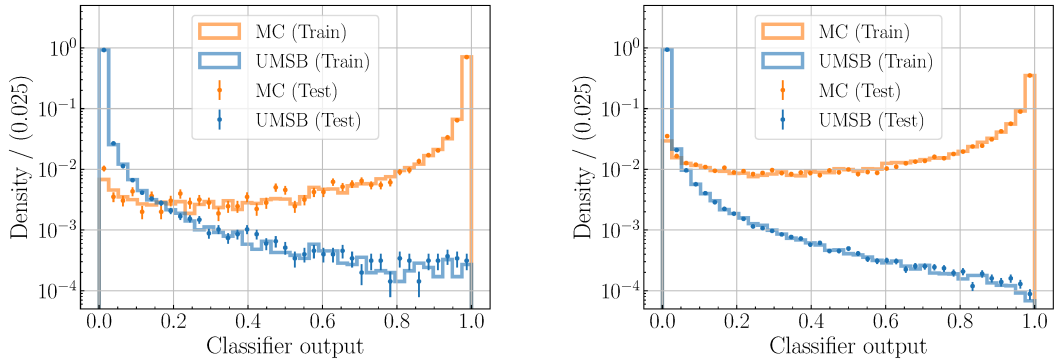
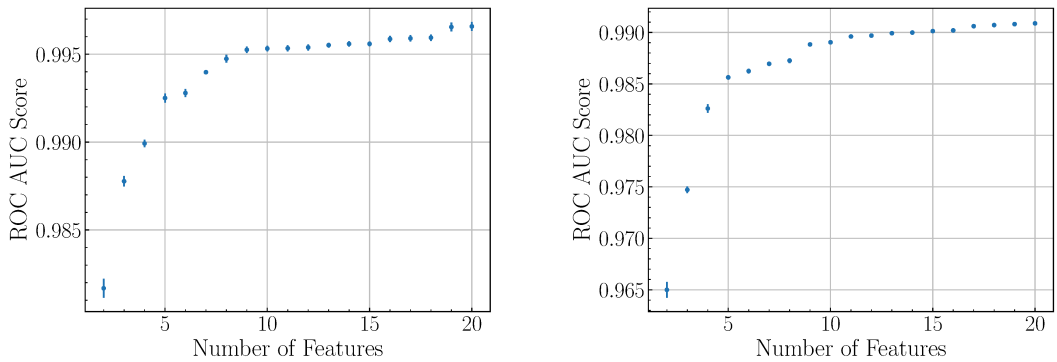


FIGURE 6.11: Comparison of the most important features for the classifier training, described in Sec. 6.3.5. The distributions of $B_s^0 \to K_S^0 K_S^0$ signal MC is given in blue, while the upper mass side band data is shown in orange. Plots on the left (right) side are based on the Run II, LL (Run II, LD) data samples. From top to bottom, the transverse momentum, p_T, decay time, τ, and impact parameter significance of the B candidate, IPχ^2, is depicted. Differences are visible between the different track types, e.g. in the latter observable.



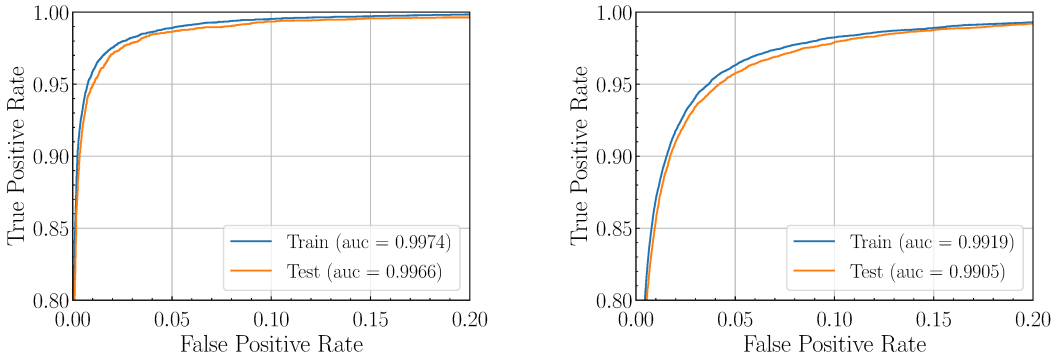
Run II, LL Run II, LD

FIGURE 6.12: Comparison between the classifier prediction performance for one of the three classifiers, validated on both, the training sample (solid lines) and test sample (dots with error bars). On the left (right) the track category LL (LD) of Run II is shown. A large deviation between training and test sample performance could indicate overfitting in the learning process.



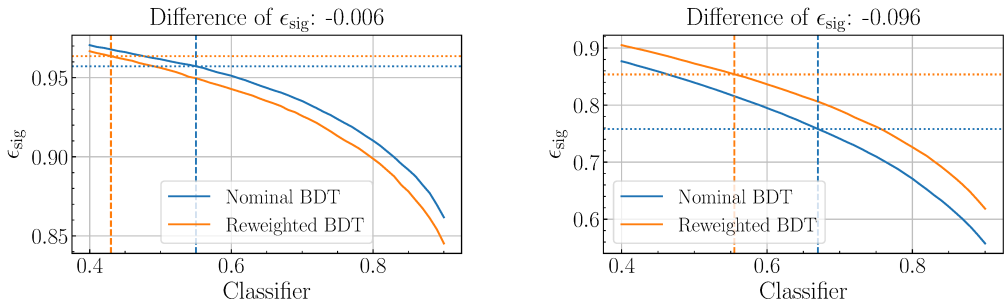
Run II, LL Run II, LD

FIGURE 6.13: Classifier performance, evaluated using the ROC AUC score metric, in dependency of the number of features in the training process. For the track category LL (left) and LD (right) in Run II, the optimal number of features is determined to be 19 and 17, respectively. Tab. 8.2 illustrates which features were used in the nominal training.



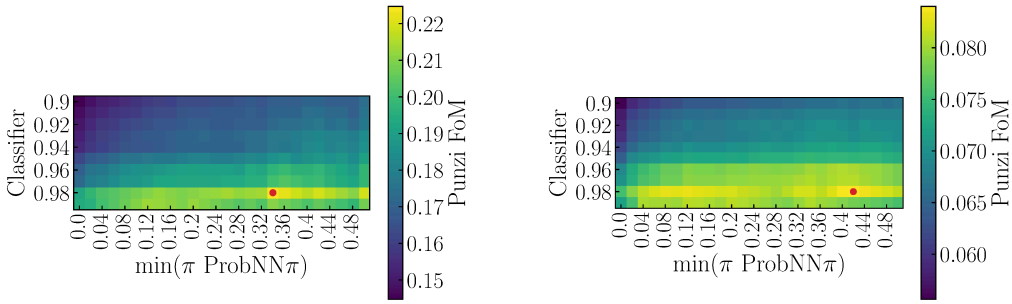
Run II, LL Run II, LD

FIGURE 6.14: Illustration of the receiver operating characteristic (ROC) curve [156], calculated separately for the train and test data sample. On the left (right) the track category LL (LD) of Run II is shown. While a small contribution of overtraining can be seen in both categories, the classifier generally performs better in the LL case.



Run II, LL Run II, LD

FIGURE 6.15: Efficiency comparison between a classifier trained with reweighted MC (orange) and non-reweighted MC (blue) for different cut-points for the Run II dataset. On the left (right) the track category LL (LD) is shown. The dashed vertical lines illustrate the cut point for the classifier, the dotted horizontal lines point to the signal efficiency. The cut point is chosen that it reduces 95 % of combinatorial in the upper mass side band.



Run II, LL Run II, LD

FIGURE 6.16: Two-dimensional cut point optimisation in the classifier output and min(πProbNNπ). As figure of merit, $\mathrm{FoM}_{\mathrm{Punzi}}(\alpha)$ (Eq. (6.14)) is used. On the left (right) the track category LL (LD) of Run II is shown. The red dot marks the optimal cut configuration.

6.4 Fit model

The signal yield is determined by an extended maximum likelihood fit to the data samples. To differentiate between signal and background components, two individual probability density functions (PDF) are introduced, where the signal shape is deduced from fits to simulated datasets. The complete PDF is given by the sum of both, scaled with the respective signal and background fractions,

$$P(m;\vec{\lambda})_{\text{tot}} = N_{\text{sig}} P_{\text{sig}}(m;\vec{\lambda}_n) + N_{\text{bkg}} P_{\text{bkg}}(m;\vec{\lambda}_m), \tag{6.16}$$

with m being the invariant mass of all final state particles and $\vec{\lambda}_n$ ($\vec{\lambda}_m$) the fit parameters of the signal (background) model. Additionally, all hypothetical K_S^0 mesons are constraint to the nominal K_S^0 mass using the decay tree fitter, further explained in Sec. 6.3.1.

By fitting the normalisation channel in the invariant mass $m(K_S^0 K^+ K^-)$, it is not possible to distinguish between $B^0 \to \phi K_S^0$ and non-ϕ $B^0 \to K_S^0 K^+ K^-$ decays. Hence, the signal yield could be polluted by these decays, causing a bias in the $\mathcal{BR}$ determination. In Sec. 6.3.4, a rough estimate about this component is given, which indicates that the proportion should be small. However, the fraction between ϕ and non-ϕ components is determined in a fit to the invariant $m(K^+K^-)$ mass.

6.4.1 Signal component

A model often used to describe invariant mass distributions in high-energy physics is the Crystal Ball function (CB) [160–162]. It consists of a Gaussian core to describe the detector resolution, assuming that the mass resolution is normally distributed, and a power-law tail to lower masses to parametrise lossy processes, like the effect of photon radiation. However, this assumption is not always true, i.e. candidates may have different uncertainties on the reconstructed mass, distorting its Gaussian core. A generalisation of the CB function provides the Hypatia PDF $I(m;\vec{\lambda})$ [163], which models unknown per-event mass resolutions via a generalized hyperbolic distribution core [164]. Furthermore, an additional CB-like tail to higher masses is attached to the Hypatia PDF, describing non-resolution effects coming from the K_S^0 mass constraint of the DTF. This two-tailed Hypatia PDF is donated as $I_2(m;\vec{\lambda})$ in the following. In its most general form $I_2(m;\vec{\lambda})$ can be expressed as

$$P_{\text{sig}}(m;\vec{\lambda}) = I_2(m;\mu,\sigma,\lambda,\zeta,\beta,a_1,n_1,a_2,n_2) \propto \begin{cases} G(m;\vec{\lambda}_G) & , -a_1 \leq \frac{m-\mu}{\sigma} \leq a_2 \\ \frac{G(\mu-a_1\sigma;\vec{\lambda}_G)}{\left(1-m/\left(n_1 R_{G',1}-a_1\sigma\right)\right)^{n_1}} & , -a_1 > \frac{m-\mu}{\sigma} \\ \frac{G(\mu-a_2\sigma;\vec{\lambda}_G)}{\left(1-m/\left(n_2 R_{G',2}-a_2\sigma\right)\right)^{n_2}} & , \quad a_2 < \frac{m-\mu}{\sigma}, \end{cases} \tag{6.17}$$

with the generalised hyperbolic distribution

$$G(m;\vec{\lambda}_G) = G(m;\mu,\sigma,\lambda,\zeta,\beta) = \tag{6.18}$$
$$((m-\mu)^2 + A_\lambda^2(\zeta)\sigma^2)^{\frac{1}{2}\lambda-\frac{1}{4}} e^{\beta(m-\mu)} K_{\lambda-\frac{1}{2}}\left(\zeta\sqrt{1+\left(\frac{m-\mu}{A_\lambda(\zeta)\sigma}\right)^2}\right).$$

Furthermore, $R_{G'}$ is a short-hand notation for the quotient of $G(m;\vec{\lambda}_G)$ and its derivative,

$$R_{G',1/2} := \frac{G(\mu - a_{1,2}\sigma;\vec{\lambda}_G)}{G'(\mu - a_{1,2}\sigma;\vec{\lambda}_G)}. \tag{6.19}$$

Finally, $K_\lambda(\zeta)$ denote the cylindrical harmonics and are also used to parametrise

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

Five parameters define the core PDF, the mean μ, the width σ, two additional degrees of freedom ζ and λ, which vary the "pointyness" of the distribution and the asymmetry parameter β. As no asymmetry is expected and thus to increase the robustness of the fits, β is fixed to zero. The remaining parameters a_1 and n_1 (a_2 and n_2) model the power-law tails to lower (upper) masses.

The parameter values are estimated by fitting the Hypatia PDF to fully selected MC samples of both, signal and normalisation channel. While all parameters are fixed to MC in the fit to $B_s^0 \to K_S^0 K_S^0$ data, the mean and width of $I_2(m;\vec{\lambda})$ is floating for the normalisation channel to compensate for possible disagreements between data and simulation. Here, the estimated MC values serve as initial values. A fit of the signal mode to $B_s^0 \to K_S^0 K_S^0$ MC is shown in Fig. 6.17.

6.4.2 Background component

As outlined in Sec. 6.3, the dominant background is combinatorial. It is modelled by an exponential function with a negative exponent,

$$P_{\text{bkg}}(m;\kappa) = \mathrm{e}^{\kappa m} \quad \text{with } \kappa < 0. \tag{6.20}$$

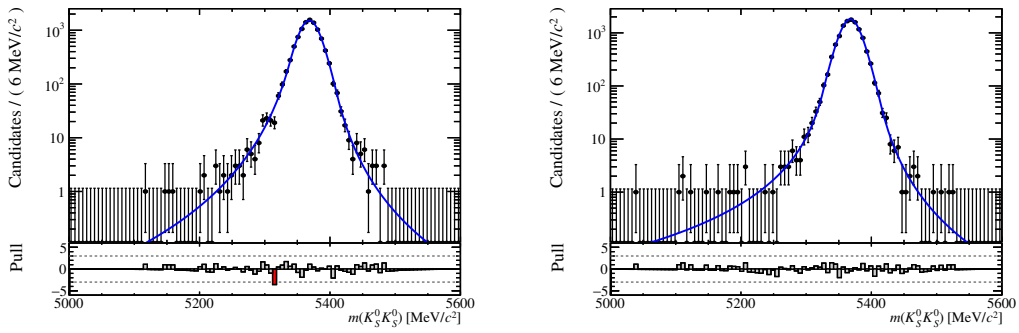


FIGURE 6.17: Result of a Hypatia PDF fit to signal MC of $B^0_s \to K^0_S K^0_S$ depicted in logarithmic scale (blue curve). On the left (right) the track category LL (LD) of Run II is shown.

6.4.3 Fit model scaling

To account for possible differences in the resolution between simulation and data in the signal channel, a scaling factor is determined with the normalisation channel $B^0 \to \phi K^0_S$. First, the signal model described above is fitted to signal MC. Additionally, the complete mass fit is applied to data, while the width and mean parameter are left floated. The relative difference between these two values are given in Tab. 6.9. While differences in the mean are negligible, the width differs by 5 to 20 %, depending on the subset. Thus, the scaling factors of the width are applied to the $B^0_s \to K^0_S K^0_S$ signal component for each run and track type category separately in the final fit.

TABLE 6.9: Relative differences between the signal component fit on $B^0 \to \phi K^0_S$ data and signal MC. The parameters mean (μ) and width (σ) are shown. As signal model, the Hypatia PDF $I_2(m; \vec{\lambda})$ is used. For the nominal fit, only the differences of σ are considered.

Scale factor	Run I, LL	Run I, LD	Run II, LL	Run II, LD	Unit
$\mu_{\text{Data}} - \mu_{\text{MC}}$	1.3 ± 1.3	7.1 ± 1.3	-1.1 ± 0.9	0.3 ± 0.9	MeV
$\frac{\sigma_{\text{Data}}}{\sigma_{\text{MC}}}$	1.21 ± 0.10	1.20 ± 0.10	1.05 ± 0.07	1.13 ± 0.06	1

6.4.4 Non-ϕ K^+K^- background in $m(K^+K^-)$

The mass shape of resonant $B^0 \to \phi K^0_S$ and non-ϕ $B^0 \to K^0_S K^+ K^-$ decays is indistinguishable in the invariant mass of $m(K^0_S K^+ K^-)$. However, the measured invariant kaon mass $m(K^+K^-)$ can be used to disentangle both contributions. The ϕ resonance is modelled by a relativistic Breit-Wigner (RBW) PDF [165], which is defined as

$$P_{\mathrm{RBW}}(m; m_\phi, m_K, \Gamma_\phi) = \frac{m_\phi^2 \Gamma_\phi A}{(m_\phi^2 - m^2)^2 + \left(\frac{m_\phi^2 \Gamma_\phi}{m} A\right)^2} \quad , \tag{6.21}$$

$$A = \left(\sqrt{\frac{m^2 - 4m_K^2}{m_\phi^2 - 4m_K^2}}\right)^3 \quad \text{with} \quad m = m(K^+K^-). \tag{6.22}$$

Here, Γ_ϕ is the decay width of the ϕ meson. The parameter m_K is the mean mass of the kaon which is fixed to its PDG value of $m_K = 493.677$ MeV [44], while m_ϕ can float in the fit. To also account for resolution effects of the detector, the RBW is numerically folded with a Gaussian distribution $G(m; \sigma)$ of width σ as resolution parameter. Hence, the full resonant model consist of two free parameters.

