Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

presented by

Dipl.-Phys. Georg Alexander Krocker

born in Ludwigshafen am Rhein, Germany

Oral examination: November 20th, 2013

GEORG ALEXANDER KROCKER

DEVELOPMENT AND CALIBRATION OF A SAME SIDE KAON TAGGING ALGORITHM AND MEASUREMENT OF THE B_s^0–$\overline{B}_s^0$ OSCILLATION FREQUENCY Δm_s AT THE LHCB EXPERIMENT

Referees:
Prof. Dr. Stephanie Hansmann-Menzemer
Prof. Dr. Norbert Herrmann

ABSTRACT

This thesis presents a so called same side kaon tagging algorithm, which is used in the determination of the production flavour of B_s^0 mesons. The measurement of the B_s^0–$\bar{B}_s^0$ oscillation frequency Δm_s in the decay $B_s^0 \to D_s^- \pi^+$ is used to optimise and calibrate this algorithm.

The presented studies are performed on a data set corresponding to an integrated luminosity of $\mathcal{L} = 1.0\,\text{fb}^{-1}$ collected by the LHCb experiment in 2011. The same side kaon tagging algorithm, based on multivariant classifiers, is developed, calibrated and tested using a sample of about 26,000 reconstructed $B_s^0 \to D_s^- \pi^+$ decays. An effective tagging power of $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2 = 2.42 \pm 0.39\%$ is achieved. Combining the same side kaon tagging algorithm with additional flavour tagging algorithms results in a combined tagging performance of $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2 = 5.13 \pm 0.54\%$.

With this combination, the B_s^0–$\bar{B}_s^0$ oscillation frequency is measured to be $\Delta m_s = 17.745 \pm 0.022\,(\text{stat.}) \pm 0.006\,(\text{syst.})\ \text{ps}^{-1}$, which is the most precise measurement to date.

ZUSAMMENFASSUNG

In dieser Arbeit wird ein sogenannter Same Side Kaon Taggingalgorithmus vorgestellt, der zur Bestimmung des Produktionsflavours von B_s^0 Mesonen dient. Die Messung der B_s^0–$\bar{B}_s^0$ Oszillationsfrequenz Δm_s im Zerfall $B_s^0 \to D_s^- \pi^+$ wird zur Optimierung und Kalibration des Algorithmus benutzt.

Die vorgestellten Studien benutzen einen Datensatz, welcher einer integrierten Luminosität von $\mathcal{L} = 1.0\,\text{fb}^{-1}$ enspricht und im Jahr 2011 am LHCb Experiment gesammelt wurde. Der Same Side Kaon Taggingalgorithmus, basierend auf multivarianten Klassifizierungsverfahren, wird auf einem Datensatz von 26,000 rekonstruierten $B_s^0 \to D_s^- \pi^+$ Zerfällen entwickelt, kalibriert und getestet. Es wird eine effektiven Taggingeffizienz von $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2 = 2.42 \pm 0.39\%$ erreicht. In Kombination mit zusätzlichen Taggingalgorithmen ergibt sich eine effektive Taggingeffizienz von $\varepsilon_{\text{eff}} = \varepsilon_{tag}(1-2\omega)^2 = 5.13 \pm 0.54\%$.

Mit Hilfe dieser Kombination wird eine B_s^0–$\bar{B}_s^0$ Oszillationsfrequenz von $\Delta m_s = 17.745 \pm 0.022\,(\text{stat.}) \pm 0.006\,(\text{syst.})\ \text{ps}^{-1}$ gemessen. Dieser Wert stellt die zur Zeit präziseste Messung dar.

CONTENTS

1 INTRODUCTION

From the discovery of cosmic rays about a hundred years ago up to the present day, the field of particle physics brought forward a plethora of discoveries and insights into the nature of fundamental particles and their interactions. In the 60s and 70s of the last century, a set of theoretical models was developed that explained those discoveries in a common framework, the Standard Model of particle physics. Since then, the Standard Model has been extensively tested and is able to describe the experimental results from collider experiments to an outstanding level of accuracy.

There are, however, some observations from cosmology that hint to shortcomings in the picture drawn by the Standard Model. Our present day universe consists predominantly of matter with almost no antimatter particles present, but cosmological models predict that an equal amount of matter and antimatter was created in the early phases of the universe. The Standard Model does not provide a mechanism by which an asymmetry between matter and antimatter of the observed size could be explained. Furthermore, the particle content of the Standard Model accounts for only 5% of the observed energy density of the universe. Another 25% of so called dark matter is needed to explain astronomical observations, for which the Standard Model does not provide a candidate.

These observations, together with other, more conceptual problems of the theory, led to the believe that physics beyond the Standard Model, *New Physics*, must exist. The Large Hadron Collider (LHC), located at CERN near Geneva, was build for the search of New Physics effects and for discovery of the Higgs Boson, the only fundamental particle predicted by the Standard Model, whose existence was not confirmed until very recently. The LHC is the most powerful particle accelerator today and is delivering proton-proton and heavy ion collisions to four major experiments, ALICE, ATLAS, CMS and LHCb. While the ALICE experiment aims to study the properties of the quark gluon plasma, an exotic state of matter, in collisions of heavy ions, the two largest experiments ATLAS and CMS are build for direct observations of potential new particles.

All New Physics models predict the existence of new, heavy particles that would also appear as virtual particles in quantum loops. The LHCb detector was specifically build for searches of New Physics in quantum loop corrections to the Standard Model, which could be indirectly detected in precision measurements of b- and c-hadron decays. These processes can be predicted with high precision and New Physics effects lead potentially to large deviations in the rates of rare decays or the amount of CP[1] violation that is observed in b- and c-hadron decays. As the particles that appear in these loop processes are only virtual, the energy scale that is probed in indirect searches for New Physics is about one order of magnitude higher than for direct observations. This, as well as the high $b\bar{b}$ and $c\bar{c}$ quark production cross section makes LHCb into an ideal tool for probing the consistency of the Standard Model.

1 where C stands for the charge conjugation transformation and P for the parity transformation

The B_s^0 meson system is an particular interesting probe in this regard. While the B^0 mesons was extensively studied in e^+e^- collisions, the heavier B_s^0 is only produced in large quantities at hadron machines. The neutral B_s^0 is periodically oscillating into its anti-particle via loop processes, and a measurement of the oscillation frequency Δm_s constrains the parameters of the theory. The measurement of CP violating parameters, *e.g.* in the decay of $B_s^0 \to J/\psi\,\phi$, $B_s^0 \to J/\psi\,\pi^+\pi^-$ and $B_s^0 \to D_s^- K^+$ yields the potential for the discovery of New Physics effects.

As B_s^0 mesons oscillate, many of the analyses in this system rely on the knowledge of the production flavour of the mesons. This information is extracted from the data by means of flavour tagging algorithms. While most flavour tagging algorithms exploit the $b\bar{b}$ pair production to extract the production flavour, in case of the B_s^0 system, a specific method of flavour tagging, the so called same side kaon tagging, is possible, that infers the flavour from the B_s^0 fragmentation process. In contrast to the situation at e^+e^- colliders, flavour tagging is in particular challenging in hadron environments due to the much higher particle densities.

This thesis presents the development and optimisation of a same side kaon tagging algorithm and its calibration in the decay $B_s^0 \to D_s^-\pi^+$, using $1\,fb^{-1}$ of data taken with the LHCb experiment in 2011. It is necessary to resolve the oscillation in the B_s^0 system to perform a calibration of this tagging algorithm using data, which provides an opportunity to measure the B_s^0 mixing frequency Δm_s.

This thesis is structured in the following way: In Chapter 2, an overview on the theoretical foundations is given. Chapter 3 introduces the experimental setup and Chapter 4 discusses the basic principles of flavour tagging. In Chapter 5, the steps necessary to extract the $B_s^0 \to D_s^-\pi^+$ decays from the data are presented and the unbinned maximum likelihood fit that is used to resolve the B_s^0 oscillation is explained.

For the development of the same side kaon tagging algorithm, it is necessary to rely on the simulation. Chapter 6 discusses the current level of agreement between the data and the simulation and the steps taken to improve this agreement. In Chapter 7, the development of a same side kaon tagging algorithm based on multivariant classifiers is described. This algorithm is calibrated using data in Chapter 8 and systematic studies on its performance are presented. In Chapter 9, an alternative method for the training of the multivariant classifier used in the same side kaon tagger is exploited, which uses the high statistics semileptonic data sample. Chapter 10 presents the measurement of the B_s^0 oscillation frequency Δm_s with a special emphasis on the impact of the same side kaon tagging algorithm that is developed in this thesis. Finally, Chapter 11 summarizes the results.

2 THEORY

The Standard Model (SM) of particle physics describes our knowledge of the fundamental particles and interactions up to the present day. The following chapter presents a short introduction on the fundamental principles of the theory and explains how features of the theory lead to an oscillation of neutral mesons. After an introduction to further observables in the flavour sector it concludes with a discussion about b production at the LHC. The topics discussed in this chapter are summarized from the discussions given in [1] and [2] on the formulation of the Standard Model and [3] on heavy flavour physics. Those articles provide an excellent overview of the topic and give an extensive list of further references.

2.1 THE STANDARD MODEL OF PARTICLE PHYSICS

The Standard Model is a quantum field theory that exploits the concept of local gauge symmetries. This means that the theory implements a set of gauge transformations under which the actual physics processes are invariant although the mathematical formulation of the theory changes. Those gauge transformations are local, *i.e.* they do depend on a specific point $X = (ct, \vec{x})$ in space time in contrast to global gauge transformations which do not depend on X.

The mathematical formulation of these gauge transformations dictates the bosonic field content of the Standard Model and the interactions between the particles, which are mediated by the bosonic force carriers. It is however important to understand that the choice of the gauge transformations in the Standard Model is not arbitrary but justified by symmetries observed in nature and the conservation of quantum numbers connected to these symmetries.

"Ordinary" matter is included in the Standard Model by means of fermion fields. Depending on the type, fermions have different quantum numbers. Two basic types of fermionic fields exist, the quarks and the leptons. There are six different quarks that can be divided in two groups, the up type quarks u, c, t and the down type quarks d, s, b. The leptons are symmetric in this regard, as there are also six leptons divided into the charged leptons e, μ, τ and the neutrinos ν_e, ν_μ, ν_τ. Both leptons and quarks can also be arranged into three generations, each containing an up and down type quark, a charged lepton and a neutrino. The corresponding anti fermions have the same mass and lifetime as the fermions but carry conjugated quantum numbers.

The interactions of quarks with each other are governed by the strong force. In the Standard Model it is described by quantum chromodynamics (QCD). The strong force couples on a quantum number called colour. Quarks can come in three different colours and their corresponding anti colours for anti quarks. The underlying gauge symmetry is described by the $SU(3)$ symmetry group. From the corresponding gauge transformation in the QCD Lagrangian, eight independent gauge fields emerge that can

be identified with the eight different gluons, which are the carriers of the strong force. QCD predicts that quarks are bound in colour neutral objects, *i.e.* in groups of three quarks of different colour or anti colour, also called baryons, or in groups of two quarks with the same colour and corresponding anti colour, called mesons. This also means that quarks are formed in $q\bar{q}$ pairs under the strong interaction.

The weak force and the electromagnetic force are unified into one theoretical framework in the Standard Model based on the $SU(2) \times U(1)$ gauge symmetry. The inclusion of the gauge transformations into the electroweak part of the Standard Model Lagrangian leads to four additional gauge bosons. The three bosons W_i correspond to the gauge fields of the $SU(2)$ part of the electroweak theory with the weak isospin T as quantum number and the gauge boson B to the $U(1)$ part that couples to the hypercharge Y. While all fermions carry some amount of the hypercharge Y, which is defined as $Y = Q - T_3$ (with Q the electric charge and T_3 the third component of the weak isospin) only left handed fermions and right handed antifermions have a weak isospin $T \neq 0$ and couple to the W_i bosons of $SU(2)$.

Up to this point in the discussion, all the fermion and boson fields in the Standard Model are massless. In fact, the inclusion of mass terms in the Standard Model Lagrangian would violate the gauge symmetry of the theory. This however does not correspond to the experimental results, which confirm masses for the charged leptons, the quarks and the $W^{\pm}$ and Z bosons. This shortcoming of the formulation of the Standard Model is cured by the introduction of an additional scalar (*i.e.* spin zero) field that couples to the fermions and bosons. The symmetry of this field can be spontaneously broken by which it acquires a non zero vacuum expectation value. This method of introducing masses to the standard model is generally referred to as Higgs mechanism and excitations of the scalar field are realized as Higgs boson. After the spontaneous symmetry breaking, the Standard Model particles can acquire mass by interacting with the Higgs field. In particular, the symmetry of the electroweak $SU(2) \times U(1)$ group is broken and the massive $W^{\pm}$ bosons emerge as linear combinations of the W_1 and W_2 bosons of the original gauge symmetry. In the same manner the massive Z boson and the massless γ photon are mixtures of the W_3 and the B field.

All Standard Model fermions and bosons have been observed, with the exception of the Higgs boson. However, a new bosonic particle has been observed by the CMS and ATLAS experiments in 2011 that so far fulfils all the requirements of a Standard Model Higgs Boson [4, 5]. Additional measurements of the quantum numbers and coupling strengths of this particle are necessary to confirm that it is compatible with the Higgs Boson.

2.2 FLAVOUR PHYSICS IN THE STANDARD MODEL QUARK SECTOR

As the gauge bosons acquire mass by their coupling to the Higgs field by the electroweak symmetry breaking, so do the fermions. This process is realised in the Standard Model by the Yukawa interactions, which manifest themselves by additional terms in the Lagrangian that couple right handed

down type quarks to left handed anti down type quarks (and equally for the up type quarks),

$$\mathcal{L}_Y^q = -\frac{v}{\sqrt{2}}(\bar{d}_L Y_d d_R + \bar{u}_L Y_u u_R) + \text{h.c.}\,, \tag{1}$$

where v is the vacuum expectation value of the Higgs field, Y_d and Y_u are the Yukawa matrices and $d_{L,R}$, $u_{L,R}$ denote the weak eigenstates of the left (index L) and right handed (index R) down and up type quarks. The Yukawa matrices are 3×3 complex matrices that are in general not diagonal. That means that the Yukawa interaction can couple quarks from different generations to each other or, more generally speaking, that the mass eigenstates (defined by the Yukawa interaction) and the weak eigenstates (defined by the weak interaction) are not necessarily the same. One can however find unitary base transformations $V_{L,d}$ and $V_{R,d}$ that transform the left and right handed weak eigenstates $d_{L,R}$ of the quark fields into mass eigenstates $\tilde{d}_L = V_{L,d} d_L$, $\tilde{d}_R = V_{R,d} d_R$ (and equivalent for up type quarks). By transforming the other terms of the Standard Model Langrangian to the mass basis of the quarks, a problem arises for terms where down type quarks couple to anti up type quarks or vice versa. This is only the case for the charged current of the weak interaction which is mediated by the $W^\pm$ bosons. Because the transformations for up and down type quarks are not necessarily the same, an additional matrix $V_{CKM} = V_{L,u} V_{L,d}^\dagger$ appears in the transformation of the Lagrangian of the charged current to the mass basis,

$$\mathcal{L}_{cc} = -\frac{g}{\sqrt{2}}(\bar{\tilde{u}}_L \gamma^\mu W_\mu^+ V_{CKM} \tilde{d}_L + \bar{\tilde{d}}_L \gamma^\mu W_\mu^- V_{CKM}^\dagger \tilde{u}_L)\,. \tag{2}$$

As the original transformation matrices $V_{L,u}$ and $V_{L,d}$ are unitary, also V_{CKM} is unitary. The matrix V_{CKM} is called Cabibbo-Kobayashi-Maskawa (CKM) matrix and gives a measure how the up type quarks are connected to the down type quarks,

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{3}$$

In principle, the entries of the CKM matrix can have arbitrary complex values under the requirement that the unitary conditions are fulfilled. If, under this assumption, the Lagrangian in Equation 2 is transformed subsequently by the charge operator C and the parity operator P,

$$\mathcal{L}_{cc}^{CP} = -\frac{g}{\sqrt{2}}(\bar{\tilde{d}}_L \gamma^\mu W_\mu^- V_{CKM}^T \tilde{u}_L + \bar{\tilde{u}}_L \gamma^\mu W_\mu^+ V_{CKM}^* \tilde{d}_L)\,, \tag{4}$$

it becomes clear that it is only invariant (*i.e.* Equations 2 and 4 are equal) under this transformation, if the CKM matrix contains no complex elements, because then $V_{CKM} = V_{CKM}^*$ and $V_{CKM}^T = V_{CKM}^\dagger$ holds.

The parameters of the CKM matrix do not emerge from the theory, they have to be determined by experiment. An arbitrary complex matrix has 18 parameters, of which 9 are absorbed, as the CKM matrix has to be unitary. Another five complex phases can be absorbed in a redefinition of the fields of the up and down type quarks. In total there are three real angles and one complex phase as independent parameters left. To make the hierarchy of the CKM matrix visible it is useful to parametrize it in terms of its free

parameters. This is achieved by the so called Wolfenstein parametrization, an expansion in the Wolfenstein parameter λ,

$$V_{CKM} = \begin{pmatrix} 1-\frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{5}$$

Putting the experimentally determined values [6] of the parameters

$$\begin{aligned} \lambda &\approx 0.23\,, & A &\approx 0.81\,, \\ \bar{\rho} = \rho(1-\frac{\lambda^2}{2}) &= 0.13\,, & \bar{\eta} = \eta(1-\frac{\lambda^2}{2}) &= 0.35\,, \end{aligned} \tag{6}$$

in the Wolfenstein parametrisation of the V_{CKM} matrix, it can be seen that the diagonal elements of the CKM matrix, that represent the transitions within one generation, are largest, followed by transitions to the neighbouring generation. Transitions between up and down type quarks from the third to the first generation or vice versa are suppressed by $\mathcal{O}(\lambda^3)$. CP violation arises for $\eta \neq 0$, it is the sole source of CP violation observed in the Standard Model.

The Standard Model as introduced before contains 19 parameters that have to be determined by the experiment: The six quark masses, the three masses of the charged leptons, the 4 parameters of the CKM matrix, the coupling constants of the strong, the weak and the electromagnetic force, a strong CP phase and the mass of the Higgs boson as well as the vacuum expectation value of the Higgs field. The majority of this parameters, *i.e.* the quark masses and the CKM parameters, are connected to the flavour sector and can partially be determined in the B meson[1] system. This makes flavour physics a powerful tool for testing the consistency of the Standard Model.

One group of such tests exploits the unitarity relations of the CKM matrix. The unitarity condition results in six relations of the form

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0\,. \tag{7}$$

As this equation describes a triangle in the complex plane with the lengths of the sides given by $V_{ud}V_{ub}^*$, $V_{cd}V_{cb}^*$ and $V_{td}V_{tb}^*$ and the angles

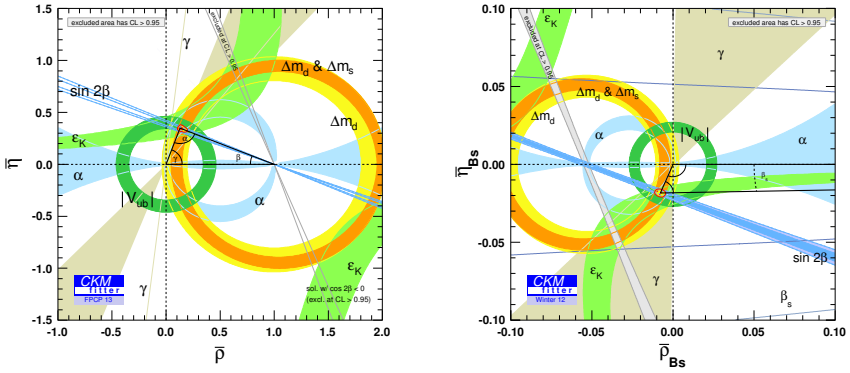
$$\alpha \equiv \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta \equiv \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma \equiv \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right), \tag{8}$$

the consistency of the CKM matrix can be tested by over constraining these equations. The status of the experimental measurements for two of the triangles, relevant in the B_s^0 and B^0 system, are shown in Figure 1, where the impressive consistency of the different results is visible.

The methods discussed in this thesis contribute in different ways to the measurements of these triangles. The same side kaon tagging algorithm improves all measurements in the B_s^0 system where the knowledge of the initial flavour is necessary. This is especially true for the measurement of the CP violating phase ϕ_s that is measured in the decay $B_s^0 \to J/\psi\,\phi$ and $B_s^0 \to J/\psi\,\pi^+\pi^-$ [7]. This phase can be related to the angle

$$\beta_s \equiv \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) \tag{9}$$

1 Throughout this thesis, the term B meson denotes mesons that contain a b quark.



(a) Unitarity triangle related to the B^0 system

(b) Unitarity triangle related to the B_s^0 system

Figure 1: Current experimental status of the unitarity triangles relevant for the B_s^0 and B_s^0 system. The single measurements are in excellent agreement. The red dashed region depicts the 95% confidence region for the apex of the triangles that is determined from the measurements. Figures are taken from [9].

from the B_s^0 triangle. Another LHCb measurement that makes use of the same side kaon tagging algorithms is the time dependent measurement of the angle γ using $B_s^0 \to D_s^- K^+$ decays [8]. Finally, the measurement of the B_s^0 oscillation frequency Δm_s from $B_s^0 \to D_s^- \pi^+$ decays constraints the length of the right side of the unitarity triangle.

2.3 THE OSCILLATION OF NEUTRAL B MESONS

Due to the structure of the charged current of the weak interaction neutral mesons exhibit an oscillation behaviour, *i.e.* there is a time dependent probability for the transition of a given flavour eigenstate, *e.g.* B_s^0, into its antiparticle, $\bar{B}_s^0$. An example of the leading order Feynman diagrams contributing to this process is given in Figure 2. The oscillation is a direct consequence of the disparity between the flavour eigenstates and the mass eigenstates of the quarks which leads to a mixing between the quark generations by the CKM mechanism. The following discussion concentrates on the oscillation in the B_s^0 system. The formulation is equivalent for all neutral mesons.

The effect of the different mass eigenstates and weak eigenstates of the quarks translates into the mesons. The B_s^0 and $\bar{B}_s^0$ can therefore be thought of as a superposition of the two mass eigenstates,

$$\begin{aligned} |B_s^0\rangle &= \frac{1}{2p}\left(|B_L\rangle + |B_H\rangle\right) \\ |\bar{B}_s^0\rangle &= \frac{1}{2q}\left(|B_L\rangle - |B_H\rangle\right), \end{aligned} \tag{10}$$

where p and q fulfil the normalisation $|p|^2 + |q|^2 = 1$ and are a measure of how much the weak and mass eigenstates are decoupled. This superposi-

tion is time dependent, *i.e.* $|B_s^0\rangle = |B_s^0(t)\rangle$, and the time dependence can be described by a phenomenological two dimensional Schrödinger equation,

$$i\frac{d}{dt}\begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} = \left(M - i\frac{\Gamma}{2}\right)\begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix}, \tag{11}$$

with the Hermitian 2×2 mass matrix M and decay width matrix Γ. The CPT theorem gives the condition that the masses and decay widths of B_s^0 and $\bar{B}_s^0$ must be equal, leading to $\Gamma_{11} = \Gamma_{22}$ and $M_{11} = M_{22}$. The mass eigenstates are the eigenvectors of the Hamiltonian $M - i\Gamma/2$. Their time evolution is given by

$$\begin{aligned} |B_H(t)\rangle &= e^{-(iM_H+\Gamma_H/2)t}|B_H\rangle\,, \\ |B_L(t)\rangle &= e^{-(iM_L+\Gamma_L/2)t}|B_L\rangle\,, \end{aligned} \tag{12}$$

where M_L and M_H are the masses of the two mass eigenstates and Γ_L and Γ_H their decay widths. Using Equations 10 and 12 one gets the time evolution of particles originally produced as B_s^0 and $\bar{B}_s^0$,

$$\begin{aligned} |B_s^0(t)\rangle &= g_+(t)|B_s^0\rangle + \frac{q}{p}g_-(t)|\bar{B}_s^0\rangle\,, \\ |\bar{B}_s^0(t)\rangle &= \frac{p}{q}g_-(t)|B_s^0\rangle + g_+(t)|\bar{B}_s^0\rangle\,. \end{aligned} \tag{13}$$

For convenience, the time dependent part is factored out in the functions $g_\pm(t)$, which show the oscillation behaviour of the pure B_s^0 and $\bar{B}_s^0$ states,

$$\begin{aligned} g_+(t) &= e^{-imt}e^{-\Gamma t/2}\left(\cosh\frac{\Delta\Gamma}{4}\cos\frac{\Delta m t}{2} - i\sinh\frac{\Delta\Gamma t}{4}\sin\frac{\Delta m t}{2}\right), \\ g_-(t) &= e^{-imt}e^{-\Gamma t/2}\left(-\sinh\frac{\Delta\Gamma}{4}\cos\frac{\Delta m t}{2} + i\cosh\frac{\Delta\Gamma t}{4}\sin\frac{\Delta m t}{2}\right). \end{aligned} \tag{14}$$

The parameters for the mass eigenstates contained in Equation 12 are substitute by the actual observables in the flavour eigenstates, *i.e.* the mass m and decay width Γ, the difference in decay width $\Delta\Gamma$ and the mass difference Δm which defines the mixing frequency. They can be translated as

$$\begin{aligned} m &= \frac{M_H + M_L}{2}\,, & \Gamma &= \frac{\Gamma_H + \Gamma_L}{2}\,, \\ \Delta m &= M_H - M_L\,, & \Delta\Gamma &= \Gamma_L - \Gamma_H\,. \end{aligned} \tag{15}$$

The mixing frequency and the decay width difference as well as the factor q/p can be directly related to the matrix elements of the M and Γ matrices, which in turn can be related to the CKM matrix elements and the other Standard Model parameters. In particular, for the mixing frequencies of the neutral B mesons one can derive

$$\Delta m_d \propto |V_{tb}V_{td}^*| \tag{16}$$

in case of the B^0 system and

$$\Delta m_s \propto |V_{tb}V_{ts}^*| \tag{17}$$

in case of the B_s^0 system, which is why their precise measurement is important to constrain the unitarity triangles.

The oscillation in the system of B^0 mesons has been first discovered by the ARGUS experiment in 1987 [10] and the measurement of the oscillation

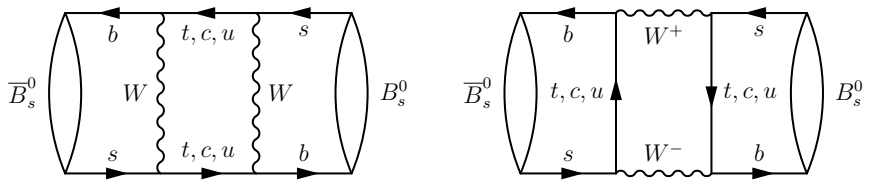


Figure 2: Leading-order diagrams for the mixing of B_s^0 mesons). Figure taken from [14].

frequency has since been improved by the BABAR and BELLE collaborations. The current world average is $\Delta m_d = 0.507 \pm 0.004\,\text{ps}^{-1}$ [6], which corresponds to about 0.12 oscillations during the lifetime of the B^0 mesons. The much faster B_s^0 oscillation has been first observed by CDF and D0 at the Tevatron in 2006 [11, 12]. Several measurements have since then been performed by the LHCb experiment. The current world best measurement results in a frequency of $\Delta m_s = 17.768 \pm 0.024\,\text{ps}^{-1}$ [13], equivalent to about four oscillation periods during one B_s^0 lifetime.

2.4 THE PRODUCTION OF b QUARKS AT THE LHC

The LHC is in particular suitable for the study of the B meson system. The main production mechanisms for $b\bar{b}$ pairs in inelastic p p collisions, *i.e.* the fusion of $q\bar{q}$ pairs or gluons, are shown in Figure 3. Cross sections for these processes can be calculated with reasonable precision from QCD [15]. Predictions for the cross sections for various processes as a function of $\sqrt{s}$ are shown in Figure 4. At the energy reached by the LHC in 2011, $\sqrt{s} = 7\,\text{TeV}$, the predicted cross section is about $\sigma_{b\bar{b}} = 300\,\mu\text{b}$. This is in good agreement with a cross section that has been measured by LHCb of $\sigma(pp \to b\bar{b}X) = 288 \pm 4 \pm 48\,\mu\text{b}$ [16].

The production mechanism of the $b\bar{b}$ quark pairs dictates the special geometry of the LHCb detector, *cf.* Chapter 3. The angular distribution of the produced b quark depends on the momentum of the original $q\bar{q}$ or gg pair. Their momentum distribution is predicted by the parton distribution functions of the proton. It is very likely that the two partons that produce the $b\bar{b}$ pair carry seizable different momentum fractions of the proton. As the energy needed for the production of the $b\bar{b}$ pair is relatively small (the mass of the b in the $\bar{M}S$ scheme is about 4.18 GeV [6] compared to the average center of mass energy of the partons which is of the order $\mathcal{O}(\sqrt{s}) = 1\,\text{TeV}$), much of the energy is available as kinetic energy and boosts the $b\bar{b}$ quark pair. Accordingly, these quarks are produced either with a small angle with respect to the beam axis in forward or backward direction, as shown in Figure 5. About 25% of the $b\bar{b}$ pairs are produced in the LHCb acceptance.

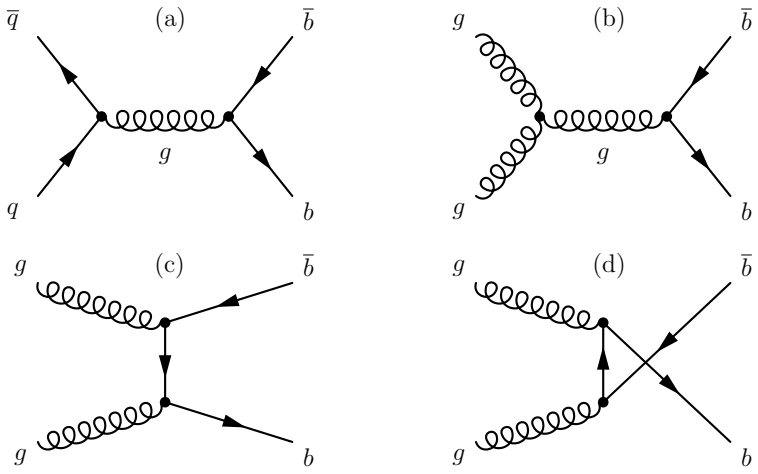


Figure 3: Leading order contributions to the production of $b\overline{b}$ quark pairs at the LHC. Figure taken from [17].

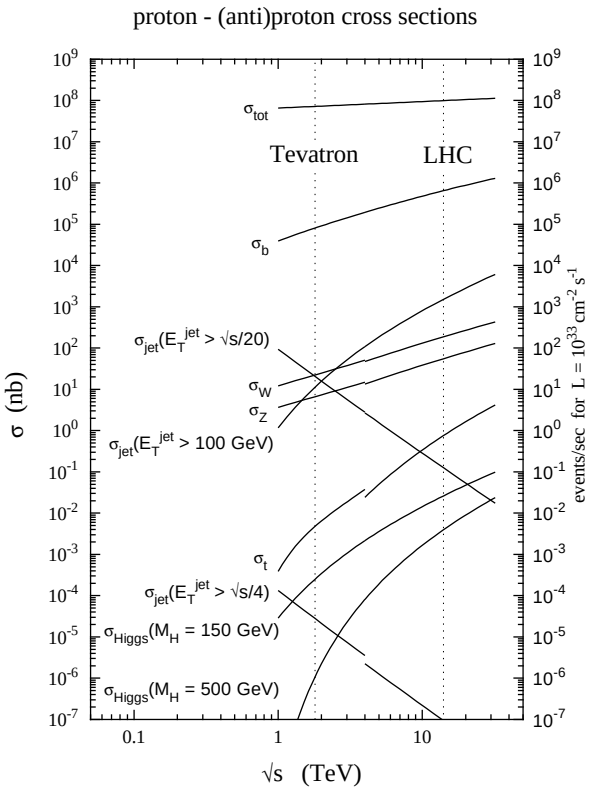


Figure 4: Cross sections for inelastic scattering and various sub processes in p p collisions versus $\sqrt{s}$. Figure taken from [18].

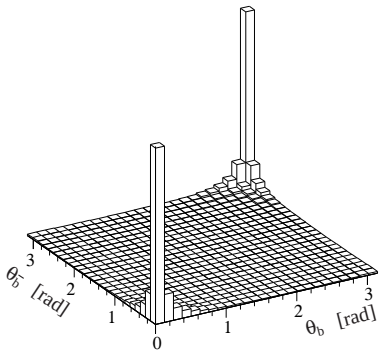


Figure 5: Correlation between the polar angle θ wrt. the z axis in the experiment for b quark and $\bar{b}$ quark produced in one p p collision. Figure taken from [19].

3 THE LHCB EXPERIMENT

The study of the rare processes in b physics and the exact determination of the quantities involved in flavour processes require a large number of B mesons. The $b\overline{b}$ cross section of up to $\sigma_{b\overline{b}} \approx 500\,\text{pb}$ in proton proton collisions at energies available at the Large Hadron Collider (LHC) opens up a large field of improvement over previous measurements. This chapter gives an overview of the experimental tools used in the study of B mesons at the LHC. After a short introduction on the key parameters of the collider the main components of the LHCb experiment and their specifications are discussed. The chapter concludes with an overview of the LHCb software framework.

3.1 THE LARGE HADRON COLLIDER

The Large Hadron Collider (LHC) is a proton - proton collider with an circumference of 27 km located at CERN near Geneva. With a design energy of $\sqrt{s} = 14\,\text{TeV}$ and a peak luminosity of $\mathcal{L} = 1 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, it is the most powerful particle accelerator today. In the nominal configuration, two beams with up to 2808 bunches with 10^{11} Protons are circulated in the machine. They are brought to collision in four interaction points every 25 ns with a crossing rate of $R_{crossing} = 40\,\text{MHz}$. The total pp cross section is about 100 mb, hence the number of interactions per bunch crossing can be calculated according to

$$N = \sigma_{tot} \int \mathcal{L} dt. \tag{18}$$

The number of actual pp collisions in a single bunch crossing is Poisson distributed with a mean number of

$$\langle n \rangle = \frac{\sigma_{tot}\mathcal{L}}{R_{crossing}} = 25 \tag{19}$$

interactions. To keep the particles on track, 1232 dipole magnets and 392 quadrupole magnets with a field strength of up to 8.3 T are needed. Such high field strengths are only possible with superconducting magnets that need to be cooled down to a temperature of 1.9 K with liquid helium.

The four major experiments ATLAS, ALICE, CMS and LHCb are located at the interaction points. ATLAS and CMS are designed as general purpose detectors conducting direct searches for heavy resonances such as the Higgs Boson or particles predicted by theories beyond the Standard Model. The ALICE detector is specialized on collisions of heavy ions which are circulated in the LHC during special run periods. LHCb is a precision experiment studying the decay of B and D mesons.

After startup and commissioning of the LHC and its experiments in 2010, the machine has performed its first full year of operation in 2011. The accelerator has not yet reached its design performance and was operated in 2011 with at a center of mass energy of $\sqrt{s} = 7\,\text{TeV}$. 1380 bunches were circulated

in the machine with a bunch spacing of 50 ns. With this configuration the LHC reached instantaneous luminosities of up to $\mathcal{L} = 3 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The analysis presented in this thesis concentrates on the data taken in 2011.

3.2 THE LHCB EXPERIMENT

The precise measurement of the decays of B and D mesons poses special requirements on the design of the detector and its operation. As discussed in the Chapter 2.4, B mesons are produced primarily in the forward and backward direction. In contrast to the typical multi purpose experiments in high energy physics, such as ATLAS and CMS, LHCb is designed as a single-arm forward spectrometer to cover the phase space of interest for the decay of B mesons. The acceptance covers an angle of 10 - 300 mrad in the bending plane of the dipole magnet and 10-250 mrad perpendicular to this plane.

A lot of measurements done at LHCb rely on the correct determination of the B meson flight distance in the laboratory system, from which the decay time can be calculated. B mesons fly on average about a centimeter from the point of their production, the primary vertex (PV), to the secondary vertex where they decay. For an optimal resolution of primary and secondary interaction point and the determination of the flight distance of the B meson, a good vertex detector close to the interaction point is vital. As a large amount of background in the experiment stems from particles coming either directly from the PV or from decays of very short lived particles, a good vertex resolution is also helpful to suppress this background.

The above stated requirements are reflected in the detector design. The detector components are aligned along the beam axis to cover the region of interest in the following order:

- The **Vertex Locator (VELO)** is directly build around the interaction point for the best possible vertex resolution followed by
- a **Ring Image Cherenkov Detector (RICH1)** that provides mainly π/K separation,
- the **Trigger Tracker (TT)** in front of the magnet,
- the **dipole magnet** that deflects charged particles and hence provides a handle for measuring their momentum,
- the main tracking system, consisting of **Inner Tracker (IT)** and **Outer Tracker (OT)**,
- a second particle identification detector **RICH2** that covers a different momentum range,
- the **calorimeter system** and finally
- the **muon chambers** for muon identification.

A schematic of the detector is shown in Figure 6. As can be seen, the origin of the LHCb coordinate system is is located in the interaction point, with the x axis horizontal, the y axis vertical and the z axis along the direction of the beam. The x-z plane defines the main deflection plain of the LHCb magnet. Further details on the design and performance of the different subdetectors

are given in the following sections. A detailed overview of the detector and its performance can be found in [20].

In addition to the special design of the detector the above stated requirements also dictate special running conditions. To ensure a clean reconstruction of the B mesons and avoid background contamination due to pileup of multiple interactions, LHCb was designed to run at a nominal instantaneous luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ compared to the nominal LHC luminosity of $1 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Studies have shown that LHCb is able to run at luminosities up to $5 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The lower luminosities are reached by defocussing the beam with special beam optics in front of and behind LHCb. This so called luminosity levelling is performed continuously by the LHC operators to ensure a constant instantaneous luminosity at LHCb.

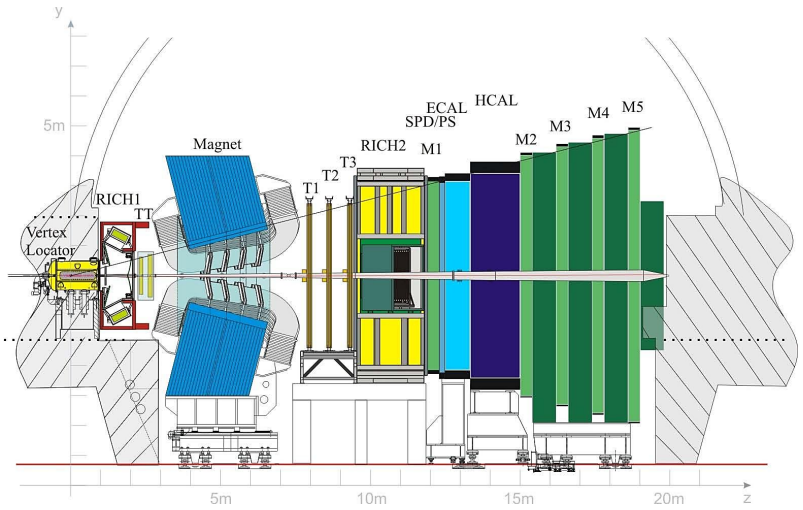


Figure 6: The main LHCb detector components, showing from left to right the Vertex Locator, the first Cherenkov detector (RICH1), the Trigger Tracker (TT), the dipole magnet, the tracking stations (T1-T3), RICH2, the firt muon station (M1), the scintillating pad detector (SPD) and preshower detector (PS), the electromagnetic (ECAL) and hadronic (HCAL) calorimeters as well as the remaining muon stations (M2-M5). [20].

3.2.1 The tracking system

The particles that traverse the LHCb detector can on their way deposit energy in the different detector systems. The aim of the tracking detectors is to measure the position of those energy depositions and provide thus input to the reconstruction of the trajectory of the particle through the detector (track), and the charge and momentum by its curvature in the magnetic field. To provide the optimal resolution of these observables the tracking system consists of several parts of which each is designed specifically for its purpose.

The magnet

Momentum and charge of the particles in the detector are measured by their curvature in the magnetic field. For this purpose LHCb uses a dipole magnet with the main field component in y direction. The integrated magnetic

field over a distance of about 10 m in z, approximately from the interaction point to the end of the tracking system, is $\int B\,dl = 4.2\,Tm$. Figure 7 shows the main component of the B field as a function of the z coordinate in the detector. The field component in x direction is negligible. The relative momentum resolution from the measurement of the curvature of the tracks is about $\sigma_p/p = 0.6\%$.

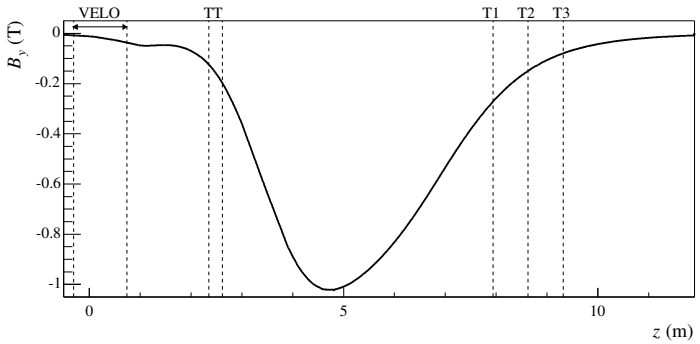


Figure 7: Primary component of the magnetic field of the LHCb dipole magnet as function of the z position in the detector. The position of the tracking detectors is indicated by dashed lines. Figure taken from [20].

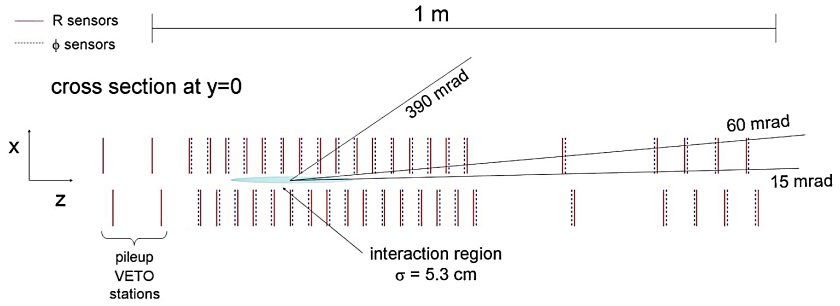
The vertex detector

The vertex detector is the detector closest to the primary pp collision in the event. It consists of two times 21 silicon detector modules placed left and right along the z axis around the PV, as shown in Figure 8a. Each module houses two different types of silicon strip sensors designed for the measurement of the coordinate r and ϕ of the track on the front and back of the sensor, *cf.* Figure 8b. The pitch size between channels varies depending on the position in the module to keep a constant occupancy of about 1% over the whole sensor.

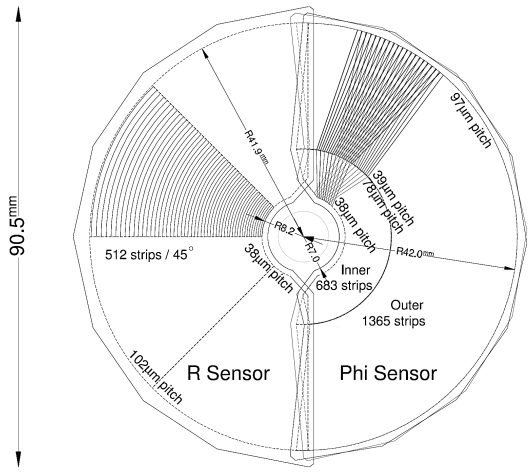
The two halves of the VELO are movable along the x axis. During injection and calibration of the beam the VELO stays in a save position about 15 cm away from the beam to avoid damage in case of beam losses. Once stable beams are achieved it gets moved to the nominal position where the active part of the sensors are about 8 mm away from the beam. The VELO is shielded from the vacuum of the beam and possible interference of the beam current with the electronics with an RF foil. This foil is only a few micron thick and about 5 mm away from the beam axis in the nominal position of the VELO. The material of the RF foil is as thin as possible to reduce multiple scattering of the particles which would reduce the spatial and momentum resolution of the reconstructed tracks.

The trigger tracker

The trigger tracker is situated about 2.5 m downstream of the interaction point directly in front of the dipole magnet. It consists of two 150 cm wide an 130 cm high tracking stations constructed using silicon microstrip sensors. Each station consists of two layers, one x- and one rotated stereo layer. This results in a (x,u,v,x) configuration of the whole detector, where the u (v) layers are rotated by $= 5$ deg (-5 deg) with respect to the y-Axis to allow for

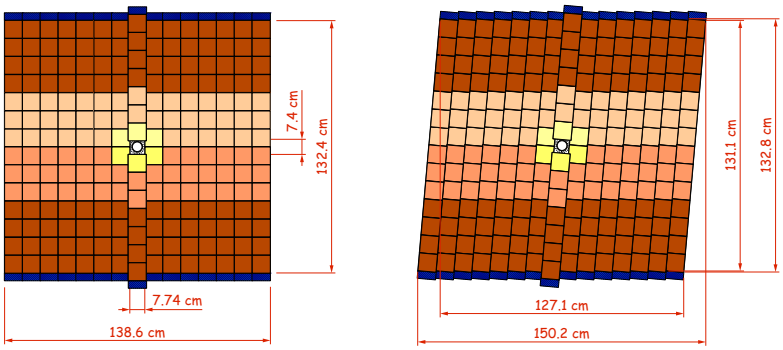


(a) Layout of the VELO, top view



(b) Single VELO sensor.

Figure 8: Schematic layout of the VELO design as viewed from the top (a) and layout of a single VELO station (b) showing r and ϕ sensors. Figure taken from [20]



(a) TT x layer. **(b)** TT stereo layer.