The main source for resonant non-ϕ contributions have been found in the decay channel of $B^0 \to K_S^0 f_0(980)$ ($f_0(980) \to K^+K^-$). The peak mass of the $f_0(980)$meson is below the mass of two kaons, $m(K^+) + m(K^-) = 987.354$ MeV, so that only a high mass tail is expected. But at the same time, the phase space is rapidly opening up, counteracting the tail and yielding a rather flat distribution, which is described by an exponential function. As an alternative model and to check for non-resonant kaon pairs, a phase space model has been implemented based on the invariant kaon mass. Nevertheless, the exponential function describes the non-ϕ contribution more accurately and is therefore chosen. In the nominal massfit, the slope parameter is Gaussian constrained to the values obtained in fits to simulated $f_0(980) \to K^+K^-$ candidates.

Instead of determining the individual yields, a fraction f_{res} between resonant and non-ϕ decay modes is introduced and left free in the fit to data. In summary, the complete PDF to describe the invariant kaon mass is defined as

$$P_{\mathrm{tot}}(m, \vec{\lambda}) \propto f_{\mathrm{res}}(P_{\mathrm{RBW}} * G)(m; m_\phi, \sigma) + (1 - f_{\mathrm{res}})\mathrm{e}^{\kappa m}. \tag{6.23}$$

6.5 Fit results

The fits to the invariant $m(K_S^0 K^+K^-)$ mass are shown in Fig. 6.18 and their results are summarised in Tab. 6.10.

As expected, the separation between signal and combinatorial background differs between the different track type samples. While the signal to background ratio on the complete mass range is the same as for the LL sample, the background component comprises a twice as large proportion on samples containing downstream tracks. Furthermore, the mass resolution σ is slightly larger for LD reconstructed candidates.

Based on these fits, ${}_s\mathcal{W}$eights have been calculated to unfold the signal from the signal and background contributions and thus get access to the invariant kaon mass, $m(K^+K^-)$, of the extracted signal component. The respective model which describes

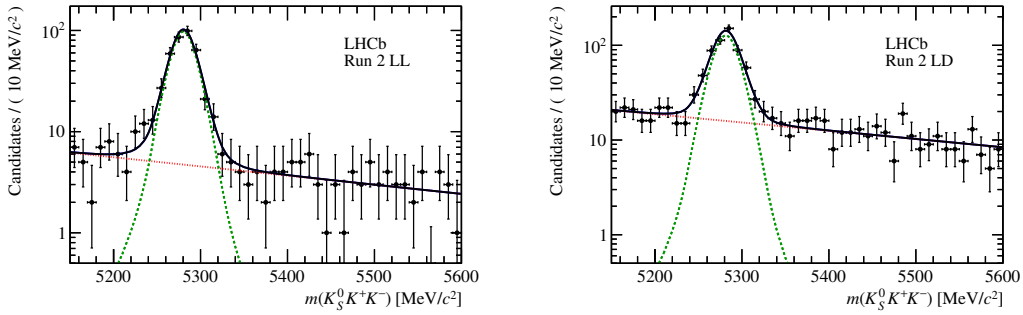


FIGURE 6.18: Fit to the invariant mass distribution of $m(K_S^0 K^+ K^-)$ of Run II data. On the left (right) the track category LL (LD) is shown. The black curve represents the complete model, the single components are given in green (dashed, Hypatia PDF, signal) and red (dotted, exponential, combinatorial background) [166].

TABLE 6.10: Fit results to the invariant mass $m(K_S^0 K^+ K^-)$ of the normalisation channel $B^0 \to \phi K_S^0$.

Parameter	Run 1, LL	Run 1, LD	Run 2, LL	Run 2, LD	Factor	Status
N_{sig}	225 ± 17	255 ± 20	358 ± 21	489 ± 27	1	Free
N_{bkg}	143 ± 15	363 ± 22	181 ± 16	613 ± 29	1	Free
a_1	2.1815	2.6957	2.2728	2.6193	1	Fixed
a_2	2.051	2.5471	2.6673	2.8297	1	Fixed
λ	1	1	1	1	1	Fixed
μ [MeV]	5282.5 ± 1.3	5288.0 ± 1.3	5280.3 ± 0.9	5281.5 ± 0.9	1	Free
n_1	3.542	2.7312	2.8186	2.3739	1	Fixed
n_2	5.4612	3.1463	3.3224	2.5843	1	Fixed
σ [MeV]	17.3 ± 1.4	17.0 ± 1.4	16.4 ± 1.0	17.0 ± 1.0	1	Free
ζ	4.6398	3.8612	2.4873	2.8668	1	Fixed
κ [MeV^{-1}]	-1.6 ± 0.7	-2.6 ± 0.4	-2.1 ± 0.6	-2.01 ± 0.33	10^{-3}	Free

this operation was explained in Sec. 6.4.4, while the fits are shown in Fig. 6.19 and fit results are given in Tab. 6.11. A small component about 15 to 25 % of non-ϕ decays is visible, depending on the data sample. The parameter N_{sig} represents the number of resonant decay modes, i.e. $B^0 \to \phi K_S^0$ candidates.

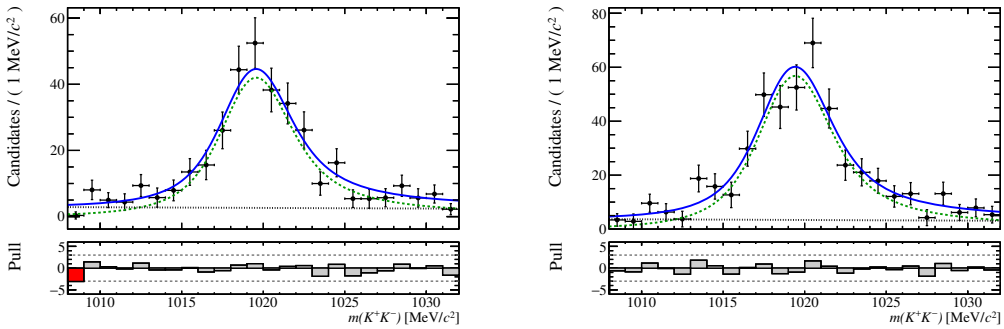


FIGURE 6.19: Fit to the invariant mass distribution of $m(K^+K^-)$ of ${}_s\mathcal{W}$eighted $B^0 \to \phi K_S^0$ data. On the left (right) the track category LL (LD) of Run II is shown. The blue curve represents the complete model, the signal components are given in green (dashed, RBW model, signal) and black (dotted, $f_0(980) \to K^+K^-$ model).

TABLE 6.11: Fit parameters of the PDF describing the reconstructed ϕ meson mass, fitted to the ${}_s\mathcal{W}$eighted invariant $m(K^+K^-)$ mass distribution in the different data subsets. The parameters μ (mean value of the relativistic Breit-Wigner function) is left free in the fit, while σ (resolution factor) and the slope κ are Gaussian constraint to values obtained from fits to simulation.

Parameter	Run 1, LL	Run 1, LD	Run 2, LL	Run 2, LD	Status
N_{sig}	186 ± 19	190 ± 23	295 ± 24	408 ± 32	Free
f_{reso}	0.83 ± 0.06	0.74 ± 0.07	0.82 ± 0.05	0.83 ± 0.05	Free
m_ϕ	1018.68 ± 0.25	1018.57 ± 0.31	1018.64 ± 0.21	1018.48 ± 0.21	Free
σ [MeV]	0.983 ± 0.012	1.13 ± 0.04	1.08 ± 0.04	1.162 ± 0.033	Gaussian constr.
κ [MeV^{-1}]	-0.008 ± 0.004	-0.006 ± 0.004	-0.007 ± 0.004	-0.007 ± 0.004	Gaussian constr.

6.5.1 Determination of the mass observable

The DTF (see Sec. 6.3.1) offers a variety of possibilities to determine the invariant mass of decay cascades. While the width of the reconstructed B_s^0 is dominated by detector resolution of the final state particles, it can be reduced by adding external constraints to the track fit. A smaller peak width is preferable since the aim is to maximize the sensitivity of the signal decay $B_s^0 \to K_S^0 K_S^0$. Four different variations of mass reconstruction were investigated:

- decay tree fit with no constraints,
- decay tree fit with primary vertex constraint (i.e. the B candidate has been constrained to originate from its associated PV),
- decay tree fit with mass constraint (which means that the mass of the K_S^0 candidate is set to its nominal value taken from [44]),
- decay tree fit with primary vertex and mass constraint.

To obtain the peak width a Hypatia function as described in Sec. 6.4.1 is fitted to simulated samples of $B_s^0 \to K_S^0 K_S^0$, whereby the individual samples are divided into year and track type categories. A detailed overview of all fit results is given in Tab. 8.1. As can be seen, their widths vary between 13 and 20 MeV. Decay chains containing downstream tracks exhibit a broader distribution. Without any constraints, this discrepancy is around 20 %, while the effect diminishes to about 10 % for constrained fits. Furthermore, the overall width grows with increasing centre-of-mass energies of 7 to 13 TeV by additional 10 %. The smallest peak widths are obtained by constraining both, the primary vertex position and K_S^0 masses. Therefore, this mass observable is used throughout this analysis. In case of the normalisation channel $B^0 \to \phi K_S^0$, only the K_S^0 candidate is constrained but not the ϕ candidate, since the natural with of the ϕ meson is larger than the detector resolution and therefore a constraint would degrade mass resolution.

6.5.2 PID corrections

Particle identification variables provide an excellent indication to distinguish different types of particles. For this analysis, the separation between pions and kaons is of particular importance, since the final state particles of both the signal and normalisation channel, consist only of these hadrons. Among others, a large part of the information on which PID variables are derived from is provided by the RICH subdetectors (see Sec. 3.2.2). An accurate simulation of the ProbNN variables is challenging. This could lead to differences between the distributions of simulation and data. Besides reweighting observables, which was done for the kinematic observables, a common method to correct such PID variables at LHCb is a data-driven resampling method. A set of clean calibration data are obtained by selecting the decay channel $D^* \to D^0 \pi^0$, where the D^0 meson decays into $K^- \pi^+$. In the selection

process, no PID information and only loose topological cuts are applied in order to maintain smooth kinematic distributions. These samples are produced centrally inside the LHCb collaboration and provided by the PID Calibration Group [167]. To be able to resample PID information from this calibration data, track kinematics and event multiplicities are modelled using an unbinned kernel density estimation (KDE) approach. Simulated PID variables are then resampled from this calibration data in dependency of the transverse momentum, p_T, the pseudo-rapidity, η, and track multiplicity, nTracks, as those observables are correlated to the PID. During the optimisation process of the selection, no single particle PID information is used for the cut but a combination. For the signal channel $B_s^0 \to K_S^0 K_S^0$, the combined variable is the minimum probability of all long track pion candidates to be an actual pion, called min(πProbNNπ). Fig. 6.20 shows both, a single probability prediction for a pion to be an actual pion and the minimum of all four long track pions, determined on simulated samples and using the resampling approach. While the former distribution features only small variations, the discrepancies become larger in the combined information.

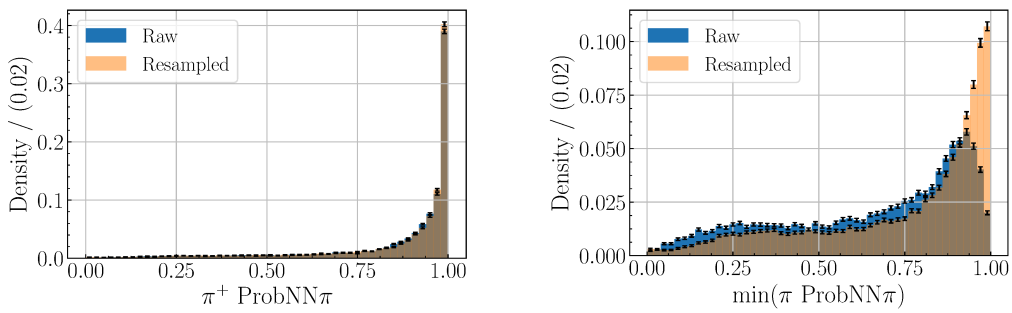


FIGURE 6.20: Comparison of ProbNNπ for a single pion (left) and min(πProbNNπ) (right) distributions taken from $B_s^0 \to K_S^0 K_S^0$ MC (raw, blue) and resampled calibration samples (resampled, orange).

6.6 Efficiency determinations

One crucial part of any branching ratio measurement is the precise determination of selection efficiencies, since these are direct inputs to the *master formula* as given in Eq. (6.4). For both decay channels and data subsamples the efficiencies are determined separately and evaluated on the corresponding MC samples. However, cross-checks on data are applied wherever possible to validate the efficiency values and to assign systematic uncertainties whenever needed, see Sec. 6.8. The total efficiency, ε_{tot}, can be separated in efficiency factors if the efficiency of a selection is to be determined with respect to the previous selection step, outlined with a vertical bar in the following equation:

$$\begin{aligned}\varepsilon_{\text{tot}} =& \varepsilon_{\text{geo}} \times \varepsilon_{\text{strip,reco|geo}} \times \varepsilon_{\text{tt|strip,reco}} \\ &\times \varepsilon_{\text{trig|tt}} \times \varepsilon_{\text{sel|trig}} \times \varepsilon_{\text{mva+pid|sel}} \times \varepsilon_{\text{tight|mva+pid}}.\end{aligned} \tag{6.24}$$

An overview of all efficiency factor values is given in Sec. 6.6.6, while the individual efficiencies are discussed in the following.

6.6.1 Geometrical acceptance

As discussed in Sec. 3.2, the probability of a b quark to hadronise inside the LHCb detector is in the order of 25 %. To reconstruct the signal decay chains, all final state particles must lie in the acceptance of the detector, reducing the aforementioned probability. To account for this, a generator level cut is introduced, so that only $B_s^0 \to K_S^0 K_S^0$ and $B^0 \to \phi K_S^0$ candidates of which all decay products lie between 0.01 and 0.40 rad relative to the beam axis, are considered. The efficiency of such requirement, ε_{geo}, is taken from the generator statistics tables provided by LHCb's simulation group.

6.6.2 Reconstruction and stripping

The efficiency $\varepsilon_{\text{strip,reco}}$ encompasses the track reconstruction, decay chain fitting and centralised loose preselection as described in Sec. 6.3.1. Although a particle trajectory lies in the acceptance of the detector, it is still possible that it is not reconstructed as track. For LHCb, the probability to reconstruct a long track pion is about 86 %, whereby hadronic interactions with detector material being most influential on the loss of efficiency. Due to their higher masses, kaons exhibit a slightly higher reconstruction efficiency of about 89 % [168]. Both efficiencies are furthermore strongly correlated to the particles' momentum.

6.6.3 Track categories

A sample splitting by track type is performed as early as possible, i.e. after stripping selection. The corresponding efficiency, ε_{tt}, is determined as fraction between LL and LD on MC samples. Therefore the equation $\varepsilon_{\text{LL}} + \varepsilon_{\text{LD}} = 1$ is fulfilled by construction.

6.6.4 Trigger efficiencies

The trigger efficiency, $\varepsilon_{\text{trig}}$, actually consists itself of three parts, based on the trigger stages L0, HLT1 and HLT2 (see Sec. 3.2.3). For detector data, at least one trigger in each stage must have a positive decision in order to store an event. In contrast, for simulated samples the trigger decision is calculated for every event independent of this requirement. All nominal trigger efficiencies are taken from simulated samples.

To validate these efficiencies a data driven approach is used. Noticed differences, especially effects not cancelling out in the ratio of Eq. (6.4), are taken as systematic uncertainty (see Sec. 6.8.5). In the year 2012, the trigger configuration keys (TCK) contained significant changes between the first to the second half of the year, whereby the aim of these changes was to improve the trigger efficiency of downstream tracks. At the same time, cuts on the flight distance of V^0 particles were tightened, leading to an efficiency loss for K_S^0 decaying into long track pions. For the decay channel $B_s^0 \to K_S^0 K_S^0$, these two modifications cancel each other in the case of the LD sample, resulting in similar trigger efficiencies. For the LL samples, the efficiency drops by about 50 %. To receive a combined efficiency for 2012, both configurations are weighted by the individual integrated luminosities recorded in the corresponding data taking period. The same approach is performed for the $B^0 \to \phi K_S^0$ samples.

6.6.5 Further selection efficiencies

The efficiency loss due to the preselection described in Sec. 6.3.2 are summarised in the efficiency ε_{sel}. After the classifier training, the modified PID variables as well as the classifier output are optimized simultaneously in two dimensions. Hence, also the efficiency $\varepsilon_{\text{mva+pid}}$ comprises the influence of both requirements. In the last selection step the mass windows of the intermediate particles are narrowed to further reduce combinatorial background, corresponding to the efficiency $\varepsilon_{\text{tight}}$. The invariant mass of L (D) K_S^0 and ϕ mesons must lie in a mass window of ± 12 MeV (± 18 MeV) around the nominal K_S^0 and ϕ meson mass, respectively. Furthermore, the last remaining multiple candidates, which have a proportion of about a per mill, are removed in this selection step by randomly choosing one candidate per event.

Finally, also the allowed trigger lines are adjusted in order to reduce systematic effects, which are caused by a misconfiguration of the trigger. It has been decided that the Topo4Body lines are omitted for all samples, while the Topo3Body lines are excluded for the Run II, LD samples only. Due to this decision, the signal efficiency is reduced by less than 5 %. These requirements are applied after the MVA training, so that the number of combinatorial background candidates is not reduced, possibly decreasing the performance of the classifier(s). An overview of the efficiencies due to the "tight" selection is given in Tab. 6.12.

6.6.6 Efficiency summary

A complete list of all selection efficiencies is given in Tab. 6.13 for the signal and in Tab. 6.14 for the normalisation channel, respectively. The geometrical efficiency increases slightly with the years for both decay modes, while the overall acceptance is better for $B^0 \to \phi K_S^0$ decays. A reason for this is that only candidates where all four pions (two pions and two kaons, respectively) which point into the LHCb detector are accepted. With higher momenta, these particles form a more focussed cone around the B meson and hence are more probable to point into the acceptance

TABLE 6.12: Efficiencies for the last selection step called "tight". It consists of tightening the mass windows of the K_S^0 candidate (and also ϕ candidate in the case of the decay $B^0 \to \phi K_S^0$) and applying a random candidate selection. The fraction of multiple candidate is also given in the table. Although the mass window is left broader for K_S^0 candidate decaying into downstream pions, the cut is less efficient.

Decay	Eff./Frac.	Run I, LL	Run I, LD	Run II, LL	Run II, LD
$B_s^0 \to K_S^0 K_S^0$	$\varepsilon_{K_S^0,\,\text{tight}}$	0.981 ± 0.003	0.84 ± 0.01	0.967 ± 0.002	0.838 ± 0.003
	$f_{\text{Mult.cand.}}$	0	0.0008	0	0.0010
$B^0 \to \phi K_S^0$	$\varepsilon_{K_S^0,\,\text{tight}}$	0.993 ± 0.001	0.867 ± 0.002	0.991 ± 0.001	0.869 ± 0.002
	$\varepsilon_{\phi,\,\text{tight}}$	0.916 ± 0.003	0.925 ± 0.002	0.901 ± 0.002	0.902 ± 0.002
	$f_{\text{Mult.cand.}}$	0.0001	0.0019	0.0001	0.0013

of LHCb altogether. While the stripping and reconstruction efficiencies seem to decrease throughout the years, it is important to see this is associated with the trigger efficiency. For example, a very loose stripping would increase the stripping efficiency, but reduce the trigger efficiency. Hence, a good compromise is the key point to optimise the stripping selection. For $B^0 \to \phi K_S^0$ decays in Run II, trigger filter requirements are directly included in the stripping. Therefore, the trigger efficiency is higher, about 50 % for 2015 and above 90 % for the year 2016. By multiplying $\varepsilon_{\text{strip, reco}}$ and $\varepsilon_{\text{trig}}$ it becomes apparent that the combined efficiency for Run II is consistently higher than in Run I, whereby the improvement ranges from a factor of two in the decay $B_s^0 \to K_S^0 K_S^0$, Run I, LL, up to almost 10 for Run I, LD in the same channel. The fraction between track types is shifted towards long tracks with higher centre-of-mass energies. This is counter-intuitive, since higher energies lead to higher momenta and therefore to K_S^0 mesons decaying at more downstream positions in the detector. The explanation for this can be again found in the stripping selection. For $B_s^0 \to K_S^0 K_S^0$ decays, the cuts in Run II are tighter for downstream decaying K_S^0 candidates than for long track K_S^0 candidates with respect to the loose MVA based decision in Run I. For the decay channel $B^0 \to \phi K_S^0$ the reason is the trigger requirement in the stripping, being biased toward long tracks. The preselection is more efficient for $B^0 \to \phi K_S^0$ decays, because many requirements are only applied on the K_S^0 mesons, but not the ϕ meson, e.g. the ΔZ cut introduced in Eq. (6.10). Different figures of merit for signal and normalisation channel explain the difference in efficiency of $\varepsilon_{\text{mva+pid}}$.

6.6.7 Combinations

All efficiencies are calculated per year and track type separately. Since the data samples are merged to Run I and Run II, the combined total efficiencies are obtained as a weighted mean of the corresponding years, with the integrated luminosity as weights,

TABLE 6.13: Efficiency overview for the signal channel $B_s^0 \to K_S^0 K_S^0$.

	$B_s^0 \to K_S^0 K_S^0$							
	2011		2012		2015		2016	
Efficiencies								
ε_{geo}	0.138 14	$\pm$ 0.000 22	0.142 64	$\pm$ 0.000 25	0.158 97	$\pm$ 0.000 35	0.1585	$\pm$ 0.0004
$\varepsilon_{\text{strip, reco}}$	0.019 75	$\pm$ 0.000 11	0.015 75	$\pm$ 0.000 09	0.012 03	$\pm$ 0.000 11	0.012 45	$\pm$ 0.000 05
ε_{LL}	0.1290	$\pm$ 0.0019	0.1183	$\pm$ 0.0018	0.1798	$\pm$ 0.0035	0.1799	$\pm$ 0.0017
$\varepsilon_{\text{trig, LL}}$	0.054	$\pm$ 0.004	0.0497	$\pm$ 0.0017	0.092	$\pm$ 0.006	0.1069	$\pm$ 0.0032
$\varepsilon_{\text{sel, LL}}$	0.757	$\pm$ 0.030	0.62	$\pm$ 0.04	0.574	$\pm$ 0.035	0.715	$\pm$ 0.014
$\varepsilon_{\text{mva+pid, LL}}$	0.674	$\pm$ 0.012	0.667	$\pm$ 0.017	0.613	$\pm$ 0.010	0.618	$\pm$ 0.004
$\varepsilon_{\text{tight, LL}}$	0.981	$\pm$ 0.004	0.946	$\pm$ 0.010	0.963	$\pm$ 0.005	0.9671	$\pm$ 0.0019
ε_{LD}	0.8710	$\pm$ 0.0019	0.8817	$\pm$ 0.0018	0.727	$\pm$ 0.004	0.7285	$\pm$ 0.0020
$\varepsilon_{\text{trig, LD}}$	0.0086	$\pm$ 0.0006	0.006 37	$\pm$ 0.000 23	0.0559	$\pm$ 0.0024	0.0698	$\pm$ 0.0013
$\varepsilon_{\text{sel, LD}}$	0.759	$\pm$ 0.029	0.68	$\pm$ 0.04	0.749	$\pm$ 0.019	0.772	$\pm$ 0.008
$\varepsilon_{\text{mva+pid, LD}}$	0.424	$\pm$ 0.011	0.490	$\pm$ 0.015	0.320	$\pm$ 0.005	0.2769	$\pm$ 0.0022
$\varepsilon_{\text{tight, LD}}$	0.819	$\pm$ 0.014	0.859	$\pm$ 0.015	0.830	$\pm$ 0.008	0.8288	$\pm$ 0.0035

TABLE 6.14: Efficiency overview for the normalisation channel $B^0 \to \phi K_S^0$.

	$B^0 \to \phi K_S^0$							
	2011		2012		2015		2016	
Efficiencies								
ε_{geo}	0.1561	$\pm$ 0.0014	0.1656	$\pm$ 0.0024	0.1695	$\pm$ 0.0016	0.1695	$\pm$ 0.0016
$\varepsilon_{\text{strip, reco}}$	0.023 51	$\pm$ 0.000 06	0.020 37	$\pm$ 0.000 07	0.014 95	$\pm$ 0.000 09	0.008 482	$\pm$ 0.000 029
ε_{LL}	0.2779	$\pm$ 0.0012	0.2629	$\pm$ 0.0015	0.2940	$\pm$ 0.0026	0.3146	$\pm$ 0.0016
$\varepsilon_{\text{trig, LL}}$	0.2511	$\pm$ 0.0021	0.381	$\pm$ 0.018	0.498	$\pm$ 0.005	0.9290	$\pm$ 0.0016
$\varepsilon_{\text{sel, LL}}$	0.8626	$\pm$ 0.0034	0.853	$\pm$ 0.005	0.909	$\pm$ 0.004	0.9197	$\pm$ 0.0017
$\varepsilon_{\text{mva+pid, LL}}$	0.9502	$\pm$ 0.0023	0.9372	$\pm$ 0.0035	0.872	$\pm$ 0.005	0.8592	$\pm$ 0.0023
$\varepsilon_{\text{tight, LL}}$	0.9035	$\pm$ 0.0032	0.913	$\pm$ 0.004	0.895	$\pm$ 0.005	0.8905	$\pm$ 0.0022
ε_{LD}	0.7221	$\pm$ 0.0012	0.7371	$\pm$ 0.0015	0.7060	$\pm$ 0.0026	0.6854	$\pm$ 0.0016
$\varepsilon_{\text{trig, LD}}$	0.2110	$\pm$ 0.0013	0.235	$\pm$ 0.008	0.4640	$\pm$ 0.0034	0.9203	$\pm$ 0.0011
$\varepsilon_{\text{sel, LD}}$	0.9355	$\pm$ 0.0016	0.9310	$\pm$ 0.0022	0.9260	$\pm$ 0.0026	0.9265	$\pm$ 0.0011
$\varepsilon_{\text{mva+pid, LD}}$	0.7374	$\pm$ 0.0030	0.739	$\pm$ 0.004	0.709	$\pm$ 0.005	0.7017	$\pm$ 0.0021
$\varepsilon_{\text{tight, LD}}$	0.8020	$\pm$ 0.0032	0.798	$\pm$ 0.004	0.769	$\pm$ 0.005	0.7681	$\pm$ 0.0023

$$\varepsilon_{\text{tot, Run I(Run II), TT}} = \frac{\varepsilon_{\text{tot, 2011(2015), TT}} \cdot \mathcal{L}_{\text{int, 2011(2015)}} + \varepsilon_{\text{tot, 2012(2016), TT}} \cdot \mathcal{L}_{\text{int, 2012(2016)}}}{\mathcal{L}_{\text{int, 2011(2015)}} + \mathcal{L}_{\text{int, 2012(2016)}}}, \tag{6.25}$$

with TT $\in$ {LL, LD}. Thus, the combined efficiencies of $B_s^0 \to K_S^0 K_S^0$ result to:

$$\begin{aligned}
\varepsilon_{\text{tot, Run I, LL}} &= (6.71 \pm 0.40) \times 10^{-6}, \\
\varepsilon_{\text{tot, Run I, LD}} &= (4.21 \pm 0.23) \times 10^{-6}, \\
\varepsilon_{\text{tot, Run II, LL}} &= (1.05 \pm 0.11) \times 10^{-5}, \\
\varepsilon_{\text{tot, Run II, LD}} &= (1.54 \pm 0.09) \times 10^{-5}.
\end{aligned}$$

While the efficiency of LD is lower as for LL in Run I, the opposite behaviour is visible for Run II. This is due to the TCK changes in 2012, which is also discussed in Sec. 6.6.4. Furthermore, the overall efficiencies are increased for Run II by a factor of two, also indicating a more efficient stripping in these years.

TABLE 6.15: Background efficiency statistics for the selection steps of $B_s^0 \to K_S^0 K_S^0$, calculated on upper and lower data side bands. After applying selection and a MVA, background contributions are almost completely rejected.

Bkg. Efficiencies	2011		2012		2015		2016	
$\varepsilon_{\text{bkg; sel, LL}}$	0.0144	$\pm$ 0.0009	0.0178	$\pm$ 0.0005	0.001 57	$\pm$ 0.000 07	0.003 11	$\pm$ 0.000 10
$\varepsilon_{\text{bkg; sel, LD}}$	0.0025	$\pm$ 0.0004	0.001 76	$\pm$ 0.000 16	0.016 95	$\pm$ 0.000 23	0.032 32	$\pm$ 0.000 30
$\varepsilon_{\text{bkg; mva+pid, LL}}$	0.000 051	$\pm$ 0.000 029	0.000 036	$\pm$ 0.000 012	0.000 15	$\pm$ 0.000 06	0.000 122	$\pm$ 0.000 016
$\varepsilon_{\text{bkg; mva+pid, LD}}$	0.000 29	$\pm$ 0.000 17	0.000 35	$\pm$ 0.000 12	0.000 013	$\pm$ 0.000 005	0.000 011	$\pm$ 0.000 001
$\varepsilon_{\text{bkg; tot, LL}}$ $[10^{-6}]$	0.7	$\pm$ 0.4	0.64	$\pm$ 0.22	0.23	$\pm$ 0.09	0.38	$\pm$ 0.05
$\varepsilon_{\text{bkg; tot, LD}}$ $[10^{-6}]$	0.7	$\pm$ 0.4	0.61	$\pm$ 0.21	0.22	$\pm$ 0.09	0.36	$\pm$ 0.05

To calculate the combined efficiencies for the normalisation channel $B^0 \to \phi K_S^0$, the same Eq. (6.25) is used, yielding

$$\begin{aligned}
\varepsilon_{\text{tot, Run I, LL}} &= (2.27 \pm 0.08) \times 10^{-4}, \\
\varepsilon_{\text{tot, Run I, LD}} &= (3.17 \pm 0.08) \times 10^{-4}, \\
\varepsilon_{\text{tot, Run II, LL}} &= (2.62 \pm 0.06) \times 10^{-4}, \\
\varepsilon_{\text{tot, Run II, LD}} &= (4.17 \pm 0.07) \times 10^{-4}.
\end{aligned}$$

The higher efficiencies of about 30 (70 in case of Run I, LD) are due to different reasons. First, the reconstruction of this decay is more straightforward, as the ϕ decays promptly so that the two kaon tracks stem directly from the $B_{(s)}^0$ meson. Secondly, a dedicated trigger line for ϕ mesons exists, increasing the probability for $B^0 \to \phi K_S^0$ decays to be stored. Finally, the combinatorial background is much less for this decay channel, allowing for looser cut requirements which consequently leads to an overall higher efficiency.

6.6.8 Background subtraction

Two selection steps are mainly responsible for the background suppression. Physical background contributions are excluded by the preselection. Moreover, the MVA rejects most of combinatorial background, i.e. randomly combined tracks. To quantify the suppression factor, a background efficiency is calculated on the lower and upper side band of the invariant and mass constraint four pion mass of 5000 to 5230 MeV and 5420 to 5600 MeV. The respective values which are based on samples after stripping and divided by years, can be found in Tab. 6.15.

Between both Runs, significant differences are visible. While the background rejection for Run I is higher for LD than LL, the opposite is the case for Run II. This is explained by the different stripping approaches, i.e. MVA for Run I and rectangular cuts for Run II, shifting the fractions of LL and LD. The total background on the other hand, $\varepsilon_{\text{bkg; tot}}$, is consistent between the track categories within a year. Moreover, the

TABLE 6.16: Normalisation constants α for the $\mathcal{BR}$ calculation, calculated after Eq. (6.4).

Dataset	$\alpha[10^{-6}]$
Run I LL	1.80 ± 0.22
Run I LD	4.01 ± 0.54
Run II LL	0.86 ± 0.12
Run II LD	0.67 ± 0.07

suppression is greater for Run II than Run I, indicating a tighter stripping selection in Run I.

6.7 Combination of the $B_s^0 \to K_S^0 K_S^0$ signal yields

To determine the total branching ratio of $B_s^0 \to K_S^0 K_S^0$, a simultaneous unbinned maximum likelihood fit is applied to all four data subsets, i.e. Run I, LL, Run I, LD, Run II, LL and Run II, LD. The mass observable chosen for this fit has a PV and mass constraint (see Sec. 6.5.1) in the range of 5000 to 5600 MeV. As described in the previous chapters, the signal model is extracted from simulated $B_s^0 \to K_S^0 K_S^0$ samples with the full selection applied. The parameters of the signal shape are fixed to the values obtained from fits to simulated samples. The background shape is floating, with the initial fit parameters set to values obtained from fits to side band of data. To fit the background shape, the signal area between 5230 MeV to 5420 MeV is cut out. In addition to the model describing the B_s^0 meson, the same signal shape is copied and shifted by the nominal $B_s^0 - B^0$ meson mass difference of $\Delta m(B_s^0, B^0) = 87.3$ MeV [169] to be able to describe a possible $B^0 \to K_S^0 K_S^0$ component. The signal yield is included as fraction, f_{B^0/B_s^0}, with respect to the signal channel $B_s^0 \to K_S^0 K_S^0$. Besides f_{B^0/B_s^0}, only the predicted branching ratio, $\mathcal{BR}_{\text{pred}}$, is shared between the samples in the fit. The latter parameter is connected to the number of fitted signal candidates, $N_{\text{sig},\,i}$, and normalisation constant, α_i, on each specific sample i via

$$\mathcal{BR}_{\text{pred}} = N_{\text{sig},\,i} \cdot \alpha_i, \tag{6.26}$$

which is a simplification of Eq. (6.4). The normalisation constants are Gaussian constrained in their uncertainties to increase the robustness of the fit. They can be found in Tab. 6.16.

This fitting procedure has been validated using pseudo datasets before the relevant mass region has been revealed (see Sec. 6.8.1. The fit results to the nominal data mass fits are given in Tab. 6.17. Each data sample with the corresponding fit projection is shown in Fig. 6.21. Consistent with expectations, the number of signal candidates in Run II is higher than in Run I. A downward fluctuation can be observed in the Run I, LL sample. As a cross-check, the branching ratios have been determined individually for Run I and Run II. The obtained values are compatible within $1.8\,\sigma$. Futhermore, a small component of $B^0 \to K_S^0 K_S^0$ is visible.