Figure 9: Design of the TT stations for x layers (a) and stereo layers (b). [20]

a two dimensional measurement of the particles position. The design of the two different layers is shown in Figure 9.

The inner tracker

The inner tracker constitutes the innermost part of the main tracking detectors. To cope with the intense particle fluxes in the inner part of the detector around the beam pipe it is build using silicon strip sensors. With this technology, a low occupancy of about 2% can be reached. The IT covers an 1.2 m wide and 0.4 m high area around the beam pipe. The main tracking detector consists of three stations along the z axis. Each station contains four IT layers build in a complementary (x,u,v,x) geometry as the TT with the same ± 5 deg stereo angle with respect to the y-axis. The layout of the IT modules is shown in Figure 10. Figure 11a shows how the IT is embedded in the middle of the main tracking detectors.

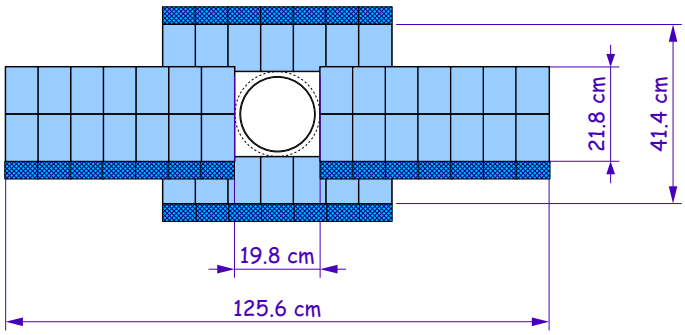
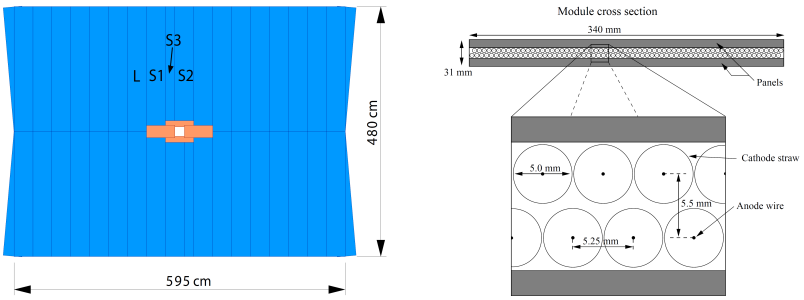


Figure 10: Design of the IT x layers. [20].

The outer tracker

The outer tracker is a drift-tube detector directly adjacent to the IT. It too consists of three stations with four layers each in a (x,u,v,x) configuration. Each OT station is 4.8 m high and 5.95 m wide, leading to an acceptance of 300 mrad in the horizontal and 250 mrad in the vertical plane. The layout of



(a) OT station with x and stereo layer.

(b) Top view on a single OT module.

Figure 11: Design of the OT stations with x layer and stereo layer (a) and top view on a single OT module (b) showing the double layer of straws. [20]

a complete station of the main tracking detector, consisting of OT and IT, is shown in Figure 11a.

The detector uses straw tubes for the detection of particles through ionization of a gas. As counting gas, a mixture of $Ar/CO_2/O_2$ is used. Each detector layer consists of 20 modules that contain two layers of straws. The straws have an inner diameter of 4.9 mm. The two straw layers inside one module are shifted by half a straw diameter for optimal spatial resolution. The layout of the straws inside the modules is shown in Figure 11b. The occupancy in the detector can be up to 10% to 15% in modules close to the beam pipe.

3.2.2 Particle identification

Different particles are produced in the primary proton - proton collision, predominantly pions, kaons, protons, electrons and muons. Together with the reconstructed track of the particle in the detector, a hypothesis of the particle type is necessary for physics analysis. To form such an hypothesis, inputs from the different particle identification subdetectors are combined in a common probability for a specific particle hypothesis. Based on this probability, a likelihood $\mathcal{L}$ is calculated. The difference in the logarithmic likelihood for a given particle hypothesis is an important variable to distinguish different particles species. To decide, *e.g.*, if a particle is more likely a kaon than a pion, the difference in the logarithmic likelihood for the kaon and pion hypothesis is used,

$$\mathrm{DLL}(K-\pi) = \ln \mathcal{L}_K - \ln \mathcal{L}_\pi \,. \tag{20}$$

If this likelihood difference $\mathrm{DLL}(K-\pi)$ is greater than zero, the particle is more likely a kaon than a pion, if it is smaller, the pion hypothesis is more likely. In the following paragraphs, the subdetectors contributing to the particle hypothesis are shortly discussed.

The Ring Image Cherenkov Detectors

Cherenkov detectors are used to differentiate between different types of charged particles. The speed of light in a medium is give by $c' = \frac{c_0}{n}$, where c_0 is the speed of light in the vacuum and n is the refractive index of the medium. The particles produced at the LHC are highly relativistic, thus their velocity v in a certain medium can be greater than c'. If this happens,

the particles emit photons in a cone around their flightpath with an opening angle of

$$\cos(\Theta_c) = \frac{1}{\beta n}, \tag{21}$$

where $\beta = v/c_0$ is the boost of the particle. By measuring the opening angle in the Cherenkov detectors, the velocity of the particle can be determined. Together with the momentum p of the particle, a mass hypothesis can be calculated and thus the particle type is determined. The range of the particle momentum over which a clear separation of the different particle types is possible depends on the refractive index n of the medium.

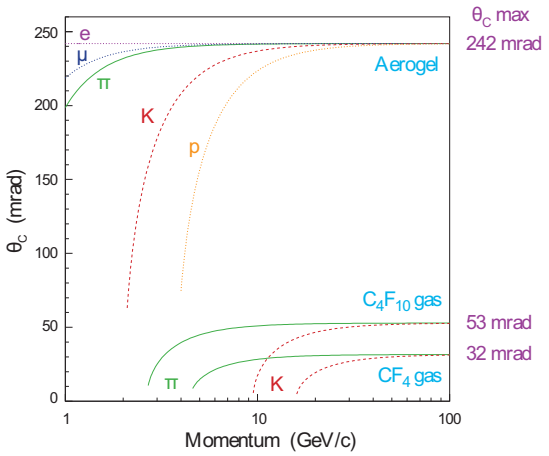
The LHCb experiment uses two Ring Image Cherenkov Detectors (RICH) with different radiator media. RICH1 in front of the dipole magnet uses aerogel and C_4F_{10} gas as radiator to distinguish low energetic particles with a momentum of 1 GeV to 60 GeV. RICH2 behind the main tracker uses CF_4 and covers a momentum range of 15 GeV to 100 GeV. The distribution of the opening angle versus the particle momentum for the different radiators used in the RICH detectors is shown in Figure 12a.

In the RICH detectors, the Cherenkov light emitted by the particles is collected by spherical and plain mirrors and reflected to hybrid photo detectors (HPDs) and photomultiplier tubes, where it is detected. An outline of the design of the two RICH detectors used by LHCb is given in Figures 12b and 12c.

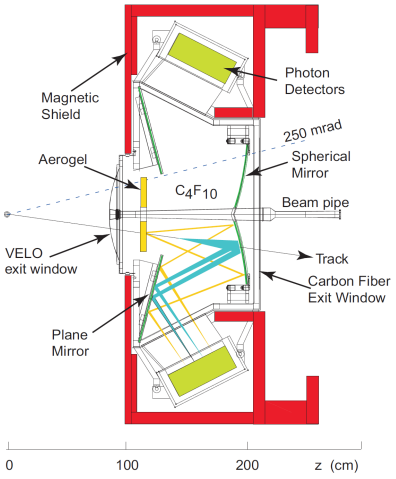
The calorimeter system

The calorimeter system is designed to detect photons and electrons in the electromagnetic calorimeter (ECAL) and neutral and charged hadrons in the hadronic calorimeter (HCAL). Particles traversing the detector interact with the material and start to produce secondary particles in hadronic or electromagnetic showers. The calorimeters are interspersed with scintillator material to detect those showers. The light that is produced by the shower particles via ionisation and excitation of the atoms in the scintillator is collected by optical fibers and guided to photomultipliers, where it is detected. As the particles are usually stopped in the calorimeter, they loose almost all their energy in the calorimeter material. The amount of light that is visible in the calorimeter is roughly proportional to the energy deposited by the particle. In positive z direction the calorimeter system is build up as follows:

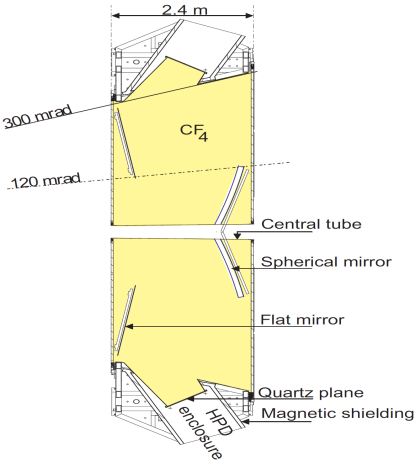
- The Scintillating Pad Detector (SPD) differentiates charged particles, which produce scintillation light in the SPD tiles, from neutral particles, which do not.
- The Preshower (PS) mainly distinguishes electrons and photons from hadrons. Together with the SPD. the material budget of the PS corresponds to 4 electromagnetic interaction lengths X_0, enough to get electrons and photons to produce an electromagnetic shower which can be detected. In contrast, the total material budget of SPD and PS corresponds to only 0.2 hadronic interaction lengths which will not cause a hadronic shower in most of the cases.
- The ECAL is build as a shashlik type calorimeter with alternating layers of lead and scintillating tiles. The total material budget of the ECAL corresponds to 25 electromagnetic interaction lengths and 1.1



(a) Opening angle vs. particle momentum



(b) RICH 1 design.



(c) RICH 2 design.

Figure 12: Design of the LHCb Cherenkov detectors showing the opening angle Θ_c observed for the different particle types depending on the radiator that is used and the particle momentum (a) as well as RICH1 (b) and RICH2 (c). [20]

hadronic interaction lengths, which is enough to stop and detect almost all electron and photons in the ECAL.

- The HCAL is a sampling type calorimeter made of iron blocks and scintillating tiles. The HCAL material amounts to 5.6 hadronic interaction lengths. The design of the HCAL cells is shown in Figure 13.

As the energy of the particle is used, in addition to information from the RICH detectors and tracking system, to form a particle hypothesis, it is crucial to achieve a reasonable energy resolution. The energy resolution for the LHCb calorimeter is

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E/\,\mathrm{GeV}}} \oplus 1.5\% \tag{22}$$

for the ECAL and

$$\frac{\sigma(E)}{E} = \frac{80\%}{\sqrt{E/\,\mathrm{GeV}}} \oplus 10\% \tag{23}$$

for the HCAL, where $\oplus$ indicates that the two terms are added in quadrature.

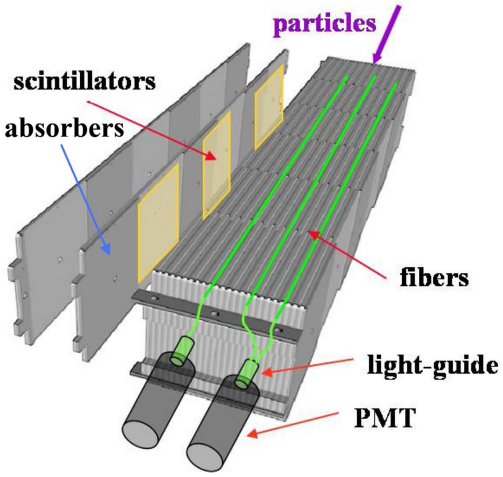
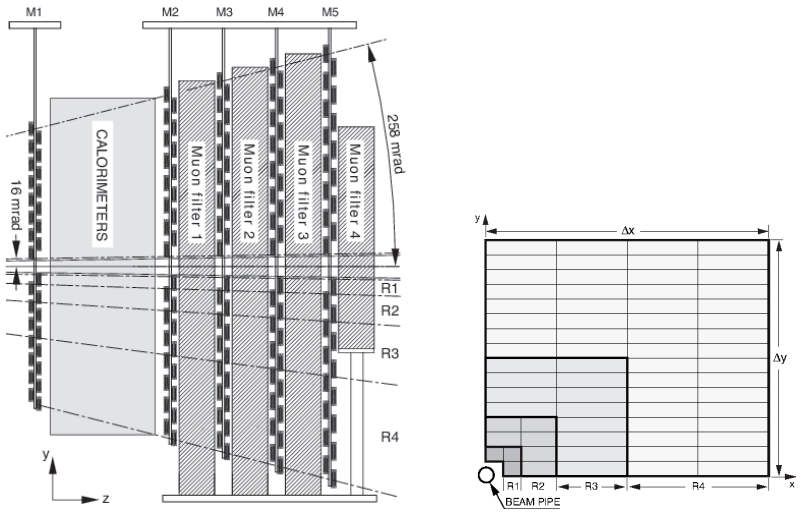


Figure 13: Design of a calorimeter cell from the LHCb HCAL. [20].

The muon chambers

As muons are not stopped in the calorimeter they can be detected by a dedicated subdetector. The muon system consists of five stations. The first station M1 is located in front of the calorimeter. The other four stations M2-M5 are located behind the calorimeter system, as is shown in Figure 14a. Between M2 to M5, 80 cm thick iron walls are placed to absorb punch-through hadrons from the calorimeter. Those walls have a high stopping power and muons therefore need a minimum momentum $p > 6\,\mathrm{GeV}$ to reach M5. The muon system uses two different detector technologies: Gas electron multipliers (GEM) near the beampipe, where the particle density is high and a high granularity is needed and multiwire proportional chambers (MWPC) in the outer regions. The segmentation of the muon chambers is



(a) Muon chambers, side view. **(b)** Granularity of muon chambers.

Figure 14: Design of the LHCb muon detector showing the position of the different muon stations (a) and the granularity of a single station (b). [20]

shown in Figure 14b, where the design of one quadrant of a single station is shown.

As many of the B meson final states interesting to LHCb contain muons, this subdetector plays an important role in both hardware and software trigger.

3.2.3 The LHCb trigger system

The LHC features interaction rates of 40 MHz in its nominal running condition, but the rate of processes interesting to the LHCb physics program is several orders of magnitude smaller. Furthermore, the ability to write events to disk is limited to about 3 kHz [21]. The task of the LHCb trigger system is therefor to analyse events recorded in the detector quickly and filter out interesting processes, that are then stored to disk for later analysis. The LHCb trigger system is divided into three parts which are subsequently run to determine the final trigger decision. The first stage, the so called Level 0 (L0), is implemented in hardware. Its aim is to reduce the original 40 MHz event rate to a rate of about 1 MHz at which a detailed analysis is possible. For that, the L0 uses information from the muon system, the calorimeter and the pile-up system in the VELO, to identify for example muons and hadrons with a high transverse momentum, that are typical for many of the B meson decays.

The first stage of the high level software trigger (HLT1) further reduces the rate of the 1 MHz output of the L0 to about 50 kHz. To do so, it performs a partial reconstruction of the event to either confirm or reject the decision of the L0.

The second stage of the high level software trigger (HLT2) finally performs a full event reconstruction at a rate of 50 kHz. At this rate the actual physics process happening in the event is reconstructed using inclusive or exclusive selections, and the final decision is taken whether or not the event is kept.

An overview of the LHCb trigger strategy is given in 15.

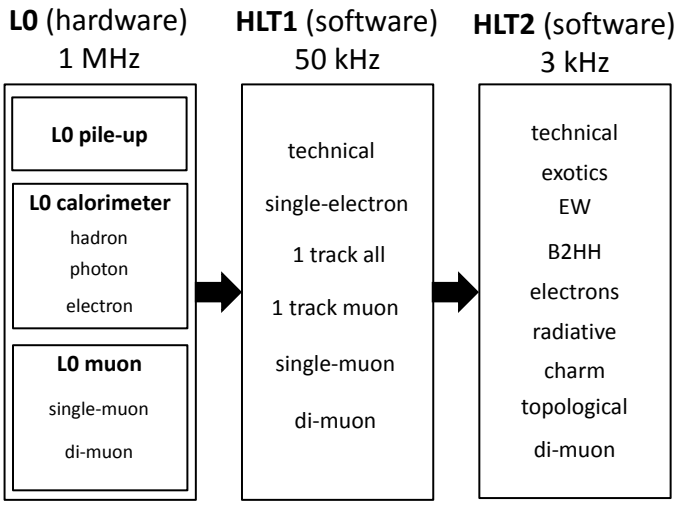


Figure 15: Overview of the LHCb trigger strategy. Figure taken from [22].

3.2.4 The LHCb software framework

The LHCb software is organized in the GAUDI framework. Several steps are taken to reconstruct the actual physics data and select interesting events or to generate simulated events.

Generation of simulated events

A detailed simulation of events is necessary for the analysis of the data that is taken in the experiment. Events are simulated using the Monte Carlo (MC) technique. Simulated pp-collisions are generated with PYTHIA v6.2 [23] using a dedicated LHCb tuning [24]. For the simulation of B meson decays, a special program, EVTGEN [25], is used. GEANT4 [26] subsequently simulates the propagation of the generated particles through the detector and the magnetic field and their interaction with the detector material. All these steps are integrated in the GAUSS software package. To simulate the detector response of the simulated particles while taking into account inefficiencies and other detector effects, the dedicated software package BOOLE is used.

Track reconstruction software

After the events have been written to disk by the trigger or simulated with GAUSS and BOOLE, the offline event reconstruction with BRUNEL is performed. This step is identical for both real data and simulated events. During the track reconstruction, several different track types can be formed depending on the algorithm. An overview over the track types is given in Figure 16.

The tracks of the highest quality are the ones that traverse the whole spectrometer and have hits in both the vertex locator and the main tracker. These so called long tracks are the main track type used in physics analysis. For the flavour tagging algorithms, tracks that have only hits in the VELO and the TT, called upstream tracks, are used additionally. Other track types consist of VELO tracks, which only have measurements in the vertex detector, downstream tracks, that have hits in the TT and main tracking stations and

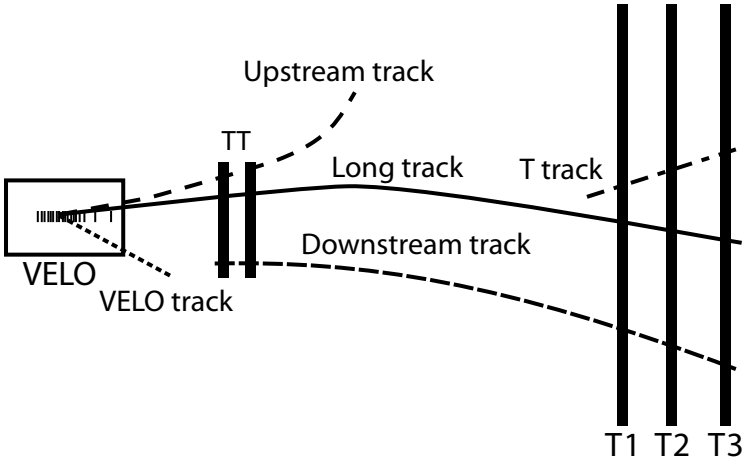


Figure 16: Overview of the LHCb tracking system and the different track types in the track reconstruction. [20].

T track segments, that consist only of measurements in the main tracking detector.

Physics analysis software

The final step of the physics analysis is carried out by a group of algorithms organized in the DaVinci package. DaVinci uses the tracks reconstructed by Brunel and combines them to subsequently reconstruct the actual B meson decay chain using inclusive and exclusive selections depending on the actual decay channel. To make the large amount of data more manageable and to save CPU time, these selections are bundled for the key physics channels in so called stripping lines. The output of these stripping lines, which contains the reconstructed B meson decays, can then be used for the analysis.

4 THE PRINCIPLES OF FLAVOUR TAGGING

For the measurement of the oscillation in the system of neutral B mesons[1] it is necessary to have a precise knowledge of the flavour of the reconstructed B meson at the time of its production and decay. The flavour at the decay can often be deduced directly from the charge of the decay products in flavour specific decays, the production flavour however is hidden by the oscillation of the neutral B mesons and has to be extracted by a special technique, called flavour tagging.

Two different approaches in the tagging of the initial flavour of B mesons are used in LHCb, which exploit particularities of the B meson production. This chapter explains in detail the physics principles behind these approaches and their implementation in dedicated algorithms. The focus is put on the flavour tagging algorithm that is studied in this thesis, the so called same side kaon tagging algorithm. The discussion continues with an introduction to the flavour tagging performance variables and a discussion about the predicted mistag probability that is provided by the tagging algorithms. A formalism for the combination of several tagging algorithms is discussed and a short overview about the tagging performance in the dataset used in this thesis is given.

4.1 INTRODUCTION TO THE FLAVOUR TAGGING PROCEDURE AND TERMINOLOGY

At the LHC, b and $\overline{b}$ quarks are predominantly produced in pairs. After their initial production, the $b\overline{b}$ quark pair hadronizes and two hadrons containing a b quark[2] are formed. One of these hadrons, the so called signal candidate, is reconstructed and used in the physics analysis while the other hadron, called opposite side (OS) B, can be used for tagging purposes. The flavour tagging algorithms exploit special features of the $b\overline{b}$ event to determine the production flavour of the signal B hadron. These algorithms can be split into two groups: Algorithms that use the decay chain of the opposite side B hadron for the determination of the signal B hadron flavour are called opposite side tagging algorithms (OST). Algorithms that exploit the fragmentation of the signal B hadron are called same side tagging algorithms (SST).

The tagging algorithms used by LHCb so far all implement a common procedure. In the first step, reconstructed particle trajectories (tracks) with sufficiently good quality for the flavour tagging are selected from all reconstructed tracks. The cuts from this preselection are listed in Table 1, *cf.* [27]. To ensure a good track quality the tracks are required to have a $\chi^2_{\text{track}}/\text{ndf} < 5$ and are either long tracks or upstream tracks. Tracks that traverse the beam pipe under a low angle are usually poorly reconstructed

1 B meson is synonym for all types of mesons that include b quarks.
2 Throughout this chapter, charge conjugation of particles is implied unless explicitly stated.

due to multiple scattering. An angle $\theta > 12\,\text{mrad}$ of the track with respect to the z axis rejects those tracks. To ensure that the tagging decision is not influenced by the signal B meson decay products, all tracks that are explicitly contained in the decay chain are vetoed and the tracks are required to be outside a cone of 5 mrad around the B and its decay products to avoid influence in case of partially reconstructed decays. A cut that requires a momentum larger than 2 GeV/c excludes low momentum particles. In addition the tracks are required to have an impact parameter significance $\text{IP}_{\text{PU}}/\sigma_{\text{IP}_{\text{PU}}} > 3$ with respect to eventual other pile-up (PU) primary vertices in the event to avoid that tracks which do not belong to the production vertex of the $b\bar{b}$ pair, but to other primary vertices in the event are used in the tagging procedure. The impact parameter (IP) of a track with respect to a vertex is given by the shortest distance between the track and the vertex, with σ_{IP} its uncertainty. Maximum momentum and transverse momentum cuts reject particles that due to kinematic constraints can not originate from B decays or from the fragmentation.

Preselection cuts	
p	$2\,\text{GeV/c} < p < 200$ GeV/c
p_T	< 10 GeV/c
$\chi^2_{\text{track}}/\text{ndf}$	< 5
track type	*long* or *upstream*
charge	± 1
θ	$> 12\,\text{mrad}$
$\lvert\Delta\phi\rvert$	$> 5\,\text{mrad}$
$\text{IP}_{\text{PU}}/\sigma_{\text{IP}_{\text{PU}}}$	> 3
other	not in the signal B decay chain

Table 1: Preselection cuts applied to select a tagging track candidate. [27]

Tracks that pass the preselection, the so called tagging track candidates, are used as input for the various tagging algorithms that provide the tagging decisions. This tagging decision, also called tag, is defined as $q_{\text{tag}} = 1$ for signal B hadrons containing a $\bar{b}$ quark and $q_{\text{tag}} = -1$ for signal B hadrons containing a b quark. If no decision is possible, the tag is $q_{\text{tag}} = 0$. In case the tagging algorithm arrives at a decision it also predicts the probability of this decision to be wrong.

4.2 TAGGING PERFORMANCE VARIABLES

The tagging decision can either be correct, wrong or the tagging algorithm is not able to make a decision at all. As only those candidates where the tagging algorithm is able to determine the initial flavour can be used in the analysis, the tagging efficiency, defined as

$$\varepsilon_{\text{tag}} = \frac{R + W}{R + W + U}\,, \tag{24}$$

effectively reduces the statistical size of the sample. Therein, R, W and U are the number of right tagged, wrong tagged and untagged events.

If the tagging algorithm gives a wrong decision about the initial flavour, the signal that is to be measured is diluted. The fraction of events in which this is the case is given by the mistag probability or mistag fraction

$$\omega = \frac{W}{R+W}. \tag{25}$$

The significance of an analysis is depending on the combination of both effects in the effective efficiency

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2 = \varepsilon_{\text{tag}} D_{\text{tag}}^2 . \tag{26}$$

It is the factor by which the statistical power of the sample is reduced due to the imperfect tagging and represents the figure of merit in an optimization of a tagging algorithm. The factor $D_{\text{tag}} = 1-2\omega$ is the tagging dilution.

4.3 THE PREDICTED MISTAG FRACTION

In addition to the tagging decision, the tagging algorithms provide an estimate of the probability of a tagging decision to be wrong using a neural network classifier. This probability can be used in the analysis to increase its significance by giving events with a better (*i.e.* more likely correct) tagging decision a larger weight. The neural network predicts the probability of a wrong tagging decision on the basis of the kinematic quantities of the tagging track candidate and further input variables connected to the event topology and signal B candidate. Variables connected to the event topology are used in the determination of the predicted mistag probability to account for correlations of the tagging performance with these variables. The neural network classifiers are trained using simulated events and differences between the simulation and the data can introduce a bias in the predicted mistag probability. To make sure that the prediction can be interpreted as a probability, it has to be calibrated. A linear dependence of the true mistag fraction ω and the predicted mistag probability η is assumed, given by the following calibration function

$$\omega = p_0 + p_1 \cdot (\eta - \langle\eta\rangle), \tag{27}$$

where p_0 and p_1 are the calibration parameters and $\langle\eta\rangle$ is the average mistag of the calibration sample. For a perfect calibration the parameters are $p_0 = \langle\eta\rangle$ and $p_1 = 1$. This parametrisation of the linear calibration function is chosen to minimize the correlations between the two calibration parameters. To check for possible non-linearities of the calibration is an important systematic study.

Because the event properties are included in the neural network, the calibration parameters are supposed to be independent of these variables and hence, the calibration can be transported from one decay channel to others. This assumption has to be checked as a systematic study.

If a per event predicted mistag probability is used, an effective mistag fraction for the whole sample can be calculated as

$$\omega_{\text{eff}} = \frac{1}{n}\sum_{i=0}^{n} \omega_i , \tag{28}$$

by summing over the predicted mistag probability for all events. With this effective mistag fraction ω_{eff}, the effective tagging efficiency ε_{eff} can be calculated.

4.4 COMBINATION OF DIFFERENT TAGGING ALGORITHMS

If multiple tagging algorithms are available, their decision and predicted mistag probabilities are combined in order to use them most efficiently. A formalism for the combination of several tagging algorithms is described in [28]. Assuming that the algorithms are uncorrelated, the combined probability of a signal B hadron to contain a b quark, $P(b)$, or $\overline{b}$ quark, $P(\overline{b})$ is given by

$$\mathcal{P}(b) = \frac{p(b)}{p(\overline{b}) + p(b)}, \qquad \mathcal{P}(\overline{b}) = 1 - \mathcal{P}(b) \tag{29}$$

with

$$p(b) = \prod_i \left(\frac{1 + q_i}{2} - q_i(1 - \omega_i) \right) \tag{30}$$

$$p(\overline{b}) = \prod_i \left(\frac{1 - q_i}{2} + q_i(1 - \omega_i) \right), \tag{31}$$

where q_i is the tagging decision of the i-th tagging algorithm and ω_i its calibrated mistag probability.

After the combination, it is decided whether the signal B hadron more likely contains a b quark ($q = 1$) or an $\overline{b}$ quark ($q = -1$) depending on whether $\mathcal{P}(b) > \mathcal{P}(\overline{b})$ or vice versa. The combined η is set to $\eta = 1 - \mathcal{P}(b)$ in case of $\mathcal{P}(b) > \mathcal{P}(\overline{b})$ or $\eta = 1 - \mathcal{P}(\overline{b})$ otherwise.

The assumption of uncorrelated tagging algorithms that is used in this formalism is justified for the combination of same side kaon and opposite side tagging algorithms, as they rely on different physics principles and employ rectangular cuts. It has to be checked that this assumption holds as a systematic study.

4.5 THE OPPOSITE SIDE TAGGING

The opposite side tagging employs the fact that b quarks are predominantly produced in quark anti-quark pairs. While one of the two quarks will hadronize into the signal B hadron, the other quark will also hadronize into a meson or baryon containing a b quark of the opposite flavour than the signal B hadron. The opposite side tagging aims to select particles originating from specific decay processes of this second b quark to determine its flavour and deduce the flavour of the signal B meson.

There are several general complications of this approach. First of all, the opposite side b event might decay outside of the acceptance of the detector or some of its decay products might not be reconstructed because of detector

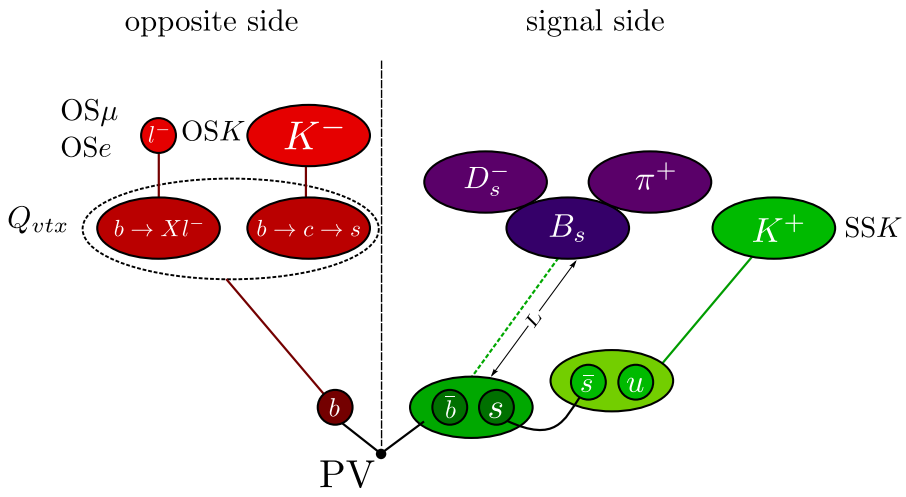


Figure 17: Schematic overview of the tagging strategy for the example of a $B_s^0 \to D_s^+\pi$ decay.

inefficiencies, in which case the information contained in this decay is lost. Second, the opposite side b quark might also hadronize into a neutral B meson and oscillate, in which case the information reconstructed from its decay is diluted, as the decay flavour and the production flavour is no longer identical. Finally, the inclusive reconstruction of the opposite side b quark decay could pick up tracks that do not belong to this decay, which would also dilute the tagging.

One key feature of the opposite site tagging algorithms is the fact that they can be easily optimized using real data, as they are supposed to behave the same regardless of the type of B hadron on the signal side. For this reason, the calibration and optimization of the tagging performance is possible using fully reconstructed decays of charged B hadrons, such as $B^+ \to J/\psi K^+$, which are available at sufficient statistics, and the optimization and calibration does not rely on simulated events.

Figure 17 presents a schematic overview of the flavour tagging strategy that is currently applied in the LHCb experiment with the OS tagging consisting of four different algorithms, the electron, muon, kaon and vertex charge tagging algorithm. A brief overview on the different taggers is given in the following paragraphs, for detailed information *cf.* [28].

Single particle tagging algorithms

The single particle tagging algorithms among the OS tagging algorithms search for charged $\mu^\pm$ or $e^\pm$ from semileptonic decays of the opposite site B hadron, such as $B^+ \to \bar{D}^0 \ell^+ \nu_\ell$ or for $K^\pm$ from $b \to c \to s$ transitions. They apply tight cuts on the particle identification to select the right particle species and reject other tracks such as kaons not coming from the B decay. Electrons, muons and kaons coming from a B decay are characterised by a high impact parameter and IP/σ_{IP} because of the long lifetime of the B mesons and a high transverse momentum, both of which are used in the final selection of the tagging candidate. If more than one tagging track candidate survives the selection, the one with the highest p_T is chosen and its charge defines the tagging decision.

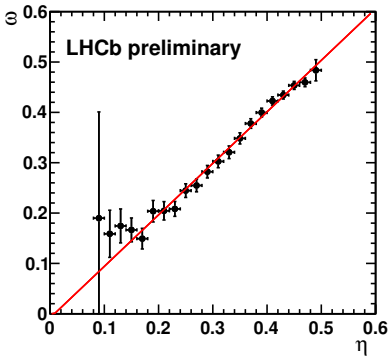


Figure 18: Measured mistag fraction ω as function of the predicted mistag probability η of the combined opposite side tagger determined from $B^+ \to J/\psi K^+$ decays in the data. Figure taken from [29].

Vertex charge tagger

The vertex charge tagger tries to inclusively reconstruct the decay vertex of the opposite side B meson. It starts by reconstructing this vertex using two track combinations of all tagging track candidates after the preselection. For those vertex seeds, the likelihood to be caused by the opposite side B decay is calculated based on the χ^2/ndf from the vertex fit, the transverse momentum, the IP and the IP/σ_{IP} of the tracks. The seed with the largest likelihood is taken and extended by adding tagging candidates that fulfil kinematic and geometrical requirements such, that they are compatible with coming from the opposite site B decay. After applying additional selection criteria on the reconstructed decay vertex that require a minimum sum of transverse momentum and IP/σ_{IP} and a maximum sum of the distance of closest approach of all the tracks in the vertex, the tagging decision is derived from the charge of the vertex. This is achieved by building the transverse momentum weighted sum of the charge of the tracks belonging to the vertex.

Performance of the opposite side tagging algorithms in the 2011 dataset

All available opposite site tagging algorithms have been optimized and calibrated using the full 2011 dataset with the signal decay $B^+ \to J/\psi K^+$ [29]. The resulting tagging performance for the individual taggers as well as for the combined opposite site decision is given in Table 2. The calibration of the predicted mistag probability from the combined opposite side decision is shown in Figure 18. The combined opposite side tagger has an efficiency of $\varepsilon_{tag} = 33.2 \pm 0.09\%$ at an effective mistag fraction of $\omega = 36.7 \pm 0.2\%$, leading to an effective efficiency of $\varepsilon_{eff} = 2.35 \pm 0.06\%$ in this channel. The tagging performance in other decay channels may vary due to their specific properties such as the momentum distribution of the B mesons or different trigger scenarios that influence the event topology.

4.6 THE SAME SIDE TAGGING ALGORITHMS

The same side tagging algorithms exploit the fragmentation process of the signal B meson to determine its production flavour. When the signal B

Individual taggers

	ε_{tag}[%]	ω_{av}[%]	ε_{eff}[%]
OS μ	5.20±0.04	30.8±0.4	0.77±0.04
OS e	2.46±0.03	30.9±0.6	0.36±0.03
OS K	17.67±0.08	39.33±0.24	0.81±0.04
OS Q_{vtx}	18.46±0.08	40.31±0.24	0.70±0.04

Combined OS tagger

	ε_{tag}[%]	ω[%]	ε_{eff}[%]
OS combination	33.2±0.09	36.7±0.2	2.35±0.06

Table 2: Measured performance of the individual opposite side taggers and the opposite side combination in the 2011 dataset with the signal decay $B^+ \to J/\psi K^+$. For the individual taggers the performance using an average mistag fraction is given. The performance of the combination is measured using per-event predicted mistag probability. Table taken from [29].

meson is created from the initial $b\bar{b}$ pair, a second quark pair, a $d\bar{d}$ or $s\bar{s}$, is needed to form either a B^0 or B_s^0 meson. While one of these quarks forms a bound state together with the b, the other is unbound and participates further in the hadronization process. With some probability a charged pion for the B^0 or kaon for the B_s^0 is formed and the charge of this particle determines the flavour of the initial B meson. The task of the same side tagging is to find this particle from the fragmentation to tag the flavour of the B^0 or B_s^0 meson.

Several facts complicate the development, optimization and calibration of the same side tagging algorithms. First, they only work correctly with B^0 and B_s^0 mesons as signal decays. As those neutral B mesons oscillate and hence their production and decay flavour are different, a fit of this oscillation is necessary to determine the tagging performance. Secondly, as it is not possible to access the information whether a particle originates from the fragmentation process in the data, one has to rely on the correct description of this process in the simulation. Furthermore, the fragmentation particle that is used to tag the flavour originates from the primary interaction point, thus it is hard to separate from the underlying event.

As the same side tagging is possible for both B^0 and B_s^0 mesons, it is implemented in two different algorithms, the same side pion and same side kaon tagging algorithms. The basic principle is the same for both algorithms, so this section will concentrate on a discussion of the same side kaon tagging algorithm, as this is the algorithm of interest for this thesis.

4.6.1 Modeling of the fragmentation process

The production of B mesons is initiated by the production of a $b\bar{b}$ quark pair, as explained in Chapter 2.4. While those processes happen at a relatively high energy and their cross sections can therefore be calculated perturbatively, it is not possible to calculate the subsequent fragmentation process in the same fashion. The fragmentation happens at a lower energy which, due to the energy dependence of the strong coupling constant α_s, introduces non-perturbative effects. It is therefore necessary to model the fragmentation process by a phenomenological model. One of the most important models, which is also used in the simulation of events with PYTHIA, is the

so called Lund string fragmentation model [30–33].
The model starts with the assumption of an initial $q\bar{q}$ pair, an $b\bar{b}$ quark pair in our case, where the two colour charges move rapidly in opposite directions. Between the two charges the so called *string* is formed, which represents the strong colour field between the two quarks. This string has a constant energy per unit length, called the *string tension* κ, which is of the order $1\,\mathrm{GeV/fm}$. Once the amount of energy which is stored in the colour field between the quarks due to their relative movement is large enough, new $q\bar{q}$ pairs are formed from the vacuum and the string is cut in half. The remaining string tension is lower, because some of the available energy was used for the creation of the new quarks. This process can be repeated until the quark pairs, which are bound by the strings, have a kinetic energy which is lower than the binding energy of the string. In this case they form hadrons.

The probability $P(m_T)$ of the creation of a new $q\bar{q}$ pair is parametrized in the Lund model as

$$P(m_T) \propto \exp\left(-\frac{\pi \cdot m_T^2}{\kappa}\right) = \exp\left(-\frac{\pi \cdot m^2}{\kappa}\right) \exp\left(-\frac{\pi \cdot p_T^2}{\kappa}\right), \qquad (32)$$

where m is the invariant mass of the $q\bar{q}$ pair and $m_T^2 = m^2 + p_T^2$ its transverse mass. This confirms the naive expectation that the probability for the creation is smaller if the quarks are heavier. The relative probabilities for the production of the quarks as implemented in the simulation are $u : d : s : c = 1 : 1 : 0.3 : 10^{-11}$. Therefore the heavy c, b and t can be neglected in the fragmentation process. With regards to the ratios of the light quarks it has to be noted that it is by no means trivial to define what the quark mass used in the calculations is, as quarks happen to only exist in bound states. Deviations in these ratios can lead to a different mixture of hadrons that are produced in the fragmentation which will influence the theoretically possible tagging efficiency in the same side tagging.

Following these assumptions, the probability of producing a kaon with the right charge for the same side kaon tagging can be estimated by evaluating possible combinations in the fragmentation. This is visualized in Figure 19. Starting from an initial B_s^0 and an additional $\bar{s}$ quark on the left, in the next step a $u\bar{u}$, $d\bar{d}$ or $s\bar{s}$ pair is created with the probabilities calculated from Equation 32. This scheme also continues further. After the second step, the lightest possible mesons are combined from the available quarks. In about 50% of the cases an excess of kaons with the right tagging charge is produced. In reality, kaons can also be produced in fragmentation steps further down the chain, but in this case the correlation with the B_s^0 mesons, *e.g.* in terms of the distance between kaon and B_s^0 meson in phase space, gets smaller. This complicates the selection of this tagging track candidates in the same side kaon tagging algorithm.

As the transverse momentum of the hadrons formed in the fragmentation is already determined by the transverse momentum of its constituent quarks (because it is set by Equation 32 during their creation), the remaining degree of freedom is its sum of energy and longitudinal momentum, $E + p_z$. The fraction z of the whole available $(E + p_z)_{string}$ that is taken by a single

hadron from the fragmentation process is determined by the so called *Lund symmetric fragmentation function*,

$$f(z) = \frac{N}{z}(1-z)^{\alpha} \exp\left(-b\frac{m_T^2}{z}\right). \tag{33}$$

The parameters α and b are model parameters that have to be determined from data. One of the pitfalls in the simulation of the fragmentation process is the tuning of these parameters. The current tunings are derived from data, using e^+ e^- collisions at the Z pole recorded by the ALEPH, DELPHI and OPAL experiment at LEP [34]. It has yet to be verified that they are valid up to the energies present at the LHC. A more in detail discussion on the topic of fragmentation models can be found in References [35, 36], that have also been used for the overview presented in this paragraph.

4.6.2 The underlying Event

In contrast to the fragmentation process, which involves the creation of the B_s^0 meson, the underlying event can be defined as the sum of all other processes in the event, besides the B_s^0 fragmentation. This includes, among other processes, reactions of the other partons in the event and jets produced by initial state or final state radiation of soft gluons. Most of the energy available in the event is given to the hard process that creates the $b\overline{b}$ pair and consequently the processes of the underlying event take place at lower energies, thus they are also not perturbatively calculable and have to be modelled phenomenologically.

These models have to be tuned on data. The current tunings, derived from measurements at the Tevatron and other experiments [24], are thus far from perfect at the much higher energies of the LHC. One approach to test and tune the models that are used to simulate the underlying event is to measure so called *minimum bias* events. Those are events selected by special triggers and selections with the minimal requirements possible (in general they require that just one track is reconstructed) to not bias the measurement. The particle production in such events is measured depending on the phase space region and corrected for detector and reconstruction effects. Finally, the measurement can be compared with different models and the model parameters can be tuned according to the measurement. The results of such a measurement performed at LHCb is shown in Figure 20. The figures compare the production of charged particles that are reconstructed as long tracks in the underlying event in bins of the transverse momentum and the pseudorapidity as measured in the data with the model used in the current LHCb simulation. From the plot it is visible that the current simulation does neither describe the observed particle density nor their phase space distribution. This description however is crucial for a correct simulation of the same side tagging since it is, as explained previously, in particular susceptible to be negatively influenced by underlying event tracks.

4.6.3 The same side kaon tagging

The goal of the same side kaon tagging is to select the $K^{\pm}$ from the fragmentation process of the B_s^0 that determines its flavour at production time. Therefore it has to distinguish this $K^{\pm}$ against all other particles from the underlying event as well as against particles coming from the opposite side B

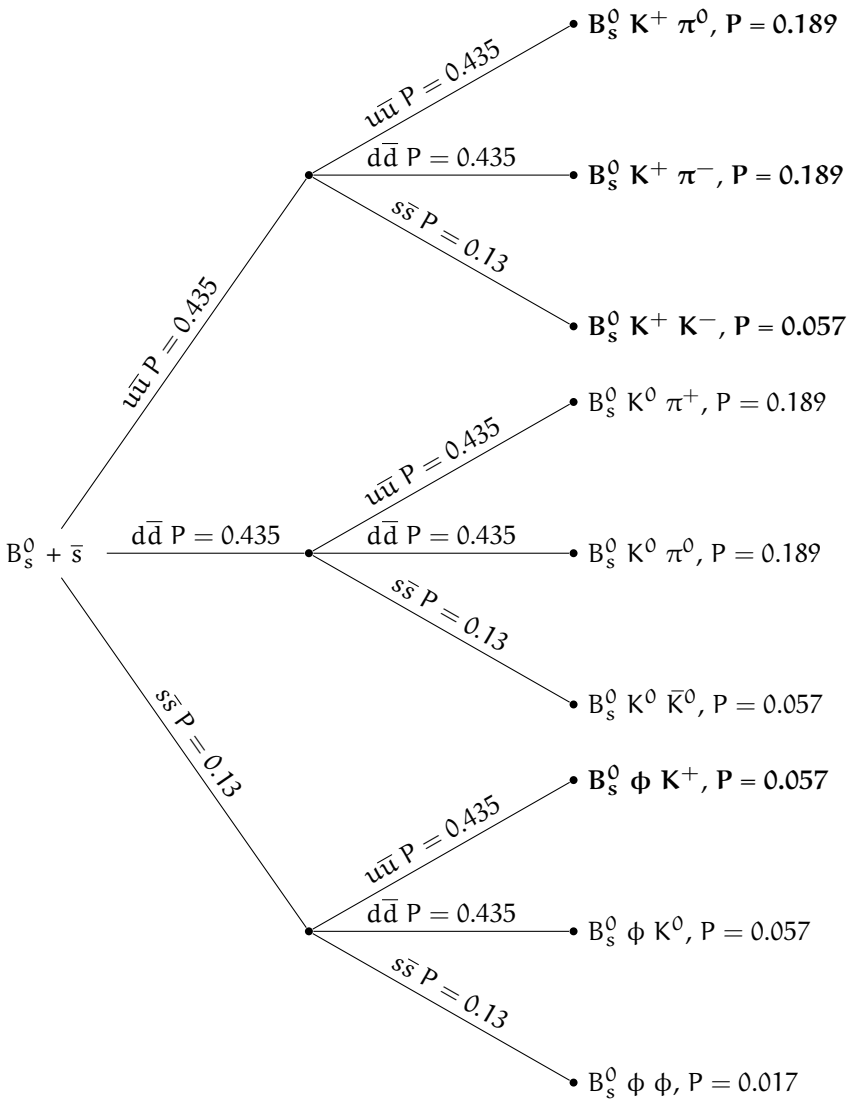
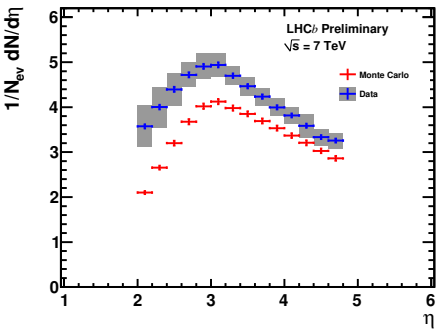
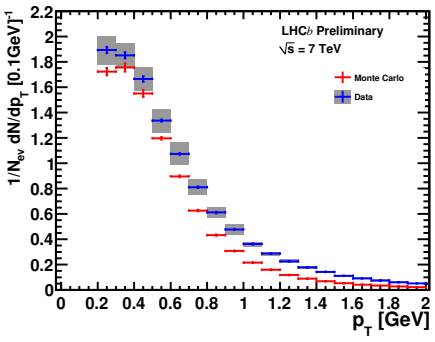


Figure 19: Schematic overview over the fragmentation process of the B_s^0 signal mesons, starting with a $B_s^0 + \bar{s}$ as initial state. The probability of producing a quark pair with a specific flavour is assumed to be $u\bar{u} : d\bar{d} : s\bar{s} = 1 : 1 : 0.3$. In about 50% of the cases, marked in bold, an excess of kaons with the right charge for SSK tagging is produced. It is assumed that only the lightest possible hadron is produced from the quark pairs. (*cf.* [37])



(a) Mean particle density as a function of η



(b) Mean particle density as a function of p_T.