TABLE 6.17: Results of the simultaneous fit to the invariant mass of $m(K_S^0 K_S^0)$.

Parameter	Run I, LL	Run I, LD	Run II, LL	Run II, LD	Factor	Status
$\mathcal{BR}_{\text{pred}}$		9.5 ± 1.9			10^{-6}	Free
f_{B^0/B_s^0}		0.30 ± 0.13			1.0	Free
N_{sig}	5.0 ± 1.2	2.4 ± 0.6	10.6 ± 2.6	13.8 ± 3.1	1.0	$\mathcal{BR}/\alpha$
N_{bkg}	10.2 ± 3.4	3.3 ± 2.1	7.5 ± 3.1	12 ± 4	1.0	Free
α	1.87 ± 0.21	3.9 ± 0.5	0.89 ± 0.11	0.69 ± 0.07	10^{-6}	Gaussian constr.

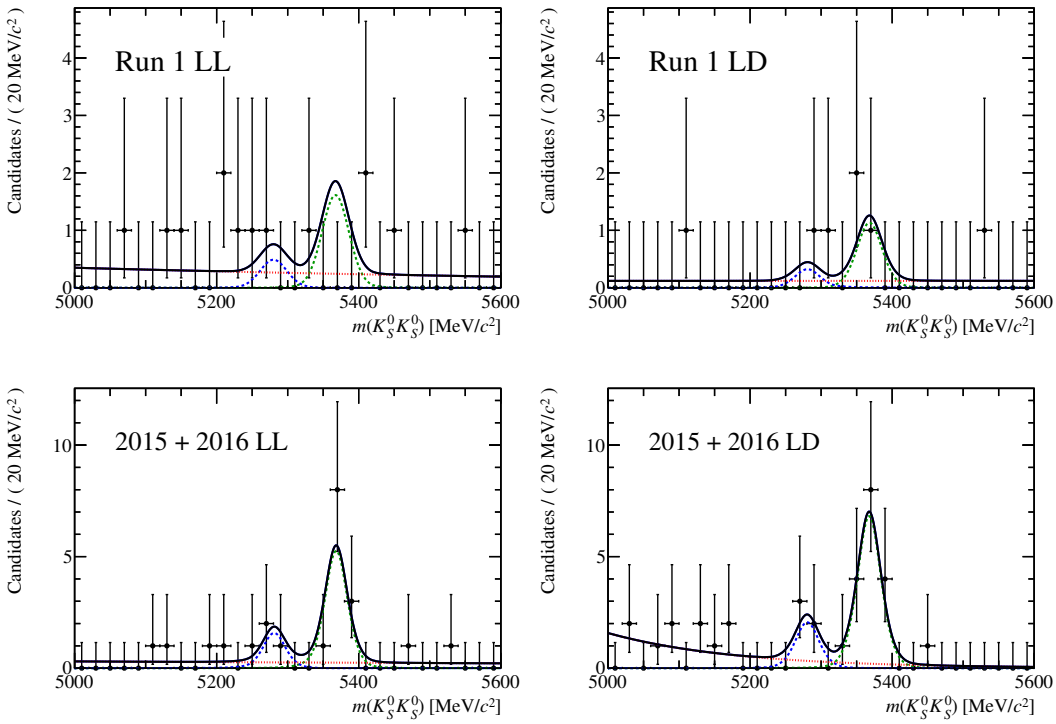


FIGURE 6.21: Fits to the invariant mass distribution $m(K_S^0 K_S^0)$ of the signal decay channel. The black (solid) curve represents the complete model, the B_s^0 signal component is given in green (dashed), the B^0 signal is given in blue (dashed) and the background component in red (dotted) [166].

6.8 Systematic uncertainties

Systematic uncertainties can be divided into two main sources. The first one is due to uncertainties on external inputs, which have already been discussed in Sec. 6.2.1. Secondly, systematic uncertainties have to be assigned if an effect is not related to the statistics of the data sample which itself is covered by the statistical uncertainty. By choosing a normalisation channel similar to the signal channel, specific systematic uncertainties are expected to cancel out in the ratio of Eq. (6.4). Nevertheless, these cancellations reach their limits at some point, so that a deeper investigation of the differences and their influences on the predicted branching ratio is necessary. For example, the final state particles are different between the decay channels $B_s^0 \to K_S^0 K_S^0$ and $B^0 \to \phi K_S^0$, i.e. four pions versus two pions and two kaons, whereby the tracking efficiencies of these particles differ. The systematic effects are estimated in a conservative way and discussed in detail in the following chapter.

6.8.1 Fit Model Validation

Theoretically, the maximum likelihood estimator provides unbiased results in the case of infinite statistics. However, in this analysis, most data samples only contain a few entries of $\mathcal{O}(10)$ candidates after the complete selection. Therefore, a study based on pseudo-experiments is performed to determine a possible intrinsic bias due to the fit procedure.

The pseudo datasets are obtained in the following way: The background candidates are duplicated from the lower and upper mass side bands. To receive an estimation of background candidates in the signal region, the background model has been integrated over the blinded area of 5230 to 5420 MeV in the invariant mass distribution. As statistical fluctuations are expected in the yields, the number of background candidates inserted in that area are not exactly the integral value, but sampled from a Poisson distribution with a rate λ of the integral. This is done individually for each data sample. The number of expected signal candidates is assessed by using Eq. (6.26), with the normalisation constants derived from Tab. 6.16. To validate the fit model, a of $\mathcal{BR}_{\text{gen}} = 10 \times 10^{-6}$ is taken, following the previous Belle measurement [136]. After the unblinding, the branching ratio is changed to the outcome of the nominal mass fit. Again, the exact number of signal candidates is drawn from a Poisson distribution with the mean of this value. In the same way, a possible $B^0 \to K_S^0 K_S^0$ distribution is added, which shares the same shape as the $B_s^0 \to K_S^0 K_S^0$ distribution, but is shifted by the nominal mass difference between a B_s^0 and B^0 meson.

In total, 1000 pseudo datasets have been generated and fitted with the nominal mass model. Fig. 6.22 shows the distribution of fitted branching ratios $\mathcal{BR}_{\text{toy, fit}}$. A small bias between the generated and fitted branching ratios can be seen, which is treated as a systematic uncertainty of $\Delta_{\text{fitbias}} = 0.036$. Furthermore, a scatter plot of $\mathcal{BR}_{\text{toy, fit}}$ against the fit uncertainty is visualised, indicating a statistical uncertainty

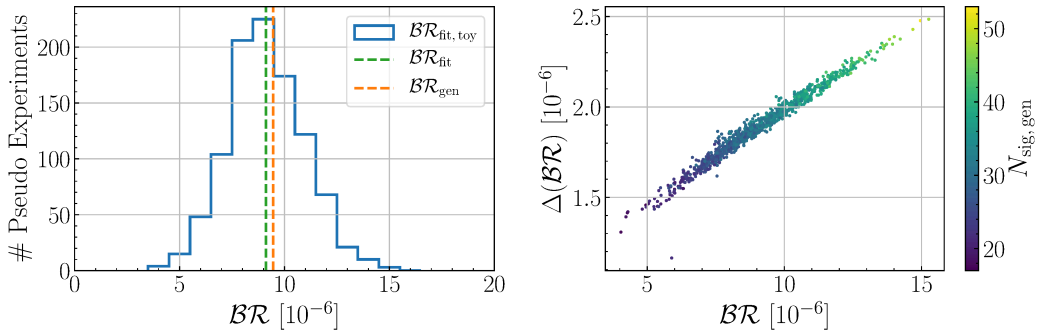


FIGURE 6.22: Summary of fits to pseudo datasets. The left plot shows the distribution of fitted branching ratios for the decay $B_s^0 \to K_S^0 K_S^0$ (blue, solid). Additionally, the generated value (orange, dashed) and the arithmetic mean of all fitted values (green, dashed) is given. A slight deviation between this two values can be seen. The right figure shows a scatter plot of the fitted $\mathcal{BR}$ and its statistical uncertainty determined directly from the fit.

in the order of 20 % for the generated branching ratio. To exclude a bias of an individual sample, the distributions of all fitted normalisation constants are also given in Fig. 6.23. An overall good agreement between generated and fitted values can be seen, while the largest deviation is found in the Run II, LD sample.

Another relevant check is to examine a possible dependency between the signal yields of $B_s^0 \to K_S^0 K_S^0$ and $B^0 \to K_S^0 K_S^0$ decays. Therefore, the fraction f_{B^0/B_s^0} is set to 0, 0.2 and 0.5, respectively, where the second value corresponds to the PDG value expectations. In brief, no dependency is observed. However, the generated and fitted fraction is off by 5.9 % when generating with f_{B^0/B_s^0} set to the value obtained from the nominal fit. To estimate the bias conservatively, a systematic uncertainty of $\Delta_{\text{fitbias},B^0} = 0.059$ is added to the measurement of the $B^0 \to K_S^0 K_S^0$ decay.

6.8.2 Fit Model Choice

Another systematic uncertainty originates from the choice of a specific model to describe the signal components. To quantify this effect, alternative PDFs are used to model the signal components in the invariant mass distributions. For the observable $m(K^+K^-)$, the ϕ meson component is described by a non-relativistic Breit-Wigner (BW) function [170],

$$P_{\text{BW}}(m; m_\phi, \Gamma_\phi) = \frac{1}{(m - m_\phi)^2 + \frac{1}{4}\Gamma_\phi^2} \quad \text{with} \quad m = m(K^+K^-), \tag{6.27}$$

convoluted with a Gaussian instead of a relativistic Breit-Wigner (RBW) function. The signal components for B_s^0 and B^0 in the four-body invariant masses $m(K_S^0 K_S^0)$ and $m(K_S^0 K^+ K^-)$ are alternatively modelled by the sum of two Crystal Ball functions with tails to both, the lower and upper mass. To assign a systematic uncertainty, two different approaches are applied. As the normalisation channel is not blinded it is

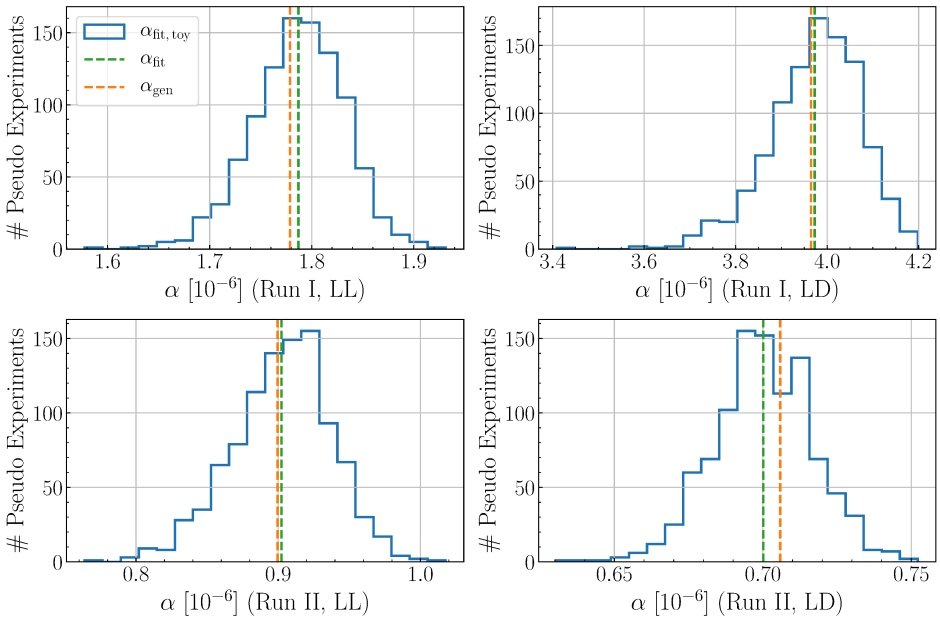


FIGURE 6.23: Distributions of the fitted normalisation constants on pseudo datasets. The corresponding sample is given in the x label description. Furthermore, the generation value (orange, dashed) and fitted mean values (green, dashed) are shown. The largest deviation between these values can be observed in Run II, LD.

TABLE 6.18: Systematic uncertainties due to the choice of a nominal fit model. The last row yields the combined uncertainties as sum in quadrature over the three individual uncertainties.

Observable	Run 1, LL	Run 1, LD	Run 2, LL	Run 2, LD
$m(K_s^0 K^+ K^-)$	0.012	0.012	0.004	0.011
$m(K^+ K^-)$	0.018	0.030	0.013	0.003
$m(K_s^0 K_s^0)$	0.006	0.006	0.006	0.006
Combined	0.022	0.033	0.015	0.013

possible to refit the invariant masses with the alternative models and determine the relative difference of the signal yields. In contrast, in case of the signal channel 1000 pseudo data samples are generated with the nominal model and fitted with the alternative model. As uncertainty, the relative difference between the averaged fitted branching ratio of the nominal and alternative model is assigned. This can be seen as most conservative estimation due to the fit model choice. Since this is a simultaneous fit in all four data samples and the branching ratio is a shared parameter, the uncertainty is added to all four samples. A summary of the systematic uncertainties is given in Tab. 6.18.

TABLE 6.19: Estimated systematic uncertainties due to the application of the multivariate analysis. The first two rows show the values in dependency of the corresponding decay channel. In the last row, these values are added in quadrature to receive a combined systematic uncertainty.

Decay/Syst.	Run I, LL	Run I, LD	Run II, LL	Run II, LD
$B_s^0 \to K_S^0 K_S^0$	0.022	0.040	0.013	0.031
$B^0 \to \phi K_S^0$	0.001	0.001	0.003	0.004
Δ_{MVA}	0.023	0.040	0.014	0.031

6.8.3 Multivariate Classification

Studies given in Sec. 6.3.5 indicate that a kinematic reweighting is entirely reasonable and is therefore carried out in the nominal classifier training. However, while this reweighting reduces the differences between data and simulated samples, it might not perfectly compensate for all inconsistencies. Therefore, a systematic uncertainty is determined to quantify this effect. This is done by calculating the signal efficiency of the BDT with and without applying kinematic weights to simulated samples. The results of this approach are shown in Tab. 6.19. Samples containing downstream tracks show larger deviations, which clearly indicate that it is more important for them to be reweighted.

6.8.4 Particle Identification

As described in Sec. 6.5.2, this analysis does not rely on PID information provided by simulated samples but resamples the values from clean calibration samples. Nevertheless, differences between the signal and normalisation channel or rather between simulated/resampled samples and data exist. This needs to be quantified. Multiple possible sources of systematic uncertainties are determined. First, the MC sample size of both decays $B_s^0 \to K_S^0 K_S^0$ and $B^0 \to \phi K_S^0$ are sufficient, which means that the resampling method might not be accurate. Thus, a possible systematic uncertainty is calculated by creating new samples using the random sampling with replacement method, where every new sample has the same number of candidates as the nominal samples. The variance in the signal efficiency after the nominal PID cut applied is then determined as systematic uncertainty. A second possible systematic uncertainty raises from the modelling of the calibration data. Thus, an alternative KDE model with a different bandwidth parameter is assumed in order to describe the underlying data. The relative differences between nominal and alternative model is taken systematic uncertainty. Additionally, a small uncertainty of 0.1 % has been assigned to the use of the sweighting method. Finally, the significant flight distance of the K_S^0 meson is not considered in the resampling process. A systematic uncertainty is calculated by reweighting the resampled ProbNNπ observable in bins of the K_S^0 flight distance and determine the relative difference of the nominal cut efficiency on

TABLE 6.20: Assigned PID systematic uncertainties. The systematic due to the K_S^0 flight distance is only relevant for long tracks, since no cuts are applied for downstream pions. Hence, the assumed systematic uncertainties are the same for the LL and LD samples for each run. In the last row, these values are added in quadrature to receive a combined systematic.

Systematic	Run 1, LL	Run 1, LD	Run 2, LL	Run 2, LD
$\Delta_{\text{PID, Modelling}}$	0.0006	0.0009	0.0015	0.0022
$\Delta_{\text{PID, MC stat.}}$	0.0032	0.0047	0.0046	0.0032
$\Delta_{\text{PID, FD}}$	0.0064	0.0064	0.0258	0.0258
$\Delta_{\text{PID, Comb}}$	0.0072	0.0080	0.0262	0.0260

"raw" simulation of $B^0 \to \phi K_S^0$ decays as well as on reweighted datasets. A summary of all systematic uncertainties due to the PID is given in Tab. 6.20. As can be seen, the uncertainty of the Run II samples is larger than in Run I and dominated by the flight distance of the K_S^0 candidates.

6.8.5 Trigger Systematics

The nominal trigger efficiencies are derived from simulated samples, where a possible inconsistency in the description could bias the result. The first trigger stage, L0, is implemented in hardware, while the HLT is software-based (see also Sec. 3.2.3). Hence, also the simulation of both triggers is taken over by different methodologies. For this reason, systematic uncertainties are determined independently for each trigger stage.