Figure 20: Mean density of charged particles reconstructed as long tracks in the event as a function of η and p_T in the data and the simulation. Error bars represent the total statistical error, shaded bands the total systematic uncertainty. Figure taken from [38].

decay. This procedure was originally implemented in a cut based algorithm that was developed using simulated events and optimized by the author of this thesis using real data. A detailed description can be found in [39, 40].

Using all tagging track candidates that pass the preselection cuts summarized in Table 1 a final selection is applied, *cf.* [39]. Cuts on the difference in log-likelihood of the particle hypothesis between kaons and protons or kaons and pions reject background of prompt protons and pions. Since the same side kaon comes from the primary fragmentation of the b quark, it is supposed to be close to the signal B_s^0 in phase space. Cuts on the maximum difference of the pseudorapidity η and the polar angle ϕ between the B_s^0 momentum and the momentum of the tagging track candidate select candidates close to the B_s^0 while additional cuts on the minimum momentum and transverse momentum reject low momentum particles from the primary interaction. A cut on the maximum IP/σ_{IP} of the tagging track candidates with respect to the signal B_s^0 production vertex assures that the tagging track candidate is coming from this vertex. A cut on the maximum difference in invariant mass $dQ = m(B_s^0 K^\pm) - m(B_s^0)$ further restricts the $K^\pm$ candidates to a phase space volume close to the B_s^0. Finally, an additional cut on the χ^2_{track}/ndf ensures a sample of tagging candidates that are well reconstructed. In case more than one tagging particle satisfies the requirements, the one with the highest transverse momentum is selected. The tag is then determined by the charge of the selected kaon. In case no suitable kaon is found the B_s^0 is considered untagged.

A detailed comparison of the distributions between the data and the simulation for the variables used in the same side kaon tagging is discussed in Chapter 6.

An optimization of the cut values was performed using the same procedure to determine the tagging performance from the fit of the B_s^0 oscillation as described in this thesis, see Chapter 5.5. The cuts have been varied in turn in order to maximise the effective tagging efficiency. The procedure is described in detail in [39]. The final cut values are summarised in Table 3.

Selection cuts	
$DLL(K-\pi) >$	3.5
$DLL(K-p) >$	-8.5
p_T [GeV/c] >	0.85
p [GeV/c] >	5.25
$\chi^2_{track}/ndf <$	3.75
$IP/\sigma_{IP} <$	4.125
$\Delta\eta <$	0.6
$\Delta\phi <$	0.825
dQ [GeV/c^2] <	1.85

Table 3: Final set of cuts derived in the optimization procedure for the same side kaon tagger. [39]

Performance of the same side kaon tagger in the 2011 dataset

The cut based same side kaon tagger has been calibrated and its performance has been measured using the full 2011 dataset with the signal decay $B_s^0 \to D_s^- \pi^+$ [39, 40]. The resulting tagging performance for the same side kaon tagger as well as for the combination with the opposite site taggers is given in Table 4. The calibration of the predicted mistag probability

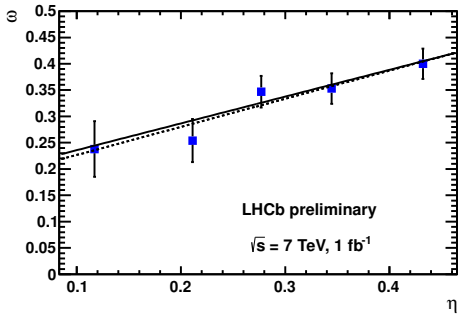


Figure 21: Measured mistag fraction ω as function of the predicted mistag probability η of the cut based same side kaon tagger determined from $B_s^0 \to D_s^- \pi^+$ decays in the 2011 dataset. The solid line is the result of the calibration using an unbinned fit of the B_s^0 oscillation. The dashed line is the result of the calibration determined from a fit of the data points, representing the average mistag fraction in bins of the predicted mistag probability. Figure taken from [39].

from the same side kaon decision is shown in Figure 21. The same side kaon tagger has an efficiency of $\varepsilon_{tag} = 15.8 \pm 0.3\%$ at an mistag fraction of $\omega = 34.1 \pm 1.5\%$, leading to an effective efficiency of $\varepsilon_{eff} = 1.6 \pm 0.3\%$. Combined with the OS taggers, the total LHCb flavour tagging has an efficiency of $\varepsilon_{tag} = 49.4 \pm 0.4\%$ at an mistag fraction of $\omega = 36.1 \pm 1.3\%$, leading to an effective efficiency of $\varepsilon_{eff} = 3.8 \pm 0.7\%$ in $B_s^0 \to D_s^- \pi^+$ events.

Individual taggers			
	ε_{tag}[%]	ω[%]	ε_{eff}[%]
SSK average	15.8±0.3	34.6±1.5	1.5±0.3
SSK per event	15.8±0.3	34.1±1.5	1.6±0.3
Combined SSK + OS tagger			
combination	49.4±0.4	36.1±1.3	3.8±0.7

Table 4: Measured performance of the same side kaon tagger and the combination of same side kaon and opposite side taggers in the 2011 dataset for $B_s^0 \to D_s^- \pi^+$ decays. The performance of the combination is measured using per-event predicted mistag probabilities. [39].

5 CALIBRATION OF THE SAME SIDE KAON TAGGER IN THE FIT OF THE B_s^0–$\bar{B}_s^0$ OSCILLATION

Unlike the opposite side tagging algorithms, the calibration of the same side kaon tagger requires a more sophisticated procedure. The reason for this lies in the physics principle that is exploited for this algorithm: it only works with B_s^0 mesons which have a neutral charge and oscillate (*cf.* Chapter 4.6). The calibration of the same side kaon tagger with B_s^0 thus requires that the oscillation is resolved which entails that the flavour of the mesons at decay time must be know, *i.e.* the B_s^0 has to decay flavour specific. The decay mode $B_s^0 \to D_s^- \pi^+$ satisfies this demands and has a high enough branching fraction to be used in this study.

This chapter describes the used dataset and the selection steps applied to extract those decays from data. The physics parameters are extracted in two steps from the selected B_s^0 candidates, using an unbinned maximum likelihood fit. In the first step, signal and background are separated by a fit of the B_s^0 mass distribution. In the second step, the parameters of the oscillation and the tagging calibration are determined in a fit of the decay time distribution of the B_s^0 candidates. This fitting procedure is an important tool for systematic studies of the same side kaon tagging algorithm and will be used extensively in the remaining chapters of this thesis, therefore details of these steps involved are discussed in the following sections.

The procedure for the fit of the B_s^0 oscillation as well as the selection used in this thesis are not the original work by the author [41, 42]. The author of this thesis and collaborators have adapted and used this tools in the measurement of the B_s^0–$\bar{B}_s^0$ oscillation frequency Δm_s and the development and calibration of the same side kaon tagger on various occasions [13, 39, 40, 42–46]. The work presented in this thesis uses the same selection and subsequent fit procedure that has been used in the measurement of the B_s^0–$\bar{B}_s^0$ oscillation with a dataset of $340\,\mathrm{pb}^{-1}$, *cf.* [43]. In contrast to this analysis, the most recent LHCb measurement [13] uses a more sophisticated selection based on multivariant techniques and additional B_s^0 decay modes, which makes a measurement of the decay time resolution more complicated. As the calibration of the same side kaon tagger relies on this input, it has been decided to use the simpler analysis procedure used in [43] for this studies.

5.1 USED DATASET AND SELECTION

The analysis presented in this thesis uses the full dataset taken by LHCb during the 2011 run of the LHC at a center of mass energy of $\sqrt{s} = 7\,\mathrm{TeV}$, which amounts to an integrated luminosity of $1\,\mathrm{fb}^{-1}$. The extraction of the signal candidates from the data involves several steps. In the first step, the events are processed in the three stages of the trigger system. Events that pass the trigger are further reduced by a stripping selection and a final

offline selection step to reach an optimal background reduction. The key principles of these selection steps are discussed in the following paragraphs.

5.1.1 The trigger selection

As described in Chapter 3.2.3, the LHCb trigger strategy consists of three stages. For this analysis the two stages of the software trigger are of particular interest. In the first stage, the selection of events containing B_s^0 candidates relies on a trigger line[1] which looks for single tracks with a good track quality from the track fit and a high transverse momentum as a key signature for B and D meson decays. A detailed description including the used selection cuts is given in [47].

In the second stage the event is required to pass the topological[2] or inclusive ϕ lines[3]. The topological lines inclusively select a wide range of different B and D meson decays with at least two charged daughters. Because most of the decays at this stage are only partially reconstructed, it does not apply strict mass cuts but relies on robust variables like the IP/σ_{IP} separation of the tracks from the long living B or D meson from the primary vertex. Cuts on the $\sum |p_T|$ of the tracks and their distance of closest approach (DOCA) further exploit the topology of the B meson decays. Depending on the specific topological line, a multivariant classifier, a so called boosted decision tree (BDT), *cf.* [48–50], is used to further suppress backgrounds.

Finally, a subset of the events used in this study are selected by the inclusive ϕ line which inclusively selects decays that contain a ϕ meson. In the case of the decay $B_s^0 \to D_s^- \pi^+$, the D_s^- can decay via a ϕ resonance. Due to the very narrow mass resonance of the ϕ this decay provides a very clean signature. Further details on the used HLT2 trigger lines are to be found in [51].

5.1.2 The stripping selection

In case an event has been triggered, the full event reconstruction is performed and all the reconstructed tracks are written to storage for later analysis. After this step, the amount of information is still too large for the individual user to perform his analysis. Therefore a set of common selections for the key physics channels are applied to the data to further reduce the amount of events. This process is called stripping and the selections are organized in so called stripping lines, which incorporate a loose set of cuts suitable for multiple analysis that use a specific decay channel. The analysis performed in the course of this thesis uses a stripping line designed for the common selection of B^0 and B_s^0 candidates which decay into $D^-\pi^+$ or $D_s^-\pi^+$ respectively with a subsequent decay of the D meson into three charged hadrons. The cuts of the stripping line are summarized in Table 5. The stripping line uses a boosted decision tree for the selection, which combines special features of B meson decays such as a high transverse momentum p_T, the primary vertex separation significance[4], which should be large as B mesons are long-lived and the sum of the χ^2/ndf of the vertex fit from

1 The name of the used trigger line in the HLT1 is Hlt1TrackAllLo.
2 The line names of the topological lines are Hlt2Topo2(3,4)BodyBBDT, Hlt2Topo2(3,4)Body.
3 The line name of the inclusive ϕ line is Hlt2IncPhi.
4 The primary vertex separation significance is defined as the distance between the two vertices divided by the uncertainty on this distance.

Cuts on B_s^0 candidate	
$m(B_s)$	$[4.75, 7.0]$ GeV/c^2
vtx χ^2	< 10
IP/σ_{IP}	< 5
cos(DIRA)	> 0.999
decay time	> 0.2 ps
BDT	> 0.05
sum of p_T of daughters	> 5 GeV/c
Cuts on D_s^- candidate	
$\lvert m(D_s) - m(D_s)_{PDG}\rvert$	< 100 MeV/c^2
vtx χ^2	< 10
primary vertex separation significance	> 36
cosDIRA	> 0
sum of p_T of daughters	> 1.8 GeV/c
For at least 1 Child:	
Track χ^2	< 3
p_T	> 500 MeV/c
p	> 5 GeV/c
Cuts on bachelor π	
Track χ^2	< 3
p_T	> 500 MeV/c
p	> 5 GeV/c
DLL($K - \pi$)	< 20
Trigger requirements	
HLT1 one track trigger	
HLT2 topological or inclusive ϕ trigger	

Table 5: Stripping selection of $B_s^0 \to D_s^- \pi^+$ candidates.

the B and D vertex, which ensures a good quality of the secondary vertices. Additionally it uses a cut on the cosine of the angle between the direction of the B meson momentum, reconstructed by the sum of the four-vectors of its daughters, and its direction of flight, given by the distance between the primary vertex and the B decay vertex (DIRA). This cut ensures that the B meson is correctly reconstructed and comes from the assigned PV. Cuts on the decay time, the IP/σ_{IP} and the χ^2 of the vertex distance with respect to the primary vertex provide a good separation of the B decay vertex from the primary vertex and suppress prompt background. The selection of the stripping line that is used in this analysis is equivalent to the one used in the most recent measurement of Δm_s published by LHCb [13].

5.1.3 The offline selection

For candidates that pass the stripping selection, a tighter selection is applied for a better background suppression. Several of the cuts already applied in the stripping selection are tightened, as, *e.g.*, the cosine of the direction angle, the IP/σ_{IP} and the particle identification requirement on the bachelor π. The bachelor π ist the pion from the $B_s^0 \to D_s^- \pi^+$ decay which is not coming from the subsequent D_s^+ decay. Additional cuts on the mass of the

D_s^+ meson and the IP/σ_{IP} of its daughter particles cut away background of randomly combined tracks and $D^\pm$ mesons which where also selected in the same stripping selection. The decay vertex of the reconstructed B_s^0 and D_s^+ are required to be clearly separated from the primary vertex.

The D_s^+ meson in the B_s^0 decay chain decays itself via different resonant decays. Those resonant decays put constraints on the phasespace of the D_s^+ daughter particles, which can be exploited to further suppress the background. The sample is split into three different categories representing $B_s^0 \to D_s^- \pi^+$ with the subsequent resonant decays of $D_s^- \to \phi\pi^-$, $D_s^- \to K^*K^-$ and events that fall in none of the two former categories, hereafter named non-resonant decays. The selection for these categories is performed sequentially to ensure that a specific candidate does only enter into one of the categories and special features of the resonances are exploited for a better background suppression.

For the $D_s^- \to \phi\pi^-$ sample an additional cut requires the invariant mass of the ϕ meson to be within 15 MeV/c^2 of its PDG value. As the ϕ has a very narrow resonance this cut provides a good separation of the signal from random combinations of tracks.

In case of decays $D_s^- \to K^*K^-$, a cut requires the invariant mass of the K^* meson to be within 50 MeV/c^2 of its PDG value. The K^* mesons features a much broader resonance, thus additional cuts on the helicity angle of the K and π from the K^* and on the particle identification of the non resonant K in the D_s^+ decay are necessary to get a clean signal. A cut on the separation of kaons and protons suppresses background from Λ_b^0 decays.

For the non-resonant sample, harder particle identification cuts on the daughters of the D_s^+ are required. An overview of the selection is given in Table 6 and the invariant mass distributions of the B_s^0 candidates passing the final selection step is shown in Figure 23 for the individual categories and in Figure 24 for the sum of all three categories. The fit of this mass distribution is explained in detail in the following chapter. The offline selection used in this thesis is equivalent to the selection that has been used for the measurement of Δm_s using a dataset of 340 pb^{-1} that is presented in [43].

5.2 FIT OF THE B_s^0 MASS DISTRIBUTION

For the fit of the invariant mass distribution of selected B_s^0 candidates, several contributions to the probability density function (PDF) are taken into account.

The signal part of the PDF is modeled using a Gaussian distribution

$$P_{sig}(m) = G(m_{B_s^0}, \sigma_m) \tag{34}$$

with the mass of the B_s^0 meson $m_{B_s^0}$ and the width σ_m as free parameters. Both mass and width are common parameters for the three categories in which the sample is split. This assumption is valid as a constraint on the D_s^+ mass is applied in the reconstruction of the B_s^0 mass used in the fit of the invariant mass distribution, thus phase space differences introduced by the resonant decays of the D_s^+ have no influence on the B_s^0 mass.

Beside the signal, several background components have to be taken into account. The background contributions can be divided roughly into two categories, physics induced background, that is caused by actual physics

Cuts on B_s^0 candidate	
p	> 2 GeV/c
IP/σ_{IP}	< 4
cos(DIRA)	> 0.9999
primary vertex separation significance	> 64
Cuts on D_s^- candidate	
p_T	> 2 GeV/c
primary vertex separation significance	> 100
$\lvert m(D_s^-) - m(D_s^-)_{PDG}\rvert$	< 30 MeV/c^2
Cuts on D_s^- daughters	
p_T	> 300 MeV/c
$\min(IP/\sigma_{IP})$	> 3
DLL$(K-\pi)$ for kaons	$>$ -10
DLL$(K-\pi)$ for pions	< 10
Cut on bachelor π^+	
DLL$(K-\pi)$ for pions	< 5
Cuts specific for $D_s^- \to \phi\pi^-$	
$\lvert m(\phi) - m(\phi)_{PDG}\rvert$	< 15 MeV/c^2
Cuts specific for $D_s^- \to K^*K^-$	
$\lvert m(K^*) - m(K^*)_{PDG}\rvert$	< 50 MeV/c^2
DLL$(K-\pi)$ for kaon with same charge as D_s^-	> 0
DLL$(K-p)$ for kaon with same charge as D_s^-	$>$ -10
Helicity angle $\lvert\cos\psi\rvert$	> 0.4
Cuts specific for non-resonant $D_s^- \to K^+K^-\pi^-$	
DLL$(K-\pi)$ for pions	< 0
DLL$(K-p)$ for kaons	> 5
DLL$(K-p)$ for kaon with same charge as D_s^-	$>$ -10

Table 6: Offline selection applied additional to the stripping selection of the $B_s^0 \to D_s^-\pi^+$ candidates.

processes similar to the signal decay and combinatorial background caused by random combinations of tracks.

The combinatorial background is the most relevant background contribution. It is empirically modeled by an exponential function in the PDF of the invariant mass,

$$P_{bkg_{comb}}(m) = e^{-m_\alpha^2 \times m} \tag{35}$$

with the parameter m_α^2. As the combinatorial background is very dependent on the specific category an independent m_α^2 and fraction $f_{bkg_{comb}}$, which is the fraction of combinatorial background relative to the physics induced background, is fitted for each of the three D_s^+ decay modes.

At invariant masses smaller than the PDG value of the B_s^0 mass, a large contribution from partially reconstructed B_s^0 decays is observed. Those are events that are similar to the signal decay mode but have one or more additional neutral particles in decay chain that are not reconstructed. Because of the missing momentum from the neutral particle the invariant mass for these B_s^0 candidates is shifted below the PDG value of the B_s^0 mass. Several physics processes with this decay topology are considered:

- $B_s^0 \to D_s^{*+}\pi^+$ where the γ or the π^0 from the D_s^{*+} decay is not reconstructed.
- $B_s^0 \to D_s^+\rho^+$ with $\rho^+ \to \pi^+\pi^0$, where the π^0 is not reconstructed.

The mass templates for these modes, shown in Figures 22a and 22b, in the following denoted as $\mathcal{P}_{N,D_s^*}(m)$ and $\mathcal{P}_{N,D_s\rho^+}(m)$, have been obtained using simulated events on the level of the MC generator, ie. no reconstruction has been applied. This is done to reach a sufficient amount of statistics, as the reconstruction efficiency of this decays in the full simulation is low due to the missing particles. After applying the kinematic cuts of the offline selection, the B_s^0 mass is obtained by adding the four-momenta of the charged particles. The resulting invariant mass is smeared with a Gaussian distribution with the same resolution as seen on data to account for detector effects. It is checked that the other cuts of the selection, namely particle identification cuts, have no influence of the shape of the distribution. The fractions of the partially reconstructed decays $f_{partial}$ and $f_{D_s^*}$ do not depend on the D_s^+ decay mode and are thus common among the three modes.

In addition to the aforementioned backgrounds there is an additional component that has to be taken into account in the fit of the invariant mass distribution, namely background caused by the misidentification of particles. In particular, the following cases are considered:

- $B^0 \to D^-\pi^+X$ where one of the π in the D decay is wrongly identified as K. Beside the decay $B^0 \to D^-\pi^+$ also partially reconstructed decays, such as $B^0 \to D^*\pi^+$ and $B^0 \to D^-\rho^+$ with neutral particles among their daughters are considered.
- $\Lambda_b^0 \to \Lambda_c^+\pi^-$ with $\Lambda_c^+ \to pK\pi$ where the p is reconstructed as K.

The mass templates for these modes, in the following denoted as $\mathcal{P}_{N,B_d}(m)$ and $\mathcal{P}_{N,\Lambda_b}(m)$, have also been obtained from the simulation on generator level, using the same procedure as discussed for the partially reconstructed decays. To account for the misidentified particle, the four-momenta of the misidentified particles have been recalculated with the misidentified mass

in the calculation of the B_s^0 invariant mass. The resulting mass templates are shown in Figures 22c and 22d. The fraction of Λ_b^0 events in the misidentified background is depending on the D_s^+ decay mode and is thus fitted independently for each mode.

Besides the signal decay mode, it is also possible for the B_s^0 to decay into the final state $B_s^0 \to D_s^- K^+$. Such events pass the selection steps if the bachelor K is misidentified as π. The resulting wrongly determined four-momentum of the K leads to a slight shift in mass compared to the nominal B_s^0 mass. Such decays show a similar decay time behaviour to the nominal $B_s^0 \to D_s^- \pi^+$ decay and are thus treated as signal. However, the $B_s^0 \to D_s^- K^+$ decay is not fully flavour specific but can decay in both $D_s^- K^+$ and $D_s^+ K^-$ which will be accounted for in the decay time part of the probability density function. In addition to the decay mode $B_s^0 \to D_s^- K^+$, partially reconstructed modes such as $B_s^0 \to D_s^{*+} K^+$ and $B_s^0 \to D_s^+ K^*$ have to be taken into account in the same manner as described above. The mass template for all this decays is shown in Figure 22e. The fraction of $B_s^0 \to D_s^- K^+$ events relative to the signal decay $B_s^0 \to D_s^- \pi^+$, $f_{D_s K}$, does not depend on the D_s^+ decay mode and is thus considered common among the three modes.

By considering all described contributions, the complete PDF describing the invariant mass of the B_s^0 candidates used in this thesis can be constructed as

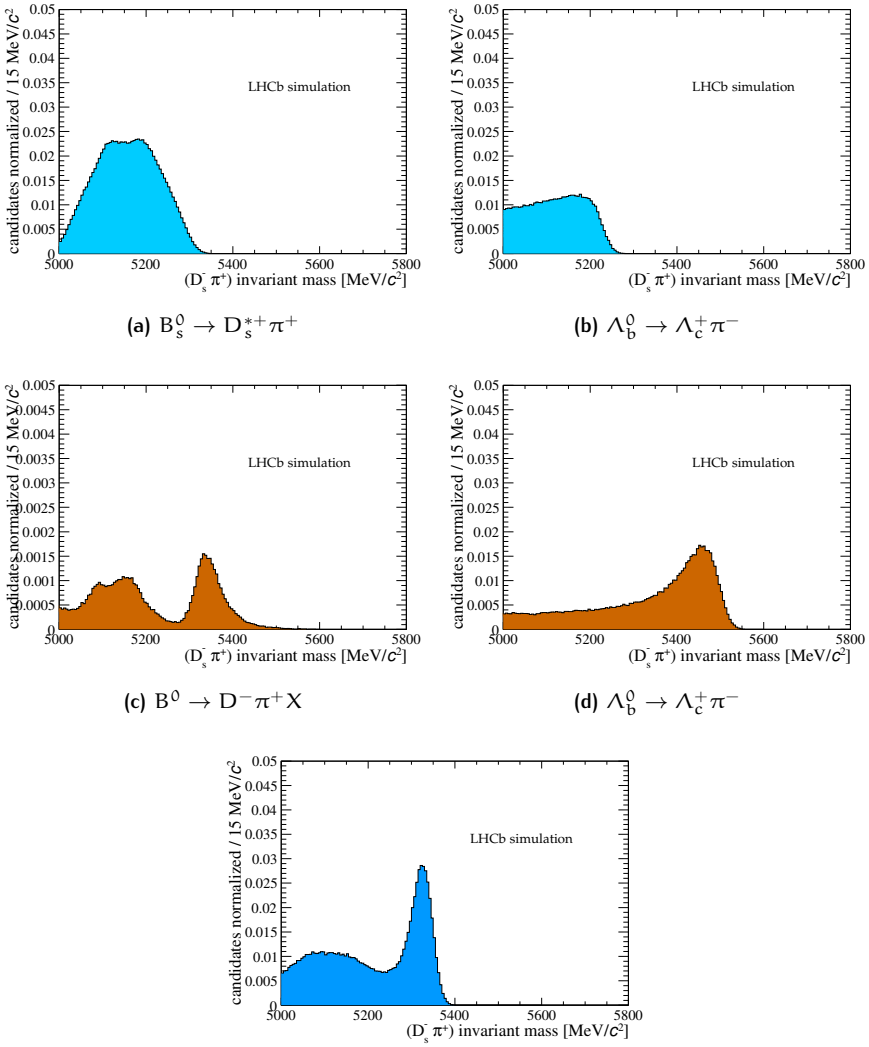
$$\begin{aligned}\mathcal{P}(m) \quad = \quad & f_{sig} \times ((1 - f_{partial}) \times (f_{D_s K} \times \mathcal{P}_{N,DsK} + (1 - f_{D_s K}) \times \mathcal{P}_{sig}(m)) \\ & + f_{partial} \times (f_{D_s^*} \times \mathcal{P}_{N,D_s^*}(m) + (1 - f_{D_s^*}) \times \mathcal{P}_{N,D_s\rho^+}(m))) \\ & + (1 - f_{sig}) \times (f_{bkg_{comb}} \times \mathcal{P}_{bkg_{comb}} + (1 - f_{bkg_{comb}}) \\ & \times (f_{\Lambda_b} \times \mathcal{P}_{N,\Lambda_b}(m) + (1 - f_{\Lambda_b}) \times \mathcal{P}_{N,B_d}(m))) \end{aligned} \tag{36}$$

The mass fit is performed in the range of $[5.00, 5.85]\,\mathrm{GeV}/c^2$. The fit results are listed in Table 7 and projections of the three individual D_s^+ decay modes and the sum of all modes are shown in Figures 23 and 24. The data sample contains 26.371 ± 223 signal B_s^0 candidates in total. The number of signal candidates in the individual modes as well as the corresponding signal fractions are listed in Table 8.

For the fits of the decay time distribution the signal and background fractions as well as the B_s^0 mass, the width of the signal mass peak and the slope of the combinatorial background are taken as input and fixed to speed up the fitting procedure.

The decay time behaviour of the partially reconstructed B_s^0 decays is difficult to describe. Due to the missing momentum from the neutral particles in the decay, the B_s^0 momentum is underestimated and likewise the decay time, as it is defined as $t = l \cdot m/p$ (with l the decay length), will be to large. The effect is comparable to semileptonic B_s^0 decays where the neutrino is missing. Furthermore, as can be seen from Figure 23, the mass fit does not describe the observed distribution perfectly in the region of the partially reconstructed decays.

The resulting improvement by adding the partially reconstructed modes to the fit of the decay time distribution was studied in [41] and was found to be negligible. Thus for the fit of the B_s^0 oscillation and tagging parameters the mass range is restricted to $[m_{B_s^0} - 3\sigma, 5.85]\,\mathrm{GeV}/c^2$.

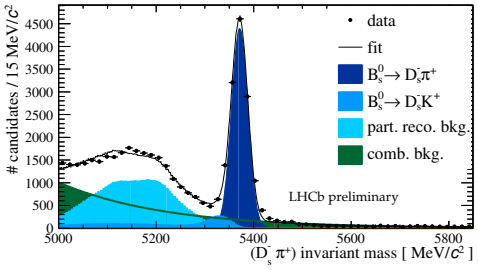


(a) $B_s^0 \to D_s^{*+}\pi^+$ **(b)** $\Lambda_b^0 \to \Lambda_c^+\pi^-$

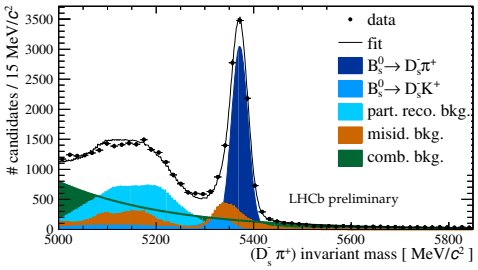
(c) $B^0 \to D^-\pi^+X$ **(d)** $\Lambda_b^0 \to \Lambda_c^+\pi^-$

(e) $B_s^0 \to D_s^-K^+$

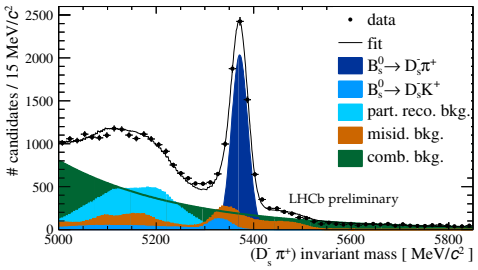
Figure 22: Mass templates obtained from generator level simulation for the different background contributions present in the B_s^0 invariant mass distribution.



(a) $B_s \to D_s(\phi\pi)\pi$



(b) $B_s \to D_s(K^*K)\pi$



(c) $B_s \to D_s\pi$ non-resonant

Figure 23: Projections of the fitted mass distributions in (a) the $B_s^0 \to D_s^-(\phi\pi^-)\pi^+$ decay, (b) the $B_s^0 \to D_s^-(K^*K^-)\pi^+$ decay and (c) the $B_s^0 \to D_s^-(K^-K^+\pi^-)\pi^+$ non resonant decay in data.[39]

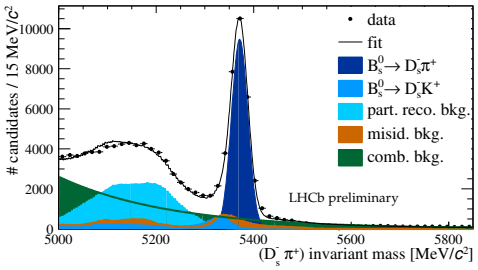


Figure 24: Projection of the sum of the 3 fitted mass distributions in $B_s^0 \to D_s^-\pi^+$ decays in the mass window [5.0,5.85] GeV/c^2.[39]

Common parameters among the 3 B_s decay modes	
m_{B_s} [MeV/c^2]	5371.27±0.15
σ_m [MeV/c^2]	16.67±0.13
$f_{partial}$	0.498±0.005
f_{Ds_K}	0.180±0.012
$f_{D_s^*}$	0.744±0.027
Parameters of the $B_s \to D_s(\phi\pi)\pi$ decay	
f_{sig}	0.676±0.008
m_α^2 [MeV^2/c^4]	0.00462±0.00007
$f_{bkg_{comb}}$	1.0 ±0.0
Parameters of the $B_s \to D_s(K^*K)\pi$ decay	
f_{sig}	0.548±0.012
$f_{bkg_{comb}}$	0.675±0.017
f_{Λ_b}	0.000±0.003
m_α^2 [MeV^2/c^4]	0.00453±0.00007
Parameters of the $B_s \to D_s(KK\pi)\pi$ non resonant decay	
f_{sig}	0.446±0.010
$f_{bkg_{comb}}$	0.764±0.016
f_{Λ_b}	0.244±0.031
m_α^2 [MeV^2/c^4]	0.00397±0.00006

Table 7: Results of the combined fit to the three $B_s^0 \to D_s^-\pi^+$ candidate mass distributions in the wide mass range in the data.

decay mode	**# signal candidates**	f_{sig}
$B_s^0 \to D_s^-(\phi\pi^-)\pi^+$	12223±134	0.81
$B_s^0 \to D_s^-(K^*K^-)\pi^+$	8402±137	0.67
$B_s^0 \to D_s^-\pi^+$ non-resonant	5746±114	0.58

Table 8: Number of B_s^0 signal candidates in data and signal fraction fitted in the mass range of $[m_{B_s} - 3\sigma, 5.85]$ GeV/c^2.

5.3 DECAY TIME ACCEPTANCE AND RESOLUTION

Naively, one would expect that the probability by which a B_s^0 candidate is reconstructed does not depend on its decay time. In reality however, direct cuts on the decay time as well as indirect cuts, such as impact parameter and pointing angle cuts bias the decay time distribution. Several of these variables are used in the trigger, in the stripping and in the offline selection. Hence, an acceptance function ϵ_t has to be introduced in the fit of the decay time distribution to account for the different reconstruction efficiencies depending on the decay time of the B_s^0 candidate.

To derive this acceptance function, simulated $B_s^0 \to D_s^- \pi^+$ events are used assuming the same trigger, stripping and offline selection cuts and the same reconstruction as used in the processing of the real data. The reconstructed decay time distribution of the B_s^0 candidates in this simulated events is divided by the expected theory distribution

$$\mathcal{P}_{theo}(t) \otimes G(0, S_{\sigma_t}) = \left(e^{-\Gamma_s t} \times \cosh(\frac{\Delta\Gamma_s}{2} \times t) \right) \otimes G(0, S_{\sigma_t}) \,. \tag{37}$$

The physics part of the decay time distribution is described by an exponential decay which is multiplied by the term $\cosh(\frac{\Delta\Gamma_s}{2} \times t)$ to account for the decay width difference of the mass eigenstates of the B_s^0. The current PDG value of $\Delta\Gamma_s/\Gamma_s = 0.150 \pm 0.020$ is used throughout this thesis. The theory distribution is convoluted with a Gaussian distribution $G(0, S_{\sigma_t})$ to account for the detector resolution. The thus observed acceptance distribution as shown in Figure 25 is parametrized by the phenomenological function

$$\epsilon_t(t) = \left(1 - e^{-\frac{(t-offset)}{\alpha_{acc}}} \right) \times (1 + \beta_{acc} t) \times \Theta(t - 0.2). \tag{38}$$

In this parametrization the Heaviside function Θ represents the cut off at $t = 0.2\,ps$ which is introduced by the stripping line, the exponential part describes the turn on and the term $(1 + \beta_{acc} t)$ is used to account for the upper decay time acceptance. The fit values for this parametrization, which are derived from the sample of simulated events, are summarized in Table 9. The parameters are fixed to the values obtained from the simulation in the fit of the decay time distribution in the data.

parameter	result
α_{acc}	0.6759 ± 0.0043 ps
offset	0.2001 ± 0.0004 ps
β_{acc}	-0.0238 ± 0.0008 ps

Table 9: Results of the decay time acceptance parameters obtained from a fit to the simulated $B_s^0 \to D_s^- \pi^+$ events.

The decay time resolution and the mistag fraction from the tagging algorithms are directly related to each other in the decay time part of the PDF used in the fit of the B_s^0 oscillation. A good knowledge of the decay time resolution is therefore crucial for the correct determination, of the tagging parameters. For the description of the decay time resolution the predicted per-event decay time uncertainty as obtained by the decay time fitter is used.

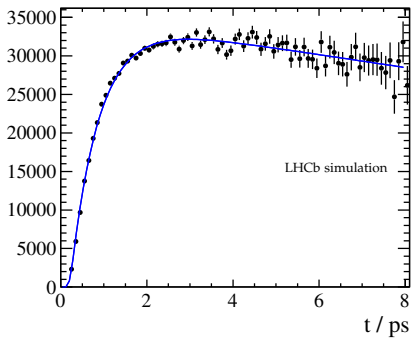


Figure 25: The acceptance function as function of the decay time for simulated $B_s^0 \to D_s^- \pi^+$ events.

The estimate of the decay time uncertainty σ_t however has to be calibrated to ensure a correct error estimate. As most of the resolution effects are caused by uncertainties of the detector alignment, which are not taken into account in the simulation, a data driven calibration is necessary to obtain a correct calibration of the decay time uncertainty. The following study has been performed by the author for the measurement of Δm_s with $340\,pb^{-1}$ of data, [43], and is redone for this thesis with the full $1\,fb^{-1}$ dataset.

To measure the decay time resolution and check if the uncertainty of the decay time from the fitter is correct, it is necessary to know the true decay time of the particles. Using data, this is only possible for prompt particles which have a decay time of zero. For the calibration procedure, fake $B_s^0 \to D_s^- \pi^+$ candidates are created from the combination of prompt D^- candidates and pions[5]. Hereby the identical selection cuts as for the $B_s^0 \to D_s^- \pi^+$ candidates described in the last section are applied to the fake B_s^0. There are, however, several selection cuts that indirectly bias the B_s^0 decay time distribution, such as the impact parameter cuts on D_s^- and on the pion, the vertex distance separation cut, the B_s^0 pointing angle and the explicit cut on the B_s^0 decay time. Those cuts have to be removed from the stripping and the selection, as their inclusion would bias the determination of the decay time resolution. A dedicated stripping selection was specifically set up for that purpose.

The decay time of the fake B_s^0 candidates is expected to be zero, as they are made up from prompt D^- mesons and prompt tracks and any deviation of this expectation is caused by resolution effects. Accordingly, the decay time resolution can be directly inferred from the width of the decay time distribution and the pull, which is calculated as the reconstructed decay time divided by the decay time uncertainty obtained from the decay time fit, can be used to measure the scaling factor needed to calibrate the per-event error estimate in the data.

The output of the stripping line also contains a combinatoric background of fake B_s^0 candidates that are made up of random combination of tracks. This background is removed by a sideband subtraction in the D^- mass. This way, the sample of fake B_s^0 candidates contains only real D^- mesons, which minimizes the differences with respect to real $B_s^0 \to D_s^- \pi^+$ decays.

5 Originally it was intended to use combinations of prompt D_s^- candidates and pions for this study but their statistics was to low for this purpose thus the kinematic and topological similar D^- was chosen instead which features a larger branching fraction.

In the distribution of the log(IP) of the D^- mesons, shown in Figure 27, a bump is visible at larger values of log(IP). By comparing this distribution to the distribution of simulated prompt and non prompt D^- mesons in this variable, Figure 26, it is visible that this bump originates from a contribution of non prompt D^- mesons to the fake B_s^0 sample. This contribution needs to be removed prior to the determination of the scale factor, as the decay time of fake B_s^0 mesons that are made up from non prompt D^- mesons can not be assumed to be zero.

The contribution of secondary D^- mesons is treated by a fit of the prompt and secondary component in the log(IP) distribution. The obtained PDFs are used to unfold the contribution of the both components in the pull of the decay time uncertainty by using the *sWeight* technique [52]. The so reweighted decay time pull is used to determine the scaling factor, which is simply the width of this distribution.

Based on the shapes of the log(IP) distributions of the prompt and secondary component from the simulation, shown in Figure 26, a Bukin PDF[6] is found as the best empirical description. Using this PDF in the fit of the log(IP) distribution of the fake B_s^0 sample on data, the weights to unfold the pull of the decay time uncertainty are extracted in two ways.

The first method lets all the parameters of the sum of the two Buckins PDFs floating in the fit. The fit projections are shown in Figure 27. Using this method results in a scaling factor of $S_{\sigma_t} = 1.382 \pm 0.008$. It can be seen from Figure 28 that the pull distributions for the fake B_s^0 mesons are fitted by a single Gaussian distribution which is accordingly chosen as the default resolution model. The pull distribution for simulated B_s^0 mesons, see Figure 31, is better fitted by a double Gaussian distribution where the fraction of the broader Gaussian is about 11% and its width about 60% larger than the width of the narrower Gaussian. Using this alternative model will be one of the systematic checks for the determination of the tagging calibration parameters and the B_s^0 oscillation frequency.

As a crosscheck, the second method uses the shapes of the log(IP) distribution as derived from the simulation directly in the weighting procedure in the fake B_s^0 sample. The only floating parameter in this case is the relative fraction of prompt and secondary component. It is expected that the PDFs for the prompt and secondary component derived from the simulation do not fit the data perfectly. Possible causes of differences include the different IP resolution on data and MC, differences in the phase space of the candidates in question and differences in the composition of the secondary component. Figure 27 shows the projection of this fit and the so obtained scaling factor $S_{\sigma_t} = 1.362 \pm 0.007$, Figure 28, is in agreement with the scaling factor obtained by the full fit on the log(IP) distribution. It can be concluded that the determination of the scaling factor using this method is reasonable

6 The Bukin function is parametrized as

$$\mathcal{P}(x; x_p, \sigma_p, \xi, \rho) = A_p \exp\left[\frac{\xi\sqrt{\xi^2+1}(x-x_1)\sqrt{2\ln 2}}{\sigma_p\left(\sqrt{\xi^2+1}-\xi\right)^2 \ln\left(\sqrt{\xi^2+1}+\xi\right)} + \rho\left(\frac{x-x_i}{x_p-x_i}\right)^2 - \ln 2\right], \tag{39}$$

where $\rho = \rho 1$ and $x_i = x_1$ for $x < x_1$ and $\rho = \rho 2$ and $x_i = x_2$ for $x \geqslant x_2$ with

$$x_{1,2} = x_p + \sigma_p\sqrt{2\ln 2}\left(\frac{\xi}{\sqrt{\xi+1}} \mp 1\right) \tag{40}$$

prompt component		
	simulation	data
x_p	-3.792±0.08	-3.732±0.002
σ_p	0.574±0.004	0.574±0.003
ξ	-0.164±0.009	-0.111±0.002
ρ_1	-0.11±0.02	-0.153±0.007
ρ_2	-0.04±0.02	-0.015±0.001
secondary component		
	simulation	data
x_p	-2.38±0.04	-2.297±0.033
σ_p	1.14±0.04	1.164±0.013
ξ	-0.0025±0.0003	-0.102±0.008
ρ_1	-0.5±0.02	-0.264±0.012
ρ_2	-1.5±0.5	-0.783±0.016

Table 10: Shape parameters of the Bukin PDF from the fit to fake B_s^0 candidates formed of prompt and secondary D^- candidates in the simulation and in the data. The parameter A_p from the PDF represents the normalisation and is omitted here.

robust against changes in the PDF used for the unfolding. The results of the fits to the log(IP) distribution in the data and the simulation is summarized in Table 10.

The decay time resolution can in principle depend on the properties of the B_s^0 decay, such as the B_s^0 momentum or transverse momentum. The sample of fake B_s^0 candidates has to be either reweighted to match the distributions of real B_s^0 or the dependence of the measured scale factor on these quantities has to be accounted for as a systematic uncertainty.

Figure 29 shows that indeed the distributions of the fake B_s^0 candidates and the real B_s^0 candidates are different. A reweighting would be necessary in multiple variables, which is not feasible. To account for possible dependencies, the scaling factor is determined in bins of these quantities instead, which is displayed for a few variables in Figure 30. In these plots, the width of the bins in the x-axis are set to have the same number of entries. In general, the scaling factor obtained as a function of the properties of the B_s^0 lies between 1.28 and 1.48 which is used as the range for evaluating the systematic uncertainties.

5.4 FIT OF THE B_s^0 DECAY TIME DISTRIBUTION

For the fit of the decay time distribution, the mass parts of the PDF and the decay time parts are combined and only the background contributions present in the restricted mass range of $[m_{B_s^0} - 3\sigma, 5.85]\,\mathrm{GeV/c^2}$ are taken into account. Those are the signal decay as well as the decay $B_s^0 \to D_s^- K^+$, the background coming from misidentified B^0 and Λ_b^0 as well as background coming from random combinations of tracks. Each contribution has to be accounted for by its own decay time PDF. The combined PDF depending on

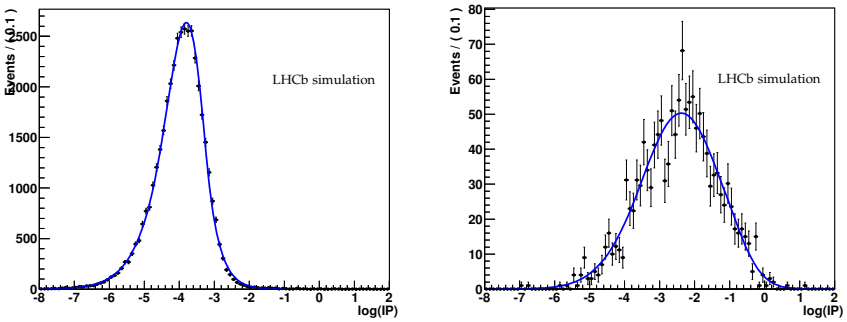


Figure 26: The log(IP) distribution of prompt MC D^- candidates (left) and secondary MC D^- candidates (right) used to reconstruct fake B_s^0 candidates. The distributions are fitted with a Buckin PDF.