L0 Hadron

Similar to the resampling method that is applied for the PID variables (see Sec. 6.5.2), efficiency tables based on calibration can be used to estimate the L0 trigger efficiency in a data-driven way for the `L0_Hadron` trigger line. To produce such tables, the same calibration data as for the PID corrections are utilised. For each individual track, the efficiency is determined in bins of the transverse energy, calculated following Eq. (6.11), and the actual hit position in the HCAL. This position is divided into *inside* and *outside* of the detector, whereby the number of modules in the inside area is doubled and thus the resolution is improved, see also Sec. 3.2.2. To receive a combined trigger efficiency, the single track efficiencies are combined using the equation

$$\varepsilon_{\text{comb}} = \varepsilon(P_1) \cdot (1 - \varepsilon(P_2)) + \varepsilon(P_2) \cdot (1 - \varepsilon(P_1)) + \varepsilon(P_1)\varepsilon(P_2), \tag{6.28}$$

where P_1 and P_2 are the oppositely charged pions (kaons) stemming from the decay $K_S^0 \to \pi^+\pi^-$ ($\phi \to K^+K^-$). At a second step, efficiencies of both K_S^0 candidates (K_S^0 and ϕ candidates) are combined exploiting the same equation. Tab. 6.21 summarises

TABLE 6.21: Trigger efficiency of the `L0_Hadron` trigger line TOS, calculated with the help of data-driven efficiency tables and separated by decay channels. The difference between those efficiencies is determined in the last column.

	$\varepsilon_{\text{comb}}$, $B_s^0 \to K_S^0 K_S^0$	$\varepsilon_{\text{comb}}$, $B^0 \to \phi K_S^0$	Difference
2011 LL	$0.212\,667 \pm 0.000\,030$	$0.275\,702 \pm 0.000\,026$	$0.063\,04 \pm 0.000\,04$
2011 LD	$0.242\,419 \pm 0.000\,014$	$0.307\,612 \pm 0.000\,018$	$0.065\,193 \pm 0.000\,023$
2012 LL	$0.195\,324 \pm 0.000\,016$	$0.258\,082 \pm 0.000\,017$	$0.062\,758 \pm 0.000\,023$
2012 LD	$0.221\,097 \pm 0.000\,007$	$0.281\,422 \pm 0.000\,012$	$0.060\,325 \pm 0.000\,014$

the obtained efficiencies. The difference between signal and normalisation channel efficiency is designated as systematic uncertainty, which is constantly around 6 %. Because no efficiency tables exist for Run II samples by the time of writing, the luminosity and track type weighted systematic uncertainties from Tab. 6.21 are also assigned to the Run II data.

HLT Trigger

For the signal and normalisation mode, different sets of trigger lines are used for the trigger selection, see Sec. 6.3.3. While for the selection of the decay channel $B_s^0 \to K_S^0 K_S^0$ only relies on the topological triggers (Topo), the channel $B^0 \to \phi K_S^0$ might be triggered by a line optimised to select $\phi \to K^+K^-$ decays (IncPhi). The nominal efficiencies are taken from simulated samples. For the Topo triggers, a possible bias is expected to cancel mostly in the ratio of Eq. (6.4). In contrast to this, an inaccurate IncPhi trigger description in MC could lead to significant systematic effects. For both, the Topo and IncPhi trigger, a systematic uncertainty is calculated separately and both are combined afterwards to obtain an overall HLT systematic.

As described in Sec. 6.3.3, the topological trigger lines rely on a combination of p_T, reconstructed (corrected) mass, DOCA, χ^2_{IP} and χ^2_{FD} of the trigger candidate. To account for a possible difference in the trigger efficiencies between both decay channels, simulated samples of $B^0 \to \phi K_S^0$ decays are reweighted to match the kinematics of $B_s^0 \to K_S^0 K_S^0$ decays in these observables. This is done in the observables p_T and χ^2_{FD} of the B candidate, since these exhibit the largest deviations. A systematic uncertainty is then assigned as the efficiency difference between the raw and reweighted distributions. Results of this procedure are given in the first column of Tab. 6.22. The largest value is seen in the Run II samples, yielding an effect of 2 %.

In the case of the IncPhi trigger, a data-driven approach is used. First, the overall benefit of the IncPhi trigger efficiency is calculated by requiring that only this trigger caused the trigger decision for the $B^0 \to \phi K_S^0$ candidate. Between the individual samples, a large difference for the efficiencies can be seen. For Run I, LL (Run I, LD) ${}_s\mathcal{W}$eighted data, the efficiency is 25 % (45 %) and for Run II, LL (Run I, LD) 5 % (10 %). Since all events need at least one positive trigger decision in order to be stored, these efficiencies hint that the topological lines improved in Run II and loose efficiency when applied to candidates containing downstream tracks.

TABLE 6.22: Relative systematics uncertainties due to high level triggers (HLT). The inclusive phi trigger (IncPhi) show the largest values, especially for Run I.

Sample	Δ_{Topo}	Δ_{IncPhi}	$\Delta_{\text{Comb, HLT}}$
Run I, LL	0.004	0.065	0.065
Run I, LD	0.000	0.106	0.106
Run II, LL	0.001	0.008	0.008
Run II, LD	0.020	0.016	0.026

6.8.6 Further systematic uncertainties

Further sources of possible systematic uncertainties are investigated and cross-checks are applied, which are summarised in this section.

The description of the signal component in the invariant mass distribution relies crucial on the information provided by simulated samples. For the signal mode, all mass shape parameters are fixed to values obtained from MC fits due to the low signal yield expected. In the case of the normalisation mode $B^0 \to \phi K_S^0$, the mean and width parameter are free in the fit, which allows to compensate for possible data and simulation disagreements. Nevertheless, the tail parameters are fixed. A systematic uncertainty due to parameter fixation is determined individually for the signal and normalisation channel. For the former, the tail parameters of the signal model are varied within the uncertainties received from simulated samples, fixed, and again fitted to the invariant mass of $m(K_S^0 K^+ K^-)$. The differences in the signal yields are taken as systematic uncertainty. For $B_s^0 \to K_S^0 K_S^0$ decays, also the mean and width parameter of the B_s^0 component are varied. One thousand pseudo-experiments are generated, fitted and the branching ratio difference between both parameter sets, the nominal and varied one, is determined. The mean branching ratio difference is taken as systematic uncertainty. Combining signal and normalisation channel, a systematic uncertainty of 2.6 % is assigned.

Hadrons can interact with nuclei of the detector material. This process generates showers of secondary particles, which are themselves absorbed by material inside the detector. Hence, these particles cannot be reconstructed by the tracking systems. Studies have shown that the hadronic interaction length, i.e. the average distance a hadron propagating through the detector before a hadronic interaction occurs, depends on the hadrons type and momentum and can differ up to 30 %. The fraction of hadronic interactions upstream of the last tracking component is about 15 % for pions and 12.5 % for kaons. Assuming that the simulation of the detector describes these interactions up to a 10 % level, a systematic uncertainty can be assigned to each particle type. For the K_S^0 mesons, this effect cancels in the efficiency ratio. Nevertheless, the two remaining pions of $B_s^0 \to K_S^0 K_S^0$ decays are replaced by kaons in $B^0 \to \phi K_S^0$ decays, yielding a small systematic uncertainty of 0.5 %.

Different vertex resolutions of data and simulated samples can also raise a systematic uncertainty. To account for this, simulated samples of $B_s^0 \to K_S^0 K_S^0$ decays are considered, where the $\Delta Z > 15\,\text{mm}$ cut is applied, see Sec. 6.3.2. Then, the actual K_S^0 candidate decay vertex position is shifted by the relative K_S^0 vertex resolution taken from simulated $B_s^0 \to K_S^0 K_S^0$ decays as well as ${}_s\mathcal{W}$eighted $B^0 \to \phi K_S^0$ data. This procedure is repeated with random combinations of decay position and vertex resolution. An inefficiency is calculated for candidates which pass the ΔZ requirement after replacement, which is then compared between the two approaches and assigned as systematic uncertainty. The obtained values are at the order of 10^{-4}, so that even when enlarging the uncertainty by a factor of 10 to create a worst-case scenario they are still negligible and thus are neglected.

Former analyses show hints of a possible misalignment in the simulation of forward VELO stations. To estimate this effect, the vertex position of K_S^0 candidates decaying at z-positions greater than 400 mm are compared. For this comparison a high statistics decay channel is needed since the number of K_S^0 candidates in the normalisation channel is too low to obtain a significant statement. Hence, $B^0 \to J/\psi K_S^0$ decays are used for this cross-check. A small relative difference of 7 % between the fraction of data and simulated candidates could be observed, which is then multiplied by the fraction of total candidates in that specific region. The resulting systematic uncertainty is 0.7 %.

Finally, a small systematic uncertainty of below 1 % is assigned to account for a misconfiguration in the trigger, only occurring in Run II of data taking.

6.8.7 Systematic uncertainty summary

A summary table of all systematic uncertainties is given in Tab. 6.23. The last row represents the combined uncertainties per sample, calculated as summation in quadrature over the individual values. To estimate the uncertainty conservatively, no possible correlation between them is considered. The uncertainties are dominated by the trigger, more specifically be the HLT for Run I and L0 for Run II. Another significant fraction is due to the chosen fit procedure to determine the branching ratio. A total relative systematic uncertainty is determined by weighting over the total number of expected signal candidates, yielding

$$\Delta(\text{syst, tot}) = 0.095. \tag{6.29}$$

Furthermore, two external uncertainties are considered. The first one stems from the external branching ratio used as input for Eq. (6.4), dominated by the $\mathcal{BR}(B^0 \to \phi K_S^0)$, and has a value of

$$\Delta(\mathcal{BR}(B^0 \to \phi K_S^0)) = 0.097. \tag{6.30}$$

Secondly, the fragmentation fraction difference between B^0 and B_s^0 mesons yields an uncertainty of

TABLE 6.23: Summary of all systematics uncertainties described in Sec. 6.8. The last row shows the combined uncertainty per data sample.

Syst. uncert.	Run 1, LL	Run 1, LD	Run 2, LL	Run 2, LD
Fit bias	0.036	0.036	0.036	0.036
Fit model choice	0.022	0.033	0.015	0.013
Fit model fixation	0.026	0.026	0.026	0.026
MVA	0.023	0.040	0.014	0.031
PID	0.007	0.008	0.026	0.026
Hardware trigger	0.063	0.062	0.063	0.062
Software trigger	0.065	0.106	0.008	0.026
Misconfig trigger	—	—	0.007	0.004
$\pi^\pm$ vs. $K^\pm$	0.005	0.005	0.005	0.005
VELO	0.008	0.008	0.008	0.008
Total	0.106	0.141	0.085	0.092

$$\Delta(f_s/f_d) = 0.051. \tag{6.31}$$

Compared to the previous Belle measurement of $B_s^0 \to K^0\overline{K}^0$ decays [136], the systematic uncertainty of this analysis is twice as large, while the external uncertainties are compatible. A better understanding of the hadronic L0 trigger and thus a reduction of its assigned uncertainty leads to systematic uncertainties being on the same level. As described in Sec. 3.3.1, the first hardware trigger stage will be replaced by a software implementation in Run III which could achieve that goal.

6.9 Significance determination

In the field of particle physics it is common to provide a significance of a measurement to be able to claim an evidence or observation. The significance, given in standard deviation σ, is a measure of the probability that the excess of events is only a statistical background fluctuation. For example, a measurement is called observation if the significance is greater than $5\,\sigma$, which corresponds to a probability of 1 to 3.5 million. To determine the significance of this measurement, two hypotheses are defined. The first hypothesis, $H_{\text{sig+bkg}}$, includes the assumption that the signal decay channel exists. It is implemented as the nominal fit model described in Sec. 6.4. Furthermore, the background only (or null-) hypothesis H_{bkg} is introduced, which assumes that no signal component exists. Specifically, this hypothesis is realised by fixing the $\mathcal{BR}$ parameter of the fit to zero. By applying maximum likelihood fits with both hypotheses, a likelihood ratio can be determined, where large values provide a stronger evidence against H_{bkg}. Taking advantage of Wilk's theorem [171], i.e. assuming that the likelihood ratio behaves like a χ^2 distribution with n degrees of

freedom, where n is the number of fixed parameters of H_{bkg} with respect to $H_{\text{sig+bkg}}$, it is possible to calculate the significance as

$$n_\sigma = \sqrt{2(NLL_0 - NLL_{\text{free}})}. \tag{6.32}$$

Here, the negative log-likelihood values NLL_0 and NLL_{free} correspond to the hypotheses H_{bkg} and $H_{\text{sig+bkg}}$, respectively. Performing this procedure for the $B_s^0 \to K_S^0 K_S^0$ component in the nominal mass fit yield a significance of $8.85\,\sigma$. Instead of the branching ratio for the signal mode, it is also possible to calculate the significance for the B^0 component by fixing the parameter f_{B^0/B_s^0} to zero. This results in a significance of 3.83 standard deviations for the $B^0 \to K_S^0 K_S^0$ decay channel.

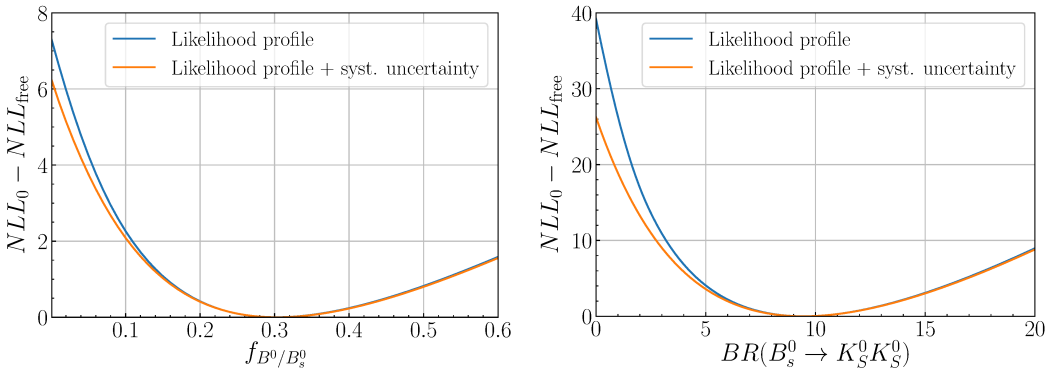


FIGURE 6.24: Profile NLL distributions of the nominal mass fit for the B^0 component (left) and B_s^0 component (right). In blue, the raw distribution is shown. In orange, the systematic uncertainties affecting the fit yields are also considered by folding a Gaussian distribution.

The aforementioned significances are solely based on the statistical uncertainties stemming directly from the fit. To include also the systematic uncertainties which influence the fit result, a profile NLL scan is performed. First, the parameter $\mathcal{BR}$ (f_{B^0/B_s^0}) is fixed to values in a range of -0.3×10^{-6} to 25×10^{-6} (0.0 to 0.8), a new fit is performed and the NLL is calculated to receive the profile shape. These NLL values are transformed into likelihoods, which are folded by a Gaussian distribution with a mean of 0 and a width equal to the relative systematic uncertainty affecting the fit results. In the case of the $B_s^0 \to K_S^0 K_S^0$ ($B^0 \to K_S^0 K_S^0$) component, this uncertainty has a value of 4.8 % (7.6 %). Finally, the likelihoods are turned back to NLLs and a new significance is calculated using the folded shapes and Eq. (6.32). The NLL profiles, both raw and folded, are shown in Fig. 6.24, where negative values are not included due to visualisation purposes. Thus, taking into account the systematic uncertainties, the significances are $7.24\,\sigma$ for $B_s^0 \to K_S^0 K_S^0$ decays and $3.54\,\sigma$ for its B^0 counterpart.

Chapter 7

Conclusion and Outlook

Today, the Standard Model of Particle Physics (SM) is the most successful and best validated description of elementary particles and their interactions. Nevertheless, it it known to be incomplete, as for example the amount of *CP* violation embedded in the SM is not able to explain the matter-antimatter asymmetry in the current universe. Many alternative theories exist, summarised under the term *New Physics*, expanding the SM to be able to describe the observed and so far inexplicable phenomena. Supersymmetry for example predicts a heavy supersymmetric partner for each elementary particle in the SM. Although the centre-of-mass energies of the LHC and other particle accelerators are not sufficient to directly produce these hypothetical particles, indirect searches are sensitive to NP effects and thus can provide evidence for their existence. Due to the ever-increasing amount of data in combination with advanced statistical methods, physics parameters can be determined more precisely, hence can help to finally answer the question "to see what holds the earth together in its innermost elements" [172].

This thesis comprises an analysis of the branching ratio in the decay channel $B_s^0 \to K_S^0 K_S^0$, where the K_S^0 mesons are reconstructed as oppositely charged pion pairs. The corresponding data are collected by the LHCb experiment in the years 2011–2012 and 2015–2016, and correspond to an integrated luminosity of 5 fb^{-1}. As normalisation mode, the decay channel $B^0 \to \phi K_S^0$ is used, with the subsequent decays of $K_S^0 \to \pi^+\pi^-$ and $\phi \to K^+K^-$. About 32 $B_s^0 \to K_S^0 K_S^0$ signal candidates have been found. The branching ratio could be determined to be

$$\begin{aligned}\mathcal{BR}(B_s^0 \to K_S^0 K_S^0) = [&9.5 \pm 1.9(\text{stat}) \pm 0.9(\text{syst}) \\ &\pm 0.9(\mathcal{BR}(B^0 \to \phi K_S^0)) \pm 0.5(f_d/f_s)] \times 10^{-6},\end{aligned}$$

where the first uncertainty is statistical, the second systematic, the third due to the external branching ratio of the decay channel $B^0 \to \phi K_S^0$ and the last due to the hadronisation probabilities. The result is compatible with SM predictions and measurements of the Belle collaboration [136] below one standard deviation. With a significance of $7.24\,\sigma$, this result is the most precise measurement of the $B_s^0 \to K_S^0 K_S^0$ branching ratio up to date, reducing the statistical uncertainty of the previous measurement by 25 %. Additionally, a hint for a $B^0 \to K_S^0 K_S^0$ component has been seen

with a significance of 3.54 standard deviations. A figure containing all data sets included in this analysis is given in Fig. 7.1.