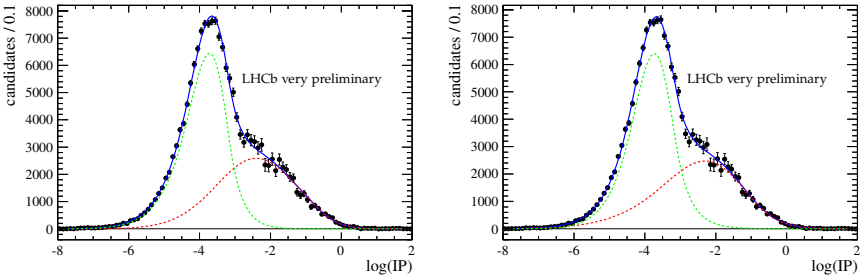


Figure 27: The log(IP) distribution of prompt D^- candidates on data fitted with shapes taken from MC (left) and floating (right).

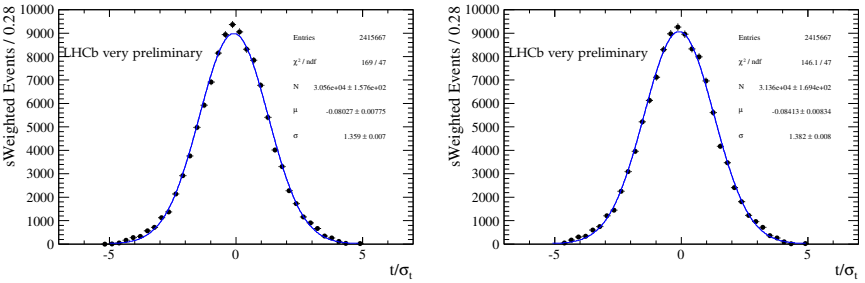


Figure 28: The pull of the decay time t sideband subtracted in the D^- mass and weighted from the log(IP) distribution of the D^- with shapes taken from MC (left) and floating(right).

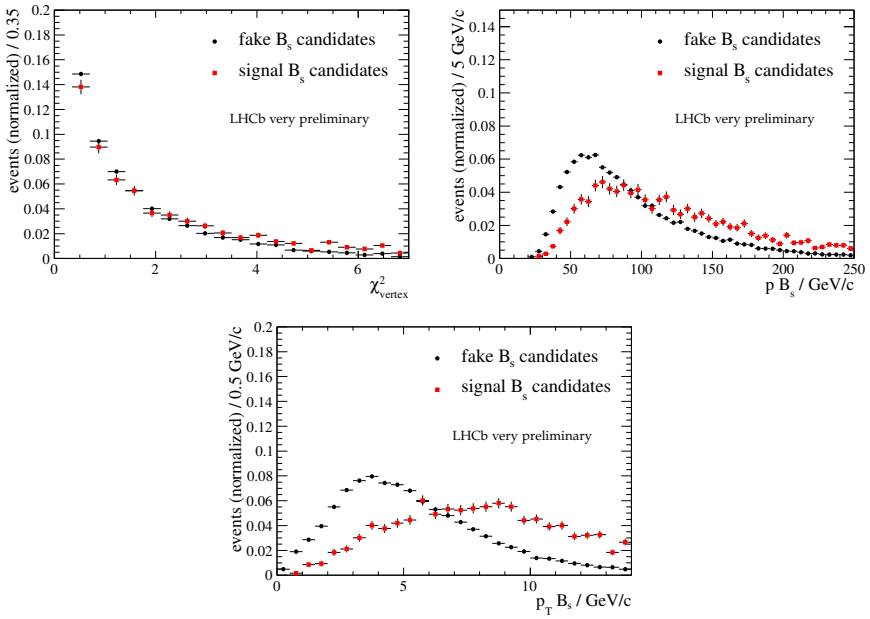


Figure 29: Distributions of fake prompt B_s^0 candidates (black circles) and sideband subtracted $B_s^0 \to D^-\pi^+$ events from 2011 data (red squares) for the B_s^0 vertex χ^2, B_s^0 momentum and transverse momentum.

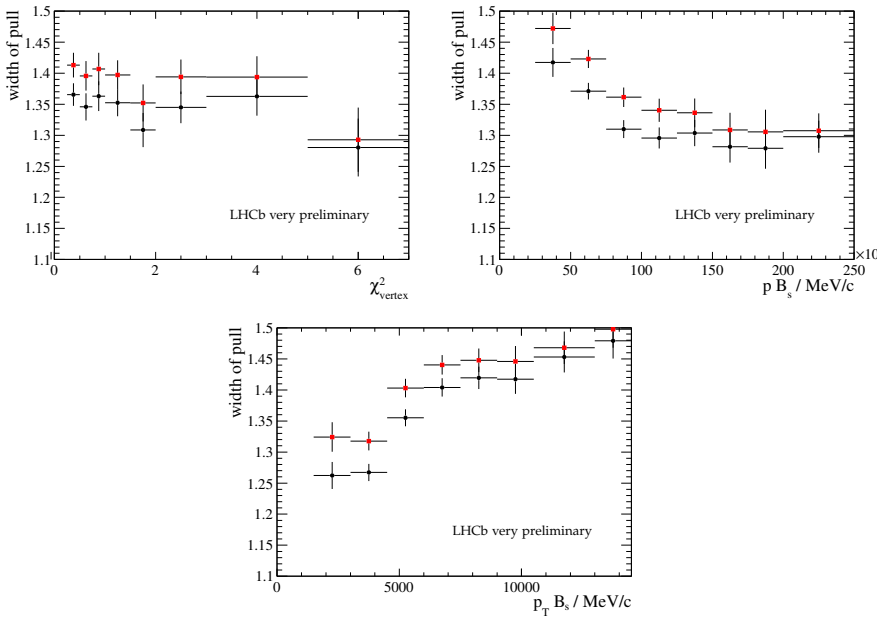


Figure 30: The width of the pull of the decay time t obtained from the prompt candidates plotted as function of: B_s^0 vertex χ^2 (top left), momentum of B_s^0 (top right) and transverse momentum of B_s^0 (bottom). The shapes for the fit in the log(IP) distribution have been taken from data (black triangles) and MC (red circles).

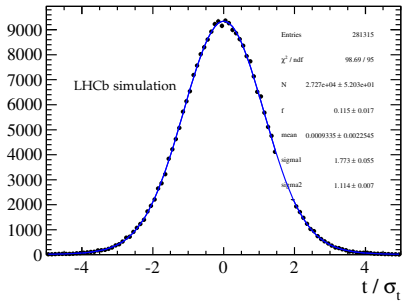


Figure 31: The pull of the decay time t for simulated B_s^0 mesons.

the mass m, the decay time t and the per event decay time resolution σ_t is then given by

$$
\begin{aligned}
\mathcal{P}(m,t,\sigma_t) \;=\; & f_{sig} \times \mathcal{P}_{sig}(t,\sigma_t) \times \mathcal{P}_{N,sig}(\sigma_t) \\
& \times \Big[f_{D_sK} \times \mathcal{P}_{N,D_sK} + (1 - f_{D_sK}) \times \mathcal{P}_{sig}(m) \Big] \\
& + (1 - f_{sig}) \\
& \times \Big[\big(f_{bkg_{comb}} \times \mathcal{P}_{bkg_{comb}}(m) \\
& \quad \times \mathcal{P}_{bkg_{comb}}(t) \times \mathcal{P}_{N,bkg_{comb}}(\sigma_t) \big) \\
& \quad + (1 - f_{bkg_{comb}}) \\
& \quad \times \big(f_{\Lambda_b} \times \mathcal{P}_{\Lambda_b}(m) \times \mathcal{P}_{\Lambda_b}(t,\sigma_t) \times \mathcal{P}_{N,sig}(\sigma_t) \\
& \quad + (1 - f_{\Lambda_b}) \\
& \quad \times \mathcal{P}_{B_d^0}(m) \times \mathcal{P}_{B_d^0}(t,\sigma_t) \times \mathcal{P}_{N,sig}(\sigma_t) \big) \Big],
\end{aligned}
\tag{41}
$$

where f_i are the fractions and $\mathcal{P}_i(m)$ the mass parts of the PDF for the signal and various background sources as discussed in Section 5.2, $\mathcal{P}_i(t,\sigma_t)$ describes the decay time parts of the PDF.

As a per event estimate for the decay time uncertainty is used, normalization factors are needed in the PDF [53], which are represented by the terms $\mathcal{P}_{N,bkg_{comb}}(\sigma_t)$ and $\mathcal{P}_{N,sig}(\sigma_t)$. They are the normalized distributions of the decay time error for the signal and background components and are derived directly from the data. The normalization of the combinatorial background part, $\mathcal{P}_{N,bkg_{comb}}(\sigma_t)$, is derived from the background in the upper mass sideband and $\mathcal{P}_{N,sig}(\sigma_t)$ is derived from background subtracted signal events. The B^0 and Λ_b^0 backgrounds can not be separated, thus it is assumed that they have the same distribution of the decay time error σ_t as the B_s^0 signal decays.

The various decay time related parts $\mathcal{P}_i(t,\sigma_t)$ of the PDF are described in the following paragraphs, [44]. Projections of the decay time fit in the three individual D_s^+ decay modes and the sum of all modes are shown in Figures 32 and 33.

5.4.1 Signal part of the decay time PDF

The signal part of the PDF that describes the decay time behaviour of B_s^0 mesons is given by

$$\mathcal{P}_{sig}(t, \sigma_t) \propto [\mathcal{P}_{theo}(t) \otimes G(0, S_{\sigma_t} \times \sigma_t)] \times \varepsilon_t(t). \tag{42}$$

The theoretical decay time distribution $\mathcal{P}_{theo}(t)$, as it is defined in Equation 37, is folded with a Gaussian distribution with mean 0 and width $S_{\sigma_t} \times \sigma_t$ to account for the decay time resolution effects. S_{σ_t} is the scale factor for the decay time error as discussed in Section 5.3. The whole term is multiplied with the acceptance function $\varepsilon_t(t)$ to account for the time dependent reconstruction efficiency of the B_s^0 mesons.

5.4.2 B^0 and Λ_b^0 part of the decay time PDF

For the decay time PDF of the B^0 and Λ_b^0 backgrounds a similar PDF as for the signal is used, assuming a simple exponential decay model and the same resolution model and scale factor S_{σ_t} as well as the same acceptance model. The resulting PDF is given by

$$\mathcal{P}_{B^0/\Lambda_b^0}(t, \sigma_t) \propto \left(e^{-\Gamma_{B^0/\Lambda_b^0} t} \otimes G(0, S_{\sigma_t} \times \sigma_t) \right) \times \varepsilon_t(t), \tag{43}$$

where Γ_{B^0/Λ_b^0} is the PDG value for the decay width of B^0 and Λ_b^0 respectively.

5.4.3 Combinatorial Background part of the decay time PDF

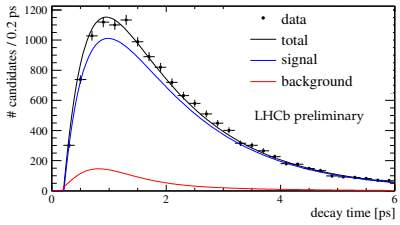
The model for the decay time PDF of the combinatorial background is derived from data using the decay time distribution of the events in the upper mass sideband $[5.5, 5.85]\,\mathrm{GeV}/c^2$, where only combinatorial background is present. The phenomenological PDF used to parametrize the observed distribution is given by a quadratic function multiplied by the sum of two exponentials,

$$\mathcal{P}_{bkg_{comb}}(t) \propto (t - a)^2 \times (f \times e^{-\alpha t} + (1 - f) \times e^{-\beta t}). \tag{44}$$

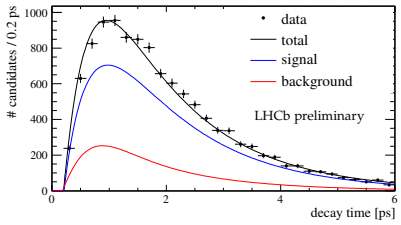
It is assumed to be the same for all three D_s^+ decay modes. A comparison of the decay time distribution of the combinatorial background in the upper mass sideband for the three D_s^+ decay modes, shown in Figure 34, supports this assumption.

5.5 EXTRACTION OF THE OSCILLATION FREQUENCY Δm_s AND OF THE SAME SIDE KAON TAGGER CALIBRATION PARAMETERS

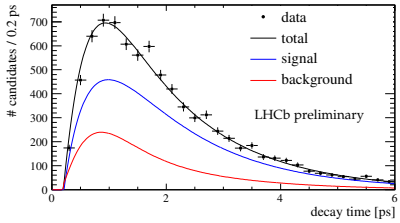
The PDF for the extraction of the oscillation frequency and the measurement of the calibration parameters for the tagging algorithms is introduced here



(a) $B_s \to D_s(\phi\pi)\pi$



(b) $B_s \to D_s(K^*K)\pi$



(c) $B_s \to D_s\pi$ non-resonant

Figure 32: Projection of the fitted decay time distribution in (a) the $B_s^0 \to D_s^-(\phi\pi^-)\pi^+$ decay, (b) the $B_s^0 \to D_s^-(K^*K^-)\pi^+$ decay and (c) the $B_s^0 \to D_s^-(K^-K^+\pi^-)\pi^+$ non resonant decay in data. [39]

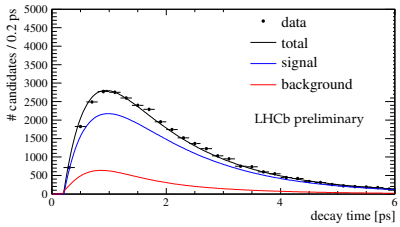


Figure 33: Projection of the sum of the 3 fitted decay time distributions in $B_s^0 \to D_s^-\pi^+$. [39]

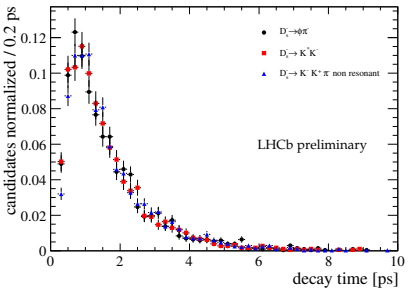


Figure 34: Decay time distributions of the combinatorial background in the upper mass sideband for the three different D_s^+ decay modes.

as part of the description of the fit procedure, *cf.* [44]. Detailed results on the calibration of the tagging algorithms are discussed in Chapter 8 and the measurement of the oscillation frequency is described in Chapter 10.

To extract the oscillation frequency Δm_s and the tagging calibration parameters, the combined mass and decay time PDF as it is presented in the previous section has to be extended to include the oscillation behaviour of the B_s^0 as well as the parts relevant to the flavour tagging. As discussed in Chapter 4, the oscillation of the B_s^0 mesons is resolved by the flavour tagging with two different sets of algorithms, the same side kaon tagging algorithm and the opposite side algorithms.

For every B_s^0 candidate the tagging algorithms give a decision on the initial flavour, which is compared to its flavour at decay time to sort the candidate into one of the following categories: B_s^0 candidates with the same flavour at production and decay (not oscillated, $q_{tag} = 1$), with different flavour at production and decay (oscillated, $q_{tag} = -1$) and those candidates where the tagging algorithm does not give a decision ($q_{tag} = 0$). As the changes in the signal part of the PDF do not depend on the specific tagging algorithm, *i.e.* opposite side or same side kaon, the tagging dependent variables are simply denoted by the index tag which can stand for either of the tagging algorithms. The PDF of the signal component is unchanged for untagged events with $q_{tag} = 0$

$$\mathcal{P}_{sig,osc}(t, \sigma_t, \eta_{tag}) \propto \mathcal{P}_{sig}(t, \sigma_t), \tag{45}$$

and for tagged events with $q_{tag} = \pm 1$, it is changes as

$$\begin{aligned}\mathcal{P}_{sig,osc}(t, \sigma_t, \eta_{tag}) \quad \propto \quad & \Big[(\mathcal{P}_{theo}(t) + q_{tag} \times (1 - 2\eta_{tag}) \times \cos(\Delta m_s \times t)) \\ & \otimes G(0, S_{\sigma_t} \times \sigma_t)\Big] \times \varepsilon_t(t),\end{aligned} \tag{46}$$

where η_{tag} is the predicted mistag of the specific tagging algorithm and Δm_s is the B_s^0 oscillation frequency. The correlation between the decay time resolution, represented by the term $G(0, S_{\sigma_t} \times \sigma_t)$ and the tagging performance, represented by $q_{tag} \times (1 - 2\eta_{tag})$ is clearly visible in the PDF.

The PDF of the misidentified $B_s^0 \to D_s^- K^+$ candidates is more evolved as they do not decay in a pure flavour specific final state. To account for this effect the time dependent CP asymmetry in this mode has to be taken into account. Details can be found in [8]. In case of the $B_s^0 \to D_s^- K^+$ decays, the theoretical expected decay time PDF changes to

$$\mathcal{P}_{theo,D_s^+K}(t) = e^{-\Gamma_s t} \times \left(\cosh(\frac{\Delta\Gamma_s}{2} \times t) - D_{eff} \sinh(\frac{\Delta\Gamma_s}{2} \times t)\right) \tag{47}$$

and the whole decay time part depending on the tagging decision is modified to

$$\mathcal{P}_{D_s^+K,osc}(t, \sigma_t, \eta_{tag}) \propto \left(\mathcal{P}_{theo,D_s^+K}(t) \otimes G(0, S_{\sigma_t} \times \sigma_t)\right) \times \varepsilon_t(t) \tag{48}$$

for untagged events and

$$\begin{aligned}\mathcal{P}_{D_s^+K,osc}(t, \sigma_t, \eta_{tag}) \quad \propto \quad & [(\mathcal{P}_{theo,D_s^+K}(t) + q_{tag} \times (1 - 2\eta_{tag}) \\ & \times (C \times \cos(\Delta m_s \times t) - S_{eff} \times \sin(\Delta m_s \times t))) \\ & \otimes G(0, S_{\sigma_t} \times \sigma_t)] \times \varepsilon_t(t)\end{aligned} \tag{49}$$

for tagged events, where C, D_{eff} and S_{eff} are the parameters describing the CP violation. They are taken as external input to the analysis and their exact definition is given in Appendix B.

As all of the B_s^0 candidates fall in one of the aforementioned three categories, depending on the decision of the tagging algorithm, the sample is split in three parts of which each contributes differently to the whole PDF. To have the overall PDF still normalized it is necessary to include an additional relative normalization factor which modifies the PDF according to

$$\mathcal{P}_{sig,osc}(t,\sigma_t,\eta) \quad \propto \quad \begin{cases} \mathcal{P}_{sig,osc} \times (1-\epsilon_{sig,tag}) & \text{if } q_{tag}=0 \\ \mathcal{P}_{sig,osc} \times 0.5\epsilon_{sig,tag} & \text{if } q_{tag}=\pm 1 \end{cases} \tag{50}$$

This normalization factor, $\epsilon_{sig,tag}$, represents the tagging efficiency.

To fit for the calibration of the predicted mistag η of a tagging algorithm, the parameter can be substituted in the PDF with the calibration function

$$\omega = p_0 + p_1 \times (\eta - \langle\eta\rangle) \tag{51}$$

as discussed in Chapter 4.3. The resulting mistag probability ω is limited in the fit by a smooth transition function to the physical range of $[0, 0.5]$. The chosen function is explained in more detail in Appendix A. If an average mistag is to be determined, the predicted mistag η is not used in the PDF and substituted by ω_{av}.

For the signal part, the following parameters are floating in the fit to the decay time distribution: The B_s^0 oscillation frequency Δm_s, the decay width Γ_s, the efficiency of the tagging algorithm $\epsilon_{sig,tag}$ limited to the physical range $[0, 1]$ and the calibration parameters p_0 and p_1. The tagging dependent parameters are different for the different tagging scenarios, *i.e.* if only the same side kaon tagging algorithm, the opposite side tagging algorithm or the combination of both is considered.

In addition to the signal part, also the background contributions to the PDF are modified if the tagging algorithms are used. Different cases for the background have to be distinguished as not all background components show the same behaviour in the presence of tagging. The opposite side tagging algorithms are sensitive to all hadrons that contain a b quark, thus they are able to resolve the slow oscillation ob the B^0 mesons, which has to be accounted for in the PDF. For the B^0 background in events tagged by the opposite side tagging algorithms, the PDF is accordingly modified as

$$\mathcal{P}_{B^0,osc,OST}(t,\sigma_t) \propto \mathcal{P}_{B^0}(t,\sigma_t) \times (1-\epsilon_{sig,OST}) \tag{52}$$

for untagged events and

$$\begin{aligned} \mathcal{P}_{B^0,osc,OST}(t,\sigma_t) \quad \propto \quad & [e^{-\Gamma_{B^0} t}(1 + q_{tag} \times (1-2\eta) \times \cos(\Delta m_{d,PDG} \times t)) \\ & \otimes G(0, S_{\sigma_t} \times \sigma_t)] \times \varepsilon_t(t) \times \epsilon_{sig,OST} \end{aligned} \tag{53}$$

for tagged events. The Λ_b^0 background does not oscillate but shows the same tagging efficiency and mistag probability as the signal when using the opposite side tagging algorithm, as it also contains a b quark.

The same side kaon tagging algorithm, on the contrary, is only sensitive to B_s^0 mesons, so it does treat all the background contributions the same. It might however also show an overall asymmetry of the background contributions, that would dilute the tagging parameters of the signal if they where not accounted for. The same reasoning applies to the combinatorial background for both same and opposite side tagged events. The PDF for all the backgrounds in the same side kaon tagging algorithm as well as for the Λ_b^0 and combinatorial background in the opposite side tagging algorithm is given by

$$\mathcal{P}_{bkg_i,osc,j}(t) \quad \propto \quad \begin{cases} \mathcal{P}_i(t,\sigma_t) \times (1-\epsilon_{bkg_i,j}) & \text{if } q_j = 0 \\ \mathcal{P}_i(t,\sigma_t) \times \omega_{bkg_i,j} \times \epsilon_{bkg_i,j} & \text{if } q_j = 1 \\ \mathcal{P}_i(t,\sigma_t) \times (1-\omega_{bkg_i,j}) \times \epsilon_{bkg_i,j} & \text{if } q_j = -1 \end{cases} \tag{54}$$

where i denotes the specific background an j the tagging algorithm. For the background part of the decay time PDF the individual tagging efficiencies ϵ_{bkg_i} and tagging asymmetries $\omega_{bkg_i,j}$ are floating in the fit.

The predicted mistag η is used in the PDF as an per-event quantity similar to the per-event error σ_t from the decay time fit. Therefor, the normalized predicted η distributions $\mathcal{P}_{N,\eta,sig,OST}(\eta_{OST})$, $\mathcal{P}_{N,\eta,bkg_{comb},OST}(\eta_{OST})$, $\mathcal{P}_{N,\eta,sig,SSKT}(\eta_{SSKT})$ and $\mathcal{P}_{N,\eta,bkg_{comb},SSKT}(\eta_{SSKT})$ for the two different tagging cases have to be multiplied to the various signal and background parts of the PDF for normalization purposes. They are derived in the same way as for the decay time error σ_t from sideband subtracted signal decays for the signal and from the high mass sideband for the background. For the physical background contributions they are assumed to be signal like in case of the opposite side tagging algorithms and background like for the same side tagging algorithms.

6 COMPARISON OF DATA AND SIMULATION

As discussed in Chapter 4.6.3, the current SSK tagging algorithm uses a cut based selection to select kaons originating from the B_s^0 fragmentation. Due to differences between the data and the simulation, this selection has been developed using simulated events, where the information whether or not a kaon originates from the fragmentation of the B_s^0 meson is available, and optimized in data in an iterative procedure. This section discusses in detail the differences between data and simulation and their impact on the performance of the SSK tagging in the simulation. First, an overview over the existing differences is given, and the resulting difference in the tagging efficiency and mistag in the simulation is discussed. Then, possible corrections for the existing simulation are introduced, and the impact on the SSK tagging performance in the simulation is illustrated.

The comparisons of the data and the simulation have been developed in parallel to the studies for a new implementation of the opposite side kaon tagger, *cf.* [54]. While the physics principles of both taggers is different, some aspects of the observed differences, especially the primary vertex multiplicity and the tagging track candidate multiplicity influence both taggers and the corrections performed to the simulation are similar.

6.1 DIFFERENCES BETWEEN DATA AND SIMULATION FOR THE SSK TAGGING

As discussed in the previous section, it is necessary to perform a fit to the oscillation of the B_s^0 mesons to extract the performance of the SSK tagger from the data. To compare these performance numbers with the simulation, the same reconstruction and selection cuts as for the data are used to select simulated $B_s^0 \to D_s^- \pi^+$ candidates. Consequently, also for the selection of the tagging candidates the same selection as for the SSK tagger in the data is used. For the extraction of the tagging performance in simulated events it is however not necessary to perform a fit of the B_s^0 oscillation as the information of the initial flavour of the B_s^0 meson is available. By simply counting the number of candidates were a tagging track candidate with the right charge, wrong charge or no tagging track candidate at all has been selected the efficiency ε_{tag}, the mistag ω and the effective tagging efficiency $\varepsilon_{eff} = \varepsilon_{tag}(1-2\omega)^2$ can be calculated.

The numbers for the tagging performance in the data and in the simulation are listed in table 11. The cut based same side kaon tagging algorithm introduced in Chapter 4.6.3 is used as an reference point. The effective tagging efficiency ε_{eff} in the simulation is about twice as high as on data. While the tagging efficiency is lower in the simulation, most notably also the mistag fraction is lower by about 8%. To better understand possible causes for this difference, the relevant input quantities to the tagging algorithm are studied. Figure 35 shows the initial distributions of several event variables such as the number of reconstructed primary vertices, the number of tag-

	ε_{tag} (%)	ω (%)	ε_{eff} (%)
SSK tagger data	15.8±0.3	34.6±1.5	1.5±0.3
SSK tagger simulation	13.68± 0.07	26.53± 0.23	3.01± 0.06

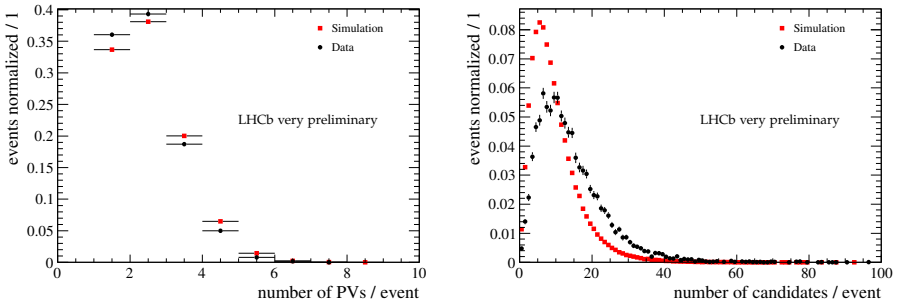
Table 11: Initial tagging performance for the cut based SSK tagger on data and simulation.

ging track candidates and the momentum and transverse momentum of the B_s^0 candidate. The distributions derived from data are sideband subtracted in the B_s^0 mass. The average number of reconstructed primary vertices per event (Figure 35a) is slightly higher in the simulation. A larger number of primary vertices and thus larger number of tagging track candidates leads to a higher mistag because there is a higher probability to chose the wrong tagging track candidate. The number of tagging track candidates (Figure 35b) shows a large deviation between the data and the simulation. The average number of tagging candidates after the preselection is 13.4 while in the simulation it is only 9.3. Figure 35c and 35d show the kinematic distributions of the B_s^0, which are rather well reproduced in the simulation. Figure 36 shows the distribution of all the variables that are used in the selection of the same side kaon tagging algorithm in the data and the simulation. The momentum (Figure 36c) and transverse momentum (Figure 36d) of the tagging tracks is higher in the data than in the simulation. This is especially problematic in case of the transverse momentum, as the track with the highest transverse momentum is later used as the final tagging track. The IP/σ_{IP} of the tagging tracks with respect to the B_s^0 production vertex (Figure 36f) is larger in the data than it is in the simulation. Finally, the χ^2_{track}/ndf (Figure 36e) is more smeared out in the data with respect to the simulation. The effect of these variables on the tagging performance is more complicated to predict. In the remainder of this chapter the aforementioned deviations are either corrected or their effect on the performance of the SSK tagger is estimated.

6.2 NUMBER OF PRIMARY VERTICES

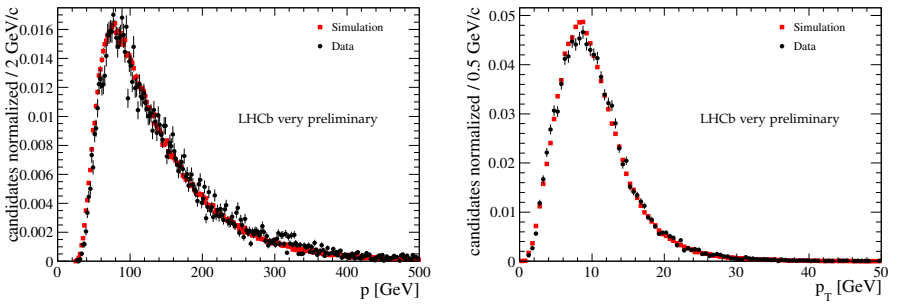
The number of primary vertices in the event and the number of tagging track candidates is correlated. Therefor, a correction is applied to the number of primary vertices first, before a correction of the number of tagging track candidates is attempted, although the differences in this variable are small, see Figure 37a.

Before the correction, there are on average 1.96 PVs per event reconstructed in the data compared to 2.05 in the simulation. The correction is performed by throwing away events in the simulation in a random way until the composition of events with different numbers of PVs is the same as in the data. This leads to a loss of 6.6% of the statistics in the simulation. After the correction, the distributions of the number of reconstructed PVs agrees per construction (Figure 37c). Figures 37b and 37d show the number of tagging track candidates before and after the PV correction. While in the data there are 13.4 candidates on average, there are 9.3 candidates in the simulation before and 9.1 after correcting for the number of PVs.



(a) Number of PVs per event.

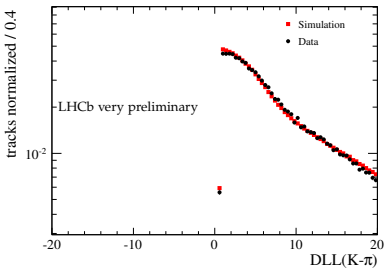
(b) Number of tagging track candidates per event



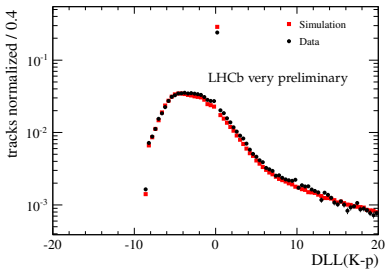
(c) B_s^0 candidate momentum.

(d) B_s^0 candidate transverse momentum.

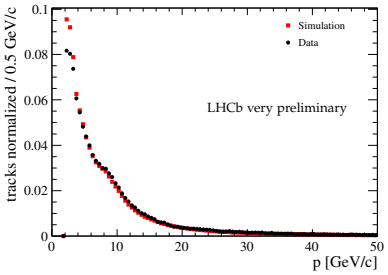
Figure 35: Initial differences between data and simulation in the number of PVs per event and in the number of tagging track candidates per event as well as for the momentum and the transverse momentum of the B_s^0 candidate. Plots normalized to the same number of B_s^0 candidates.



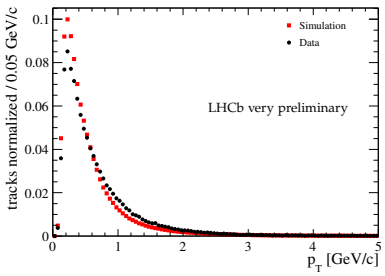
(a) $DLL(K-\pi)$ of tagging track candidates.



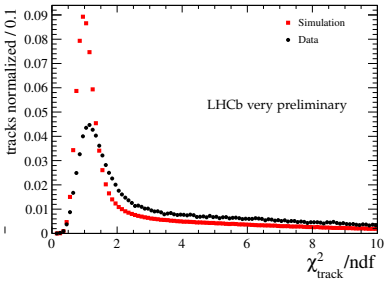
(b) $DLL(K-P)$ of tagging track candidates.



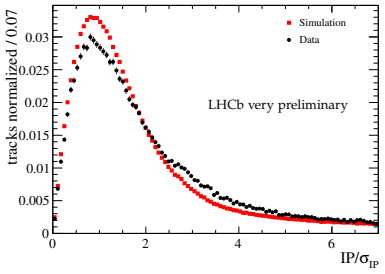
(c) Momentum of tagging track candidates.



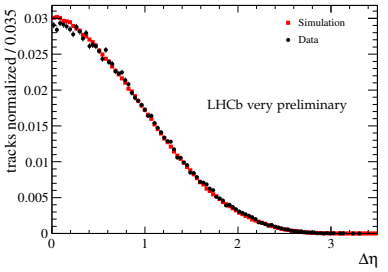
(d) Transverse momentum of tagging track candidates.



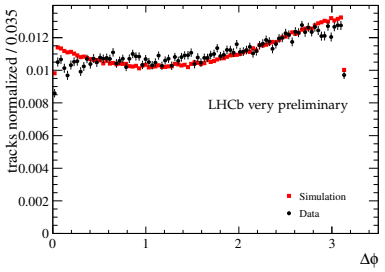
(e) χ^2_{track}/ndf of tagging track candidates.



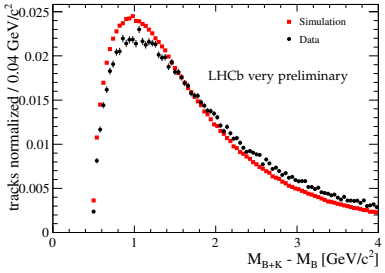
(f) IP/σ_{IP} of tagging track candidates.



(g) $\Delta\eta$ of tagging track candidates and B^0_s.

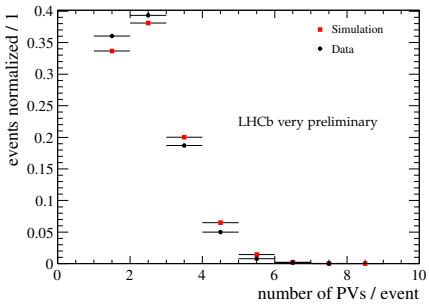


(h) $\Delta\phi$ of tagging track candidates and B^0_s.

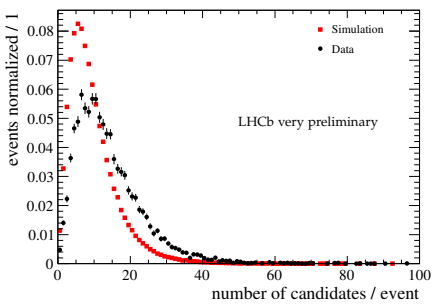


(i) $M_{B+K} - M_B$ of tagging track candidates.

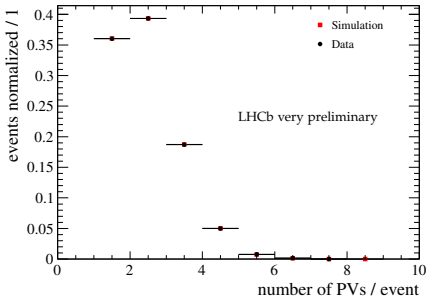
Figure 36: Initial differences between the data and the simulation of tagging track quantities relevant to the selection in the SSK tagger. Plots normalized to the same number of tagging track candidates.



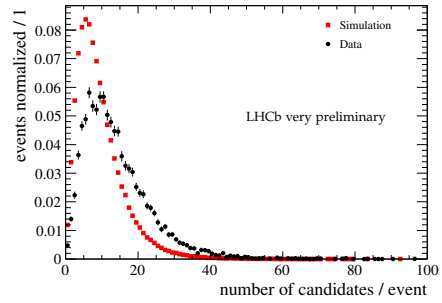
(a) Number of PVs per event before correction.



(b) Number of tagging track candidates per event before correction.



(c) Number of PVs per event after correction.



(d) Number of tagging track candidates per event after correction.

Figure 37: Number of PVs and tagging track candidates before and after correction for the difference in the number of PVs distribution in the data and the simulation. Plots are normalized to the same number of B_s^0 candidates.

Table 12 lists the influence of this correction on the tagging performance in the simulation. As the correction is very small also the effect on the tagging performance is very small, resulting in an relative 0.7% larger effective tagging efficiency εD^2 in the simulation after the correction. For all further corrections, the MC dataset with the corrected number of reconstructed primary vertices is used.

6.3 TRACK MULTIPLICITY

By comparing the distributions of the data and the simulation, a large discrepancy in the number of tagging track candidates is observed. The ad-

change	$\Delta\varepsilon_{tag}$ (%)	$\Delta\omega$ (%)	$\Delta\varepsilon_{eff}$ (%)
absolute	-0.01±0.10	0.05±0.34	-0.02±0.08
relative	-0.07 ± 0.72	0.19± 1.28	-0.66± 2.82

Table 12: Absolute and relative change of the average tagging performance of the SSK tagger in the simulation by applying the correction for the observed difference in the number of PVs. The numbers show the average tagging performance before subtracted by the performance after the correction.

ditional tracks in the data can in principle stem from two different origins: the B_s^0 meson fragmentation and the underlying event (UE), ie. from soft QCD processes. On the other hand, the fragmentation of the B_s^0 meson produces only few tracks per event, and they are in specific regions of the variables shown in Figure 36, thus additional tracks would be seen primarily in these regions (a detailed comparison of the distributions of the fragmentation tracks and the underlying event tracks can be found in Chapter 7.2). It is therefor unlikely that the observed large difference in the multiplicity can be explained by missing fragmentation tracks in the simulation. The simulation of soft QCD processes however is challenging, as discussed in Chapter 4.6.2. It might be, that the UE processes are not correctly simulated and thus the MC does not reproduce the track multiplicity and kinematic distributions observed in the data.

The soft QCD processes are separated from the B_s^0 fragmentation and should thus not show any charge correlation with the B_s^0 flavour[1]. A track from the underlying event which is selected by the SSK tagger has therefor a mistag probability of 0.5. Consequently, a higher multiplicity of UE tracks in the data which are potentially selected by the SSK tagger would increase the mistag probability. To test the sized of this effect a correction for the track multiplicity is performed. Tagging track candidates are added to the simulated events until the distribution of the number of tagging track candidates agrees with the data.

By comparing the momentum and transverse momentum of the tagging tracks before the correction (Figure 36c and 36d) one can observe that there are also differences in the kinematical distributions. Therefor, as a second goal, the observed difference in the kinematic distribution is corrected in the same step. The tracks that are added are taken from the same simulated sample. To not bias the simulated events, the following requirements are made to the tracks that are added:

- The tracks are chosen in order to make up for the observed differences in the kinematical distributions of tagging track candidates in the data compared to the simulation.
- The tracks must fulfill the same preselection and PID cuts as all other tagging track candidates.
- The tracks must not come from the fragmentation or the decay of the B_s^0 or the opposite side B meson.
- The tracks must come from a different event than the one they are added to.
- The tracks must come from events which have the same number of reconstructed primary vertices as the events they are added to.
- Each track can only be added once.

To calculate the number of tracks that are added, the simulated dataset is split into five different categories corresponding to 1, 2, 3, 4 or $\geqslant 5$ reconstructed PVs. From the mean number of tagging track candidates present

1 In fact they may show a charge asymmetry due to the fact that the initial state at the LHC is not charge neutral. The charge asymmetry in the initial state, represented by the valence quarks in the proton, can lead to an asymmetry in the production of certain particle species in the event, *cf.* [55]. This however is unrelated to the initial flavour of the signal B_s^0.

in the data and in the simulation in each of these categories, the fraction of missing tracks

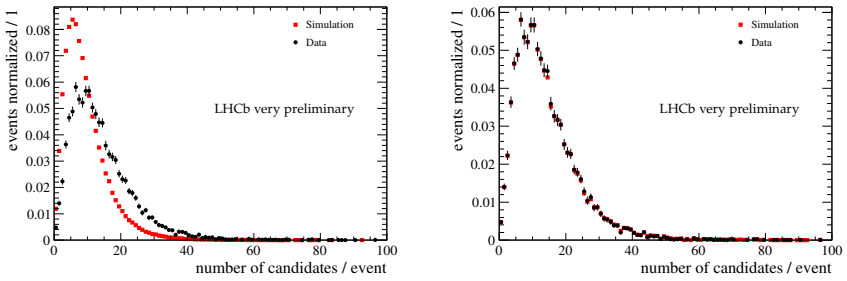
$$f = \frac{\mu_{data} - \mu_{MC}}{\mu_{MC}} \tag{55}$$

is calculated, where μ_{data} and μ_{MC} are the mean number of tagging tracks present in the data and the simulation respectively. The values for f are 0.39, 0.54, 0.52, 0.40 and 0.28 for the five categories. From the fraction of missing tracks, the number of added tracks for a given event is calculated as

$$n_{add} = f \cdot n_{tracks}, \tag{56}$$

where n_{tracks} is the number of tagging track candidates. Only an integer number of tracks can actually be added to the event. To account for that, the final number is determined by throwing a gaussian distributed random number with mean $\mu = n_{add}$ and width $\sigma = n_{add}/3$ and round that number to the nearest integer. Figure 38 shows the distributions of the number of tagging track candidates in the data and the simulation before (38a) and after (38b) the correction. After the correction, the two distributions are in very good agreement. The comparison of Figures 41 and 42 shows that the agreement also holds when the sample is split according to the different number of reconstructed primary vertices.

To correct the phase space distribution, the added tracks are chosen in such a way that they make up for the observed difference. For that purpose the difference of the two-dimensional $p - \eta$ distribution of the tagging tracks in the data and the simulation is taken to generate the phasespace of the missing tracks, see Figure 39. The tracks that are added in the track multiplicity correction are then chosen in such a way that they populate this phasespace. Figure 43 gives an overview of the kinematic distributions of the tagging track candidates before and after the track multiplicity correction. While there are still differences before adding the missing tracks, these differences shrink once the correction is applied. However, there are remaining differences and several effects complicate the correction of the kinematical distributions. First, there are regions in the phasespace which are only sparsely populated by tracks in the simulation but show still differences between the data and the simulation that have to be corrected in the track adding. Due to the sparse population, it can be impossible to find tracks to add to these regions. Second, when looking at the difference of the two-dimensional $p - \eta$ distribution of the tagging tracks in the data and the simulation, see Figure 39c, some regions of phasespace show a negative weight, ie. tracks would have to be subtracted in order to match the distributions. As this effect is only small, this is not done here. Phasespace bins with negative weights are treated as zero. Finally, the kinematic distributions are corrected on average and correlations within the event are not treated. The effect of such correlations is visible comparing the kinematic distributions of the tagging track candidate with the highest transverse momentum in the data and the simulation after the overall discrepancies have been corrected, see Figure 40. While the kinematic distributions of all tagging track candidates agree, there are still differences visible in the distributions for the tagging track candidate with the highest p_T, which determines the tagging decision in the cut based same side kaon tagger. These remaining differences partially explain the still persisting differences in the tagging performance.



(a) Number of tagging track candidates before correction.

(b) Number of tagging track candidates after correction.

Figure 38: Number of tagging track candidates in the data and the simulation before and after the correction of the track multiplicity. Plots are normalized to the same number of B_s^0 candidates.

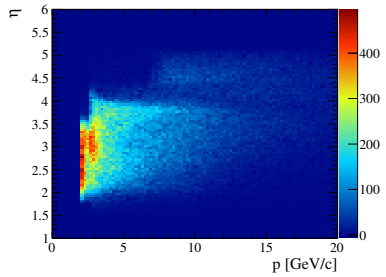
change	$\Delta\varepsilon_{tag}$ (%)	$\Delta\omega$ (%)	$\Delta\varepsilon_{eff}$ (%)
absolute	-6.66±0.11	-7.99±0.32	1.07±0.08
relative	-48.65 ± 0.82	-30.17± 1.23	35.31± 2.67

Table 13: Absolute and relative change of the average tagging performance of the SSK tagger in the simulation by applying the correction for the observed difference in the number of tagging track candidates. The numbers show the average tagging performance before subtracted by the performance after the correction.

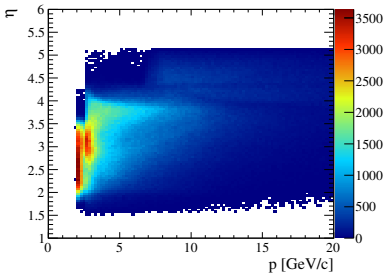
The effect of the track multiplicity correction on the tagging performance in the simulation is shown in Table 13. The correction has an significant impact on both the tagging efficiency ε_{tag}, which is increased by an absolute 6.66 %, as well as on the mistag fraction ω, which rises by an absolute 7.99 %. The change in the efficiency originates from the fact that due to the larger multiplicity there are now more events where at least one track passes the SSK tagging selection, while the change in the mistag comes from the fact that the added tracks are solely UE tracks that feature no tagging information, *i.e.* they have a mistag probability of 50 %. In total, this correction amounts to a decrease in the effective tagging power $\varepsilon_{tag}D^2$ of an absolute 1.07 % which corresponds to an relative loss of 35.3 %.

6.4 IMPACT PARAMETER RESOLUTION

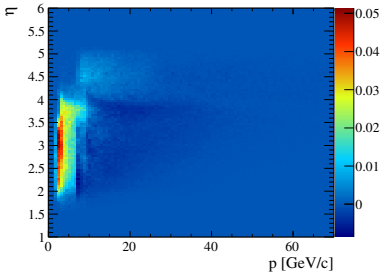
The kaons used for the tagging in the SSK tagger are produced in the fragmentation of the B_s^0, and for all practical matters one can consider this process to happen instantaneously, which means that the kaon originates from the same primary vertex as the B_s^0. A cut on the IP/σ_{IP} rejects tracks from the opposite side B decay and badly reconstructed tagging track candidates. As can be observed in Figure 36f, this quantity shows significant deviations in the simulation from the data. This deviations have their origin in a remaining misalignment of the detector and other higher order effects of material interaction, which are not properly simulated in the MC.