Furthermore, this analysis is the first containing b-hadrons decaying solely into V^0 particles at LHCb and can be seen as a benchmark measurement for the experiment. Due to the topology of $B_s^0 \to K_S^0 K_S^0$ decays and the four pion final state in a hadronic environment, a high fraction of combinatorial background is observed. Thus, this measurement relies heavily on statistical methods like the application of multivariate methods in order to be able to select a signal component. Nevertheless, the reweighting and resampling procedures based on calibration samples reduce the discrepancies in the observables between data and simulated samples, so that the systematic uncertainties are small and this analysis is still statistically limited.

In the course of this work, new trigger lines were developed specialised on $H_b \to K_S^0 K_S^0$ decays, which were implemented in the 2017 trigger configuration key of the LHCb software. By making use of them, it is also possible to account for the track category where both K_S^0 mesons decay into downstream pions, increasing the statistics by about 40 %. Hence, a statistical uncertainty of below 10 % is expected when including the full Run II data set and all track types.

To be able to process the demanding running conditions of Run III, i.e. the increased centre-of-mass energy of 14 TeV and the higher instantaneous luminosity by a factor of five with respect to Run II, the LHCb experiment receives a major upgrade during the Long Shutdown 2 period. While parts of the subdetectors are updated, the current trigger system is completely replaced by a full software trigger, allowing a readout at the nominal collision rate of 40 MHz. Thus, in the years 2021–2023 more than 20 fb^{-1} of data are to be recorded. Assuming that the sensitivities and efficiencies are similar to those derived from this analysis, the statistical uncertainty for $\mathcal{BR}(B_s^0 \to K_S^0 K_S^0)$ is expected the be reduced by a factor of four with this data set. Furthermore, an observation of the decay channel $B^0 \to K_S^0 K_S^0$ could be feasible.

Finally, the ideas and techniques developed in this work will help to measure similar decay channels with unprecedented precision and hence form the basis for new discoveries.

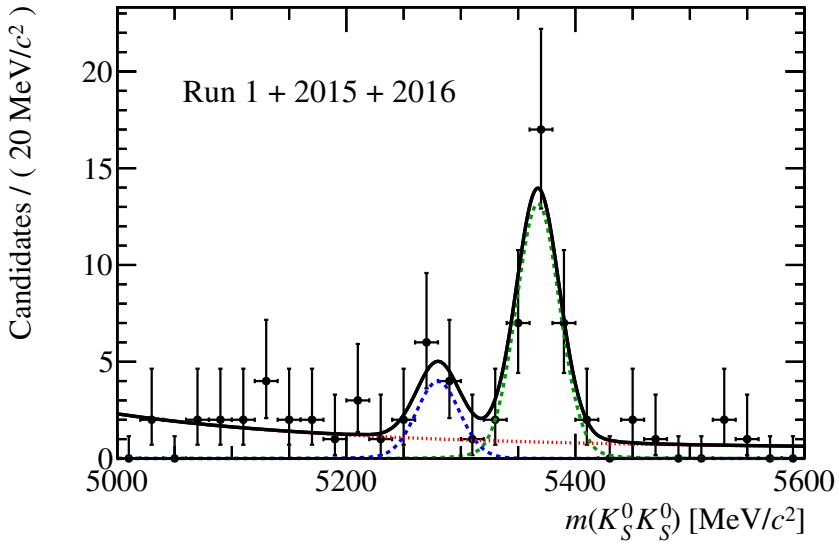


FIGURE 7.1: Invariant mass distribution of $m(K_S^0 K_S^0)$ after full selection, including candidates from the four data samples Run I LL, Run I LD, Run II LL and Run II LD. The black (solid) curve represents a projection the complete model, the $B_s^0 \to K_S^0 K_S^0$ signal component is given in green (dashed), the $B^0 \to K_S^0 K_S^0$ signal is given in blue (dashed) and the combinatorial background component in red (dotted) [166].

Chapter 8

Appendix

Supplementary Figures

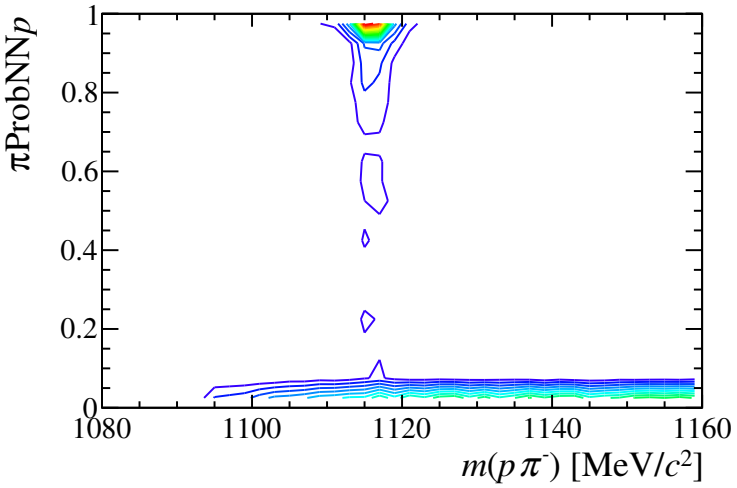
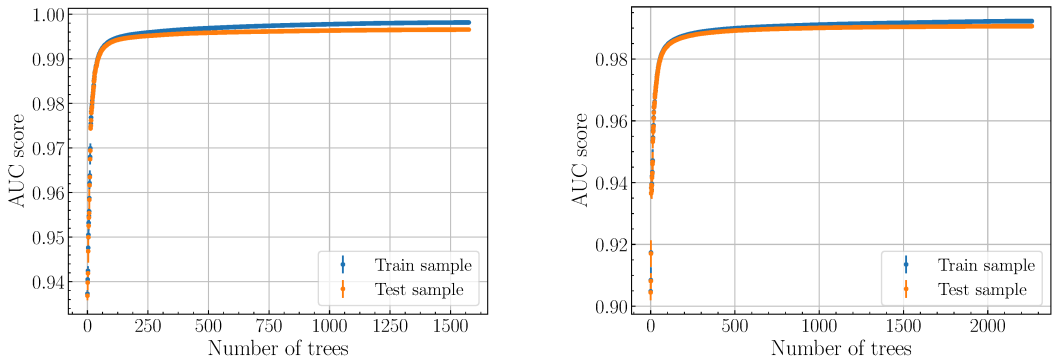
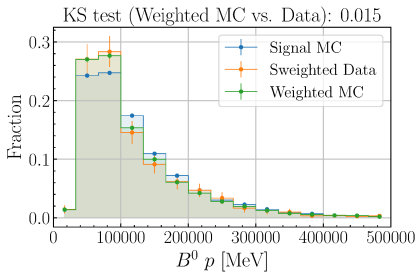


FIGURE 8.1: Two dimensional contour plot of the invariant pion-proton mass against the probability of the replaced pion candidate to be proton. A clear accumulation of events is visible at the nominal Λ meson mass and high ProbNNp probabilities.

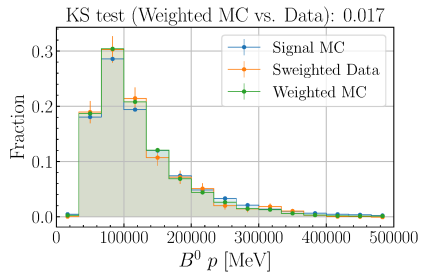


Run II, LL Run II, LD

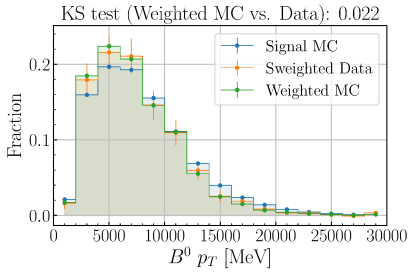
FIGURE 8.2: Comparison of learning curves (i.e. the performance gain by adding more trees in the BDT algorithm) for the train and test sample. On the left (right) the track category LL (LD) of Run II is shown. As performance metric, the ROC AUC score is used. A small contribution of overtraining is visible in both plots.



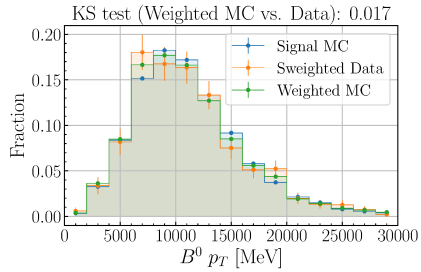
Run II, LL



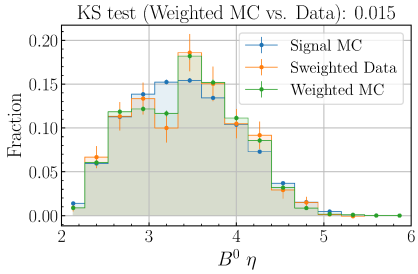
Run II, LD



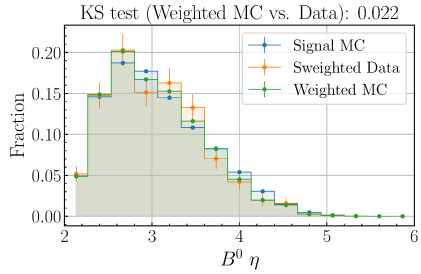
Run II, LL



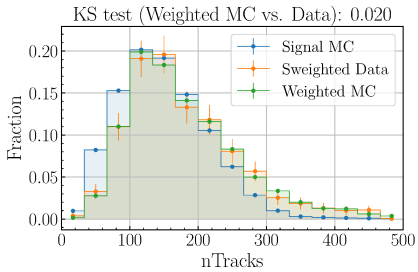
Run II, LD



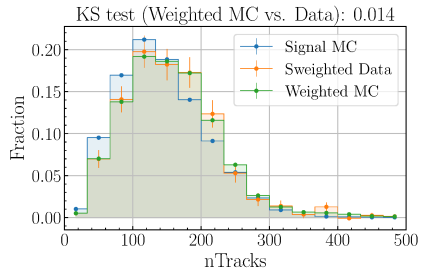
Run II, LL



Run II, LD



Run II, LL



Run II, LD

FIGURE 8.3: Comparison between distributions from Run II weighted simulation (green), "raw" simulation (blue) and ${}_s\mathcal{W}$eighted data (orange) of the decay channel $B^0 \to \phi K^0_S$. From top to bottom, the variables are the B momentum, the B transverse momentum, the pseudo-rapidity of the B and the track multiplicity in the event `nTracks`. On the left (right) the track category LL (LD) is shown. The Kolmogorow Smirnow (KS) test in the title of each plot is an indicator for how compatible these distributions are and calculated between weighted simulation and ${}_s\mathcal{W}$eighted data. Smaller numbers denote better agreement.

Supplementary Tables

TABLE 8.1: Fitted signal widths of the decay channel $B_s^0 \to K_S^0 K_S^0$, determined by the DTF algorithm (Sec. 6.3.1). As model, the Hypatia function as defined in Sec. 6.4.1 has been used. The results are given in MeV/c^2. From left to right, the following constraints are applied: No constraint, primary vertex constraint, mass constraint, primary vertex and mass constraint.

Year	Sample	Free	PV constr.	Mass constr.	PV + mass constr.
2011	LL	14.8 ± 1.0	15.7 ± 0.7	14.8 ± 0.7	13.3 ± 0.4
2012	LL	16.3 ± 0.7	16.1 ± 0.7	14.3 ± 0.6	14.5 ± 0.5
2015	LL	16.2 ± 0.6	15.9 ± 0.4	16.1 ± 0.6	15.2 ± 0.4
2016	LL	16.6 ± 0.4	15.59 ± 0.19	15.22 ± 0.19	14.85 ± 0.13
2011	LD	17.55 ± 0.25	16.1 ± 0.4	16.44 ± 0.25	14.67 ± 0.26
2012	LD	18.83 ± 0.34	16.77 ± 0.24	17.5 ± 0.4	15.21 ± 0.28
2015	LD	19.3 ± 0.4	17.7 ± 0.5	17.9 ± 0.4	15.72 ± 0.27
2016	LD	19.29 ± 0.17	17.71 ± 0.11	18.13 ± 0.19	15.94 ± 0.10

TABLE 8.2: Features included in the multivariate classification process for the decay channel $B_s^0 \to K_S^0 K_S^0$, split by dataset. A description of the features is given in Sec. 6.3.5. Numbers express the rank of the corresponding feature. A ✗ denotes that the feature was either not available in the data set or excluded due to the low performance gain.

	Run I, LL	Run I, LD	Run II, LL	Run II, LD
B_s^0 τ [ns]	3	1	16	9
B_s^0 FD [mm]	8	4	19	17
B_s^0 IP [mm]	5	✗	12	✗
B_s^0 η	✗	✗	10	12
$\log_{10}(B_s^0$ p_{T} [MeV])	6	2	5	3
$\log_{10}(B_s^0$ Vertex $\chi^2)$	4	7	7	11
$\log_{10}(B_s^0$ IP $\chi^2)$	2	3	1	1
$\log_{10}(B_s^0$ FD $\chi^2)$	✗	11	6	15
$\min(\Delta Z)$ [mm]	✗	✗	15	14
K_S^0[L] IP [mm]	✗	10	✗	13
K_S^0[D] IP [mm]	✗	5	✗	5
K_S^0[L] FD [mm]	✗	9	✗	16
K_S^0[D] FD [mm]	✗	8	✗	6
$\min(K_S^0$ FD) [mm]	✗	✗	9	✗
$\max(K_S^0$ FD) [mm]	✗	✗	13	✗
$\min(K_S^0$ IP) [mm]	✗	✗	11	✗
$\max(K_S^0$ IP) [mm]	✗	✗	17	✗
$\log_{10}(K_S^0$[L] IP $\chi^2)$	✗	6	✗	7
$\log_{10}(K_S^0$[D] IP $\chi^2)$	✗	✗	✗	2
$\log_{10}(K_S^0$[D] FD $\chi^2)$	✗	12	✗	4
$\log_{10}(K_S^0$[L] Vertex $\chi^2)$	✗	✗	✗	8
$\log_{10}(K_S^0$[D] Vertex $\chi^2)$	✗	✗	✗	10
$\log_{10}(\min(K_S^0$ IP $\chi^2))$	1	✗	2	✗
$\log_{10}(\max(K_S^0$ IP $\chi^2))$	✗	✗	4	✗
$\log_{10}(\min(K_S^0$ FD $\chi^2))$	7	✗	3	✗
$\log_{10}(\max(K_S^0$ FD $\chi^2))$	✗	✗	8	✗
$\log_{10}(\max(K_S^0$ Vertex $\chi^2))$	✗	✗	14	✗
$\log_{10}(\min(K_S^0$ Vertex $\chi^2))$	✗	✗	18	✗

Bibliography

[1] E. Rutherford. *The scattering of alpha and beta particles by matter and the structure of the atom*. Phil. Mag. Ser. 6 **21** (1911). DOI: 10.1080/14786440508637080.

[2] S. L. Glashow. *Partial-symmetries of weak interactions*. Nuclear Physics **22**.4 (1961) 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[3] A. Salam and J. Ward. *Electromagnetic and weak interactions*. Physics Letters **13**.2 (1964) 168–171. DOI: 10.1016/0031-9163(64)90711-5.

[4] S. Weinberg. *A Model of Leptons*. Phys. Rev. Lett. **19** (21 1967) 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[5] F. Abe *et al. Observation of top quark production in $\overline{p}p$ collisions*. Phys. Rev. Lett. **74** (1995) 2626–2631. DOI: 10.1103/PhysRevLett.74.2626.

[6] K. Kodama *et al. Observation of tau neutrino interactions*. Phys. Lett. **B504** (2001) 218–224. DOI: 10.1016/S0370-2693(01)00307-0.

[7] S. Chatrchyan *et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. Phys. Lett. **B716** (2012) 30–61. DOI: 10.1016/j.physletb.2012.08.021.

[8] 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) 1–29. DOI: 10.1016/j.physletb.2012.08.020.

[9] E. Corbelli and P. Salucci. *The Extended Rotation Curve and the Dark Matter Halo of M33*. Mon. Not. Roy. Astron. Soc. **311** (2000) 441–447. DOI: 10.1046/j.1365-8711.2000.03075.x.

[10] D. Huterer and M. S. Turner. *Prospects for probing the dark energy via supernova distance measurements*. Phys. Rev. **D60** (1999) 81301. DOI: 10.1103/PhysRevD.60.081301.

[11] A. D. Sakharov. *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*. Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[12] K. Abe *et al. Search for proton decay via $p \to \nu K^+$ using 260 kilo · year data of Super-Kamiokande*. Phys. Rev. D **90** (7 2014) 072005. DOI: 10.1103/PhysRevD.90.072005.

[13] R. A. McKee and J. P. Stark. *Violation of thermal equilibrium in a thermomigration experiment*. Phys. Rev. B **11** (4 1975) 1374–1381. DOI: 10.1103/PhysRevB.11.1374.