(a) p vs. η distribution of tagging track candidates in the data.

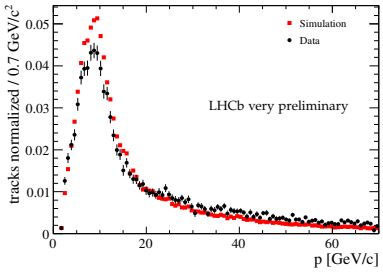


(b) p vs. η distribution of tagging track candidates in the simulation.

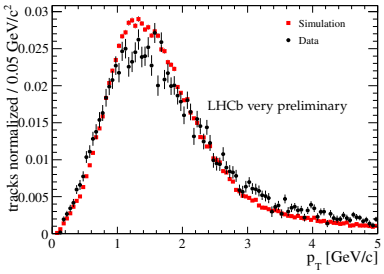


(c) Difference of the p vs. η distributions of tagging track candidates in the data and the simulation.

Figure 39: p vs. η distribution of tagging track candidates in the data, the simulation and the relative difference showing the phasespace of the missing tagging track candidates.

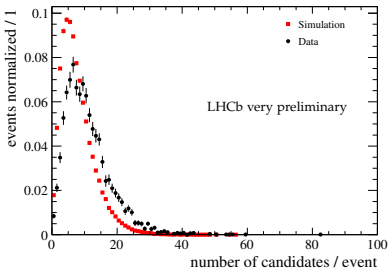


(a) Momentum of tagging track candidates with highest p_T after correction.

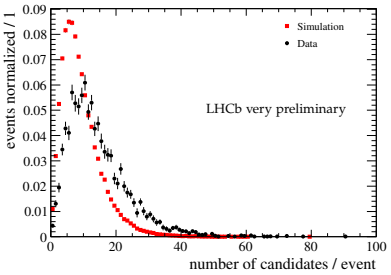


(b) Transverse momentum of tagging track candidates with highest p_T after correction.

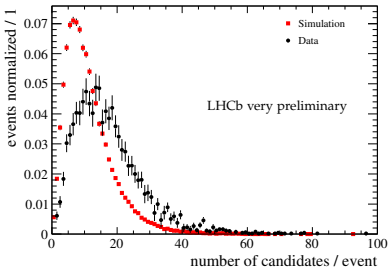
Figure 40: Kinematical distribution of tagging track candidates with the highest transverse momentum after the track multiplicity correction. Plots are normalized to the same number of tagging track candidates.



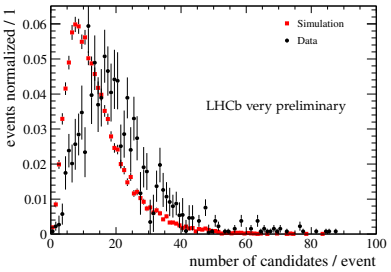
(a) Number of tagging track candidates before correction n=1 PV.



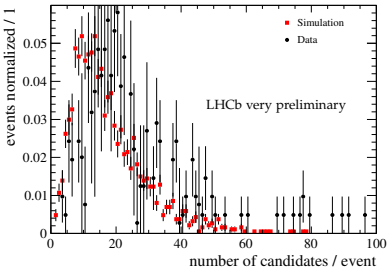
(b) Number of tagging track candidates before correction n=2 PV.



(c) Number of tagging track candidates before correction n=3 PV.

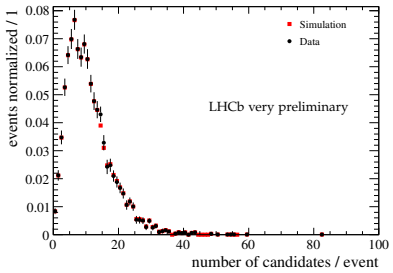


(d) Number of tagging track candidates before correction n=4 PV.

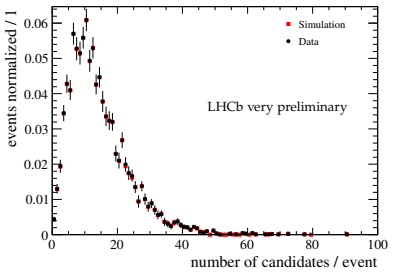


(e) Number of tagging track candidates before correction n=5 PV.

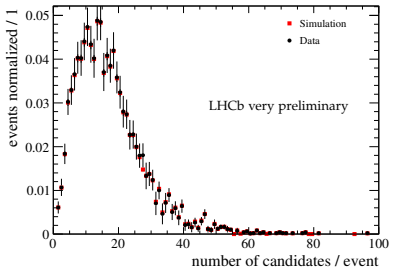
Figure 41: Number of tagging track candidates before the track multiplicity correction for different numbers of reconstructed primary vertices. Plots are normalized to the same number of events.



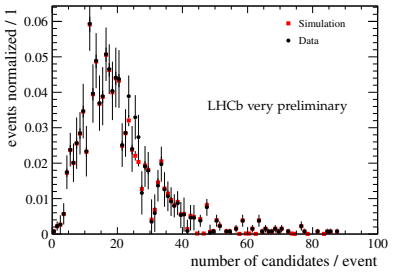
(a) Number of tagging track candidates per event after correction n=1 PV.



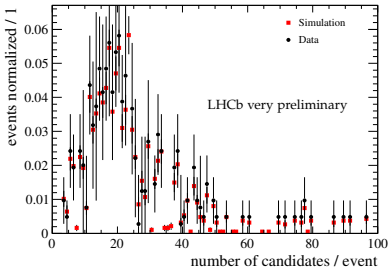
(b) Number of tagging track candidates per event after correction n=2 PV.



(c) Number of tagging track candidates per event after correction n=3 PV.

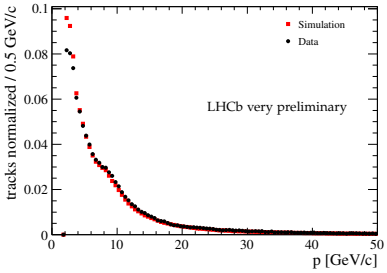


(d) Number of tagging track candidates per event after correction n=4 PV.

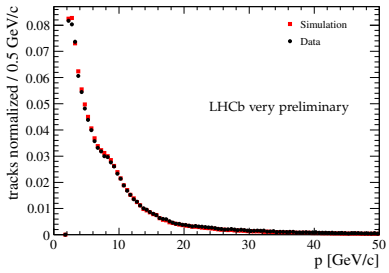


(e) Number of tagging track candidates per event after correction n=5 PV.

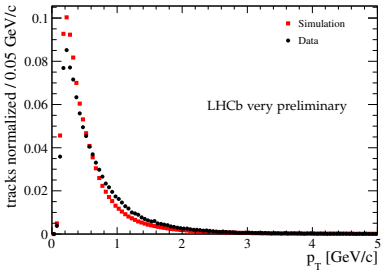
Figure 42: Number of tagging track candidates after the track multiplicity correction for different numbers of reconstructed primary vertices. Plots are normalized to the same number of events.



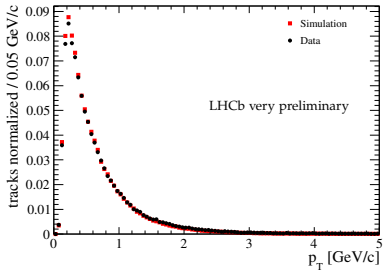
(a) Momentum of tagging track candidates before correction.



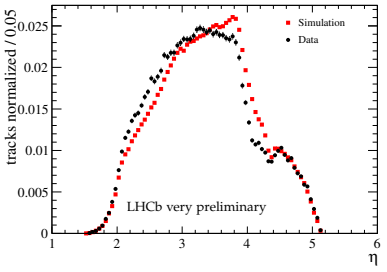
(b) Momentum of tagging track candidates after correction.



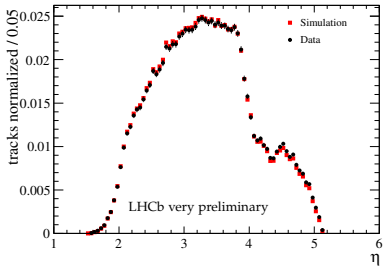
(c) Transverse momentum of tagging track candidates before correction.



(d) Transverse momentum of tagging track candidates after correction.



(e) η distribution of tagging track candidates before correction.



(f) η distribution of tagging track candidates after correction.

Figure 43: Kinematical distribution of tagging track candidates before and after the track multiplicity correction. Plots are normalized to the same number of tagging track candidates.

change	$\Delta\varepsilon_{tag}$ (%)	$\Delta\omega$ (%)	$\Delta\varepsilon_{eff}$ (%)
absolute	0.45±0.11	-0.01±0.30	0.05±0.07
relative	2.21 ± 0.56	-0.03± 0.86	2.55± 3.61

Table 14: Absolute and relative change of the average tagging performance of the SSK tagger in the simulation by applying the smearing of the impact parameter. The numbers show the average tagging performance before subtracted by the performance after the correction.

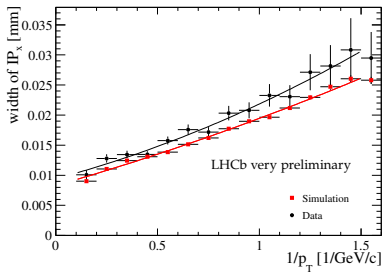
To correct for this difference, a smearing of the IP of the tagging track candidate is applied. For the smearing, the impact parameter is split into its x and y component IP_x and IP_y. As the resolution effects are dependent on the momentum of the particle, both components are fitted by a Gaussian distribution in bins of the transverse momentum p_T to get their width. Figure 44 shows the results of these fits in bins of the transverse momentum p_T in both the data and the simulation. The shape of the p_T dependence of the width is parametrized by a second order polynomial, in the following called $IP_{x(data)}(p_T)$ and $IP_{x(MC)}(p_T)$. To account for the difference between the data and the simulation, the width difference

$$\Delta\sigma_x(p_T) = \sqrt{IP^2_{x(data)}(p_T) - IP^2_{x(MC)}(p_T)} \tag{57}$$

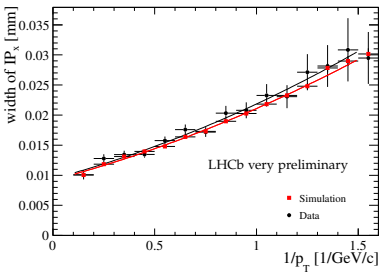
is calculated from this parametrization. To get the smearing for a given track, a random number, $\sigma_{x,random}$, distributed according to a Breit Wigner function is drawn with a mean of zero and a width $\Delta\sigma_x(p_T)$. The same applies for the y component of the impact parameter. The smeared impact parameter of the track is then given by

$$IP_{smeared}(p_T) = \sqrt{(IP_x(p_T) + \sigma_{x,random}(p_T))^2 + (IP_y(p_T) + \sigma_{y,random}(p_T))^2}. \tag{58}$$

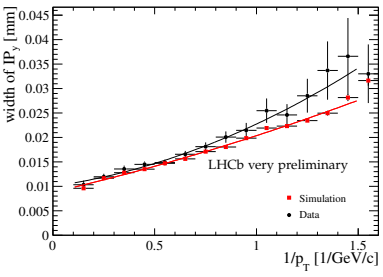
After the smearing, the fitted widths of the IP_x and IP_y distributions as a function of the transverse momentum of the particle agree, see Figure 44. Figure 45 shows the IP_x and IP_y distributions summed over all transverse momenta before and after the correction. The impact of the smearing procedure on the IP and IP/σ_{IP} is shown in Figure 46. Although there are still differences visible, the overall agreement - especially for the IP/σ_{IP} that is used in the selection of the SSK tagger - has improved. Table 14 summarises the effect of this correction on the tagging performance. The selection of the SSK tagger requires that the IP/σ_{IP} of a track is smaller than 4.125. As the distribution in the simulation was narrower than the one in the data, more tracks passed this cut in the simulation. The correction of the impact parameter leads to an absolute 0.05 % higher effective tagging power which corresponds to an relative change of 2.55 %.



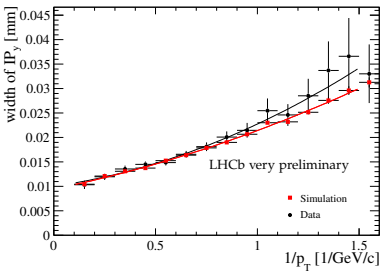
(a) IP_x resolution before smearing.



(b) IP_x resolution after smearing.

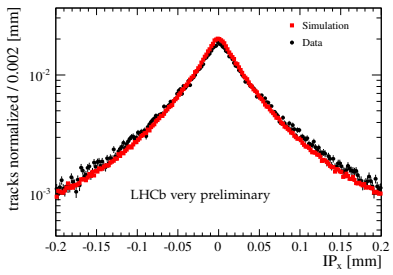


(c) IP_y resolution before smearing.

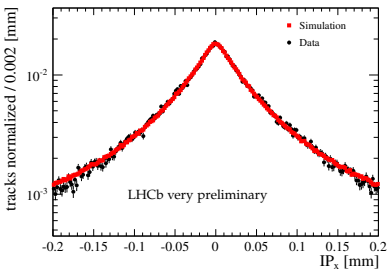


(d) IP_y resolution after smearing.

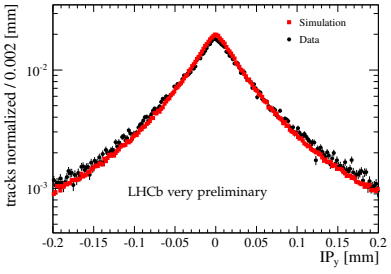
Figure 44: x and y component of the IP resolution as function of transverse momentum p_T of tagging track candidates before and after the smearing.



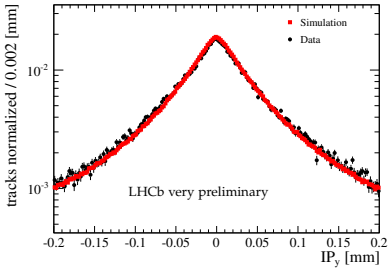
(a) IP_x before smearing.



(b) IP_x after smearing.

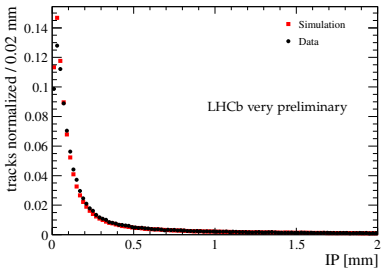


(c) IP_y before smearing.

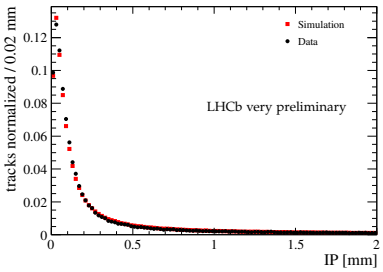


(d) IP_y after smearing.

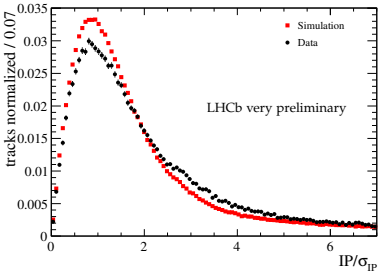
Figure 45: IP_x and IP_y of tagging track candidates before and after the smearing. Plots are normalized to same number of tagging candidates.



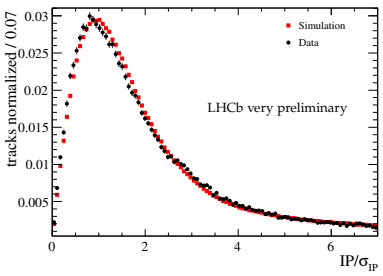
(a) IP of tagging track candidates before smearing.



(b) IP of tagging track candidates after smearing.



(c) IP/σ_{IP} of tagging track candidates before smearing.



(d) IP/σ_{IP} of tagging track candidates after smearing.

Figure 46: IP and IP/σ_{IP} of tagging candidates before and after the smearing. Plots are normalized to same number of tagging track candidates.

6.5 χ^2_{track}/ndf QUALITY OF TAGGING TRACK CANDIDATES

After the correction of the PV and track multiplicity, the kinematic distributions of the tagging track candidates and the impact parameter distribution, the χ^2_{track}/ndf obtained from the track fit of the tagging candidates (see Figure 36e) is the variable with the largest uncorrected difference between the data and the simulation. The distribution in the data has a less prominent peak and consequently more events in the tail toward higher χ^2_{track}/ndf. The reason for this difference are the material distribution and misalignement, which are not properly described in the simulation. To select tracks of good quality and reject misreconstructed tracks, the SSK tagger requires a $\chi^2_{track}/ndf < 3.75$ in the selection. Due to the aforementioned differences this cut has a different efficiency in the data than it has in the simulation where it rejects less tagging track candidates.

As the χ^2_{track}/ndf distribution has a rather complicated shape it is not possible to apply a smearing as used for the impact parameter to account for the differences. Instead, another approach is taken to estimate the effect of the discrepancies. To estimate the efficiency of a given χ^2_{track}/ndf cut in the data and the simulation respectively the cumulative distribution is taken,

$$F_{\chi^2_{track}/ndf}(x) = \int_0^x PDF_{\chi^2_{track}/ndf}(x)dx. \tag{59}$$

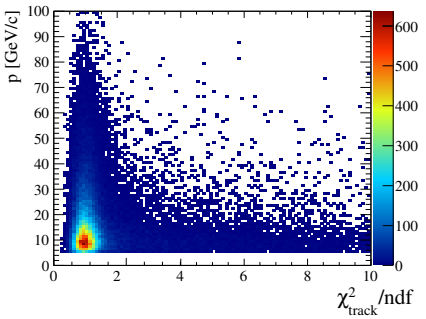


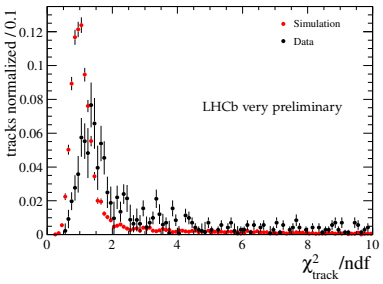
Figure 47: Correlation between χ^2_{track}/ndf and momentum p of tagging track candidate.

bin [GeV/c^2]	$\langle p \rangle$ [GeV/c]	ϵ_{Data} [%]	χ^2_{track}/ndf simulation
$0 \leqslant p < 9$	6.714 ± 0.010	78.58	1.45
$9 \leqslant p < 14$	10.369 ± 0.013	85.94	1.75
$14 \leqslant p < 26$	17.850 ± 0.029	89.34	1.75
$26 \leqslant p$	38.178 ± 0.140	92.23	1.95

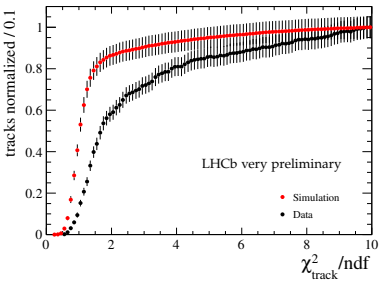
Table 15: Efficiency of a $\chi^2_{track}/ndf < 3.75$ cut in the data and the corresponding cut values in the simulation for four different bins in the momentum of the tagging track candidates.

The value of the cumulative distribution then represents directly the efficiency for a given maximum cut value χ^2_{track}/ndf. As the distribution in the simulation is much narrower than in the data, the cut in the simulation has to be tighter to reach the same efficiency. To get this tighter cut value, first the efficiency for a given cut value in the data is taken from the cumulative distribution. Then, the cut value with the same efficiency is searched from the cumulative distribution for the simulation. All other cuts besides the χ^2_{track}/ndf cut are already applied prior to this procedure to minimize correlations. The size of the effects that cause the difference between the data and the simulation are potentially dependent on the momentum of the particle (see Figure 47). To account for this dependence, the estimation of the different cut values for the simulation is performed in four bins of tagging track momentum. Table 15 summarizes the efficiency of a $\chi^2_{track}/ndf < 3.75$ cut in the data and the corresponding cut value in the simulation for the four different momentum bins. The distributions of the χ^2_{track}/ndf as well as the cumulative distributions are shown in Figure 48.

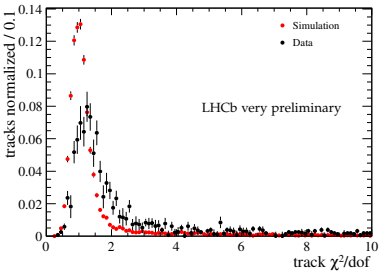
To estimate the influence of the χ^2_{track}/ndf distributions in the data and the simulation, the cuts that were obtained by the above mentioned procedure are applied in the simulation. Different cut efficiencies have an influence especially on the tagging efficiency, which is lower due to this correction by relative $6.58 \pm 0.57\%$ in the simulation and also the mistag fraction ω is lower because more badly reconstructed tagging candidates are rejected, see Table 16. The effective efficiency is unchanged because both effects cancel. It should be kept in mind that this study only provides an estimate of the size of the effect of the differences between the χ^2_{track}/ndf distributions observed in the data and the simulation as the χ^2_{track}/ndf can potentially be correlated to other variables than the momentum.



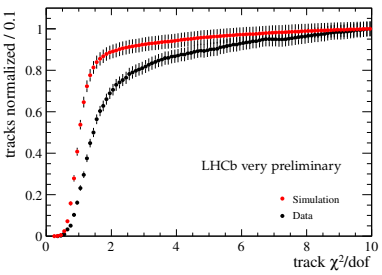
(a) χ^2_{track}/ndf distribution for $0\,GeV/c \leqslant p < 9\,GeV/c$



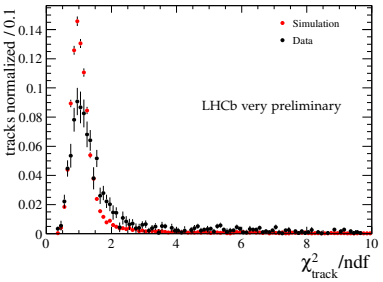
(b) Cumulative χ^2_{track}/ndf for $0\,GeV/c \leqslant p < 9\,GeV/c$



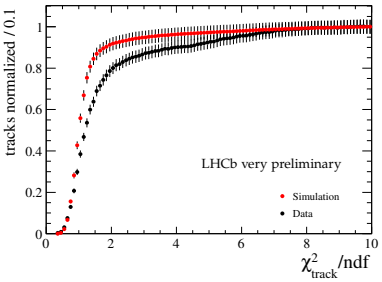
(c) χ^2_{track}/ndf distribution for $9\,GeV/c \leqslant p < 14\,GeV/c$



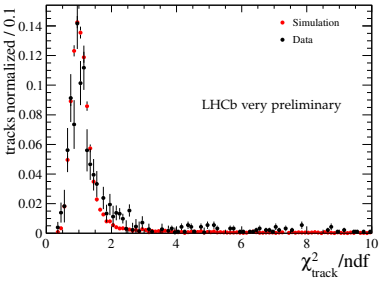
(d) Cumulative χ^2_{track}/ndf for $9\,GeV/c \leqslant p < 14\,GeV/c$



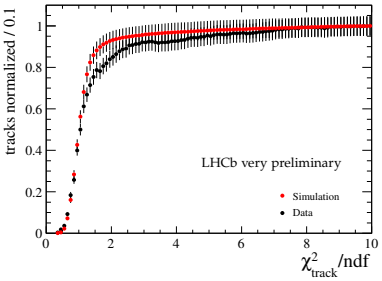
(e) χ^2_{track}/ndf distribution for $14\,GeV/c \leqslant p < 26\,GeV/c$



(f) Cumulative χ^2_{track}/ndf for $14\,GeV/c \leqslant p < 26\,GeV/c$



(g) χ^2_{track}/ndf distribution for $26\,GeV/c \leqslant p$



(h) Cumulative χ^2_{track}/ndf for $26\,GeV/c \leqslant p$

Figure 48: Comparison of the χ^2_{track}/ndf distribution and its cumulative distribution for tagging track candidates in the data and the simulation for the four different momentum bins.

change	$\Delta\varepsilon_{tag}$ (%)	$\Delta\omega$ (%)	$\Delta\varepsilon_{eff}$ (%)
absolute	1.31±0.11	0.53±0.30	0.00±0.07
relative	6.58±0.57	1.54±0.88	0.00±3.70

Table 16: Absolute and relative change of the average tagging performance of the SSK tagger in the simulation by applying different χ^2_{track}/ndf in the simulation to account for the observed differences between the data and the simulation. The numbers show the average tagging performance before subtracted by the performance after the correction.

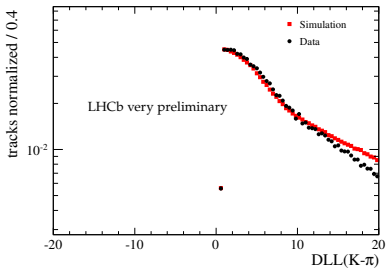
	ε_{tag} (%)	ω (%)	ε_{eff} (%)
SSK tagger data	15.8±0.3	34.6±1.5	1.5±0.3
SSK tagger simulation initial	13.68±0.07	26.53±0.23	3.01±0.06
PV correction	13.69±0.07	26.48±0.24	3.03±0.06
multiplicity and kinematic correction	20.35±0.08	34.47±0.21	1.96±0.05
IP smearing	19.90±0.08	34.48±0.21	1.91±0.05
χ^2_{track}/ndf (final correction)	18.58±0.08	33.95±0.22	1.91±0.05

Table 17: Average tagging performance for the cut based SSK tagger in the data and the simulation for different stages of the correction.

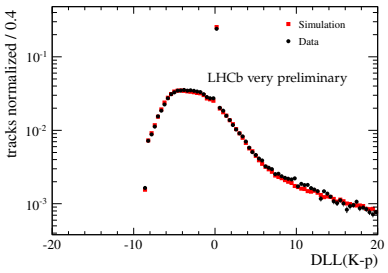
6.6 SUMMARY THE DIFFERENCES BETWEEN THE DATA AND THE SIMULATION

Table 17 gives a summary of the changes in the tagging performance in the simulation introduced by the corrections that are discussed in this chapter. Several effects are corrected: The primary vertex multiplicity is adapted and tracks are added to the simulated events to account for the difference in the multiplicity of the tagging track candidates. These tracks are chosen in such a way that they improve the agreement in the kinematic distributions of the tagging track candidates in the data and the simulation. The impact parameter is smeared to make up for a worse resolution in the data and the effects of the different χ^2_{track}/ndf distribution are evaluated. The largest modification is introduced by the correction of the tagging track candidate multiplicity. After all corrections, the agreement of the tagging performance in the data and the simulation has improved and the mistag fraction agrees within 0.5 standard deviations. The tagging efficiency is too high by relative 15%. Figure 49 shows the agreement between the distributions of the variables used in the SSK tagger after the corrections. The agreement is improved with respect to the situation before the correction, see Figure 36 (besides for the χ^2_{track}/ndf distribution which was not corrected but uses a different cut value in the simulation).

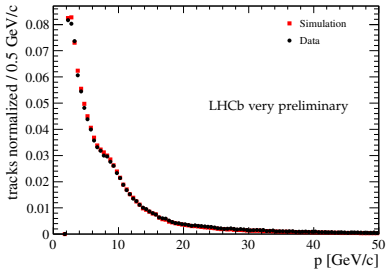
The resulting agreement between the data and the simulation is not perfect. Possible remaining discrepancies stem from untreated correlations within the event or a wrongly described fragmentation. Those effects require a tuning of the initial steps of the simulation procedure and are beyond the scope of this work. The correction steps mentioned in this chapter are - besides the treatment of the χ^2_{track}/ndf - the foundation of the following improvement of the SSK tagging algorithm.



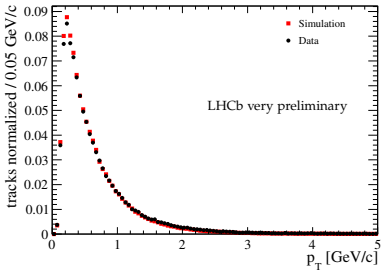
(a) $DLL(K-\pi)$ of tagging track candidates.



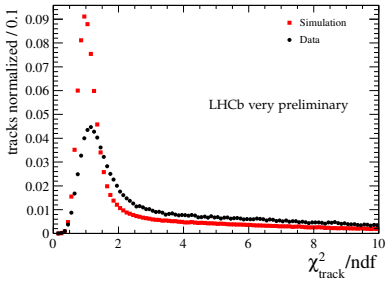
(b) $DLL(K-P)$ of tagging track candidates.



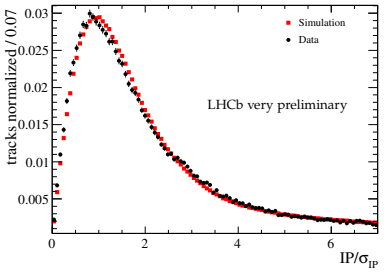
(c) Momentum of tagging track candidates.



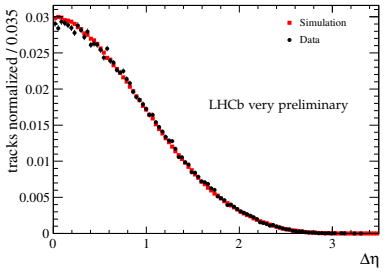
(d) Transverse momentum of tagging track candidates.



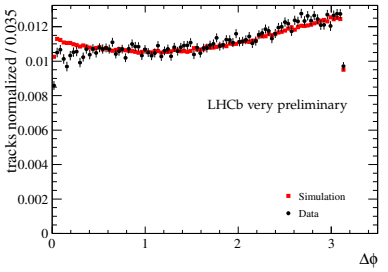
(e) χ^2_{track}/ndf of tagging track candidates.



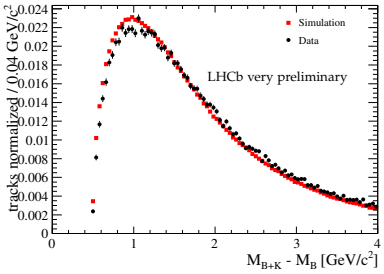
(f) IP/σ_{IP} of tagging track candidates.



(g) $\Delta\eta$ of tagging track candidates and B_s^0.



(h) $\Delta\phi$ of tagging track candidates and B_s^0.



(i) $M_{B+K} - M_B$ of tagging track candidates.

Figure 49: Comparison between data and simulation of tagging track quantities relevant to the selection in the SSK tagger after all corrections to the simulation. Plots normalized to the same number of tagging track candidates.

7 SAME SIDE KAON TAGGING USING NEURAL NETWORKS

The default same side kaon tagging algorithm selects a single tagging track candidate for its decision, not taking into account if the decision was unambiguous or if other tagging track candidates in the event show a quality comparable to the track that is chosen. This strategy is not optimal if more than one tagging track candidate passes the selection.

To improve on this strategy, the cut based selection of the fragmentation tracks is substituted by a neural network classifier. The output of this neural network gives a probability of the fragmentation track likeliness for a given tagging candidate.

After this first neural network classifier, all tagging track candidates with a high probability to come from the B_s^0 fragmentation are combined in a second neural network to make the final tagging decision for a B_s^0 meson and give an estimate on the mistag probability. This approach has the advantage that it covers cases where multiple tracks with similar quality but different charge are present.

This chapter starts with a general introduction on the architecture of the neural network classifiers used in the development of a new SSK tagging algorithm. The first neural network classifier, which is used to discriminate fragmentation tracks from underlying event tracks is presented and the combination of these tracks in a second neural network classifier to determine the final tag decision and mistag probability is discussed. The chapter concludes with a study on the probabilistic interpretation of the neural network output as a mistag probability and the application of the new same side tagging algorithm to simulated events.

7.1 NEURAL NETWORK ARCHITECTURE

An artificial neural network is a self-learning algorithm that combines several input variables to one output variable for the purpose of discriminating between two hypotheses. The basic building block of every neural network is a neuron. A neuron combines several input variables using either a linear or non-linear response function to one output variable with, Sigmoid, hyperbolic tangent or radial functions as examples of the possible response functions. Each input variable can have a specific weight that defines its importance in the combination. By combining several neurons and use multiple layers of neurons one can reach a high complexity. Figure 50 depicts the layout of a simple Multi Layer Perceptron (MLP) neural network algorithm with one input, hidden and output layer. A MLP is characterised by the fact that it is a feedforward artificial neural network, ie. the flow of information in the algorithm is only going in one direction from the input layer to the output layer and there are no feedback loops. The mathematical formulation of this neural network can be given in the following way.

In the first layer, each neuron has only one input variable on which a variable transformation is performed. Several variable transformations are

possible, e.g. linear transformations, principal component analysis or Gaussian transformations. The variable transformations are performed to adjust the very different shapes of the different input variables x_i in such a way that they are more similar and comparable. It is also possible to use a linear decorrelation transformation on the input variables of the neural network. For the purpose of this thesis, two transformations are important: linear transformations and decorrelations. Linear transformations simply use a function

$$f^{(1)}(h) = ah + b \tag{60}$$

on the input variable h with the parameters a, b chosen in such a way that the transformed variable

$$y_i^1 = f^{(1)}(x_i) \tag{61}$$

is constrained to the range between -1 and 1. The upper index 1 on the transformation function $f^{(i)}$ and the transformed variables y_i^1 stands for the first layer of the neural network.

If linear correlations are present between different input variables, this can negatively influence the learning process. To resolve the linear correlations that are found between the input variables, decorrelation transformations, so called whitening filters, are applied. They work in the following way, cf [56]. If C is the covariance matrix of a given variable vector x one can calculate the square root of the covariance matrix C' my means of diagonalising the symmetric covariance matrix (*i.e.* the matrix C is symmetric if $C = C^T$) with a symmetric matrix S

$$D = S^T C S \rightarrow C' = S^T \sqrt{D} S. \tag{62}$$

The decorrelated variable vector x' is then given by multiplying x by the inverse of C'

$$x' \rightarrow (C')^{-1} x. \tag{63}$$

Several variable transformations can be used in conjunction, eg. by first applying a linear transformation to constrain them all to the interval $[-1, 1]$ and then use a decorrelation to get rid of linear correlations between the input variables.

The output of the so transformed input variables y_i^1 is then used as input for the neurons in the second layer of the neural network, the so called hidden layer. The output y_j^2 of a neuron of the hidden layer is calculated as

$$y_j^2 = f^{(2)}(w_{0j}^1 + \sum_i w_{ij}^1 y_i^1) \tag{64}$$

with j is the index of neurons in the hidden layer, i the index of the transformed input variables, w_{ij}^1 the weight for the specific variable, w_{0j}^1 the bias weight and

$$f^{(2)}(h) = \tanh(h) \tag{65}$$

the response function of the neuron, in the case of the neural networks discussed in this thesis a hyperbolic tangent function.

The output of all the neurons of the hidden layer, y_j^2, are combined into a single output neuron in the same fashion. The final output y_1^3 of the neural network is then given by

$$y_1^3 = f^{(3)}(w_{01}^2 + \sum_i w_{ij}^2 y_i^2). \tag{66}$$

If a sigmoid transformation function is used as response function of the output neutron,

$$f^{(3)}(h) = \frac{1}{1+e^{-h}}, \tag{67}$$

the output of the neural network is constrained to the interval $[0,1]$ and can be interpreted as Bayesian probability [57].

The development of a neural network can be viewed as a two step procedure. In the first step, a supervised training is performed by using a sample of signal and background events to determine the weights that are used in the neural network. In this phase the classification of a given event is known to the algorithm. In an ideal case with perfect separation of the two hypotheses, *e.g.* signal and background, one hypothesis would give a response of 0, the other of 1. The training procedure uses an error function to estimate agreement of the network response for a given event $y_{NN,a}$ with the ideal response $\hat{y}_a \in \{0,1\}$,

$$E(\mathbf{x}_1,\ldots,\mathbf{x}_N\|\mathbf{w}) = \sum_{a=1}^{N} E_a(\mathbf{x}_a\|\mathbf{w}) = \sum_{a=1}^{N} \frac{1}{2}\left(y_{NN,a} - \hat{y}_a\right)^2. \tag{68}$$

In this context, the vectors $\mathbf{x}_a = (x_1,\ldots,x_{n_{var}})_a$ with $a = 1,\ldots,N$ represent the vectors of input variables for the N training events and $\mathbf{w}$ is the set of weights used in the neural network. Several training methods are available. In this study, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is used, cf. [58–61]. It is based on the minimisation of second derivatives of the error function to adapt the weights. This particular learning method is the recommended method, if a probabilistic interpretation of the output of the neural network is desired [57]. A detailed explanation of the algorithm can be found in [56].

After the first training step, in the second step, this training is tested by using a sample of known composition to which the trained neural network is applied. This step is necessary to check for possible overtraining and to make sure that the determination of weights from the first step has converged. Overtraining happens, when the neural network does not only exploit features of signal and background events which are general to the sample and every subset of it, but also such features which are specific to a given subset of events used in the training, ie. statistical fluctuations. In the third step the neural network is finally applied to a dataset of unknown composition to discriminate between the two hypotheses.

All of the above mentioned algorithms are implemented in the TMVA package [56] which is used for the studies presented in this work.

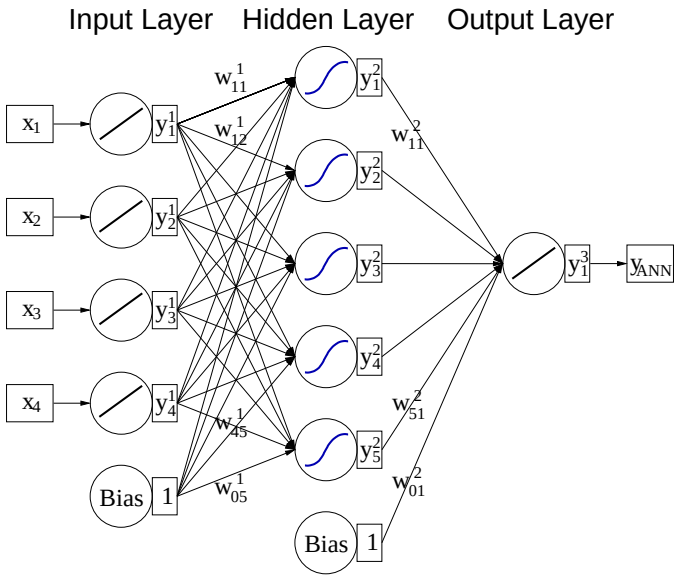


Figure 50: Structure of a simple Multi Layer Perceptron neural network with one hidden layer. The transformation functions shown in the representation of the neurons are for illustration only. Figure taken from [56]

7.2 SELECTION OF FRAGMENTATION TRACKS

The first neural network is used to separate fragmentation tracks from underlying event tracks. To distinguish between the two classes of tracks in the training process, the information from the MC generator on their origin is used in the training process.

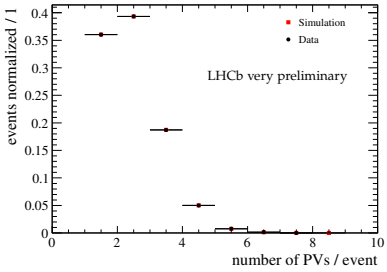
7.2.1 Neural network input variables

The input variables used for this neural network can be divided in four categories.

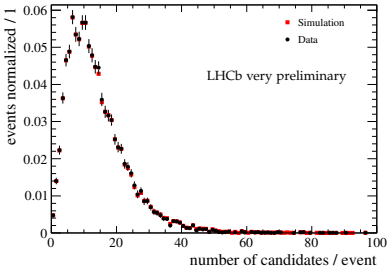
General event variables

These variables give some information about general properties of the event such as the number of primary vertices, the number of tagging track candidates after the preselection and the transverse momentum of the B_s^0 candidate. The number of tagging track candidates after the preselection is linked to the probability of a given candidate to originate from the fragmentation as the more tracks are available, the less probable it is to chose the fragmentation tracks. The same is true for the number of primary vertices as the number of tagging track candidates increases with a higher primary vertex multiplicity. The B_s^0 transverse momentum is linked by the fragmentation to the transverse momentum of the fragmentation tracks and is therefor correlated to the quality of the tagging track candidates.

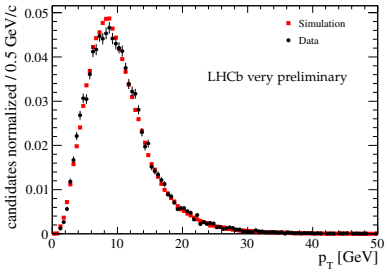
Figures 51a, 51b and 51c show comparisons of the distribution of all tagging track candidates between the data and the simulation after the corrections introduced in Chapter 6. Figure 52 shows the distribution of these event quantities for tagging track candidates originating from the fragmentation, the signal, and tagging track candidates coming from the underlying event, which are background in the neural net training.



(a) Number of PVs.

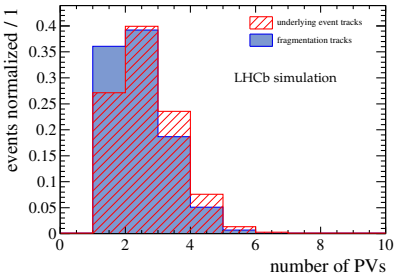


(b) Number of tagging track candidates.

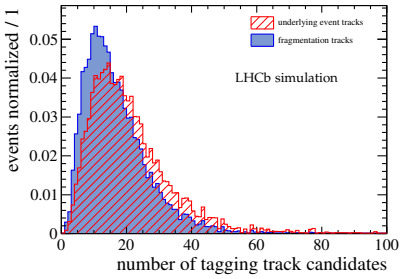


(c) B_s^0 candidate transverse momentum.

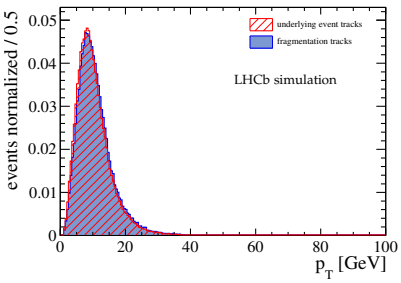
Figure 51: Comparison of distributions of general event variables for tagging track candidates in the data and the simulation. Plots are normalized to same number of events.



(a) Number of PVs.

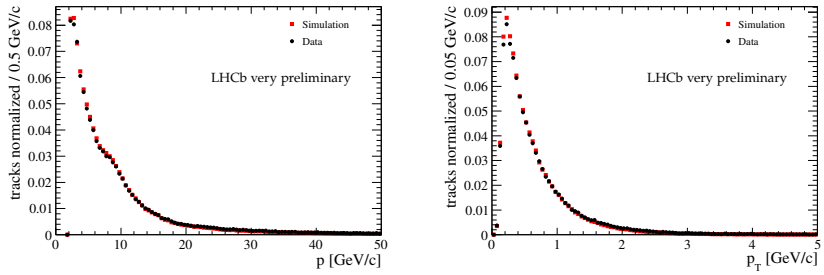


(b) Number of tagging track candidates.



(c) B_s^0 candidate transverse momentum.

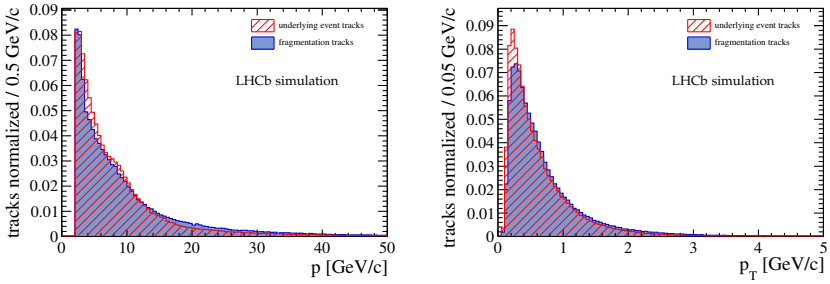
Figure 52: Comparison of distributions of general event variables for fragmentation tracks and underlying event tracks. Plots are normalized to same number of events.



(a) Momentum of tagging track candidates.

(b) Transverse momentum of tagging track candidates.

Figure 53: Comparison of distributions of kinematical variables for all tagging track candidates in the data and the simulation. Plots are normalized to same number of tagging track candidates.



(a) Momentum of tagging track candidates.

(b) Transverse momentum of tagging track candidates.

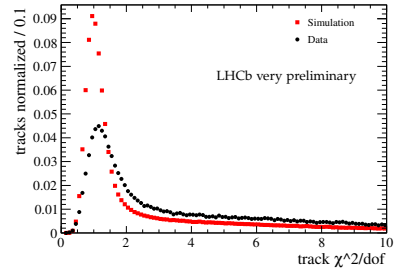
Figure 54: Comparison of distributions of kinematical variables for fragmentation tracks and underlying event tracks. Plots are normalized to same number of tagging track candidates.

Kinematical quantities

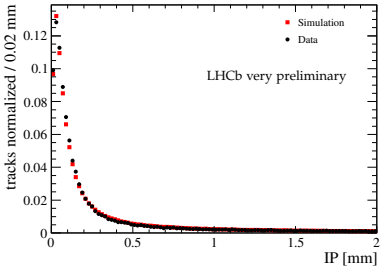
The momentum and transverse momentum are used to describe the kinematics of the tagging track candidate. Figures 53a and 53b show comparisons of the distribution of all tagging candidates between the data and the simulation after the corrections introduced in Chapter 6. Figure 54 shows the distributions for the fragmentation tracks and the underlying event tracks.

Quality variables

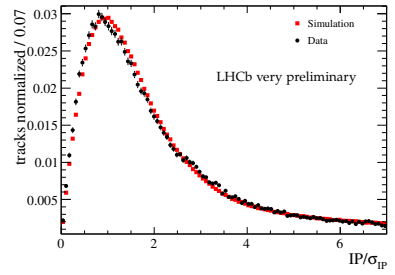
These variables describe the quality of the track reconstruction of the tagging candidate, ie. the $\chi^2_{\text{track}}/\text{ndf}$, as well as the agreement of the track with the hypothesis that it originates from the same primary vertex as the B^0_s, namely its impact parameter and the $\text{IP}/\sigma_{\text{IP}}$ to the primary vertex and the $\text{IP}_{\text{PU}}/\sigma_{\text{IP}_{\text{PU}}}$ to the nearest pile up vertex. The comparison of the distributions for tagging track candidates in the data and in the simulation for these variables are shown in Figures 55a, 55b, 55c and 55d, Figure 56 summarises the distributions for the fragmentation tracks and the underlying event tracks.



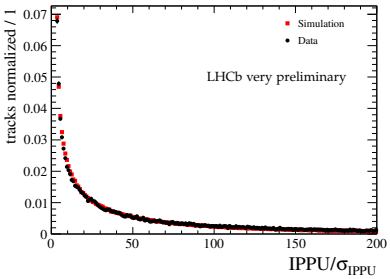
(a) $\chi^2_{\text{track}}/\text{ndf}$ of tagging track candidates.



(b) IP of tagging track candidates.

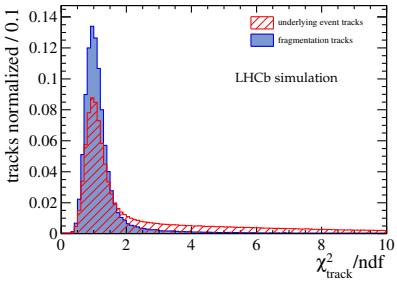


(c) $\text{IP}/\sigma_{\text{IP}}$ of tagging track candidates.

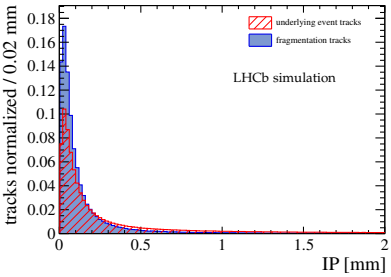


(d) $\text{IP}_{\text{PU}}/\sigma_{\text{IP}_{\text{PU}}}$ of tagging track candidates.

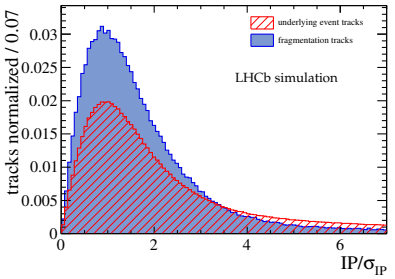
Figure 55: Comparison of distributions of quality variables for all tagging track candidates in the data and the simulation. Plots are normalized to same number of tagging track candidates.



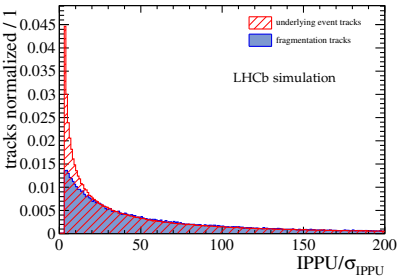
(a) $\chi^2_{\text{track}}/\text{ndf}$ of tagging track candidates.



(b) IP of tagging track candidates.



(c) $\text{IP}/\sigma_{\text{IP}}$ of tagging track candidates.



(d) $\text{IP}_{\text{PU}}/\sigma_{\text{IP}_{\text{PU}}}$ of tagging track candidates.

Figure 56: Comparison of distributions of quality variables for fragmentation tracks and underlying event tracks. Plots are normalized to same number of tagging track candidates.

Variables exploiting the fragmentation process

These are variables that exploit special features of the fragmentation process, as for example the distance between the B_s^0 candidate and the tagging track candidate in the phase space. The variables taken into account are:

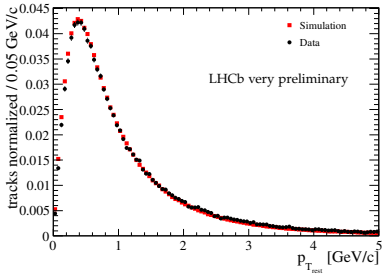
- The transverse momentum of the tagging track candidate in the B_s^0 rest frame.
- The transverse and longitudinal momentum relative to the B_s^0. The longitudinal momentum of the tagging track candidate relative to the B_s^0 is given by $p_{L_{rel}} = \vec{p}_{B_s^0} \cdot \vec{p}_{track} / \left|\vec{p}_{B_s^0}\right|$ and the relative transverse momentum by $p_{T_{rel}} = \sqrt{p_{track}^2 - p_{L_{rel}}^2}$.
- The distance in the polar angle ϕ and the pseudorapidity η between tagging track candidate momentum and the B_s^0 momentum as well as the radial distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.
- The difference between the invariant mass of the sum of the four momenta of the B_s^0 and the tagging track candidate and the invariant mass of the B_s^0, $\Delta Q = M_{B+K} - M_B$.
- The angle between the momenta of tagging track candidate and the B_s^0, $\rho = \arccos\left(\vec{p}_{B_s^0} \cdot \vec{p}_{track} / \left|\vec{p}_{B_s^0}\right| \cdot |\vec{p}_{track}|\right)$.
- The rank[1] in $\Delta\eta$, relative longitudinal and relative transverse momentum among all tagging track candidates in the event.

Figures 57 and 59 show the comparison of these distributions in the simulation and the data. In Figures 58 and 60 the distributions of background tracks and fragmentation tracks in the simulation are depicted.

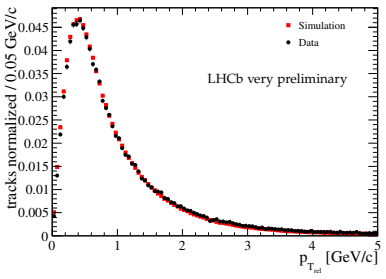
7.2.2 Training procedure

The aim of the training procedure is to find a set of variables and a configuration (ie. define a number of neurons and hidden layers) of the neural network to reach the best separation power between fragmentation tracks and underlying event tracks. In an ideal case, adding information to a neural network by using additional input variables or combining existing information in a new way with additional neurons or hidden layers inside the network can only increase the performance of the network. In reality however this increases the complexity of the network by requiring additional weights in the combination of the variables that have to be extracted from the training sample. A network with N input variables and one hidden layer with $N + M$ neurons has $(N + 1) \cdot (N + M)$ weights for the hidden layer and $(N + M + 1)$ weights for the output layer, thus adding additional input variables increases the number of weights by a noticeable amount. In a real world example such as the networks discussed in this thesis, with $N = 10$ input variables and one hidden layer with $2N - 1$ neurons, the total number of weights is 220 while for $N = 11$ it is 264 and for $N = 12$ it is 312. As there is only a finite number of training events available also the number of weights that can be determined from these events is limited and

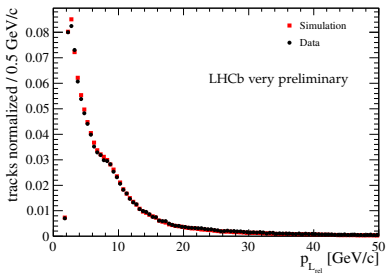
1 For the calculation of the rank of a tagging track candidate in a specific quantity, *e.g.* $\Delta\eta$, all tagging track candidates are sorted according to $\Delta\eta$. The tagging track candidate with the smallest $\Delta\eta$ has rank 0, the second smallest rank 1 and so on.



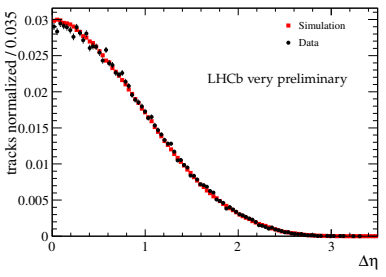
(a) Transverse momentum of tagging track candidate in the B_s^0 rest frame.



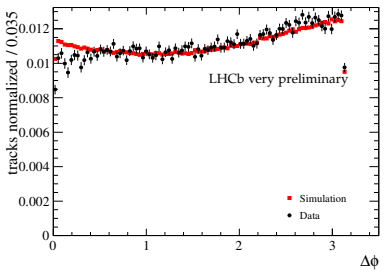
(b) Relative transverse momentum of tagging track candidate wrt. B_s^0.



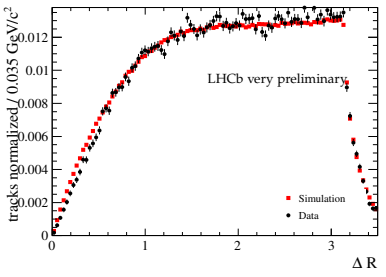
(c) Relative longitudinal momentum of tagging track candidate wrt. B_s^0.



(d) $\Delta\eta$ of tagging track candidates and B_s^0.

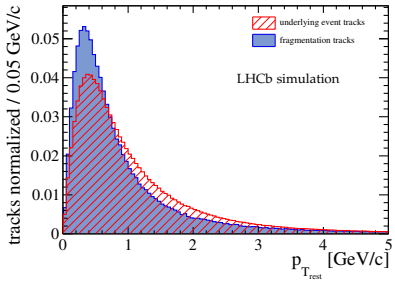


(e) $\Delta\phi$ of tagging track candidates and B_s^0.

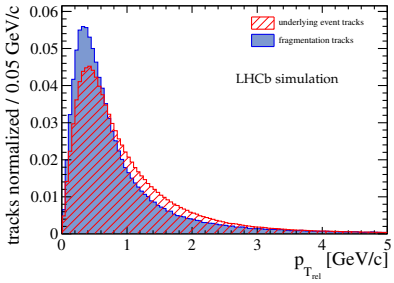


(f) ΔR of tagging track candidate wrt. B_s^0.

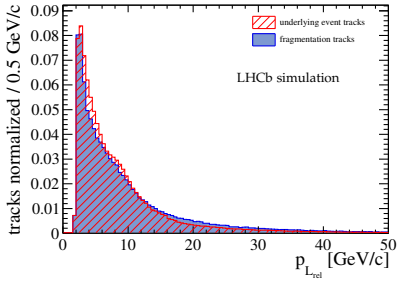
Figure 57: Comparison between data and simulation of tagging track quantities related to the fragmentation process. Plots normalized to the same number of tagging track candidates.



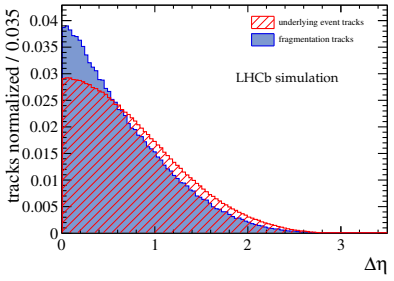
(a) Transverse momentum of tagging track candidate in the B_s^0 rest frame.



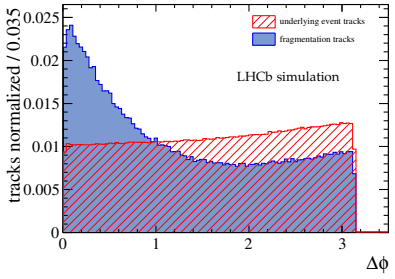
(b) Relative transverse momentum of tagging track candidate wrt. B_s^0.



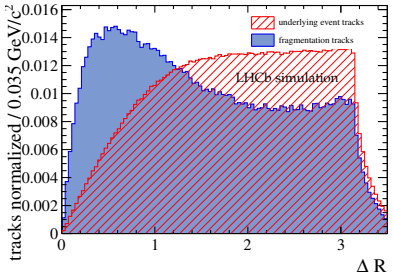
(c) Relative longitudinal momentum of tagging track candidate wrt. B_s^0.



(d) $\Delta\eta$ of tagging track candidates and B_s^0.

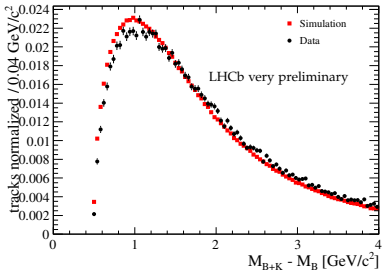


(e) $\Delta\phi$ of tagging track candidates and B_s^0.

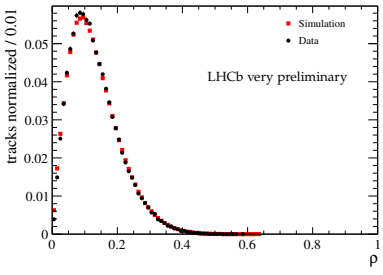


(f) ΔR of tagging track candidate wrt. B_s^0.

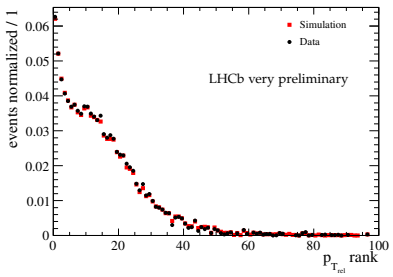
Figure 58: Comparison of distributions of fragmentation related variables for fragmentation tracks and underlying event tracks. Plots are normalized to same number of tagging track candidates.



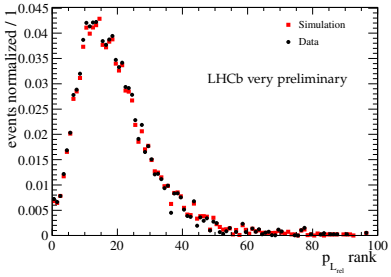
(a) $M_{B+K} - M_B$ of tagging track candidates.



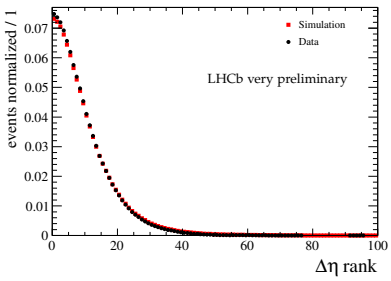
(b) Angle between the momenta of tagging track candidate and B_s^0.



(c) $p_{T_{rel}}$ rank of tagging track candidate.

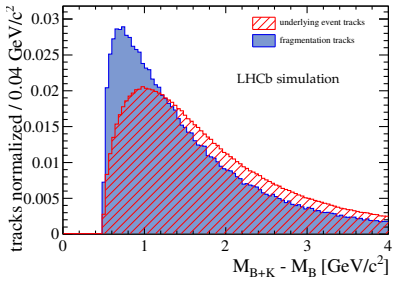


(d) $p_{L_{rel}}$ rank of tagging track candidate.

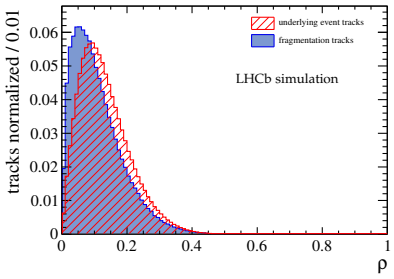


(e) $\Delta\eta$ rank of tagging track candidate.

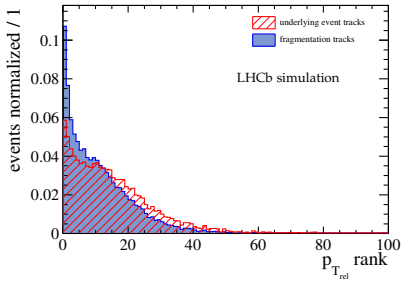
Figure 59: Comparison between data and simulation of further tagging track quantities related to the fragmentation process. Plots normalized to the same number of tagging track candidates.



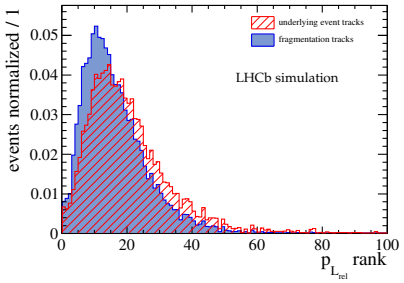
(a) $M_{B+K} - M_B$ of tagging track candidates.



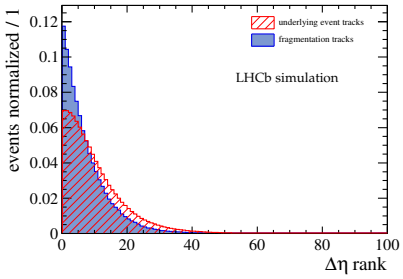
(b) Angle between the momenta of tagging track candidate and B_s^0.



(c) $p_{T_{rel}}$ rank of tagging track candidate.



(d) $p_{L_{rel}}$ rank of tagging track candidate.



(e) $\Delta\eta$ rank of tagging track candidate.

Figure 60: Comparison of distributions of further fragmentation related variables for fragmentation tracks and underlying event tracks. Plots are normalized to same number of tagging track candidates.

thus the separation power between signal and background that is reached by the neural network will converge or in the worst case even decrease by increasing the complexity.

A naive estimation is performed to asses the available statistics for the determination of the weights: In the given case, the training is performed with about 75000 fragmentation tracks. To not bias the training of the neural network it is important to use a comparable amount of underlying event tracks as background sample. Additional samples of about the same size are used for both signal and background for testing the neural network training. For a network as described above with $N = 10$ input variables about 340 events for both signal and background are left over to determine a given weight, which seems sensible. However, one has to keep in mind that an exact estimate of the necessary training statistics and the optimal complexity of the neural network classifier is not easy to asses as it depends also on the PDFs for signal and background in the dataset and other factors. In case there are too little events to determine the individual weights the training is very susceptible to overtraining or convergence problems. In this case, the network exploits no longer the general differences between signal and background but is sensitive to statistical fluctuations of the particular training sample.

To find an optimal set of input variables, the following procedure is applied in the training of the first neural network. The training starts with a given set of input variables that are derived from variables used in the cut based same side kaon tagging algorithm. These variables are the number of tagging candidates after the preselection, the number of primary vertices in the event, the transverse momentum of the B_s^0 candidate, the momentum and transverse momentum of the tagging track candidate, its IP/σ_{IP} to the B_s^0 production vertex, the χ^2_{track}/ndf from the track fit as well as the difference in the polar angle ϕ and the pseudorapidity η between tagging track candidate momentum and the B_s^0 momentum.

The difference in the polar angle ϕ has a very good separation power, cf. Figure 58e, therefore an additional cut is placed which requires the track to have a $\Delta\phi < 1.5$. This reduces the amount of fragmentation tracks by 39.6% and the amount of background tracks by 65.2% and helps the training of the neural network to concentrate on the more subtle differences between the two samples. Figure 61 shows the $\Delta\phi$ distribution of fragmentation kaons with the right charge for tagging and the wrong charge for tagging in simulated events. For $\Delta\phi > 1.5$ the kaons with the wrong charge dominate.

The selection of the input variables is then performed by an iterative procedure. In step 0 a neural network is trained with the aforementioned set of input variables and its separation power $\langle S^2 \rangle$ and the discrimination significance σ_S are calculated. The separation of the neural network output y is defined by the integral

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy, \tag{69}$$

where $\hat{y}_S(y)$ and $\hat{y}_B(y)$ are the signal and background PDFs of y. In the extreme cases $\langle S^2 \rangle$ is zero for identical signal and background shapes and one for distributions with no overlap, ie. a perfect separation. The discrimination significance is defined as

$$\sigma_S = \frac{\bar{y}_S - \bar{y}_B}{y^2_{S_{RMS}} + y^2_{B_{RMS}}}, \tag{70}$$

with the difference in the means $\bar{y}_S$ and $\bar{y}_B$ of the signal and background distribution of the neural network output divided by the quadratic sum of their root-mean-squares [56]. These quantities are determined by TMVA and calculated using the independent test sample.

In step 1 of the training procedure, the influence of all of the variables listed in Section 7.2.1 on the discrimination power of the neural network is evaluated. For each of the variables that are already used in step 0, new networks are trained that use all the input variables from the last step but the one under consideration. This checks if due to correlations among variables or limited statistics a better performance can be achieved by removing one of the variables. An overview of the separation power and discrimination of these $n-1$ configurations for step 0 is given in Table 18. These numbers give an indication of the importance of the individual variables.

For all the input variables that are not used in step 0, new networks are trained that use all the input variables from the previous step and the variable under consideration. This checks if by adding information, a better performance can be achieved. For each of the newly trained networks, the separation power $\langle S^2 \rangle$ is compared with the network of step 0. The configuration with the highest $\langle S^2 \rangle$ is the starting point for the next iteration step. If one or more networks have the same $\langle S^2 \rangle$ the one with the highest discrimination significance σ_S is taken. The procedure ends once no further improvement is measurable.

Table 19 summarises the results from the iterative training. The final configuration of the neural network contains the number of tagging candidates after the preselection, the number of primary vertices in the event, the transverse momentum of the B_s^0 candidate, the momentum and transverse momentum of the tagging candidate, its IP and IP/σ_{IP} to the B_s^0 production vertex, the χ^2_{track}/ndf from the track fit, the relative transverse momentum of the tagging candidate with respect to the B_s^0 candidate as well as the difference in the polar angle ϕ and the pseudorapidity η between tagging candidate momentum and the B_s^0 momentum. Figure 62a shows the response of the NN separately for fragmentation tracks and underlying event tracks for both the training and the test sample. A good discrimination is visible between the background peaking at zero and the signal. The distributions of the training sample and the test sample are in good agreement, which serves as an indication that the NN is not overtrained. Figure 62b shows the evolution of the neural network training over the number of training cycles. The estimator, which shows the agreement of the actual PDFs for signal and background with the desired response of the neural network is constant over a large number of training cycles which shows that the training procedure has converged. As the training procedure relies solely on simulated events, it is necessary to check the behaviour of the final neural network for real events. Figure 62c shows that the NN response for all tagging candidates in the simulation and in the data is very similar but shows some remaining differences that can be attributed to the differences in the input variables that were discussed in Chapter 6. The configuration described in this section will be the input to the further steps in the development of a neural network based same side kaon tagging algorithm.

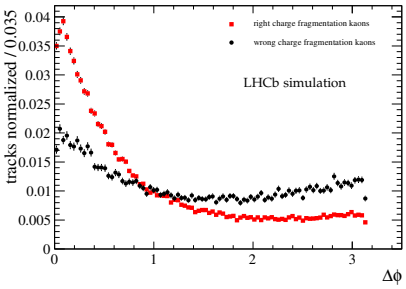
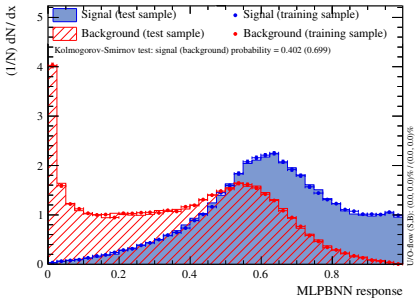
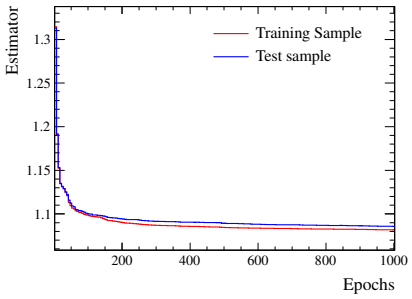


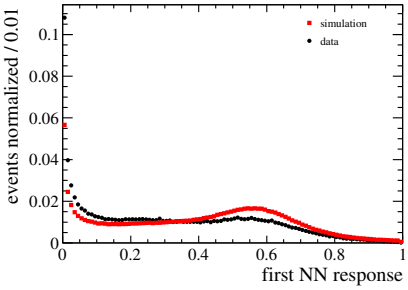
Figure 61: Distribution in $\Delta\phi$ of kaons coming from the B_s^0 fragmentation that have the same correct charge for tagging (red) or wrong charge (black) in simulated $B_s^0 \to D_s^- \pi^+$ event.



(a) Distribution of NN response for fragmentation tracks (signal) and UE tracks (background).



(b) Convergence of NN training.



(c) NN response in the data and the simulation.

Figure 62: Verification plots for the training of the first neural network. Shown are the response for fragmentation tracks and underlying event tracks in the simulation (a), the convergence of the training estimator (b) and the comparison of the response of the NN in the data and the simulation (c).

$n-1$ variable	$\langle S^2 \rangle$	σ_S
number of tracks	0.231	0.772
number of PVs	0.243	0.799
p_T B_s^0	0.243	0.799
p_T track	0.236	0.784
p track	0.239	0.791
$\Delta\eta$	0.231	0.772
$\Delta\phi$	0.226	0.762
IP/σ_{IP}	0.196	0.699
χ^2_{track}/ndf	0.195	0.697

Table 18: Separation power $\langle S^2 \rangle$ and separation significance σ_S of the first neural network, if one of the variables in the initial configuration of step 0 is left out (n-1).

step	network configuration	$\langle S^2 \rangle$	σ_S
0	number of tracks, number of PVs, p_T B_s^0, p_T track, p track, $\Delta\eta$, $\Delta\phi$, IP/σ_{IP}, χ^2_{track}/ndf	0.243	0.801
1	number of tracks, number of PVs, p_T B_s^0, p_T track, p track, $p_{T_{rel}}$, $\Delta\eta$, $\Delta\phi$, IP/σ_{IP}, χ^2_{track}/ndf	0.245	0.803
2	number of tracks, number of PVs, p_T B_s^0, p_T track, p track, $p_{T_{rel}}$, $\Delta\eta$, $\Delta\phi$, IP/σ_{IP}, IP, χ^2_{track}/ndf	0.251	0.815

Table 19: Network configuration, separation power $\langle S^2 \rangle$ and separation significance σ_S for different steps in the iterative training of the first neural network.

7.3 COMBINATION OF TAGGING TRACK CANDIDATES AND ESTIMATION OF MISTAG PROBABILITY

The neural network that is presented in the previous section distinguishes tracks originating from the fragmentation of the B_s^0 candidate from other tracks in the event. The next step is to combine the potential fragmentation tracks to form a tagging decision. Several options to do so come to mind, of which the simplest and most straight forward would be to simply take the tagging track candidate with the highest probability to be a fragmentation track to determine the flavour. This however has several pitfalls. As stated in Chapter 4.6.1 it is possible that more than one track was produced in the fragmentation process, thus by choosing only one track it is not guaranteed to arrive at the correct decision even if the track was really coming from the fragmentation. Furthermore it is also desirable to have an estimate of the probability of the tag being wrong. This is not to be confused with the output of the first neural network which does not contain charge information and thus does not give any information about the flavour.

Instead, a different approach is chosen to account for these problems. All tagging track candidates are sorted according to the output from the first neural network. To cut away obvious background tracks and to ensure a low correlation between the same side kaon tagger and the opposite side taggers, especially the OS kaon tagger[2], it is required additionally that the output of the first neural network for a tagging track candidate is larger

2 The possible correlations of the SSK tagger with the OS taggers is discussed in detail in Chapter 8.

than 0.65. This requirement is chosen in such a way, that for about 50% of the B_s^0 signal candidates in data there is at least one tagging track candidate passing it. Figure 63 shows the number of tagging track candidates per event that survive these cuts for all B_s^0 candidates in the data and the simulation. On average there are 1.42 tracks per event left over in the data and 1.57 tracks in the simulation.

Based on this selection, multiple tracks are then combined in a second neural network with their response from the first neural network (called nn_1 in the following) and the difference in the logarithmic likelihood of the particle to be consistent with the hypothesis of being a kaon versus a pion, $DLL_{K-\pi}$ as input variables. To gain information about the flavour of the B_s^0 candidate, these quantities are multiplied with the charge of the track, ie. $q \times DLL_{K-\pi}$ and $q \times nn_1$. The resulting input variables are shown in Figure 64.

General event properties might influence the probability of a track to originate from the fragmentation and likewise lead to a different mistag probability. In an event with, *e.g.*, 10 tracks in addition to the B_s^0 decay and one primary vertex it is more likely to choose correctly the one fragmentation kaon than it is in an event with a higher multiplicity. To ensure that this is reflected in the predicted mistag probability the second neural network includes in addition the number of tagging track candidates that survive the cut on the first neural network, the number of primary vertices and the transverse momentum of the B_s^0 candidate. These variables have no predictive power on the flavour of a B_s^0 candidate but via the correlations of these variables with the tagging track quantities, the neural net can learn to give the right mistag probability, taking into account the general structure of the event.

The second neural network is trained with simulated B_s^0 candidates and about 80000 B_s^0 and $\overline{B}_s^0$ each are used in the training process. To increase the available training statistics in order to make it easier for the neural network to learn the discrimination between the two B_s^0 flavours, the relative amount of training to test statistics has been changed in such a way that 90% (about 72000) of the candidates are used for the training and 10% (about 8000) are used for testing. This leads to a larger statistical uncertainty in the test sample but a more stable learning process.

Tests with different input variables for the second neural network have been performed, considering only the best and, if available, up to four additional tagging track candidates. If multiple input tracks are considered, the input variables are set to zero if less tracks are available[3]. Table 20 summarises the separation power and the discrimination significance of the different neural network configurations, showing that configurations with two, three and four input tracks perform very similar. The configuration with up to three input tracks was chosen because it features the best discrimination significance on all events.

Figure 65a shows the response of the second neural network for candidates produced as B_s^0 and candidates produced as $\overline{B}_s^0$ for both the training and test sample. The distributions feature a good discrimination between the two flavours especially in the tails. Figure 65b shows the evolution of the neural network training over the number of training cycles. The offset between the test and the training sample can be attributed to the larger sta-

3 As an example, if up to three tagging track candidates are considered in the specific neural network configuration, $q \times nn_1$ and $q \times DLL(K-\pi)$ are set to zero for the second and third track in events with only one tagging track candidate.

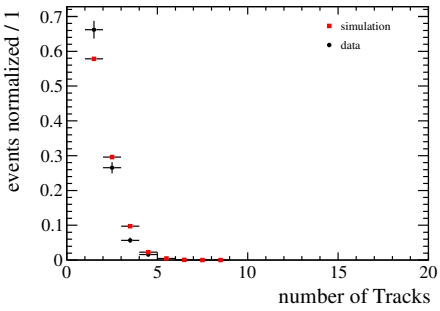


Figure 63: Comparison of the number of tagging track candidates after the cut on the output of the first NN in the data and the simulation. Plot normalized to the same number of events.

tistical error of the test sample due to the splitting, however the test sample performs more optimal than the training sample. The plot shows that the training has converged and is stable.

For the cut based SSK tagger, as described in Chapter 4.6.3, the final tagging decision was derived from the charge of the tagging track candidate with the highest transverse momentum after the selection cuts. In contrast to that, the tagging decision for the neural network based SSK tagger is derived based on the response of the second neural network (called nn_2 in the following). Figure 65a shows that candidates with an response larger than 0.5 have more likely been produced as B_s^0 while those with an response smaller 0.5 have more likely been $\bar{B}_s^0$. Consequently, the particle is assumed to be produced as

- B_s^0 if $nn_2 > 0.5$ (tag $+1$) and
- $\bar{B}_s^0$ if $nn_2 < 0.5$ (tag -1).

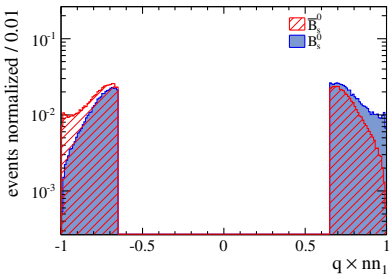
Likewise, the predicted mistag probability is derived from the response of the second neural network. The predicted mistag probability η is given as

- $\eta = 1 - nn_2$ for candidates tagged as B_s^0 and
- $\eta = nn_2$ for candidates tagged as $\bar{B}_s^0$.

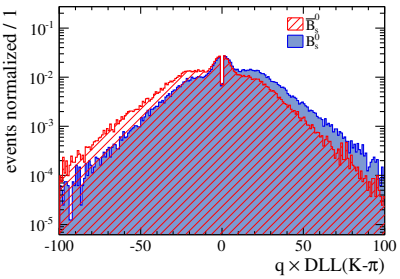
It has to be shown that the so defined mistag probability is correctly defined and does indeed reflect the probability of the tag of a given candidate to be wrong. The validity of this approach will be shown in the following section for simulated B_s^0 candidates. The application of the neural network based same side kaon tagger to the data will be discussed in Chapter 8.

network configuration	$\langle S^2 \rangle$	σ_S
one track	0.063	0.365
two tracks	0.067	0.374
three tracks	0.067	0.377
four tracks	0.067	0.373
five tracks	0.066	0.373

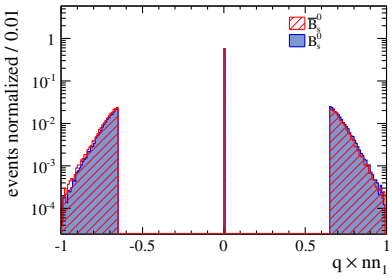
Table 20: Separation power $\langle S^2 \rangle$ and separation significance σ_S for different configurations of the second neural network.



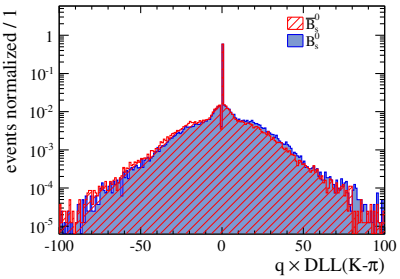
(a) $q \times nn_1$ best tagging track candidate.



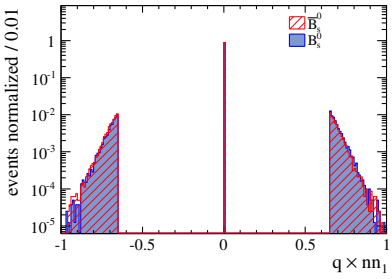
(b) $q \times DLL(K-\pi)$ best tagging track candidate.



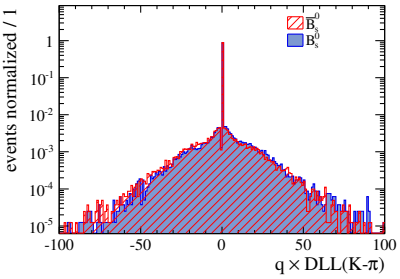
(c) $q \times nn_1$ 2^{nd} best tagging track candidate.



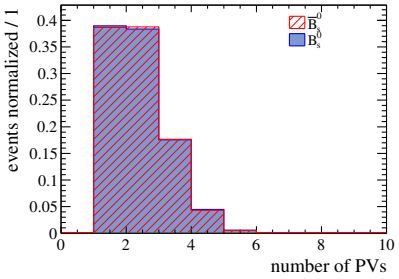
(d) $q \times DLL(K-\pi)$ 2^{nd} best tagging track candidate.



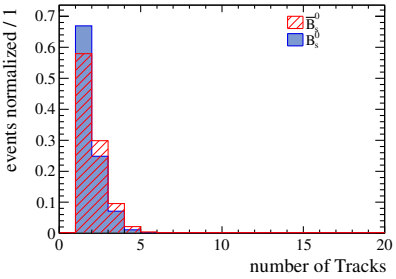
(e) $q \times nn_1$ best tagging track candidate.



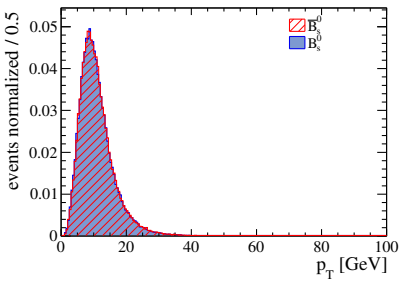
(f) $q \times DLL(K-\pi)$ 3^{rd} best tagging track candidate.



(g) Number of PVs.

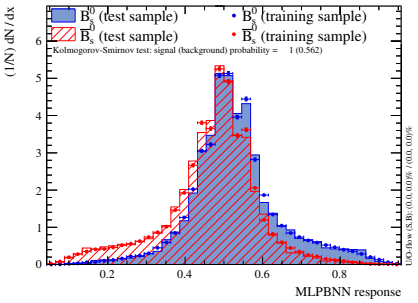


(h) Number of tagging track candidates.

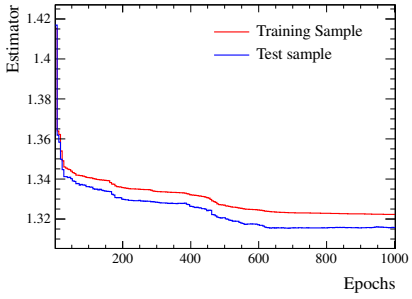


(i) B_s^0 candidate transverse momentum.

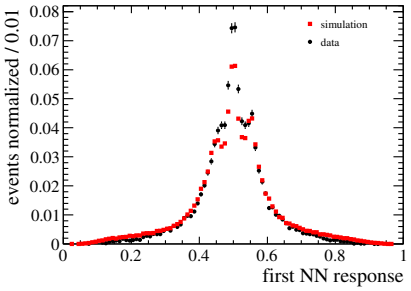
Figure 64: Input variables to the training of the second neural network comparing the distributions for B_s^0 and $\bar{B}_s^0$ candidates. Plots normalized to the same number of events.



(a) Distribution of NN response for B_s^0 and $\bar{B}_s^0$ events.



(b) Convergence of NN training.



(c) NN response in the data and the simulation.

Figure 65: Verification plots for the training of the second neural network. Shown are the response for B_s^0 and $\bar{B}_s^0$ in the simulation (a), the convergence of the training estimator (b) and the comparison of the response of the NN in the data and the simulation (c).

7.4 PROBABILISTIC INTERPRETATION OF THE NEURAL NET OUTPUT

As stated in the last section, the output of the second neural network is interpreted as a probability and the predicted mistag probability η is derived from it as defined above. This interpretation is correct, as the output of the neural network is normalized in such a way that it represents the probability of an event with a given response nn_2 to be signal (to be a B_s^0 in the case discussed here), *i.e.*

$$nn_2 = \frac{N_{B_s^0}(nn_2)}{N_{B_s^0}(nn_2) + N_{\bar{B}_s^0}(nn_2)}, \tag{71}$$

where $N_{B_s^0}(nn_2)$ and $N_{\bar{B}_s^0}(nn_2)$ is the number of B_s^0 and $\bar{B}_s^0$ candidates with a given response nn_2. This dependence is also shown in Figure 66a. Accordingly, the predicted mistag probability for all events can be viewed as the total probability to get a certain η if it is not known whether the particle is a B_s^0 or $\bar{B}_s^0$,

$$P(\eta | B_s^0 \cup \bar{B}_s^0) = P(\eta)\,. \tag{72}$$

As the tagging is to be used in CP analyses it is desirable that the predicted mistag probability can also be interpreted for a sample that only contains B_s^0 or $\bar{B}_s^0$. In this case, however, asking for the mistag probability η' leads to a conditional probability,

$$P(\eta' \cap B_s^0) = P(\eta' | B_s^0) \cdot P(B_s^0) \tag{73}$$

and

$$P(\eta' \cap \bar{B}_s^0) = P(\eta' | \bar{B}_s^0) \cdot P(B_s^0)\,. \tag{74}$$

Consequently, a total mistag probability can be calculated if both $P(\eta' \cap B_s^0)$ and $P(\eta' \cap \bar{B}_s^0)$ are known,

$$P(\eta') = P(\eta' \cap B_s^0) + P(\eta' \cap \bar{B}_s^0) = P(\eta' | B_s^0) \cdot P(B_s^0) + P(\eta' | \bar{B}_s^0) \cdot P(B_s^0)\,. \tag{75}$$

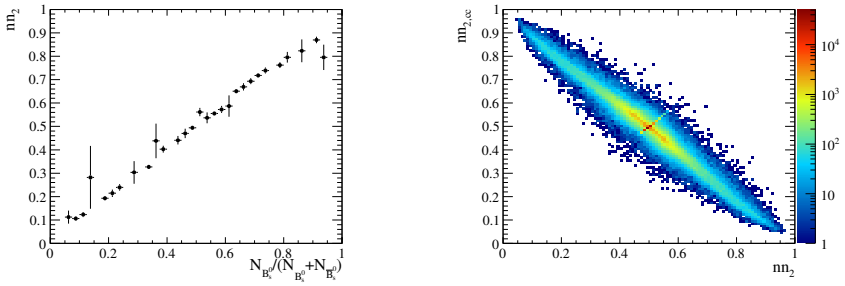
Assuming the same amount of B_s^0 and $\bar{B}_s^0$, one can set

$$P(B_s^0) = P(\bar{B}_s^0) = \frac{1}{2}, \tag{76}$$

and hence follows

$$P(\eta') = \frac{P(\eta' | B_s^0) + P(\eta' | \bar{B}_s^0)}{2}\,. \tag{77}$$

For a specific candidate, the conditional probabilities for a given η' under the assumption of the particle beeing a B_s^0 or $\bar{B}_s^0$, *i.e.* $P(\eta' | B_s^0)$ and $P(\eta' | \bar{B}_s^0)$,



(a) Probabilistic interpretation of nn_2.

(b) nn_2 vs. $nn_{2,cc}$.

Figure 66: Verification plots for the probabilistic interpretation of the second neural network output. Shown are the response of the neural network nn_2 vs the fraction of B_s^0 (signal) events for the given output (a) and the comparison of the second neural network output nn_2 with the output if the input variables are charge conjungated, $nn_{2,cc}$ (b).

can be evaluated by flipping the charge signs of the input variables of the second neural network. In this case Equation 77 simplifies to

$$P(\eta\prime) = \frac{\eta + (1-\eta_{cc})}{2}\,, \tag{78}$$

where η_{cc} stands for the predicted mistag probabilities with the charge conjugated input variables.

It should be noted that this calculation is necessary in case there is a difference in the behaviour of the neural network for the two flavours expected, *i.e.* $P(\eta\prime|B_s^0) \neq P(1-\eta\prime|\bar{B}_s^0)$, which manifests itself in an asymmetry for the PDFs of the neural network output for B_s^0 and $\bar{B}_s^0$, see Figure 65a. In case of the same side kaon tagger, such an asymmetry is expected due to the different reconstruction efficiencies for positive and negative charged kaons. In case the output of the neural network is symmetric for the two flavours, the joint probability becomes

$$P(\eta\prime \cap B_s^0) = P(\eta\prime) \cdot P(B_s^0) \tag{79}$$

and Equation 75 simplifies to

$$P(\eta\prime) = P(\eta\prime) \cdot P(B_s^0) + P(\eta\prime) \cdot P(\bar{B}_s^0) = P(\eta\prime) \cdot \frac{1}{2} + P(\eta\prime) \cdot \frac{1}{2} = P(\eta\prime)\,. \tag{80}$$

The redefined $\eta\prime$, *cf.* Equation 78, is used for all further calculations. As a crosscheck, Figure 66b shows the dependence of the charge sign conjugated neural network response to the normal neural network response for all B_s^0 and $\bar{B}_s^0$ candidates. The dependence is linear, meaning that applying the formalism discussed in this section there is no change in the calibration or tagging performance on the whole sample.

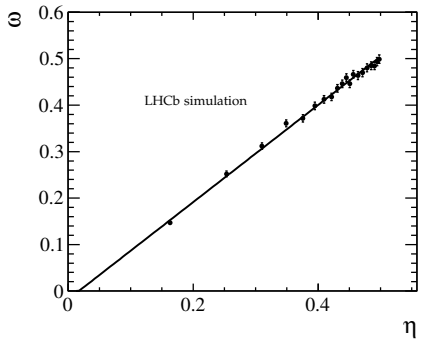


Figure 67: True mistag fraction ω in bins of the predicted mistag probability η in simulated events. The solid line is the result of a fit of the calibration function to the data points.

7.5 APPLICATION OF THE NEURAL NETWORK SAME SIDE KAON TAGGING ALGORITHM ON SIMULATED EVENTS

To check if the neural network same side kaon tagging algorithm developed in this chapter gives the correct tagging decision and predicted mistag probability, it is applied to the sample of simulated events. The whole sample is split into 20 categories according to the predicted mistag η of the candidates. The categories are chosen in such a way that they feature the same number of B_s^0 signal events. In each of the categories the true mistag fraction ω is determined by using the MC information on the true flavour of the reconstructed B_s^0 meson and counting the number of right and wrong decisions. Figure 67 shows the mean of the predicted mistag η in each of the bins versus the true mistag fraction ω in this bin. The plot is fitted with the linear function

$$\omega_{\text{true}} = p_0 + p_1 \times (\eta - \langle\eta\rangle) \tag{81}$$

where p_0 and p_1 are the calibration parameters that are left floating in the fit and $\langle\eta\rangle$ is the average predicted mistag probability of all events. As the second neural network, from which the predicted mistag η is derived, was trained on the same sample, the calibration should give the ideal values of $p_0 = \langle\eta\rangle$ and $p_1 = 1$. This assumption is confirmed, see Table 21 for a list of the fitted values.

The tagging efficiency in the simulated sample is $\varepsilon_{tag} = (57.82 \pm 0.10)\%$, the effective mistag fraction $\omega_{eff} = (37.08 \pm 0.17)\%$ and the effective tagging power $\varepsilon_{eff} = (3.86 \pm 0.10)\%$. Compared to the default cut based same side kaon tagger in the simulation this is an improvement of $\Delta\varepsilon_{eff} = (1.95 \pm 0.11)\%$.

parameter	result
p_0	0.415±0.002
p_1	1.04±0.02
$\langle\eta\rangle$	0.414

Table 21: Fit results from the calibration of the output of the second NN in the simulation.

8 APPLICATION OF THE NEURAL NETWORK BASED SAME SIDE KAON TAGGING ALGORITHM TO THE DATA

Despite the corrections that are applied to the simulation, as explained in Chapter 6, some differences between the data and the simulation persist. It is therefore necessary to apply the neural network based same side kaon tagging algorithm, that was described in the previous chapter, to the data in order to measure its performance and calibrate the predicted mistag fraction. This is done in this chapter by using the fit to the B_s^0 oscillation that has been described in Chapter 5.5. Several sources of systematic uncertainties and biases on the calibration parameters that are obtained in this way are studied and the applicability of this calibration in other decay modes is checked. To obtain the best possible tagging efficiency, the same and the opposite side tagging algorithms have to be combined. The chapter concludes with the description of the performance and the calibration of this combination and systematic studies concerning the correlation between same and opposite side tagging algorithms.