[14] R. Aaij *et al. Observation of CP violation in charm decays* (2019). arXiv: 1903.08726 [hep-ex].

[15] L. Evans and P. Bryant. *LHC Machine*. JINST **3** (2008) S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[16] A. A. Alves Jr. *et al. The LHCb Detector at the LHC*. JINST **3** (2008) S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[17] A. J. Buras *et al. Can we reach the Zeptouniverse with rare K and $B_{s,d}$ decays?* Journal of High Energy Physics **2014**.11 (2014) 121. DOI: 10.1007/JHEP11(2014)121.

[18] M. Ciuchini, M. Pierini, and L. Silvestrini. *$B_s^0 \to K^{(*)0}\overline{K}^{(*)0}$ decays: The Golden channels for new physics searches*. Phys. Rev. Lett. **100** (2008) 031802. DOI: 10.1103/PhysRevLett.100.031802.

[19] A. R. Williamson and J. Zupan. *Two body B decays with isosinglet final states in SCET*. Phys. Rev. **D74** (2006). [Erratum: Phys. Rev.D74,03901(2006)] 014003. DOI: 10.1103/PhysRevD.74.014003,10.1103/PhysRevD.74.03901.

[20] M. Beneke and M. Neubert. *QCD factorization for $B \to PP$ and $B \to PV$ decays*. Nuclear Physics B **675**.1 (2003) 333–415. DOI: https://doi.org/10.1016/j.nuclphysb.2003.09.026.

[21] A. Ali *et al. Charmless nonleptonic B_s decays to PP, PV, and VV final states in the perturbative QCD approach*. Phys. Rev. D **76** (7 2007) 074018. DOI: 10.1103/PhysRevD.76.074018.

[22] J.-J. Wang *et al. $\overline{B}_s^0 \to K\pi, KK$ decays and effects of the next-to-leading order contributions*. Phys. Rev. D **89** (7 2014) 074046. DOI: 10.1103/PhysRevD.89.074046.

[23] Y. Grossman and P. Tanedo. *Just a Taste: Lectures on Flavor Physics. Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics (TASI 2016)*. 2018 109–295. DOI: 10.1142/9789813233348_0004.

[24] I. C. Brock and T. Schörner-Sadenius. *Physics at the Terascale*. Wiley-VCH Verlag GmbH & Co. KGaA, 2011.

[25] G. C. Branco, L. Lavoura, and J. P. Silva. *CP Violation*. Oxford Science Publications, 1999.

[26] K. G. Wilson. *Confinement of quarks*. Phys. Rev. D **10** (8 1974) 2445–2459. DOI: 10.1103/PhysRevD.10.2445.

[27] R. Aaij *et al. Amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays*. Phys. Rev. **D95**.1 (2017) 012002. DOI: 10.1103/PhysRevD.95.012002. eprint: 1606.07898.

[28] R. Aaij *et al. Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays*. Phys. Rev. Lett. **118**.2 (2017) 022003. DOI: 10.1103/PhysRevLett.118.022003.

[29] 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. DOI: 10.1103/PhysRevLett.115.072001.

[30] W. Ochs. *The Status of Glueballs*. J. Phys. **G40** (2013) 043001. DOI: 10.1088/0954-3899/40/4/043001.

[31] M. B. Robinson *et al. A Simple Introduction to Particle Physics. Part I - Foundations and the Standard Model* (2008). arXiv: 0810.3328 [hep-th].

[32] A. Pich. *Quantum chromodynamics. High energy physics. Proceedings, 2nd European School, Sorrento, Italy, August 29-September 1, 1994. Vol. 1, 2*. 1995 157–207. arXiv: hep-ph/9505231 [hep-ph].

[33] V. I. Borodulin, R. N. Rogalev, and S. R. Slabospitsky. *CORE: COmpendium of RElations: Version 2.1* (1995). arXiv: hep-ph/9507456 [hep-ph].

[34] N. Cabibbo. *Unitary Symmetry and Leptonic Decays*. Phys. Rev. Lett. **10** (12 1963) 531–533. DOI: 10.1103/PhysRevLett.10.531.

[35] M. Kobayashi and T. Maskawa. *CP-Violation in the Renormalizable Theory of Weak Interaction*. Progress of Theoretical Physics **49**.2 (1973) 652–657. DOI: 10.1143/PTP.49.652. eprint: /oup/backfile/content_public/journal/ptp/49/2/10.1143/ptp.49.652/2/49-2-652.pdf. http://dx.doi.org/10.1143/PTP.49.652.

[36] L. Wolfenstein. *Parametrization of the Kobayashi-Maskawa Matrix*. Phys. Rev. Lett. **51** (21 1983) 1945–1947. DOI: 10.1103/PhysRevLett.51.1945.

[37] C. Jarlskog. *Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal* CP *Nonconservation*. Phys. Rev. Lett. **55** (10 1985) 1039–1042. DOI: 10.1103/PhysRevLett.55.1039.

[38] K. A. Olive *et al. Review of Particle Physics*. Chin. Phys. **C38** (1914) 090001. DOI: 10.1088/1674-1137/38/9/090001.

[39] J. e. a. Charles. *CKM Fitter group*. **C41** (2005). [Online; accessed 28. May 2019] 1–131. http://ckmfitter.in2p3.fr/www/html/ckm_main.html.

[40] T. Damour. *The Theoretical significance of G*. Measur. Sci. Tech. **10** (1999) 467–469. DOI: 10.1088/0957-0233/10/6/309.

[41] C. S. Wu *et al. Experimental Test of Parity Conservation in Beta Decay*. Phys. Rev. **105** (4 1957) 1413–1415. DOI: 10.1103/PhysRev.105.1413.

[42] V. Weisskopf and E. Wigner. *Berechnung der natürlichen Linienbreite auf Grund der Diracschen Lichttheorie*. Zeitschrift für Physik **63**.1 (1930) 54–73. DOI: 10.1007/BF01336768.

[43] V. Weisskopf and E. Wigner. *Über die natürliche Linienbreite in der Strahlung des harmonischen Oszillators*. Zeitschrift für Physik **65**.1 (1930) 18–29. DOI: 10.1007/BF01397406.

[44] M. Tanabashi *et al. Review of particle physics*. Phys. Rev. **D98** (2018) 030001. DOI: 10.1103/PhysRevD.98.030001.

[45] Y.-T. Duh *et al. Measurements of branching fractions and direct CP asymmetries for $B \to K\pi$, $B \to \pi\pi$ and $B \to KK$ decays*. Phys. Rev. **D87**.3 (2013) 031103. DOI: 10.1103/PhysRevD.87.031103.

[46] J. P. Lees *et al. Measurement of CP Asymmetries and Branching Fractions in Charmless Two-Body B-Meson Decays to Pions and Kaons*. Phys. Rev. **D87**.5 (2013) 052009. DOI: 10.1103/PhysRevD.87.052009.

[47] T. A. Aaltonen *et al. Measurements of Direct CP-Violating Asymmetries in Charmless Decays of Bottom Baryons*. Phys. Rev. Lett. **113**.24 (2014) 242001. DOI: 10.1103/PhysRevLett.113.242001.

[48] R. Aaij *et al. First observation of CP violation in the decays of B_s^0 mesons*. Phys. Rev. Lett. **110**.22 (2013) 221601. DOI: 10.1103/PhysRevLett.110.221601. arXiv: 1304.6173 [hep-ex].

[49] J. H. Christenson *et al. Evidence for the 2π Decay of the K_2^0 Meson*. Phys. Rev. Lett. **13** (1964) 138–140. DOI: 10.1103/PhysRevLett.13.138.

[50] *The Nobel Prize in Physics 1980*. Nobel Media AB 2018 (2018). NobelPrize.org.

[51] R. Aaij *et al. Measurement of the flavour-specific CP-violating asymmetry a_{sl}^s in B_s^0 decays*. Phys. Lett. **B728** (2014) 607–615. DOI: 10.1016/j.physletb.2013.12.030.

[52] R. Aaij *et al. Measurement of CP violation in $B^0 \to J/\psi K_S^0$ and $B^0 \to \psi(2S)K_S^0$ decays*. JHEP **11** (2017) 170. DOI: 10.1007/JHEP11(2017)170.

[53] L. Canetti, M. Drewes, and M. Shaposhnikov. *Matter and antimatter in the universe*. New Journal of Physics **14**.9 (2012) 095012. DOI: 10.1088/1367-2630/14/9/095012.

[54] M. J. Reid *et al. Motions of Galaxies in the Local Group and Beyond: an Astro2010 Science White Paper* (2009). arXiv: 0902.3932 [astro-ph.GA].

[55] M. A. Taylor *et al. Observational Evidence for a Dark Side to NGC 5128'S Globular Cluster System*. The Astrophysical Journal **805**.1 (2015) 65. DOI: 10.1088/0004-637x/805/1/65.

[56] K. Abe *et al. Evidence of electron neutrino appearance in a muon neutrino beam*. Phys. Rev. D **88** (3 2013) 032002. DOI: 10.1103/PhysRevD.88.032002.

[57] *The Nobel Prize in Physics 2015*. Nobel Media AB 2018 (2018). NobelPrize.org.

[58] R. Aaij *et al. Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays*. Phys. Rev. Lett. **113** (2014) 151601. DOI: 10.1103/PhysRevLett.113.151601.

[59] R. Aaij *et al. Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$*. Phys. Rev. Lett. **111** (2013) 191801. DOI: 10.1103/PhysRevLett.111.191801.

[60] S. P. Martin. *A Supersymmetry primer* (1997). [Adv. Ser. Direct. High Energy Phys.18,1(1998)] 1–98. DOI: 10.1142/9789812839657_0001.

[61] *LEP Design Report Vol.1* (1983). https://cds.cern.ch/record/98881.

[62] *LEP Design Report: Vol.2. The LEP Main Ring* (1984). https://cds.cern.ch/record/102083.

[63] C. Ilgner *et al. The Beam Conditions Monitor of the LHCb Experiment* (2010). arXiv: 1001.2487 [physics.ins-det].

[64] C. Lefèvre. *The CERN accelerator complex. Complexe des accélérateurs du CERN*. 2008. https://cds.cern.ch/record/1260465.

[65] G. Aad *et al. The ATLAS Experiment at the CERN Large Hadron Collider*. JINST **3** (2008) S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[66] S. Chatrchyan *et al. The CMS Experiment at the CERN LHC*. JINST **3** (2008) S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[67] K. Aamodt *et al. The ALICE experiment at the CERN LHC*. JINST **3** (2008) S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[68] O. Adriani *et al. The LHCf detector at the CERN Large Hadron Collider*. JINST **3** (2008) S08006. DOI: 10.1088/1748-0221/3/08/S08006.

[69] J. Pinfold *et al. Technical Design Report of the MoEDAL Experiment* (2009).

[70] G. Anelli *et al. The TOTEM experiment at the CERN Large Hadron Collider*. JINST **3** (2008) S08007. DOI: 10.1088/1748-0221/3/08/S08007.

[71] J. M. Campbell, J. W. Huston, and W. J. Stirling. *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*. Rept. Prog. Phys. **70** (2007) 89. DOI: 10.1088/0034-4885/70/1/R02.

[72] M. Calvi, O. Leroy, and M. Musy. *Flavour tagging algorithms and performances in LHCb* (2007).

[73] R. Aaij *et al. Performance of the LHCb Vertex Locator*. JINST **9** (2014) P09007. DOI: 10.1088/1748-0221/9/09/P09007.

[74] 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) 330–340. DOI: 10.1016/j.physletb.2011.04.031.

[75] R. Aaij *et al. Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*. Phys. Rev. Lett. **118**.19 (2017) 191801. DOI: 10.1103/PhysRevLett.118.191801. arXiv: 1703.05747 [hep-ex].

[76] 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** (2015) 68–72. DOI: 10.1038/nature14474.

[77] *Angle correlation in the $b\bar{b}$ production at the LHC*. https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html. [Online; accessed 5. December 2018].

[78] R. Aaij *et al. LHCb Detector Performance*. Int. J. Mod. Phys. **A30**.07 (2015). DOI: 10.1142/S0217751X15300227.

[79] P. R. Barbosa-Marinho *et al. LHCb VELO (VErtex LOcator): Technical Design Report*. Tech. rep. Geneva: CERN, 2001. http://cds.cern.ch/record/504321.

[80] J. Gassner, M. Needham, and O. Steinkamp. *Layout and Expected Performance of the LHCb TT Station*. Tech. rep. LHCb-2003-140. Geneva: CERN, 2004. http://cds.cern.ch/record/728548.

[81] P. R. Barbosa-Marinho *et al. LHCb inner tracker: Technical Design Report*. Tech. rep. Geneva: CERN, 2002. http://cds.cern.ch/record/582793.

[82] P. R. Barbosa-Marinho *et al. LHCb outer tracker: Technical Design Report*. Tech. rep. Geneva: CERN, 2001. http://cds.cern.ch/record/519146.

[83] M. Adinolfi *et al. Performance of the LHCb RICH detector at the LHC*. Eur. Phys. J. **C73** (2013) 2431. DOI: 10.1140/epjc/s10052-013-2431-9.

[84] N. Brook *et al. LHCb RICH1 Engineering Design Review Report*. Tech. rep. LHCb-2004-121. Geneva: CERN, 2005. http://cds.cern.ch/record/897981.

[85] M. Adinolfi *et al. LHCb RICH 2 engineering design review report*. Tech. rep. LHCb-2002-009. Geneva: CERN, 2002. http://cds.cern.ch/record/691478.

[86] S. Amato *et al. LHCb RICH: Technical Design Report*. Tech. rep. Geneva: CERN, 2000. http://cds.cern.ch/record/494263.

[87] S. Amato *et al. LHCb calorimeters: Technical Design Report*. Tech. rep. Geneva: CERN, 2000. http://cds.cern.ch/record/494264.

[88] P. R. Barbosa-Marinho *et al. LHCb muon system: Technical Design Report*. Tech. rep. Geneva: CERN, 2001. http://cds.cern.ch/record/504326.

[89] *LHCb muon system: addendum to the Technical Design Report*. Tech. rep. Geneva: CERN, 2003. http://cds.cern.ch/record/600536.

[90] *LHCb muon system: second addendum to the Technical Design Report*. Tech. rep. Geneva: CERN, 2005. http://cds.cern.ch/record/831955.

[91] R. Aaij *et al. The LHCb Trigger and its Performance in 2011*. JINST **8** (2013) P04022. DOI: 10.1088/1748-0221/8/04/P04022.

[92] J. Albrecht *et al. Performance of the LHCb High Level Trigger in 2012*. J. Phys. Conf. Ser. **513** (2014) 012001. DOI: 10.1088/1742-6596/513/1/012001.

[93] *LHCb trigger schemes*. http://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html. [Online; accessed 11. December 2018].

[94] O. Callot and S. Hansmann-Menzemer. *The Forward Tracking: Algorithm and Performance Studies*. Tech. rep. LHCb-2007-015. Geneva: CERN, 2007. https://cds.cern.ch/record/1033584.

[95] M. Clemencic *et al. Recent developments in the LHCb software framework gaudi*. Journal of Physics: Conference Series **219**.4 (2010) 042006. http://stacks.iop.org/1742-6596/219/i=4/a=042006.

[96] G. Barrand *et al. GAUDI - The software architecture and framework for building LHCb data processing applications. Proceedings, 11th International Conference on Computing in High-Energy and Nuclear Physics (CHEP 2000)*. 2000 92–95. http://chep2000.pd.infn.it/abst/abs_a152.htm.

[97] *The Moore Project*. http://lhcbdoc.web.cern.ch/lhcbdoc/moore/. [Online; accessed 12. December 2018].

[98] *The Brunel Project*. http://lhcbdoc.web.cern.ch/lhcbdoc/brunel/. [Online; accessed 12. December 2018].

[99] *The DaVinci Project*. http://lhcbdoc.web.cern.ch/lhcbdoc/davinci/. [Online; accessed 12. December 2018].

[100] M. Clemencic *et al. The LHCb Simulation Application, Gauss: Design, Evolution and Experience*. Journal of Physics: Conference Series **331**.3 (2011) 032023. http://stacks.iop.org/1742-6596/331/i=3/a=032023.

[101] I. Belyaev *et al. Handling of the generation of primary events in Gauss, the LHCb simulation framework*. Journal of Physics: Conference Series **331**.3 (2011) 032047. http://stacks.iop.org/1742-6596/331/i=3/a=032047.

[102] O. Callot. *FastVelo, a fast and efficient pattern recognition package for the Velo*. Tech. rep. LHCb-PUB-2011-001. LHCb. Geneva: CERN, 2011. https://cds.cern.ch/record/1322644.

[103] T. Sjöstrand *et al. An introduction to PYTHIA 8.2*. Computer Physics Communications **191** (2015) 159–177. DOI: https://doi.org/10.1016/j.cpc.2015.01.024.

[104] I. Beiyaev *et al. Handling of the generation of primary events in Gauss, the LHCb simulation framework. IEEE Nuclear Science Symposuim Medical Imaging Conference*. 2010 1155–1161. DOI: 10.1109/NSSMIC.2010.5873949.

[105] T. Gleisberg *et al. Event generation with SHERPA 1.1*. JHEP **02** (2009) 007. DOI: 10.1088/1126-6708/2009/02/007.

[106] D. J. Lange. *The EvtGen particle decay simulation package*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462**.1 (2001) 152–155. DOI: https://doi.org/10.1016/S0168-9002(01)00089-4.

[107] P. Golonka and Z. Was. *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*. Eur. Phys. J. **C45** (2006) 97–107. DOI: 10.1140/epjc/s2005-02396-4.

[108] S. Agostinelli *et al.* *Geant4—a simulation toolkit*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**.3 (2003) 250–303. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[109] J. Allison *et al.* *Geant4 developments and applications*. IEEE Transactions on Nuclear Science **53**.1 (2006) 270–278. DOI: 10.1109/TNS.2006.869826.

[110] *The DaVinci Project*. http://lhcbdoc.web.cern.ch/lhcbdoc/boole/. [Online; accessed 12. December 2018].

[111] M. Lamont. *LHC Performance in Run 2 and Beyond*. PoS **LeptonPhoton2015** (2016) 001. DOI: 10.22323/1.245.0001.

[112] *The LHCb Run Database*. https://lbrundb.cern.ch. [Online; accessed 14. December 2018].

[113] *HL-LHC High Luminosity Large Hadron Collider: The HL-LHC project*. http://hilumilhc.web.cern.ch/about/hl-lhc-project. [Online; accessed 17. December 2018].

[114] S. Roiser, C. Bozzi, and Bozzi. *The LHCb Software and Computing Upgrade towards LHC Run 3*. J. Phys. Conf. Ser. **1085**.3 (2018) 032049. DOI: 10.1088/1742-6596/1085/3/032049.

[115] *LHCb Trigger and Online Upgrade Technical Design Report*. Tech. rep. CERN-LHCC-2014-016. LHCB-TDR-016. 2014. https://cds.cern.ch/record/1701361.

[116] *LHCb Tracker Upgrade Technical Design Report*. Tech. rep. CERN-LHCC-2014-001. LHCB-TDR-015. 2014. https://cds.cern.ch/record/1647400.

[117] *LHCb VELO Upgrade Technical Design Report*. Tech. rep. CERN-LHCC-2013-021. LHCB-TDR-013. 2013. https://cds.cern.ch/record/1624070.

[118] P. Bhat. *Multivariate Analysis Methods in Particle Physics*. Annual Review of Nuclear and Particle Science **61** (2011) 281–309. DOI: 10.1146/annurev.nucl.012809.104427.

[119] L. Breiman *et al.* *Classification and regression trees*. Chapman and Hall/CRC, 1984. ISBN: 0412048418, 9780412048418.

[120] L. Breiman. *Random Forests*. Machine Learning **45**.1 (2001) 5–32. DOI: 10.1023/A:1010933404324.

[121] Y. Freund and R. E. Schapire. *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*. Journal of Computer and System Sciences **55**.1 (1997) 119–139. DOI: https://doi.org/10.1006/jcss.1997.1504.

[122] J. H. Friedman. *Greedy function approximation: A gradient boosting machine.* Ann. Statist. **29**.5 (2001) 1189–1232. DOI: 10.1214/aos/1013203451.

[123] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Teubner Verlag (1998).

[124] F. James and M. Roos. *Minuit - a system for function minimization and analysis of the parameter errors and correlations*. Computer Physics Communications **10**.6 (1975) 343–367. DOI: https://doi.org/10.1016/0010-4655(75)90039-9.

[125] M. Pivk and F. R. Le Diberder. *SPlot: A Statistical tool to unfold data distributions*. Nucl. Instrum. Meth. **A555** (2005) 356–369. DOI: 10.1016/j.nima.2005.08.106.

[126] *Kolmogorov–Smirnov Test*. Springer New York, 2008 283–287. DOI: 10.1007/978-0-387-32833-1_214.

[127] *Critical Values for the Two-sample Kolmogorov-Smirnov test* (2013). [Online; accessed 04. June 2019]. https://web.archive.org/web/20130613002106/http://www.soest.hawaii.edu/wessel/courses/gg313/Critical_KS.pdf.

[128] A. Abashian *et al. The Belle Detector*. Nucl. Instrum. Meth. **A479** (2002) 117–232. DOI: 10.1016/S0168-9002(01)02013-7.

[129] S. Baek *et al.* $B_s^0 \to K^+K^-$ *and* $B_s^0 \to K^0\overline{K}^0$ *Decays within Supersymmetry*. JHEP **12** (2006) 019. DOI: 10.1088/1126-6708/2006/12/019.

[130] C.-H. Chen. *Analysis of* $B^0_{(s)} \to KK$ *decays in the PQCD*. Phys. Lett. **B520** (2001) 33–40. DOI: 10.1016/S0370-2693(01)01117-0.

[131] A. J. Buras and L. Silvestrini. *Nonleptonic two-body B decays beyond factorization*. Nucl. Phys. **B569** (2000) 3–52. DOI: 10.1016/S0550-3213(99)00712-9.

[132] B. Bhattacharya *et al. Measuring* β_s *with* $B_s \to K^{0(*)}\overline{K}^{0(*)}$ *– a Reappraisal*. Phys. Lett. **B717** (2012) 403–408. DOI: 10.1016/j.physletb.2012.09.054.

[133] S. L. Glashow, J. Iliopoulos, and L. Maiani. *Weak Interactions with Lepton-Hadron Symmetry*. Phys. Rev. D **2** (7 1970) 1285–1292. DOI: 10.1103/PhysRevD.2.1285. https://link.aps.org/doi/10.1103/PhysRevD.2.1285.

[134] T. Inami and C. S. Lim. *Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes* $K_L \to \mu^+\mu^-$, $K^+ \to \pi^+\nu\overline{\nu}$ *and* $K^0 \leftrightarrow \overline{K^0}$. Prog. Theor. Phys. **65** (1981). [Erratum: Prog. Theor. Phys.65,1772(1981)] 297. DOI: 10.1143/PTP.65.297.

[135] C.-C. Peng *et al. Search for* $B_s^0 \to hh$ *Decays at the* $\Upsilon(5S)$ *Resonance*. Phys. Rev. **D82** (2010) 072007. DOI: 10.1103/PhysRevD.82.072007.

[136] B. Pal *et al. Observation of the decay* $B_s^0 \to K^0\overline{K}^0$. Phys. Rev. Lett. **116**.16 (2016) 161801. DOI: 10.1103/PhysRevLett.116.161801. arXiv: 1512.02145 [hep-ex].

[137] B. Aubert *et al. Observation of* $B^+ \to \overline{K}^0K^+$ *and* $B^0 \to K^0\overline{K}^0$. Phys. Rev. Lett. **97** (17 2006) 171805. DOI: 10.1103/PhysRevLett.97.171805.

[138] Y. Nakahama *et al. Measurement of Time-Dependent CP-Violating Parameters in* $B^0 \to K_S^0K_S^0$ *decays*. Phys. Rev. Lett. **100** (2008) 121601. DOI: 10.1103/PhysRevLett.100.121601.

[139] R. Aaij *et al. First observation of the decay* $B_s^0 \to K_S^0K^*(892)^0$ *at LHCb*. JHEP **01** (2016) 012. DOI: 10.1007/JHEP01(2016)012.

[140] G. A. Cowan, D. C. Craik, and M. D. Needham. *RapidSim: an application for the fast simulation of heavy-quark hadron decays*. Comput. Phys. Commun. **214** (2017) 239–246. DOI: 10.1016/j.cpc.2017.01.029.

[141] R. Aaji *et al. Measurement of b hadron production fractions in 7 TeV pp collisions*. Phys. Rev. D **85** (3 2012) 032008. DOI: 10.1103/PhysRevD.85.032008.

[142] LHCb collaboration. *Updated average f_s/f_d b-hadron production fraction ratio for 7 TeV pp collisions*. LHCb-CONF-2013-011 (2013).

[143] R. Aaij *et al. Measurement of b-hadron fractions in 13 TeV pp collisions* (2019). arXiv: 1902.06794 [hep-ex].

[144] W. D. Hulsbergen. *Decay chain fitting with a Kalman filter*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **552**.3 (2005) 566–575. DOI: https://doi.org/10.1016/j.nima.2005.06.078.

[145] T. Basar. *A New Approach to Linear Filtering and Prediction Problems*. IEEE, 2001. DOI: 10.1109/9780470544334.ch9.

[146] F. Archilli *et al. Performance of the Muon Identification at LHCb*. JINST **8** (2013) P10020. DOI: 10.1088/1748-0221/8/10/P10020.

[147] P. Koppenburg. *Statistical biases in measurements with multiple candidates* (2017). arXiv: 1703.01128 [hep-ex].

[148] A. Martin Sanchez, P. Robbe, and M.-H. Schune. *Performances of the LHCb L0 Calorimeter Trigger*. Tech. rep. LHCb-PUB-2011-026. Geneva: CERN, 2012. https://cds.cern.ch/record/1407893.

[149] V. V. Gligorov, C. Thomas, and M. Williams. *The HLT inclusive B triggers*. Tech. rep. LHCb-PUB-2011-016. LHCb-INT-2011-030. Geneva: CERN, 2011. https://cds.cern.ch/record/1384380.

[150] A. Puig. *The LHCb trigger in 2011 and 2012*. Tech. rep. LHCb-PUB-2014-046. Geneva: CERN, 2014. http://cds.cern.ch/record/1970930.

[151] M. Williams *et al. The HLT2 Topological Lines*. Tech. rep. LHCb-PUB-2011-002. CERN-LHCb-PUB-2011-002. Geneva: CERN, 2011. https://cds.cern.ch/record/1323557.

[152] A. M. Dendek. *Deep Neural Nets and Bonsai BDTs in the LHCb patern recognition* (2017). https://cds.cern.ch/record/2255842.

[153] R. Aaij *et al. Amplitude analysis of the decay $\overline{B}^0 \to K^0_S\pi^+\pi^-$ and first observation of the CP asymmetry in $\overline{B}^0 \to K^*(892)^-\pi^+$. Amplitude analysis of the decay $\overline{B}^0 \to K^0_S\pi^+\pi^-$ and first observation of the CP asymmetry in $\overline{B}^0 \to K^*(892)^-\pi^+$*. Phys. Rev. Lett. **120**.26 (2017). http://cds.cern.ch/record/2299123.

[154] J. Podolanski and R. Armenteros. *III. Analysis of V-events*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **45**.360 (1954) 13–30. DOI: 10.1080/14786440108520416.

[155] P. Bhat. *Multivariate Analysis Methods in Particle Physics*. Annual Review of Nuclear and Particle Science **61** (2011) 281–309. DOI: 10.1146/annurev.nucl.012809.104427.

[156] T. Fawcett. *ROC Graphs: Notes and Practical Considerations for Researchers*. Machine Learning **31** (2004) 1–38.

[157] R. Bellman, R. Corporation, and K. M. R. Collection. *Dynamic Programming*. Rand Corporation research study. Princeton University Press, 1957. ISBN: 9780691079516. https://books.google.it/books?id=wdtoPwAACAAJ.

[158] A. Rogozhnikov. *Reweighting with Boosted Decision Trees*. http://arogozhnikov.github.io/2015/10/09/gradient-boosted-reweighter.html. [Online; accessed 28. May 2019.

[159] G. Punzi. *Sensitivity of searches for new signals and its optimization*. *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*. Ed. by L. Lyons, R. Mount, and R. Reitmeyer. 2003 79. arXiv: physics/0308063.

[160] J. E. Gaiser. *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'*. PhD thesis. SLAC, 1982. http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=slac-r-255.html.

[161] M. Oreglia *et al. A Study of the Reaction $\psi' \to \gamma\gamma J/\psi$*. Phys. Rev. **D25** (1982) 2259. DOI: 10.1103/PhysRevD.25.2259.

[162] T. Skwarnicki. *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*. PhD thesis. Cracow, INP, 1986. http://www-library.desy.de/cgi-bin/showprep.pl?DESY-F31-86-02.

[163] D. Martínez Santos and F. Dupertuis. *Mass distributions marginalized over per-event errors*. Nucl. Instrum. Meth. **A764** (2014) 150–155. DOI: 10.1016/j.nima.2014.06.081.

[164] O. Barndor-nielsen. *Exponentially Decreasing Distributions for the Logarithm of Particle Size*. Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences **353** (1977). DOI: 10.1098/rspa.1977.0041.

[165] R. A. Kycia and S. Jadach. *Relativistic Voigt profile for unstable particles in high energy physics*. J. Math. Anal. Appl. **463**.2 (2018) 1040–1051. DOI: 10.1016/j.jmaa.2018.03.065.

[166] R. Aaji *et al. Search for $B_s^0 \to K_S^0 K_S^0$ decays*. Phys. Rev. D (2019). In preparation.

[167] Anderlini, L. and Contu, A and Jones, C.R. and Malde, S. and Mueller, D. *The PIDCalib package*. LHCb-PUB-2016-021 (2016).

[168] R. Aaij *et al. Measurement of the track reconstruction efficiency at LHCb*. JINST **10**.02 (2015) P02007. DOI: 10.1088/1748-0221/10/02/P02007.

[169] R. Aaij *et al. Observation of the decay $B_s^0 \to \psi(2S)K^+\pi^-$*. Physics Letters B **747** (2015) 484–494. DOI: https://doi.org/10.1016/j.physletb.2015.06.038.

[170] A. R. Bohm and Y. Sato. *Relativistic resonances: Their masses, widths, lifetimes, superposition, and causal evolution*. Phys. Rev. **D71** (2005) 085018. DOI: 10.1103/PhysRevD.71.085018.

[171] S. S. Wilks. *The large-sample distribution of the likelihood ratio for testing composite hypotheses*. Ann. Math.Stat. **9** (1938) 60–62. DOI: 10.1214/aoms/1177732360.

[172] J. W. von Goethe. *Johann Wolfgang Goethe 'Faust', Der Tragödie Erster Teil. Erläuterungen und Dokumente*. Reclam, Ditzingen, 2001.

Acknowledgements

Eine Dissertation niederzuschreiben ist eine einsame und mühselige Sache. Glücklicherweise ist der Weg dorthin, also das Studium sowie die Doktorandenzeit, das genau Gegenteil davon. Viele Menschen haben mich im Laufe der letzten zehn Jahre, die ich an der TU Dortmund verbrachte, geprägt. Den wichtigsten Personen davon möchte ich in den folgenden Passagen einige Worte widmen.

Zunächst danke ich meinem Doktorvater Prof. Dr. Bernhard Spaan, der mir die Chance gegeben hat, an seinem Lehrstuhl nicht nur meine Bachelor- und Masterarbeit zu schreiben, sondern im Anschluss auch noch die Möglichkeit zur Promotion gegeben hat. Die Arbeit an Ihrem Lehrstuhl hat mich in der Zeit vor allem auf dem Feld des wissenschaftlichen Arbeitens vorangebracht und als sehr lehrreich erwiesen.

Prof. Dr. Kevin Kröninger möchte ich dafür danken, dass er für meine Dissertation als Zweitgutachter zur Verfügung stand. Komplettiert wurde meine Prüfungskommission durch Prof. Dr. Markus Betz und Dr. Ulf Berges, denen ich für meine faire Disputation danke.

Große Teile der Analyse habe ich nicht alleine, sondern mit meinem Kollegen und Freund Dr. Moritz Demmer bestritten. Nicht immer wurden unsere Ideen mit derselben Begeisterung aufgenommen, mit der wir sie präsentierten. Dennoch trotzten wir dem Widerstand und erzielten am Ende ein herausragendes Ergebnis mit unserer Arbeit. Für die gemeinsame Zeit am Lehrstuhl, sowie für die Zeit danach, danke ich dir sehr.

Dem gesamten E5-Lehrstuhl danke ich für die angenehme Atmosphäre sowie die kritischen und alltags-physikalischen Diskussionen, die vor allem bei den zahlreichen Kaffeepausen stattfanden. Nicht zuletzt werde ich die etwas einseitige Wasserbombenschlacht, die bei E5 zum Promotionsritual dazugehört, mit Sicherheit niemals vergessen. Besonderen Dank gilt an dieser Stelle auch Dr. Julian Wishahi, der mich vor allem zu Beginn meiner Doktorarbeit als Postdoc kräftig unterstützt hat.

Während meines gesamten Studiums haben mich Dr. Vanessa Müller und Dr. Alex Birnkraut begleitet. Gemeinsam Übungszettel rechnen, Feedback für gehaltene Vorträge geben, zusammen die Erfolge feiern – dies waren nur einige der Aktivitäten, an die ich mich gerne zurückerinnere. Für euren Beistand und die schöne Zeit danke ich euch.

For all the good advice and helpful discussions I have to thank the whole BnoC Working Group, especially the Review Committee consisting of Dr. Thomas Latham and Dr. Louis Henry, with whom I had almost daily contact during the last year of my PhD.

Besonderer Dank gilt meiner Familie, natürlich insbesondere meinen Eltern, die mich während des gesamten Studium und Promotionsprojektes unterstützt und nie aufgehört haben, an mich zu glauben.

Zuletzt möchte ich meiner Freundin Jana danken, die nun seit mehr als zehn Jahren an meiner Seite ist. Vor allem in den stressigeren Phasen der Promotion half sie mir, den Mut nicht zu verlieren und mit neuer Energie an die gegebenen Herausforderungen heranzutreten.