8.1 CALIBRATION OF THE PREDICTED MISTAG PROBABILITY

The calibration parameters of the predicted mistag probability of the same side kaon tagging algorithm are extracted in two ways. The parameters can be, on the one hand, determined directly in the unbinned maximum likelihood fit by using the calibration function as described in Equation 51 as mistag fraction in the signal PDF. The calibration parameters p_0 and p_1 are floating in the fit, the predicted per event mistag probability η is given by the neural network of the same side kaon tagging algorithm as described in the previous chapter and the average predicted mistag $\langle\eta\rangle$ is determined from all sideband subtracted B_s^0 signal candidates.

Alternatively, instead of the calibration function, an average mistag fraction ω is used in the unbinned fit to the data. The sample is binned with increasing predicted mistag probability η in six bins of the predicted mistag probability from the same side kaon tagging algorithm and the average mistag fraction is derived in each of these categories. The average predicted mistag probability in each of the bins is taken as bin center. A linear fit of the calibration function is then performed to the so obtained data points.

The two methods should give identical results. The first method, that determines the calibration parameters directly from the unbinned fit, is intrinsically more precise as it uses all the available information. This method is used as default for the determination of the calibration parameters and the tagging performance. The second method however gives a measure for the linearity of the calibration, as this can be checked by comparing the data points to the linear calibration function. The difference between the two fit methods is taken as a systematic uncertainty.

	p_0	p_1	$\langle\eta\rangle$
unbinned fit	0.408±0.009	0.86 ± 0.12	0.430 (fixed)
fit in bins	0.402± 0.009	0.77±0.11	0.430 (fixed)
unbinned fit, calibration applied	0.408±0.009	1.00±0.14	0.408 (fixed)

Table 22: Results of the fit of the calibration parameters for an direct unbinned fit of p_0 and p_1 and of a fit in bins of the predicted mistag probability η. The last line shows the result for the fit if the calibration is already applied.

	$\epsilon(\%)$	$\omega(\%)$	$\epsilon D^2(\%)$	$\langle\eta\rangle(\%)$
$0 \leqslant \eta < 0.358$	6.9±0.2	29.1±1.9	1.21±0.28	29.415±0.111
$0.358 \leqslant \eta < 0.416$	7.1±0.2	35.9±2.4	0.56±0.20	39.164±0.031
$0.416 \leqslant \eta < 0.442$	6.8±0.2	41.2±2.5	0.21 ±0.12	43.055±0.014
$0.442 \leqslant \eta < 0.458$	6.7±0.2	43.5±2.5	0.11±0.09	45.001±0.008
$0.458 \leqslant \eta < 0.476$	7.2±0.2	41.4±2.5	0.21±0.13	46.690±0.010
$0.476 \leqslant \eta$	14.8±0.3	44.9±1.6	0.16±0.10	48.987±0.008
sum of categories	49.4±0.5	40.3±5.8	2.46±0.41	-

Table 23: Results of the fit for the average mistag fraction ω in bins of the predicted mistag probability η.

The results for the two calibration methods are summarized in Table 22. They are in agreement with each other. The results for the fit for an average mistag fraction in bins of the predicted mistag probability is summarized in Table 23. The datapoints from these fits, overlaid with the calibration functions for the two calibration methods, are shown in Figure 68. The distribution of the predicted mistag probability η before and after the calibration is shown in Figure 69.

Once the calibration has been obtained, it can be applied to the predicted mistag probability from the neural network and again a fit for the calibration parameters can be performed as a validation. The calibration should then yield the ideal values, that is $p_0 = \langle\eta\rangle$ and $p_1 = 1$. The results of this check for the nominal calibration parameters derived from the unbinned fit are also summarized in Table 22 and agree with the expectation.

With the calibrated predicted mistag probability, the tagging performance of the tagging algorithm can be derived. This is done by summing the calibrated predicted mistag probability over all sideband subtracted B_s^0 signal candidates, as explained in Chapter 4.2. The so obtained effective tagging power is $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2 = 2.42 \pm 0.39\%$ with a tagging efficiency of $\varepsilon_{\text{tag}} = 49.5 \pm 0.4\%$, which corresponds to an increase of about 56% in effective tagging power compared with the previous cut based tagging algorithm, *cf.* Chapter 4.6.3.

8.2 SYSTEMATIC STUDIES

The sources of systematic uncertainties and biases on the determination of the calibration parameters that are studied in the following section can be coarsely divided into three categories.

- **Systematic effects due to charge and detector asymmetries:** These effects originate from the fact that particles and anti particles interact

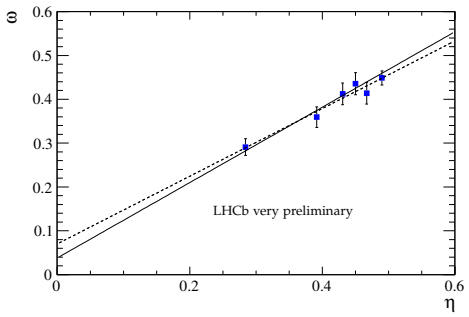


Figure 68: Average mistag fraction ω in bins of predicted mistag probability η. The solid line is the result of the unbinned fit for the calibration parameters p_0 and p_1. The dashed line is the result of a linear fit to the data points.

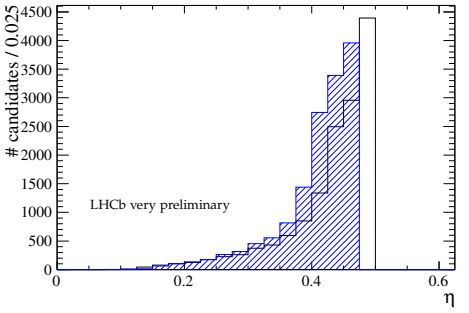


Figure 69: Distribution of the predicted mistag probability from the neural network based same side kaon tagging algorithm for sideband subtracted $B_s^0 \to D_s^- \pi^+$ signal candidates before (black line) and after (blue shaded area) calibration.

differently in the detector due to, *e.g.* different cross sections for material interaction. Furthermore, they are bend into different halves of the detector by the magnetic field. As the detector material is not distributed symmetric, the magnetic field can introduce additional asymmetries. Both effects can influence the tagging algorithms and by doing so unwanted asymmetries can be introduced in physics analyses. The following effects are studied in this regard: Dependence of the calibration parameters on the initial flavour, tagging decision and magnetic field polarity.

- **Systematic effects from the analysis procedure:** The extraction of the tagging performance depends on a reasonable understanding of the decay time behaviour, accordingly the choice of the resolution model or uncertainties in the determination of the scale factor for the decay time error can influence the calibration parameters. Also, choices in the fit procedure such as the parametrization of the mass PDF and asymmetries in the tagging of the background might have a small influence.

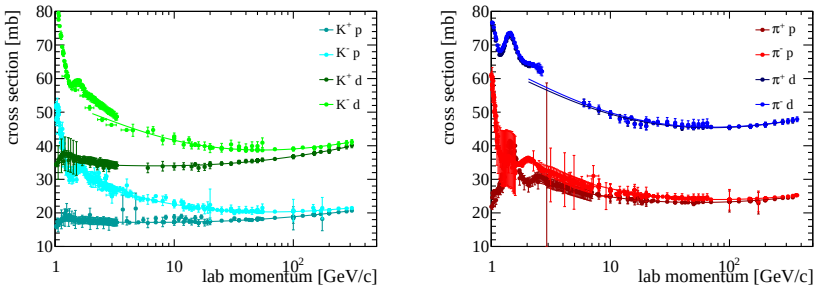
- **Systematic effects caused by the event topology and running conditions:** Quantities like the multiplicity in the event, number of reconstructed primary vertices and the B_s^0 momentum influence the tagging performance and can in principle also influence the calibration parameters. Furthermore different running conditions, such as instantaneous luminosity, present in different run periods could influence the calibration. Although they are already taken into account in the training of the second neural network, a check for remaining dependencies is performed.

8.2.1 Influence of the K^+/K^- detection asymmetry on the calibration

It is well established that charged kaons have a momentum dependent charge asymmetric cross section with matter [6]. The result of such a cross section measurement depending on the particle momentum performed by the COMPAS group is shown in Figure 70. Compared to the cross section asymmetry for $\pi^\pm$, large differences can be observed between K^+ and K^-. The tagging candidates have an average momentum of around 9 GeV. At this momentum, the cross section for K^- with protons is about 35% larger than for K^+ and the difference for $K^\pm$ with deuterium is about 24%. Up to the tracking stations, the material a kaon has to traverse is of the order of 0.15 to 0.2 hadronic interaction lengths. Detection asymmetries between K^+ and K^- of up to 10% are therefore expected.

As the same side kaon tagging algorithm uses charged kaons for the determination of the tagging decision, this effect can lead to a difference in tagging performance depending on the initial flavour of the B_s^0. If the K^+ from the B_s^0 fragmentation is less likely to be absorbed than the K^- from the $\bar{B}_s^0$ fragmentation, the mistag fraction should be higher for the $\bar{B}_s^0$ than for the B_s^0.

If the influence of this kaon detection asymmetry on the calibration of the predicted mistag probability is known it can be used as correction in the physics analyses. As the initial flavour of the B_s^0 is not directly accessible in the data it can not be used for the individual determination of the calibration parameters for B_s^0 and $\bar{B}_s^0$, however, the two flavours can in principle be separated statistical in the unbinned likelihood fit. As the current statistics



(a) Kaon cross section

(b) Pion cross section

Figure 70: Momentum dependent cross section of charged $K^{\pm}$ and $\pi^{\pm}$ with protons and deuterium determined by the COMPAS experiment. Figure taken from [6].

for this separation is too low, two different methods are used to estimate the effect of the $K^{\pm}$ detection asymmetry on the calibration instead, using simulated events and prompt D_s^+ in the data.

K^+/K^- *asymmetry in simulated events*

In order to extract the impact of the different material interaction on the tagging algorithms in the simulation, one has to check first if the charge dependent absorption of the kaons is modeled correctly. A data driven method of determining the kaon asymmetry in the LHCb detector has been performed, which is in detail described in [62], using $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ and $D^{\pm} \to K_S^0\pi^{\pm}$ decays.

The combined kaon pion asymmetry is given by

$$A_{K\pi} = \frac{\epsilon_{K^+\pi^-} - \epsilon_{K^-\pi^+}}{\epsilon_{K^+\pi^-} + \epsilon_{K^-\pi^+}}, \tag{82}$$

with the efficiencies measured as

$$\frac{\epsilon_{K^+\pi^-}}{\epsilon_{K^-\pi^+}} = \frac{N(D^- \to K^+\pi^-\pi^-)}{N(D^+ \to K^-\pi^+\pi^+)} \times \frac{N(D^+ \to K_S^0\pi^+)}{N(D^- \to K_S^0\pi^-)}, \tag{83}$$

to account for the additional asymmetry from the two charged pions. As the pion asymmetry is small compared to the kaon asymmetry it can be neglected. The results of this measurement are shown as function of the kaon momentum in Figure 71. From the comparison of the measured K^+/K^- asymmetry in the data and the simulation, it can be seen that the kaon asymmetry is underestimated in the simulation. In order to correct this underestimation in the simulated B_s^0 sample, negative kaons are removed on a random basis in the tagging process. The momentum dependence of the K^+/K^- asymmetry is taken into account in the correction. It is ensured that the simulated sample contains equal amounts of magnet up and down events such that magnetic field effects are averaged out.

After this correction, the sample of simulated $B_s^0 \to D_s^-\pi^+$ events is split according to the initial flavour of the B_s^0 mesons and the calibration of the predicted mistag probability in both subsamples is independently determined. The calibration is plotted in Figure 72 and the results of the difference between both subsamples are listed in Table 24. The calibrations result in $\Delta p_0 = p_{0,B_s^0} - p_{0,\bar{B}_s^0} = -0.020 \pm 0.004$ and $\Delta p_1 = p_{1,B_s^0} - p_{1,\bar{B}_s^0} = -0.01 \pm 0.03$, which corresponds as expected to a higher mistag probability

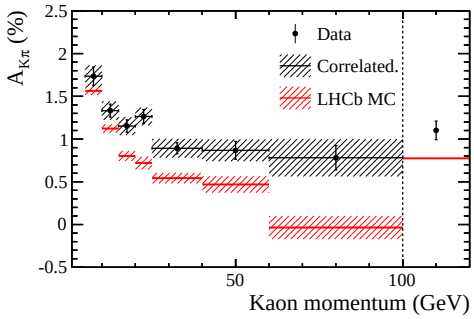


Figure 71: Measured $K^{\pm}\,\pi^{\mp}$ asymmetries as a function of kaon momentum, compared to predictions from MC (red line). The last bin shows the average of the asymmetry over the whole momentum range. Figure taken from [62].

for particles tagged by negative kaons than for particles tagged by positive kaons.

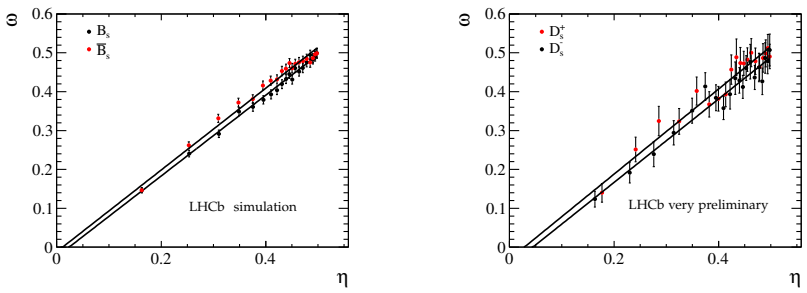
Influence of K^+/K^- asymmetry from prompt D_s^+ meson decays

Another possibility to determine the influence of the kaon detection asymmetry on the calibration of the predicted mistag probability is to study the tagging behaviour in prompt D_s^+ meson decays. As D_s^+ mesons, like B_s^0 mesons, contain a strange quark they show the same fragmentation behaviour as discussed in Chapter 4.6.1 for the B_s^0 mesons. The different masses of charm and bottom quarks can, however, cause differences in the kinematic properties of the fragmentation particles, therefor this study is used only as a cross check. In contrast to the neutral B_s^0 meson, the D_s^+ meson is charged and therefore does not oscillate, consequently its production and decay flavour are identical.

A sample of 100.000 prompt $D_s^+ \to \phi\pi^+$ decays are used to measure the calibration parameters for the two different production flavours of D_s^- (corresponding to B_s^0) and D_s^+ (corresponding to $\bar{B}_s^0$). The D_s^+ transverse momentum is reweighted to match the observed B_s^0 transverse momentum distribution. This is done to account for some of the differences that are expected in the kinematic distributions of the fragmentation kaons. As the initial flavour is known, the calibration is performed in the same way as for the $B_s^0 \to D_s^-\pi^+$ decays in the simulation using the neural network trained on the simulated $B_s^0 \to D_s^-\pi^+$ sample. The calibration is plotted in Figure 72 and the results are summarized in Table 24. The calibrations result in $\Delta p_0 = p_{0,D_s^-} - p_{0,D_s^+} = -0.026 \pm 0.011$ and $\Delta p_1 = p_{1,D_s^-} - p_{1,D_s^+} = -0.02 \pm 0.10$, which corresponds to a higher mistag probability for particles tagged by negative kaons.

Determination of the systematic uncertainty caused by the K^+/K^- detection asymmetry

The results of the two methods described in the previous section are compatible, the result obtained from the simulated B_s^0 sample is taken as a systematic uncertainty. The corresponding result is a deviation of $\Delta p_0 = p_{0_{B_s^0}} - p_{0_{\bar{B}_s^0}} = -0.020 \pm 0.004$ and $\Delta p_1 = p_{1_{B_s^0}} - p_{1_{\bar{B}_s^0}} = -0.01 \pm 0.03$. The deviation in the calibration parameter p_0 corresponds to a smaller mistag probability for particles tagged by positively charged kaons, which confirms



(a) Simulated B_s^0 events

(b) Data $D_s^{\pm}$ events

Figure 72: Calibration parameters depending on the initial flavour extracted from B_s^0 simulation, corrected for $K^{\pm}$ detection asymmetry and prompt D_s^+ from data.

	Δp_0	Δp_1
from simulation	-0.020± 0.004	-0.01±0.03
from prompt D_s^+	-0.026±0.011	-0.02±0.10

Table 24: Deviations on the p_0 and p_1 parameters for initial flavour B_s^0 and $\bar{B}_s^0$ determined with different methods.

the expectations. The deviation has to be corrected for in the specific physics analysis. If no correction is possible, it has to be considered as systematic uncertainty of the calibration of the predicted mistag probability. In the later case, the deviation is added to the other systematic uncertainties in quadrature.

8.2.2 Influence of additional charge asymmetries on the calibration

To check for an influence of additional charge asymmetries on the calibration, the sample is split into candidates tagged as B_s^0 (with $q_{tag} = 1$) and $\bar{B}_s^0$ (with $q_{tag} = -1$) and events taken with the magnetic field direction in the detector in positive y direction (magnet up) and negative y direction (magnet down).

The results of events tagged as B_s^0 and $\bar{B}_s^0$ are summarized in Table 25. A difference of 0.6σ in p_0 and 1.5σ in p_1 is observed between the two categories.

The calibration parameters determined separately for events taken with magnet configuration up and down is summarized in Table 26. The size of the two samples is not identical, the magnet up sample contains about $0.4\,fb^{-1}$ while the magnet down sample contains $0.6\,fb^{-1}$. A difference of 1.1σ in p_1 is observed between the two categories and no difference in p_0 has been found.

As no significant charge asymmetry for different tagging decisions or different magnetic field polarities is observed no systematic uncertainty is assigned.

	p_0	p_1	$\langle\eta\rangle$
tagged as B_s^0	0.413±0.013	0.68 ± 0.17	0.430
tagged as $\bar{B}_s^0$	0.403± 0.013	1.05±0.17	0.430

Table 25: Results of the unbinned fit for calibration parameters p_0 and p_1 for candidates tagged as B_s^0 and $\bar{B}_s^0$ respectively.

	p_0	p_1	$\langle\eta\rangle$
magnet up	0.407±0.015	0.72 ± 0.19	0.4284
magnet down	0.408± 0.012	0.95±0.16	0.4311

Table 26: Results of the unbinned fit for calibration parameters p_0 and p_1 for magnet up and magnet down data.

8.2.3 Influence of the analysis procedure on the calibration

Fit related systematic effects

It has been shown that the calibration using the unbinned fit and the fit in bins of the predicted mistag probability yield slightly different results, see Table 22. The difference between the two fit methods is considered as a systematic uncertainty.

To test for possible nonlinearities, an additional fit is performed with a second order polynomial as calibration function. The calibration coefficient for the quadratic term is however compatible with 0 within its error, so a linear correlation is valid.

Systematic effects related to the decay time uncertainty

The scaling factor, by which the decay time uncertainty has to be multiplied to give the correct estimate, is determined using prompt D candidates and random tracks to simulate fake B_s^0 candidates, *cf.* Chapter 5.3. The scaling factor has a systematic variation in the range between $[1.28, 1.48]$ due to kinematic and topological differences between the fake and the real B_s^0 candidates. To estimate the systematic uncertainty on the calibration parameters caused by this variation, two fits with the two extreme values of the scaling factors are performed. The results of these fits are listed in Table 27. The absolute difference between the default scaling factor and these results, which amounts to, $\sigma_{S,p_0} = 0.003$ and $\sigma_{S,p_1} = 0.03$, is taken as systematic uncertainty.

In simulated events, a double Gaussian resolution model for the decay time has been found to be more optimal than the Gaussian model used in this analysis, as is shown in Figure 31. To test the effect of a different resolution model on data, a fit is performed using the same parametrisation that was found to be optimal in the simulation, that is a double Gaussian with fraction of the broader Gaussian of about 11% and the width of the broader Gaussian 60% larger then the width of the inner one. A scaling factor of 1.27 is found for this model. The results of the tagging calibration are listed in Table 27 and are found to be in agreement with the default calibration parameters. Thus no systematic uncertainty is assigned.

	p_0	p_1	$\langle\eta\rangle$
decay time resolution			
proper time scaling factor 1.28	0.411±0.009	0.83 ±0.12	0.4302
proper time scaling factor 1.48	0.405± 0.010	0.89±0.13	0.4302
alternative resolution model	0.409±0.009	0.85 ±0.12	0.4302
decay time acceptance			
acceptance $\beta \cdot 1.1$	0.408±0.009	0.86 ±0.12	0.4302
acceptance $\beta \cdot 0.9$	0.408±0.009	0.86 ±0.12	0.4302
acceptance $\alpha \cdot 1.1$	0.408±0.009	0.85 ±0.12	0.4302
acceptance $\alpha \cdot 0.9$	0.409±0.009	0.85 ±0.12	0.4302

Table 27: Results of the unbinned fit for calibration parameters p_0 and p_1 with an under and overestimated event by event proper time calibration factor and an alternative proper time calibration model and different decay time acceptance models.

Systematic effects related to the decay time acceptance

To check the influence of the decay time acceptance parametrisation on the calibration parameters, the acceptance parameters α, describing the turn on, and β, describing the upper decay time acceptance (*cf.* Chapter 5.3) are varied by a relative 10%. The according changes in the acceptance function are illustrated in Figures 73 and 74. The results for the calibration parameters are listed in Table 27. In case of the parameter β, no effect is observed. In case of the parameter α, the variation leads to an effect of 0.001 on p_0 and 0.01 on p_1, which is considered in the combination of the total systematic uncertainty. Alternatively, the effect of a wrongly determined acceptance parametrisation on the calibration parameters could be studied using simulated pseudo experiments. As the effect of changing the acceptance parameters is small in the data, this is not done here and the results of the aforementioned studies are taken as an estimate of the corresponding systematic uncertainty.

Systematic effects related to the decay time PDF

For the decay width difference $\Delta\Gamma_s$, the current PDG value of $\Delta\Gamma_s/\Gamma_s = 0.150 \pm 0.020$ is taken as a constant in the fit. To account for possible deviations of this quantity, $\Delta\Gamma_s$ is varied within its uncertainty and the calibration parameters are determined. No deviation from the nominal calibration parameters is observed.

Systematic effects related to the mass fit

To test for possible systematic effects introduced by the choice of the mass model, a fit with a double Gaussian mass model for the signal component is performed instead of the nominal single Gauss. The fit results from the mass fit are listed in Table 28. The different mass fit results in a deviation of 0.002 in p_0 and 0.02 in p_1 with respect to the nominal calibration, *cf.* Table 29.

Additionally, the effect of a wrongly determined signal to background ratio is tested by varying the signal fraction with the default mass model by $\pm 3\sigma$. The observed deviation, listed in Table 29, is 0.002 on p_0 and 0.02 on p_1.

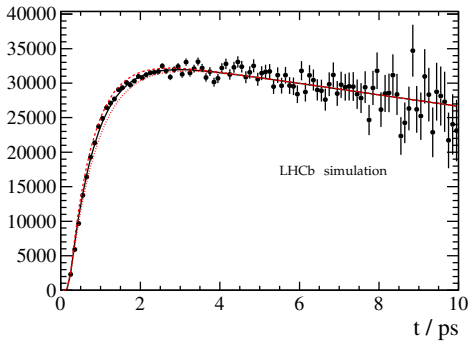


Figure 73: Illustration of the variation of the acceptance parameter α for the study of systematic effects introduced by the decay time acceptance. The solid line shows the default acceptance, the dotted line the case $\alpha \cdot 1.1$ and the dashed line the case $\alpha \cdot 0.9$.

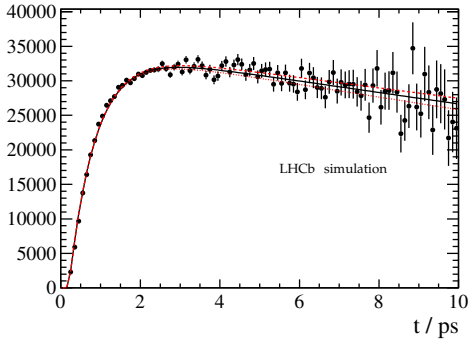


Figure 74: Illustration of the variation of the acceptance parameter β for the study of systematic effects introduced by the decay time acceptance. The solid line shows the default acceptance, the dotted line the case $\beta \cdot 1.1$ and the dashed line the case $\beta \cdot 0.9$.

Common parameters among the 3 B_s decay modes	
m_{B_s}	5370.88±0.15
$\sigma_{m,1}$ [MeV/c^2]	15.43±0.26
$\sigma_{m,2}$ [MeV/c^2]	37.42±3.59
f_{Gauss}	0.844±0.025
$f_{partial}$	0.511±0.005
f_{Ds_K}	0.149±0.013
$f_{D_s^*}$	0.699±0.028
Parameters of the $B_s \to D_s(\phi\pi)\pi$ decay	
f_{sig}	0.700±0.009
m_α^2 [MeV$^2/c^4$]	0.00465±0.00007
$f_{bkg_{comb}}$	1.0 ±0.0
Parameters of the $B_s \to D_s(K^*K)\pi$ decay	
f_{sig}	0.579±0.012
$f_{bkg_{comb}}$	0.704±0.019
f_{Λ_b}	0.000±0.003
m_α^2 [MeV$^2/c^4$]	0.00453±0.00007
Parameters of the $B_s \to D_s(KK\pi)\pi$ non resonant decay	
f_{sig}	0.470±0.012
$f_{bkg_{comb}}$	0.791±0.018
f_{Λ_b}	0.238±0.037
m_α^2 [MeV$^2/c^4$]	0.00393±0.00006

Table 28: Results of the combined fit to the three $B_s^0 \to D_s^- \pi^+$ candidate mass distributions in the wide mass range in the data using a double Gaussian model for the signal component.

As introducing the broader double Gaussian mass model automatically leads to a higher signal fraction, it has the same effect as varying the signal fractions by $+3\sigma$. The two effects are taken as one to avoid double counting. The net systematic uncertainty by a wrongly determined signal fraction is 0.002 on p_0 and 0.02 on p_1, which is considered in the total systematic uncertainty.

The fraction of $B_s^0 \to D_s^- K^+$ decays is difficult to determine as these events are within the B_s^0 mass peak. To check for a possible effect on the calibration parameters, two calibrations are performed where the fraction of these events is assumed to be half and double the value determined in the mass fit. The calibration parameters for these cases are listed in Table 29. A deviation of 0.001 in p_0 and 0.01 in p_1 is observed and taken into account in the calculation of the total systematic uncertainty.

Alternatively, the effect of a wrongly determined parametrisation of the signal mass PDF and a wrong determination of the signal and background fractions on the calibration parameters could be studied using simulated pseudo experiments. As the observed effects are small in the data, this is not done here and the results of the aforementioned studies are taken as an estimate of the systematic uncertainty.

Check of the tagging behaviour of the background

Overall tagging asymmetries in the background could influence the determination of the calibration parameters. To account for such asymmetries, the tagging efficiency and an overall tagging asymmetry for each background

	p_0	p_1	$\langle\eta\rangle$
mass model			
double Gaussian	0.410±0.009	0.84 ±0.12	0.430
signal fraction			
$f_{sig} + 3\sigma$	0.410±0.009	0.84 ±0.12	0.430
$f_{sig} - 3\sigma$	0.406±0.009	0.88 ±0.12	0.430
$B_s^0 \to D_s^- K^+$ **fraction**			
$f_{D_s^+K} \cdot 2$	0.407±0.009	0.87 ±0.12	0.430
$f_{D_s^+K} \cdot 0.5$	0.409±0.009	0.86 ±0.12	0.430

Table 29: Results of the unbinned fit for calibration parameters p_0 and p_1 with a double Gaussian model for the signal part of the mass PDF as well as different fractions for signal and $B_s^0 \to D_s^- K^+$ decays.

	ϵ	ω
combinatorial background	50.5±0.7	50.6± 1.0
misidentified Λ_b^0 background	30.9±6.7	37.1± 14.8
misidentified B^0 background	53.4±2.8	49.2± 3.3

Table 30: Results for the tagging efficiencies and asymmetries of the background components in the unbinned fit for calibration parameters p_0 and p_1.

component is left floating in all fits for the calibration parameters, see Chapter 5.5. The results in the fit for the default calibration parameters are listed in Table 30. The observed tagging asymmetries are compatible with 0.5, therefore no effect of the background on the calibration is expected.

8.2.4 Stability of the calibration for different run conditions and event topologies

Several event quantities could influence the tagging performance and hence the calibration of the predicted mistag probability. To study these effects, the sample is split into different parts according to these quantities. The calibration is determined for each of the subsamples and compared to the default set of calibration parameters to study possible variations. The quantities considered in this study are the track multiplicity, the number of reconstructed primary vertices and the B_s^0 transverse momentum. These quantities are already taken into account in the training of the neural network and the predicted mistag probability should account for correlations with the tagging performance. This assumption, however, has to be checked. As the running conditions can also influence the calibration parameters, a splitting into different run periods is studied in addition.

A higher number of primary vertices and a higher number of reconstructed tracks in the event might cause a higher mistag probability due to the increased probability to use underlying event tracks in the tagging process. If not correctly accounted for in the neural network, this might also be visible in the calibration parameters p_0 and p_1. Tables 31 and 32 summarise the results for the determination of the calibration parameters in bins of these quantities. By comparing these numbers among each other and with the default calibration parameters, no significant deviation is found so

	p_0	p_1	$\langle\eta\rangle$
1 PV	0.403±0.014	0.68 ±0.19	0.430
2 PVs	0.406± 0.015	1.09±0.20	0.432
$\geqslant$ 3 PVs	0.416±0.020	0.81 ±0.26	0.427

Table 31: Results of the unbinned fit for calibration parameters p_0 and p_1 for different PV multiplicities in the event.

	p_0	p_1	$\langle\eta\rangle$
$<$ 60 tracks	0.395±0.040	0.79 ±0.43	0.413
60 – 120 tracks	0.408± 0.015	0.97±0.20	0.428
$\geqslant$ 120 tracks	0.408±0.012	0.81 ±0.16	0.433

Table 32: Results of the unbinned fit for calibration parameters p_0 and p_1 for different track multiplicities.

no systematic uncertainty is assigned. The largest discrepancies between two subsamples amount to 0.5σ in p_0 and 1.5σ in p_1 in case of the primary vertex multiplicity and 0.3σ in p_0 and 0.6σ in p_1 for the track multiplicity.

For the transverse B_s^0 momentum, the sample is split up in three categories of approximately the same size. The results are summarised in Table 33 and show no significant deviation and no systematic uncertainty is assigned. The largest observed difference between two subsamples is 0.8σ in p_0 and 1σ in p_1.

Different run periods can correspond to different running conditions like, *e.g.* instantaneous luminosities, and different alignment scenarios, *etc.*. To check for these effects the sample is split into three different run periods that are each intersected by a technical stop. The results are summarised in Table 34 and show no significant deviation from the default calibration parameters. The largest observed difference between two subsamples is 1.4σ in p_0 and 1.8σ in p_1.

8.2.5 Combination of systematic effects

The relevant systematic effects studied in this chapter are combined in a total systematic uncertainty as listed in Table 35 by adding them in quadrature. For the total systematic uncertainty effects from the initial flavour of the B_s^0 mesons, the fit method, the decay time scaling factor and the decay time resolution model are considered. The total systematic uncertainty is $\sigma_{p_0} = 0.021$ and $\sigma_{p_1} = 0.10$. If the specific physics analysis in which this tagging algorithm is used accounts for the difference in the calibration de-

	p_0	p_1	$\langle\eta\rangle$
$p_{T_{B_s^0}} \leqslant 8\,\text{GeV/c}$	0.420±0.018	1.08 ±0.32	0.442
$8\,\text{GeV/c} < p_{T_{B_s^0}} \leqslant 14\,\text{GeV/c}$	0.403± 0.013	0.71±0.18	0.431
$p_{T_{B_s^0}} \geqslant 14\,\text{GeV/c}$	0.401±0.018	0.97 ±0.20	0.415

Table 33: Results of the unbinned fit for calibration parameters p_0 and p_1 for different B_s^0 transverse momenta.

Run number	Date	p_0	p_1	$\langle\eta\rangle$
< 94386	April - June	0.428±0.017	1.11±0.21	0.428
94387 - 100256	July - August	0.394±0.017	0.56±0.23	0.430
> 100257	September - November	0.401±0.016	0.91±0.21	0.432

Table 34: Results of the calibration parameters in different run periods.

	σ_{p_0}	σ_{p_1}
initial flavour	0.020	–
fit method	0.006	0.09
decay time resolution	0.003	0.03
decay time acceptance	0.001	0.01
mass model/signal fraction	0.002	0.02
$B_s^0 \to D_s^- K^+$ fraction	0.001	0.01
total systematic uncertainty		
quadratic sum	0.021	0.10
partial quadratic sum	0.007	0.10

Table 35: Results of the total systematic uncertainty of the calibration parameters.

pending on the initial flavour of the B_s^0, this contribution should be removed and the total systematic uncertainty amounts to $\sigma_{p_0} = 0.007$ and $\sigma_{p_1} = 0.10$.

The final set of calibration parameters to be applied to the predicted mistag probability of the neural network based same side kaon tagging algorithm is

$$\begin{aligned} p_0 &= 0.408 \pm 0.009(\text{stat.}) \pm 0.007(\text{syst.}) \\ p_1 &= 0.86 \pm 0.12(\text{stat.}) \pm 0.10(\text{syst.}) \\ \langle\eta\rangle &= 0.430\,. \end{aligned}$$

8.3 APPLICATION OF THE CALIBRATION TO OTHER DECAY MODES

Although the calibration parameters have been determined in $B_s^0 \to D_s^- \pi^+$ events, the same side kaon tagging algorithm is supposed to be used in all B_s^0 channels. For this to be possible it is necessary that the calibration parameters are independent of the specific decay channel. To check this assumption, the calibration parameters are reevaluated in simulated $B_s^0 \to J/\psi\,\phi$ events and compared to the default calibration parameters as measured on the simulated $B_s^0 \to D_s^- \pi^+$ sample. A sample of 220.000 simulated events is used, applying the same selection and trigger requirements as used in the measurement of the CP violating phase ϕ_s described in [63].

For a more extreme case, also the calibration obtained from the prompt D_s^+ sample in the data, that is used for the systematic studies on the $K^\pm$ detection asymmetry, are compared. The calibration parameters extracted from this sample can not be compared directly to the calibration in the simulated $B_s^0 \to D_s^- \pi^+$ and $B_s^0 \to J/\psi\,\phi$ samples. The prompt D_s^+ sample is not corrected for possible differences between the data and the simulation

	p_0	p_1	$\langle\eta\rangle$
$B_s^0 \to D_s^- \pi^+$ simulation	0.415±0.002	1.04±0.02	0.414
$B_s^0 \to J/\psi\,\phi$ simulation	0.419±0.002	0.99±0.03	0.414
prompt $D_s^+ \to \phi\pi$ data	0.409±0.005	1.09±0.05	0.414

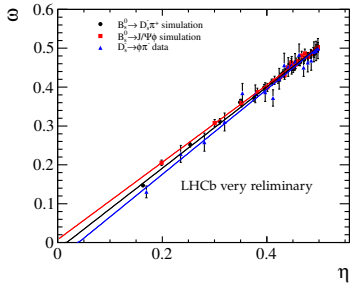
Table 36: Results of the fit for calibration parameters p_0 and p_1 for different B_s^0 and prompt D_s^+ decay modes.

such as, *e.g.*, the multiplicity of tagging track candidates, that can have an influence on the calibration. On the other hand, differences in the fragmentation, such as a different phase space occupied by the fragmentation kaons, might also cause deviations in the calibration of the predicted mistag probability. While there is some some deviation with respect to the B_s^0 decays expected, the prompt D_s^+ sample can serve nevertheless as an cross check and indication if the calibration can be transported between different decay modes.

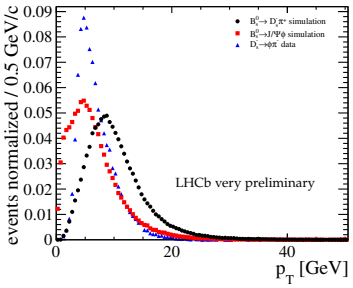
Figure 75b compares the transverse momentum distributions of the B_s^0 or D_s^+ from the three channels. The B_s^0 and D_s^+ meson decays are triggered by the topological hadron trigger lines, see Chapter 5.1.1. Those triggers require a high transverse momentum, consequently the transverse momentum of the B_s^0 or D_s^+ is in general higher as for the $B_s^0 \to J/\psi\,\phi$ decay which is triggered by the muons from the J/ψ decay. The momentum spectrum of the D_s^+ is on average lower than that of the B_s^0 in $B_s^0 \to D_s^- \pi^+$ decays because of the lower threshold energy needed to produce a D_s^+ meson.

The transverse momentum of the fragmentation kaons is correlated to the transverse momentum of the B_s^0 or D_s^+, which is in turn correlated to the mistag probability. Figure 75c shows the distributions of the predicted mistag probability for all three decays. The average predicted mistag probability $\langle\eta\rangle$ in the three modes is different due to the different trigger and selection and the kinematic differences in the production of B_s^0 and D_s^+. It can bee seen that the average predicted mistag probability $\langle\eta\rangle$ is the smallest in $B_s^0 \to D_s^- \pi^+$ decays, with the highest B_s^0 transverse momentum, followed by the D_s^+ and the $B_s^0 \to J/\psi\,\phi$ decays. For the determination of the calibration, the D_s^+ sample has been reweighted to match the B_s^0 transverse momentum.

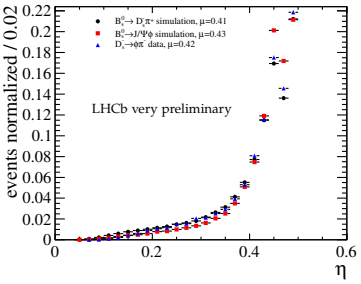
To compare the calibration parameters in the three different modes, the calibration function $\omega = p_0 + p_1 \cdot (\eta - \langle\eta\rangle)$ is fitted to the distribution of the average mistag fraction in bins of the predicted mistag probability using the same average predicted mistag probability $\langle\eta\rangle = 0.4140$ for all three modes. Figure 75a shows the three calibrations from the three different modes overlaid and the corresponding calibration parameters are compared in Table 36. The p_0 and p_1 values of the $B_s^0 \to D_s^- \pi^+$ and the $B_s^0 \to J/\psi\,\phi$ decays are consistent among each other which confirms that the calibration parameters obtained from the $B_s^0 \to D_s^- \pi^+$ channel can be transported to other channels. The calibrated mistag probability provides a good estimate of the real mistag fraction despite the difference between the decays. The p_0 parameter of the calibration extracted from the prompt D_s^+ decays shows a 1.1σ deviation to the $B_s^0 \to D_s^- \pi^+$ channel and a 1.9σ deviation with respect to the $B_s^0 \to J/\psi\,\phi$. These differences can be explained with the remaining differences between the data and the simulation and the differences in the B_s^0 and D_s^+ fragmentation.



(a) Comparison of calibration



(b) Comparison of transverse momentum



(c) Comparison of predicted mistag probability

Figure 75: Comparison of the calibration of the predicted mistag probability in different modes (a) and comparison of the transverse momentum of the reconstructed particles (b) and their predicted mistag probability distribution (c).

8.4 COMBINATION OF SAME SIDE KAON AND OPPOSITE SIDE TAGGING ALGORITHMS

To reach the best possible tagging performance, the same side kaon tagging algorithm is combined with the opposite side tagging algorithms. In order to perform this combination, several steps are necessary. First, the calibration of the combined opposite side tagging algorithms is checked. Then, the correlation between the same and the opposite side tagging algorithms is studied. Finally, the tagging algorithms are combined and the calibration of the combined predicted mistag probability and the performance of the combined tagging decision is determined.

8.4.1 Check of the opposite side tagging calibration in $B_s^0 \to D_s^- \pi^+$

The predicted mistag probability of the combined opposite side tagging has already been calibrated in the $B^+ \to J/\psi K^+$ decay channel , *cf.* Chapter 4.5, and checked in control channels. The expected calibration parameters determined in the decay $B^+ \to J/\psi K^+$ are $p_0 = 0.392 \pm 0.002(\text{stat.}) \pm 0.009(\text{syst.})$, $p_1 = 1.035 \pm 0.021(\text{stat.}) \pm 0.012(\text{syst.})$ and $\langle\eta\rangle = 0.391$ [29]. Minor deviations of the calibration parameters can be expected, as the trigger composition of each sample is different and influences the correlation among the individual OS tagging algorithms and accordingly also the combination.

To account for this effect, the calibration parameters are determined on the $B_s^0 \to D_s^- \pi^+$ sample. The fit results are $p_0 = 0.411 \pm 0.011$, $p_1 = 1.127 \pm 0.115$ and $\langle\eta\rangle = 0.385$. The difference to the expected calibration pa-

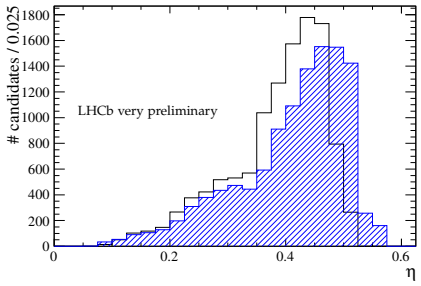


Figure 76: Distribution of the predicted mistag probability from the opposite side tagging combination for sideband subtracted $B_s^0 \to D_s^- \pi^+$ signal candidates before (black line) and after (blue shaded area) calibration.

rameter in p_0 amounts to 2.4σ, p_1 is compatible with the expectation. The corresponding tagging efficiency is $\varepsilon_{tag} = 40.20 \pm 0.30\%$ and the effective tagging power is $\varepsilon_{tag} = 2.74 \pm 0.39\%$. Due to the deviation observed in the p_0 calibration parameter the calibration parameters measured in $B_s^0 \to D_s^- \pi^+$ are used for the combination.

8.4.2 Correlation of same side kaon and opposite side tagging

The same side kaon tagging algorithm and the opposite side tagging algorithms are combined using the formalism presented in Chapter 4.4. A necessary condition for this combination is that the individual tagging algorithms are sufficiently uncorrelated. A low correlation between the SSK and the opposite side tagging algorithms can be expected due to the different physics principles that are exploited. The OS tagging algorithms use tracks originating from long lived B decays while the SSK tagging algorithm uses prompt tracks from the fragmentation. Both types of algorithms use impact parameter based variables to reduce background originating from the respectively other source and are therefore using mostly disjunct samples of tagging candidates.

In order to ensure this condition, three different types of correlations between the neural network based same side kaon tagging algorithm and the opposite side tagging algorithms are calculated.

The correlation of the predicted mistag probabilities η_i and η_j of two tagging algorithms i, j is given by

$$\rho(\eta_i, \eta_j) = \frac{\mathrm{cov}(\eta_i, \eta_j)}{\sigma_{\eta_i} \sigma_{\eta_j}}, \tag{84}$$

where

$$\mathrm{cov}(\eta_i, \eta_j) = E\big[(\eta_i - E[\eta_i])(\eta_j - E[\eta_j])\big] \tag{85}$$

is the covariance of the predicted mistag probabilities for tagging algorithm i and j with $E[\eta_i]$ the expected value of η_i and σ_{η_i} and σ_{η_j} their standard deviation.

The correlation between the tagging efficiencies tests if two tagging algorithms mostly tag the same events. It is given by

$$\rho^{tag}(i, j) = \frac{\varepsilon_{tag}^{i\&j}}{\varepsilon_{tag}^{i} \varepsilon_{tag}^{j}} - 1, \tag{86}$$

where $\varepsilon_{tag}^{i\&j}$ is the fraction of events where both tagging algorithms give a decision and ε_{tag}^{i} and ε_{tag}^{j} is the tagging efficiency of the individual tagging algorithms. This quantity can have values greater than 1 if one tagging decision fully includes the other one, as is the case for the combined opposite side decision and the individual opposite side tagging algorithms.

Finally, in those events where both tagging algorithms have given a decision, the correlation of these decisions d_i and d_j to be the same is calculated as

$$\rho(d_i, d_j) = \frac{N_{sametag}/N_{alltag}}{P_{same}} - 1 = \frac{N_{d_i=d_j\neq 0}/N_{d_i,d_j\neq 0}}{\eta_i\eta_j + (1-\eta_i)(1-\eta_j)} - 1\,. \tag{87}$$

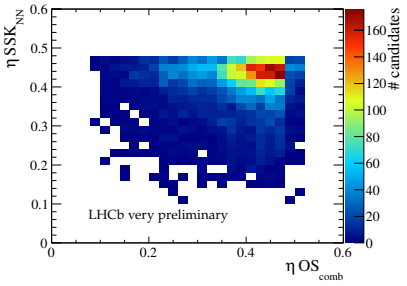
In this case, $N_{d_i,d_j\neq 0}$ is the number of events where both tagging algorithms give any decision unequal to zero and $N_{d_i=d_j\neq 0}$ is the fraction of event where both tagging algorithms give the same decision unequal to zero. The results of the calculated correlations of the neural network based same side kaon tagging algorithm with the individual opposite side tagging algorithms, the cut based same side kaon tagging algorithm and the combined opposite side decision is given in Table 37. Ideally, if no correlation is present, the calculated correlations should be approximately zero. It can be seen that this is the case for the correlation of the neural network based same side kaon tagging algorithm with all individual opposite side tagging algorithms as well as with the opposite side combination. It is therefore justified to use the combination discussed in Chapter 4.4. The correlations of the predicted mistag probabilities of the neural network based SSK and the opposite side tagging algorithms are also visualised in Figure 77, where their distributions are plotted. The distributions confirm the conclusion that no correlations between the same and the opposite site tagging algorithms is present.

The correlation numbers for the neural network based same side kaon tagging algorithm and the cut based same side kaon tagging algorithm serve as a cross check. The two tagging algorithms are not combined, as the neural network based same side kaon tagging algorithm supersedes the cut based one. The high correlation, which can be also seen in Figure 77, is expected, as both tagging algorithms exploit the same physics principle and use mostly the same variables.

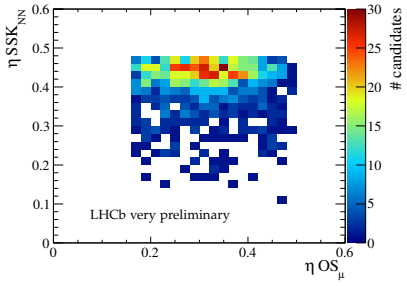
8.4.3 Calibration and performance of the combined same side kaon and opposite side tagging algorithms

After the combination, the calibration parameters and the performance of the combined tagging decision and predicted mistag probability are determined in the same way as for the individual tagging algorithms in a fit of the B_s^0 mixing. As the tagging algorithms are in principle uncorrelated it is expected that the combination of the individually calibrated predicted mistag probabilities is already an accurate prediction for the mistag probability. Likewise, the effective tagging power is expected to be the sum of the tagging power of the individual tagging algorithms.

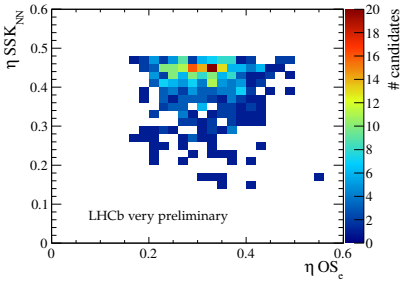
The results from the calibration process are listed in Table 38 and shown in Figure 78. The distribution of the predicted mistag probability of the combination for sideband subtracted $B_s^0 \to D_s^- \pi^+$ signal candidates is shown in Figure 79. As for the calibration of the neural network based same side kaon tagging algorithm, two different methods are used to extract the calibration parameters from the unbinned fit of the B_s^0 oscillation and by determining



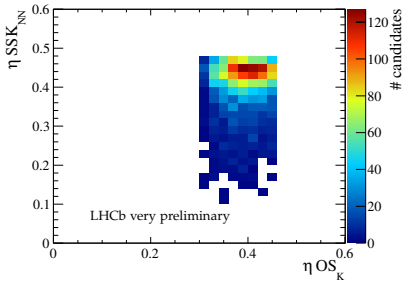
(a) NN based SSK tagging algorithm *vs.* combined OS tagging algorithm



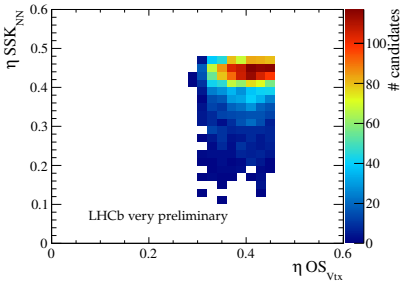
(b) NN based SSK tagging algorithm *vs.* OS μ tagging algorithm



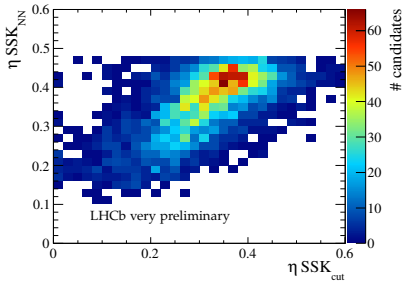
(c) NN based SSK tagging algorithm *vs.* OS electron tagging algorithm



(d) NN based SSK tagging algorithm *vs.* OS kaon tagging algorithm



(e) NN based SSK tagging algorithm *vs.* OS vertex tagging algorithm



(f) NN based SSK tagging algorithm *vs.* cut based SSK tagging algorithm

Figure 77: Correlation of predicted mistag probability for the neural network based same side kaon tagging algorithm with several other tagging algorithms.

	OS μ	OS e	OS K	OS Vtx	SS old	SS NN
Correlation between predicted mistag probabilities $\rho(\eta_i, \eta_j)$						
OS comb	0.74	0.63	0.51	0.54	0.03	0.02
OS μ	-	-0.11	0.14	0.34	-0.08	-0.01
OS e	-	-	0.10	0.38	-0.10	-0.03
OS K	-	-	-	0.41	0.18	0.04
OS Vtx	-	-	-	-	0.06	0.03
SS old	-	-	-	-	-	0.63
Correlation between tagging efficiencies $\rho^{tag}(i, j)$						
OS comb	1.47	1.47	1.47	1.47	0.06	0.03
OS μ	-	-0.28	0.28	0.6	0.01	0.00
OS e	-	-	0.42	0.63	0.23	0.04
OS K	-	-	-	0.79	0.06	0.04
OS Vtx	-	-	-	-	0.09	0.03
SS old	-	-	-	-	-	0.73
Correlation between decisions $\rho(d_i, d_j)$						
OS comb	0.75	0.76	0.73	0.81	-0.01	-0.02
OS μ	-	-0.33	0.11	0.49	-0.01	-0.05
OS e	-	-	0.13	0.32	-0.04	0.04
OS K	-	-	-	0.48	-0.00	-0.04
OS Vtx	-	-	-	-	-0.01	-0.02
SS old	-	-	-	-	-	0.67

Table 37: Correlation of the NN based SSK tagging algorithm with other tagging algorithms.

	p_0	p_1	$\langle\eta\rangle$
unbinned fit	0.388±0.007	0.98 ± 0.08	0.391 (fixed)
fit in bins	0.387± 0.008	0.98±0.09	0.391 (fixed)

Table 38: Results of the fit of the calibration parameters for an direct unbinned fit of p_0 and p_1 and of a fit in bins of the predicted mistag probability η for the combination of the opposite side and the neural network based same side kaon tagging algorithm.

the average mistag fraction in bins of the predicted mistag probability. The results from the two calibration methods are in agreement. The results for the calibration parameters agree with the expectation $p_0 - \langle\eta\rangle = 0$ within 0.37σ and $p_1 = 1$ within 0.25σ. The combination is therefore considered to be properly calibrated.

The effective tagging efficiency determined from the $B_s^0 \to D_s^- \pi^+$ candidates is is $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2 = 5.13 \pm 0.54\%$ at a tagging efficiency of $\varepsilon_{\text{tag}} = 66.4 \pm 0.4\%$. This effective tagging efficiency is consistent with the expectation that the tagging power of the combination is the sum of the performances of the individual tagging algorithms, $\varepsilon_{\text{eff,OS}} = 2.74 \pm 0.39\%$ and $\varepsilon_{\text{eff,SSK}} = 2.42 \pm 0.39\%$, in case no correlation is present.

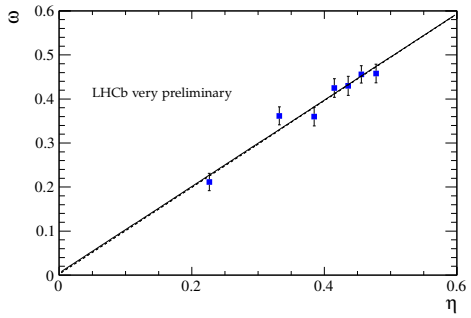


Figure 78: Average mistag fraction ω in bins of predicted mistag probabilities η for the combination of the opposite side and neural network based same side kaon tagging algorithm. The solid line is the result of the unbinned fit for the calibration parameters p_0 and p_1. The dashed line is the result of a linear fit to the data points.

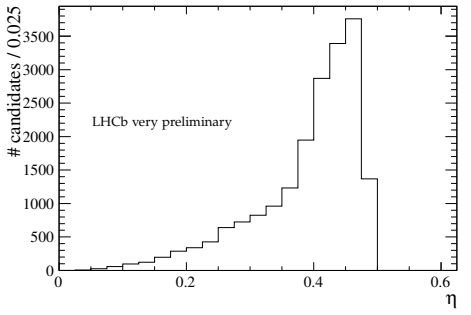


Figure 79: Distribution of the predicted mistag probability from the combination of opposite side tagging and the neural network based same side kaon tagging algorithm for sideband subtracted $B_s^0 \to D_s^- \pi^+$ signal candidates.

9 TRAINING OF THE NEURAL NETWORK BASED SAME SIDE KAON TAGGING ALGORITHM USING DATA

The neural network based same side kaon tagging algorithm, introduced in Chapter 7 and calibrated using data in Chapter 8 provides a substantial improvement over the cut based same side kaon tagging algorithm, that is currently used as default in the LHCb flavour tagging algorithms. There is however one drawback in the development process of this tagging algorithm: The training of the neural networks fully relies on simulated events and a sophisticated process is needed to remove differences between the data and the simulations as far as possible. For the first neural network, that separates fragmentation tagging track candidates from underlying event tracks, this is necessary, as the fragmentation is not accessible in the data. For the second network, which separates B_s^0 from $\bar{B}_s^0$ candidates, a training on data is possible if a statement on the flavour is possible from other sources. The calibration performed in Chapter 8 ensures that the SSK tagging algorithm trained on simulated events works correctly in the data, regardless of remaining differences between the simulation and the data. In view of these differences in the output of the first neural network, as well as for the particle identification variables that are used in the second neural network, a training that uses data might yield in a better performance. On the other hand, if the differences are only present in the first neural network, a training of the second neural network will not improve the overall performance of the tagging algorithm.

A possible solution for a training of the second neural network that does not rely on the simulation is discussed in this chapter, the usage of semimuonic B_s^0 decays where the opposite side tagging algorithms are used to gain information on the flavour of the B_s^0. Ideas for an additional approach using prompt D_s^+ decays for the training of the tagging algorithm are discussed in Appendix D.

9.1 USE OF SEMIMUONIC B_s^0 DECAYS FOR THE TRAINING OF THE NEURAL NETWORK

In contrast to the $B_s^0 \to D_s^- \pi^+$ decays used in the previous chapter for the calibration of the neural network based same side kaon tagging algorithm in the data, semimuonic $B_s^0 \to D_s^- \mu^+ \nu_\mu X$ decays are only partially reconstructed, because the neutrino is not detected. Due to the missing momentum from the neutrino, the B_s^0 mass and lifetime are not correctly determined which complicates their use in a measurement of the B_s^0 oscillation.

On the other hand, semileptonic $B_s^0 \to D_s^- \ell^+ \bar{\nu}_\ell$ decays feature a larger branching fraction of $7.9 \pm 2.4\%$ [6] compared to the full hadronic $B_s^0 \to D_s^- \pi^+$ mode with $(3.04 \pm 0.23) \cdot 10^{-3}$ and due to the muon in the decay chain, the semimuonic decays can be triggered by the muon trigger lines

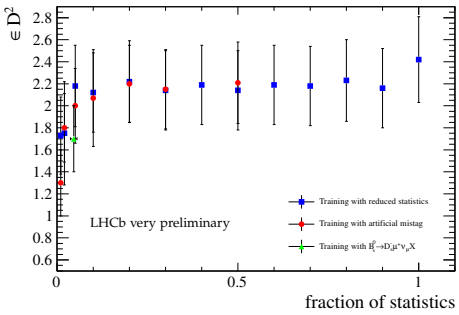
that have a high efficiency. Due to this features, the yield in the semimuonic decay modes is higher than in fully hadronic B_s^0 decays.

To use the semimuonic B_s^0 decays in the data for the training of the second neural network, their flavour is deduced on a statistical basis from the opposite side tagging algorithms. The decision of the opposite side tagging algorithms however is not always correct. In the $B^+ \to J/\psi K^+$, *e.g.*, a mistag fraction of $\omega = 36.7 \pm 0.2\%$ is measured. If a specific event is not correctly tagged, a wrong flavour is used in the training of the neural network. For this approach to work, the wrong information given to the neural network must not cause any bias in the training but may only correspond to a loss in statistics that is available to the neural network to learn from. If this is the case, the effect of the wrong information caused by the mistag fraction of the OS tagger can be compensated by increasing the training sample size.

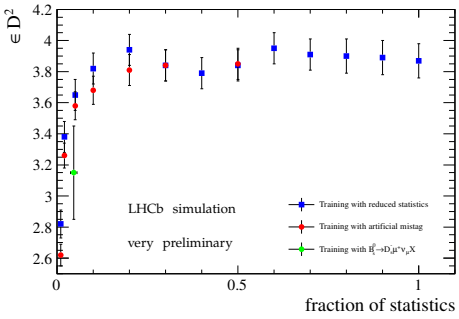
To study the effect of a wrong information about the flavour of a B_s^0 candidate on the training procedure of the neural network, simulated events are used where the production flavour is precisely known. An artificial mistag is introduced by randomly assigning the wrong production flavour in a fraction ω of events. After the training process, the performance of the neural network based same side kaon tagging algorithm is measured using the fit of the B_s^0 oscillation in the data and also directly on the simulated events. The mistag fractions that are studied are chosen in such a way, that they correspond to a loss in ε_{eff} of 50%, 60%, 70%, 80%, 90%, 95% and 99% with respect to the total available statistics in the simulation. A tagging efficiency of 100% is assumed to profit optimally from the available statistics. The correspondence between ω and the tagging efficiency is then given by $\omega = (1 - \sqrt{\varepsilon_{\text{eff}}/\varepsilon_{\text{tag}}})/2$. The results of this study are plotted in Figure 80. It can be seen that both in the data and the simulation, the performance of the tagging algorithm if a mistag is introduced is constant up to a certain mistag fraction where it drops off.

To complement these results, a study is performed where the true flavour is used in the network training, but the fraction of events used in the training is reduced with respect to the total available statistics. The performance of the so trained tagging algorithms is also shown in Figure 80. It can be seen that the results are in agreement with using a fraction ω of wrongly tagged events corresponding to the same loss in statistics in the training.

The results of these two studies can be understood in view of the training process of the neural network as discussed in Chapter 7.1. In the training process, the weights of the network are adapted in such a way that the best possible separation between signal and background is reached. The necessary information for this process is extracted from the training sample. If the statistics of the training sample is reduced less information about the events is available and at a certain point, characterised by the drop off in Figure 80, the network is not able to reach an optimal separation between B_s^0 and $\bar{B}_s^0$. On the other hand, introducing a mistag in the events used for training is reducing the available information in the same way. Under this assumption, the effect of a mistag in the training sample can be balanced by using more statistics for training and it is possible to use semimuonic $B_s^0 \to D_s^- \mu^+ \nu_\mu X$ decays for a data based training of the second neural network of the same side kaon tagging algorithm, provided enough statistics is available.



(a) Effective tagging power in the data



(b) Effective tagging power in the simulation

Figure 80: Effective tagging power of the neural network based same side kaon tagging algorithm measured in the data (a) and the simulation (b). The second neural network has been trained with different amounts of training events (blue circles), with the full training statistics but an artificial mistag (red squares) using simulated events and on semimuonic B_s^0 decays reconstructed in the data using the opposite side tagging algorithm for the information on the production flavour (green triangles).

9.1.1 Training with $B_s^0 \to D_s^- \mu^+ \nu_\mu X$ decays on data

To test the hypotheses discussed in the last section, a training is performed using $B_s^0 \to D_s^- \mu^+ \nu_\mu X$ decays reconstructed on data. To reach a better momentum resolution, the momenta are corrected for the momentum loss due to the neutrino by an average scaling factor which is derived from the simulation. To remove combinatoric background, the events are sideband subtracted in the D_s^+ invariant mass distribution. About 690.000 reconstructed signal candidates are available in the $1\,fb^{-1}$ dataset used in this analysis.

This dataset is reduced by several factors. Due to the cut on the output of the first NN, only in 50% of the cases at least one track is selected as input for the second NN. The training on data requires the event to be tagged by the opposite side tagging algorithms, which is the case in 33% of the events. The mistag fraction of the opposite side tagging algorithms reduces the statistical power of the events by $(1-2\omega)^2$. As there is no measurement available for the mistag fraction of the opposite side tagging algorithms in this specific decay channel, the same mistag fraction as for the $B^+ \to J/\psi K^+$, discussed in Chapter 4.5 is assumed, $\omega = 36.7 \pm 0.2\%$. The statistical power usable for the training corresponds to at least 7.300 events where the true flavour is known or to about 4.6 percent of the statistics available for training in the simulation. It should be noted, that additional background originating from the combination of prompt D_s^+ and random muons is not removed from the sample. Due to the D_s^+, this background component has the opposite tagging decisions than the B_s^0 which leads to an additional reduction of the statistical size of the sample due to the mistag that is introduced.

A training of the second neural network for the same side kaon tagging algorithm is performed using this dataset. The predicted mistag probability from the opposite side tagging algorithms is used as a weight in the training to improve the statistical power. Events with a low predicted mistag probability contribute more to the training than events with a high mistag probability. With the trained neural network, the performance of the same side kaon tagging algorithm is determined, using simulated events and the fit of the oscillation in the data. The resulting performances are plotted in Figure 80. With the current statistical precision the results agree with the expectation derived from the studies using simulated events but a conclusive statement about additional gains in the performance of the tagger are not possible. It can be concluded that that a factor of $3-5$ more statistics is needed to be at least competitive to the training procedure on simulated events or to conclude if a training on data improves the performance of the SSK tagging algorithm.

10 MEASUREMENT OF THE B_s^0–$\bar{B}_s^0$ MIXING FREQUENCY Δm_s

While the necessity to resolve the B_s^0 oscillation to calibrate the same side kaon tagging algorithm complicates its usage, it at the same time opens up the possibility for the measurement of the mixing frequency Δm_s in the B_s^0 system. This mixing frequency can be related to fundamental Standard Model parameters, see Chapter 2.3, and its measurement therefore provides an important test for the consistency of the Standard Model.

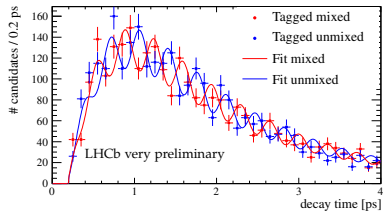
This chapter discusses the extraction of the B_s^0 mixing frequency Δm_s from the data. The comparison of the results and the statistical sensitivities of Δm_s in several flavour tagging scenarios gives informations on the performance of the tagging algorithms. Several sources of systematic uncertainties on the measurement of Δm_s are discussed. The chapter concludes with a comparison of the result with a recent published LHCb measurement [13] and a discussion on the prospects for further measurements of Δm_s.

10.1 COMPARISON OF Δm_s IN DIFFERENT FLAVOUR TAGGING SCENARIOS

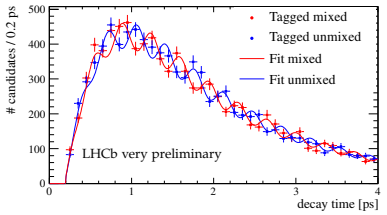
The measurement of the oscillation frequency Δm_s is performed using the fit procedure described in Chapter 5.5. For the extraction of the oscillation frequency, the tagging parameters for signal and background, the parameters of the decay time distribution of the combinatorial background, the decay width Γ_s and the oscillation frequency Δm_s are varied in the fit of the B_s^0 decay time distribution. Three different flavour tagging scenarios are compared using only the neural network based same side kaon tagging algorithm, only the combined opposite side tagging algorithms and a combination of both. These scenarios correspond to the ones discussed in the Chapter 8 for the calibration of these tagging algorithms with $B_s^0 \to D_s^- \pi^+$ decays on the data.

The results of the three independent fits are listed in Tables 39, 40 and 41. The values of the measured oscillation frequency in the three scenarios, $\Delta m_s^{SSKT} = 17.724 \pm 0.031\,ps^{-1}$, $\Delta m_s^{OST} = 17.762 \pm 0.029\,ps^{-1}$ and $\Delta m_s^{combined} = 17.745 \pm 0.022\,ps^{-1}$ are in agreement. As discussed in Chapter 4.2, the statistical sensitivity on the mixing frequency directly depends on the tagging power. This is reflected in the statistical error of Δm_s in the different scenarios. The combined opposite site tagging algorithm has an effective tagging efficiency of $\varepsilon_{tag} = 2.74 \pm 0.39$ corresponding to a statistical error on Δm_s of $\sigma_{\Delta m_s} = 0.029$, compared to $\sigma_{\Delta m_s} = 0.031$ in case of the neural network based same side kaon tagging algorithm with an effective tagging efficiency of $\varepsilon_{tag} = 2.42 \pm 0.39$. The combination of both tagging approaches gives the best statistical sensitivity on the mixing frequency, with an uncertainty of $\sigma_{\Delta m_s} = 0.022$. This result is used as the default value for Δm_s.

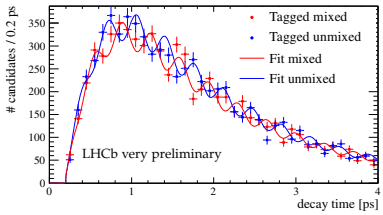
Figure 81 compares the decay time distributions of events that are tagged as mixed, *i.e.* they have a different production and decay flavour, and un-



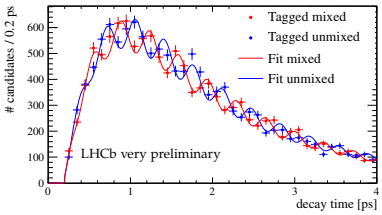
(a) Cut based SSK tagging algorithm



(b) NN based SSK tagging algorithm



(c) Combined OS tagging algorithms



(d) NN based SSK and combined OS tagging algorithms

Figure 81: Decay time distributions for events tagged as mixed and unmixed and corresponding projections of the fit PDF in a window of $\pm 3\sigma$ around the fitted B_s^0 mass. The plots correspond to different tagging scenarios, for the cut based same side kaon tagging algorithm (a), the neural network based same side kaon tagging algorithm (b), the combined opposite side tagging algorithm (c) and the combination of opposite side and neural network based same side kaon tagging algorithm (d).

mixed, *i.e.* they have the same production and decay flavour, in a window of $\pm 3\sigma$ around the fitted B_s^0 mass, for the different flavour tagging scenarios. The oscillation pattern is clearly visible in all the plots. For comparison reasons, also the decay time distribution for events tagged by the cut based same side kaon tagging algorithm is shown, see Figure 81a. When compared to the new neural network based same side kaon tagging algorithm, Figure 81b, the clear improvement in the resolution of the oscillation introduced by the new algorithm is visible.

10.2 SYSTEMATIC STUDIES ON THE MEASUREMENT OF Δm_s

Systematic uncertainties that influence the measurement of the B_s^0 oscillation frequency Δm_s can originate from an imprecise understanding of the decay time behaviour, or decay time scale of the B_s^0 signal candidates, or problems in the extraction of the signal from the mass distribution. The systematic effects considered in this analysis are based on the systematic studies performed for the published LHCb measurement of Δm_s [13]. All systematic uncertainties are given relative to the nominal result for the oscillation frequency, $\Delta m_s = 17.745 \pm 0.022$, which has been obtained using the combination of the neural network based same side kaon tagging algorithm and the combined opposite side tagging algorithms.

parameter	fit value and uncertainties
Δm_s [ps^{-1}]	17.7241 ± 0.0312
Γ_s [ps^{-1}]	0.6453 ± 0.0043
$p0_{SSKT}$	0.408 ± 0.009
$p1_{SSKT}$	0.86 ± 0.12
$\epsilon_{sig,SSKT}$	0.4951 ± 0.0038
$\omega_{bkg_{comb},SSKT}$	0.506 ± 0.010
$\epsilon_{bkg_{comb},SSKT}$	0.5053 ± 0.0072
$\omega_{B^0,SSKT}$	0.492 ± 0.033
$\epsilon_{B^0,SSKT}$	0.534 ± 0.028
$\omega_{\Lambda_b^0,SSKT}$	0.37 ± 0.15
$\epsilon_{\Lambda_b^0,SSKT}$	0.309 ± 0.067
Parameters of the combinatorial background decay time distribution	
a	0.0773 ± 0.0143
α	2.7288 ± 0.0976
β	1.0786 ± 0.0395
f	0.97670 ± 0.00339

Table 39: Fit results of all parameters varied in the fit of the B_s^0 oscillation if only the neural network based same side kaon tagging algorithm is used.

parameter	fit value and uncertainties
Δm_s [ps^{-1}]	17.7615 ± 0.0285
Γ_s [ps^{-1}]	0.6511 ± 0.0040
$p0_{OST}$	0.41095 ± 0.00935
$p1_{OST}$	1.127 ± 0.109
$\epsilon_{sig,OST}$	0.40195 ± 0.00295
$\omega_{bkg_{comb},OST}$	0.49371 ± 0.00912
$\epsilon_{bkg_{comb},OST}$	0.54218 ± 0.00700
Parameters of the combinatorial background decay time distribution	
a	0.0718 ± 0.0156
α	2.768 ± 0.103
β	1.0731 ± 0.0432
f	0.98047 ± 0.00313

Table 40: Fit results of all parameters varied in the fit of the B_s^0 oscillation if only the combined opposite side tagging algorithms are used.

parameter	fit value and uncertainties
Δm_s [ps^{-1}]	17.7454 ± 0.0217
Γ_s [ps^{-1}]	0.6450 ± 0.0043
$p0_{Comb}$	0.38808 ± 0.00745
$p1_{Comb}$	0.9810 ± 0.0831
$\epsilon_{sig,Comb}$	0.66408 ± 0.00358
$\omega_{bkg_{comb},OST}$	0.49369 ± 0.00893
$\epsilon_{bkg_{comb},OST}$	0.54014 ± 0.00692
$\omega_{bkg_{comb},SSKT}$	0.5033 ± 0.0102
$\epsilon_{bkg_{comb},SSKT}$	0.51073 ± 0.00735
$\omega_{B^0,SSKT}$	0.4911 ± 0.0308
$\epsilon_{B^0,SSKT}$	0.5238 ± 0.0252
$\omega_{\Lambda_b^0,SSKT}$	0.456 ± 0.107
$\epsilon_{\Lambda_b^0,SSKT}$	0.3211 ± 0.0513
Parameters of the combinatorial background decay time distribution	
a	0.0737 ± 0.0151
α	2.751 ± 0.101
β	1.0738 ± 0.0424
b	0.97958 ± 0.00321

Table 41: Fit results of all parameters varied in the fit of the B_s^0 oscillation if the combination of neural network based same side kaon tagging algorithms and opposite side tagging algorithms is used.

10.2.1 Systematic effects related to the decay time

Most of the systematic effects related to the decay time that are studied for the measurement of Δm_s are already discussed discussed in Chapter 8.2.3.

To check for possible influences of the acceptance parametrisation, the acceptance parameters α for the turn on and β for the upper lifetime acceptance are varied by a relative $\pm 10\%$ and the influence on Δm_s is measured. No change in the oscillation frequency has been found due to this variation, see Table 42.

The scale factor of the decay time uncertainty is varied within the range of its systematic uncertainty, $[1.278, 1.48]$ and a deviation of $\Delta(\Delta m_s) = \pm 0.0006\,\text{ps}^{-1}$ has been found, see Table 42. A systematic error of 0.001 is assigned and taken into account in the combined systematic uncertainty.

The alternative double Gaussian resolution model for the decay time resolution is tested and results in a deviation of $\Delta(\Delta m_s) = -0.0003\,\text{ps}^{-1}$. This deviation corresponds to less than 2% of the statistical uncertainty and is therefore negligible in the combined systematic uncertainty.

The value for the decay width difference $\Delta\Gamma_s$, which is constant in the fit and set to the current PDG value, is varied within its statistical uncertainty to estimate possible influences on the determination of Δm_s. A deviation of $\Delta(\Delta m_s) = \pm 0.0002\,\text{ps}^{-1}$ is observed, see Table 42. This deviation corresponds to less than 1% of the statistical uncertainty and is therefore negligible in the combined systematic uncertainty.

For the measurement of the oscillation frequency, two additional systematic effects play an important role, the precision with which the scale for the measurement of the z position and the momentum of a track is known. Those effects represent a scaling of the decay time, as it is measured from

the momentum and the decay length of a particle, and directly correspond to an uncertainty on the oscillation frequency.

The following associated systematic uncertainties are recommended by the LHCb tagging and alignment group. The decay length of a particle is measured by the distance of the primary interaction point and its decay vertex. The precision with which this distance is known depends on the precision of the position measurement of the tracks in the detector. As the decay products of the B_s^0 are boosted in z direction the uncertainty in this direction is the dominating effect. The highest accuracy of the measurement is reached in the VELO. Two effects limit the precision of the measurement in the VELO. The length of the detector is only known up to a certain precision. The size of this uncertainty on the overall scale of the VELO has been determined in a survey at the time of the assembly to be $\sigma_{\text{survey}} = 0.01\%$. The second effect regards the accuracy by which the absolute position of the detector modules is known. It is determined by comparing the z position of a module from the track based alignment and the survey data. This uncertainty in the absolute module position amounts to $20\,\mu m$ and is divided by the spread in the position of the first hit of the tracks used in this analysis in the detector to get the relative size of the uncertainty. This is done to account for the fact that not all tracks hit the same module and thus do not all have the same uncertainty in their position measurement. The position of the first measurement of the tracks used in this analysis is shown in Figure 82, the spread is given by the root mean square (RMS) of the distribution. The resulting uncertainty from the alignment is than given by

$$\sigma_{\text{alignment}} = \frac{20\,\mu m}{100 mm} = 0.02\% \,. \tag{88}$$

The two effects are added in quadrature, giving a total systematic uncertainty on the z scale of $\sigma_{\text{zscale}} = 0.022\%$, which translates directly into an uncertainty of Δm_s of $\pm 0.004\,\text{ps}^{-1}$.

The momentum is measured by the curvature of the track in the magnetic field. Two effects can influence the measurement of the momentum, the position measurements of the hits on the track, which are limited by the alignment of the detector elements and the knowledge of the magnetic field. Most contributions are already accounted for in the reconstruction software and the remaining uncertainty on the scale of the particle momentum is 0.15%.

The momentum scale uncertainty of the B_s^0 decay products influences both the determination of the mass and the momentum of the B_s^0 candidates. As the quotient of both quantities appears in the calculation of the decay time, the effects cancel to some extend. A calculation of the net effect of the momentum scale uncertainty of the B_s^0 decay products on the quotient m/p of the B_s^0 is not trivial for a four body decay. Therefore, the size of the systematic uncertainty is determined from the simulation. The three momenta of the daughter particles are scaled by a factor of 1.0015 and the mass and the momentum of the B_s^0 are determined by four vector addition. This is valid, as no mass constraints on the D_s^+ are used in the determination of the decay time of the B_s^0. The resulting impact on the decay time of the B_s^0 is shown in Figure 83. The relative uncertainty of $\Delta t/t = 0.022\%$ on the decay time translates directly into an uncertainty of $\pm 0.004\,\text{ps}^{-1}$ on the oscillation frequency Δm_s.

	$\Delta(\Delta m_s)\,[ps^{-1}]$
decay time resolution	
proper time scaling factor 1.28	-0.0006
proper time scaling factor 1.48	+0.0006
alternative resolution model	-0.0003
decay time acceptance	
acceptance $\beta \cdot 1.1$	0
acceptance $\beta \cdot 0.9$	0
acceptance $\alpha \cdot 1.1$	0
acceptance $\alpha \cdot 0.9$	0
variation of $\Delta\Gamma_s$	
$\Delta\Gamma_s = 0.13 \cdot \Gamma_s$	+0.0002
$\Delta\Gamma_s = 0.17 \cdot \Gamma_s$	-0.0002
scale dependencies	
z scale dependence	$\pm$ 0.004
momentum scale dependence	$\pm$ 0.004

Table 42: Absolute differences of the fitted value for the oscillation frequency Δm_s with an under and overestimated event by event proper time calibration factor, an alternative proper time calibration model and different decay time acceptance models as well as uncertainties introduced by the z scale and momentum scale dependence.

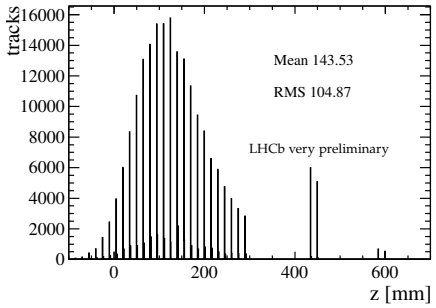


Figure 82: Position along the z axis of the first hit of all B_s^0 decay particles used in the determination of Δm_s.

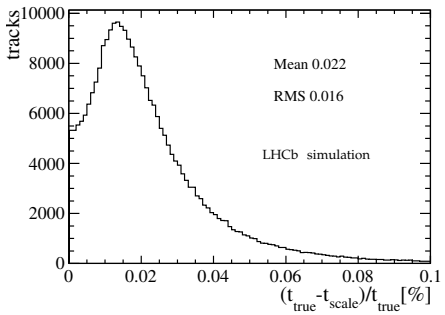


Figure 83: Relative difference of default B_s^0 decay time and the decay time if the momentum of the decay particles is scaled by 1.0015.

10.2.2 Systematic effects related to the mass

The extraction of signal and background from the mass distribution can also influence the determination of the oscillation frequency Δm_s. If, *e.g.*, the fraction of slowly oscillating B^0 background is not properly determined, the determination of the B_s^0 oscillation frequency is biased. To study potential biases, several tests are performed in the same fashion as discussed in Chapter 8.2.3.

The mass model is changed to the double Gaussian model, with the parameters listed in Table 28. The fit of the B_s^0 oscillation frequency with this mass model leads to a deviation of $\Delta(\Delta m_s) = -0.0028\,\mathrm{ps}^{-1}$. As this deviation is significant it is studied in more detail using simulated pseudo experiments. For this, simulated events are generated according to the PDF described in Chapter 5 with fixed input parameters. The values of these input parameters are taken from the fits to the data for the mass part of the PDF as well as for the tagging performances. As input for the lifetime and the decay width, the current PDG values are taken. For the oscillation frequency, a value of $\Delta m_s = 17.77\,\mathrm{ps}^{-1}$ is used. To asses the impact of a wrongly determined mass model on the determination of the oscillation frequency, three different sets of pseudo experiments are used. For each set, 2.500 experiments with the same number of signal and background events as observed on data are generated and fitted. As a baseline, the same single gaussian mass model is used in the generation and in the fit. Furthermore, two sets of pseudo experiments are studied where a single gaussian mass model was used in the generation and a double gaussian mass model was used in the fit and vice versa. The distribution of the measured values of Δm_s is shown in Figure 84. The observed deviation in the mean value of the Δm_s distribution in those two cases with respect to the baseline scenario is less than $0.001\,\mathrm{ps}^{-1}$ which is negligible to the total systematic uncertainty, accordingly no uncertainty is assigned.

To study a possible influence of an incorrectly determined signal fraction, this fraction is changed by $\pm 3\sigma$ of its statistical uncertainty. The effect of a larger signal fraction on the oscillation frequency is $\Delta(\Delta m_s) = -0.0002\,\mathrm{ps}^{-1}$ and of a smaller signal fraction $\Delta(\Delta m_s) = +0.0001\,\mathrm{ps}^{-1}$. This corresponds to less then 1% of the statistical uncertainty on Δm_s and is therefore not considered in the combination of the total systematic uncertainty.

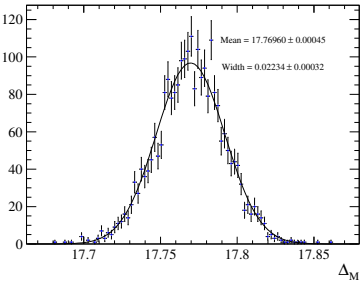
Doubling the fraction of wrongly identified $B_s^0 \to D_s^- K^+$ results in a shift of $\Delta(\Delta m_s) = -0.0012\,\mathrm{ps}^{-1}$ while assuming only half the measured fraction leads to a shift of $\Delta(\Delta m_s) = +0.0007\,\mathrm{ps}^{-1}$. A systematic uncertainty on the determination of Δm_s due to this effect of $\sigma = 0.001\,\mathrm{ps}^{-1}$ is assigned and taken into account in the total systematic uncertainty.

10.2.3 Combination of systematic uncertainties on the B_s^0 oscillation frequency Δm_s

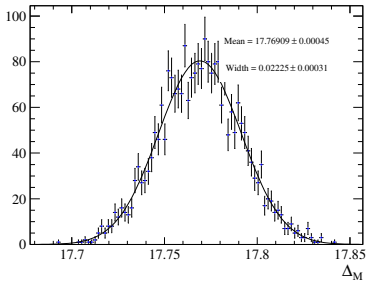
The systematic uncertainties that are discussed in the previous paragraphs are added in quadrature for the determination of the total systematic uncertainty on the oscillation frequency Δm_s. The relevant systematic effects originate from the uncertainty on the z scale and the momentum scale, the scaling factor for the decay time resolution and the fraction of $B_s^0 \to D_s^- K^+$ decays, see Table 44. The combined systematic uncertainty amounts to $\sigma_{\mathrm{syst.}} = \pm 0.006\,\mathrm{ps}^{-1}$.

	$\Delta(\Delta m_s)\,[ps^{-1}]$
mass model	
double Gaussian	-0.0005
signal fraction	
$f_{sig} + 3\sigma$	-0.0002
$f_{sig} - 3\sigma$	+0.0001
$B_s^0 \to D_s^- K^+$ **fraction**	
$f_{D_s^+ K} \cdot 2$	-0.0012
$f_{D_s^+ K} \cdot 0.5$	+0.0007

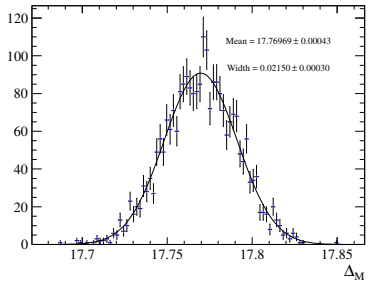
Table 43: Results of the unbinned fit the B_s^0 oscillation frequency Δm_s with a double Gaussian model for the signal part of the mass PDF from pseudo experiments as well as different fractions for signal and $B_s^0 \to D_s^- K^+$ decays.



(a) Default mass model



(b) Single Gaussian generated, double Gaussian fitted



(c) Double Gaussian generated, single Gaussian fitted

Figure 84: Results for the oscillation frequency Δm_s from pseudo experiments with three different mass models: (a) The default mass model where a single Gaussian has been used in the generation and the fit, (b) a single Gaussian has been used in the generation and a double Gaussian in the fit and (c) a double Gaussian has been used in the generation and a single Gaussian has been used in the fit. 5000 datasets of the size of the observed dataset have been generated for each scenario.

	$\Delta(\Delta m_s)$ ps^{-1}
z scale	0.004
momentum scale	0.004
decay time resolution	0.001
$B_s^0 \to D_s^- K^+$ fraction	0.001
total systematic uncertainty	
quadratic sum	0.006

Table 44: Results of the total systematic uncertainty of the B_s^0 oscillation frequency Δm_s.

10.3 COMPARISON OF THE MEASUREMENT OF Δm_s WITH RECENT LHCB RESULTS AND PROSPECTS FOR FURTHER STUDIES

10.4 PROSPECTS FOR FURTHER STUDIES

With the analysis presented in this thesis, the B_s^0 oscillation frequency is measured in decays of $B_s^0 \to D_s^- \pi^+$, using three different decay modes of the D_s^+, namely $D_s^- \to \phi\pi^-$, $D_s^- \to K^* K^-$ and non resonant D_s^- decays. The final result is

$$\Delta m_s = 17.745 \pm 0.022\,(\text{stat.}) \pm 0.006\,(\text{syst.})\ \text{ps}^{-1}\,.$$

The most recent published LHCb measurement [13] reports a value of

$$\Delta m_s = 17.768 \pm 0.023\,(\text{stat.}) \pm 0.006\,(\text{syst.})\ \text{ps}^{-1}\,.$$

The later measurement was performed on the same dataset of $1\,\text{fb}^{-1}$ taken by the LHCb experiment in 2011, however several important differences exist between the two analyses. While the analysis performed in this thesis concentrates on the measurement of the calibration parameters, the published LHCb analysis uses a more sophisticated selection and additional D_s^+ meson decay modes to improve the sensitivity on Δm_s. The selection used in the published LHCb analysis uses a boosted decision tree with variables that bias the decay time, which makes the measurement of the scaling factor for the decay time error more complicated. As the measurement of the calibration parameters for the predicted mistag probability of the same side kaon tagging algorithm relies on the correct determination of this scaling factor, it was decided to not use the updated selection for the sacrifice of a lower yield. This also applies to the additional decay modes of the D_s^+, for which no measurement of the scaling factor for the decay time error was available.

The different selection results in a higher signal yield of about 7% in the common D_s^+ meson decay modes and a lower background level in case of the published analysis. The overlap of events in the two analyses in the common modes is about 70%. Two additional decay modes of $D_s^- \to K^-\pi^+\pi^-$ and $D_s^- \to \pi^-\pi^+\pi^-$ are used, increasing the yield by an additional 22%. Overall, the published result uses about 30% more statistics compared to the analysis presented in this thesis.

However, the statistical precision on the B_s^0 oscillation frequency of the two measurements is comparable as the analysis done in this thesis profits

from the improved same side kaon tagging algorithm that is presented in the previous chapters and has not been available at the time the official LHCb result was published

In view of the aforementioned differences in the flavour tagging and the selection it can be concluded that the two measurements are about 50% uncorrelated and are consistent within their statistical and systematic uncertainties.

By combining the benefits of the two analysis, *i.e.* the improved selection from the published Δm_s measurement and the neural network based same side kaon tagging algorithm presented in this thesis, and using the additional data sample of $2\,fb^{-1}$ taken by the LHCb experiment in 2012, it can be expected that the statistical uncertainty can be improved by at least a factor two, making it comparable to the systematic uncertainty.

11 CONCLUSION

This thesis presented the development and calibration of a same side kaon tagging algorithm and the measurement of the B_s^0–$\bar{B}_s^0$ oscillation frequency Δm_s.

The development of the same side kaon tagging algorithm relies on the correct simulation of the fragmentation process and the underlying event. By comparing the simulation and the data and step by step correcting for differences in the primary vertex multiplicity, the tagging track candidate multiplicity and other discrepancies, it has been shown that the majority of the difference in the tagging performance can be explained. The by far largest correction originates from the wrongly simulated particle multiplicity in the underlying event, which accounts for about 50% of the effect on the tagging efficiency and can explain all the differences seen in the mistag fraction.

The corrected sample of simulated $B_s^0 \to D_s^- \pi^+$ events has been used for the development of a same side kaon tagging algorithm which is based on a two step procedure. In the first step, by exploiting a neural network, for each tagging track candidate the probability to come from the B_s^0 fragmentation is calculated based on its kinematic properties, track quality variables and properties of the reconstructed B_s^0 meson. Tagging track candidates with a significant probability to come from the B_s^0 fragmentation are passed to a second neural network, which calculates the probability for a specific production flavour based on these tagging track candidates and general event variables.

The implementation of the same side kaon tagging algorithm has been calibrated using an unbinned maximum likelihood fit of the B_s^0–$\bar{B}_s^0$ oscillation with about $26,000$ $B_s^0 \to D_s^- \pi^+$ decays reconstructed from the data and its performance has been determined. The calibration parameters are found to be

$$\begin{aligned} p_0 &= 0.408 \pm 0.009(\text{stat.}) \pm 0.007(\text{syst.}) \\ p_1 &= 0.86 \pm 0.12(\text{stat.}) \pm 0.10(\text{syst.}) \\ \langle \eta \rangle &= 0.4302\,. \end{aligned}$$

The effective tagging power has been determined to be $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2 = 2.42 \pm 0.39\%$ with a tagging efficiency of $\varepsilon_{\text{tag}} = 49.5 \pm 0.4\%$. This corresponds to an increase of about 56% in effective tagging power compared with a previous implementation of the same side tagging algorithm. It has been shown, that the calibrated same side kaon tagging algorithm can be combined with the opposite side tagging algorithms and that the combine predicted mistag probability is still calibrated.

In the final part of the thesis, the B_s^0–$\bar{B}_s^0$ oscillation frequency has been determined using the combination of the new same side kaon tagging algorithm and the opposite side tagging algorithms. The measured value is

$$\Delta m_s = 17.745 \pm 0.022\,(\text{stat.}) \pm 0.006\,(\text{syst.})\ \text{ps}^{-1}\,,$$

which is the most precise measurement to date. It has been shown that the measurement profits immensely from the improved same side kaon tagging

algorithm. By combining the improvements in the tagging with the more sophisticated selection procedure developed for the published LHCb result on the B_s^0–$\bar{B}_s^0$ oscillation, and taking into account the additional data sample of $2\,fb^{-1}$ taken in 2012, future improvements in the measurement precision of Δm_s are expected.

Appendices

A CP VIOLATING PARAMETERS IN $B_s^0 \to D_s^- K^+$

The following explanation of the CP violating parameters in the decay time PDF for $B_s^0 \to D_s^- K^+$ events has been taken from [44]. The time-dependent CP violation of the decay $B_s^0 \to D_s^- K^+$ has to be taken into account in the combined decay time PDF. The PDF for this decay is described in Equation 49. It includes the CP violating parameters C, D_{eff} and S_{eff}. They are defined as follows

$$C = \frac{1-\lambda^2}{1+\lambda^2} \tag{89}$$

with λ being the ratio of the two Feynman diagrams of $B_s^0 \to D_s^- K^+$ and $B_s^0 \to D_s^+ K^-$. It is set to $\lambda = 0.372$.

D_{eff} and S_{eff} depend on the charge of the Bachelor $q_f = +1, -1$ for K^+ and K^- respectively.

$$D_{eff} = \frac{1}{2}\left((1+q_f)D_f + (1-q_f)D_{\bar{f}}\right) \tag{90}$$

$$S_{eff} = \frac{1}{2}\left((1+q_f)S_f + (1-q_f)S_{\bar{f}}\right) \tag{91}$$

with the CP-violation parameters D_f, $D_{\bar{f}}$, S_f and $S_{\bar{f}}$

$$D_f = \frac{2\lambda\cos\left(\Delta - (\gamma+\phi_s)\right)}{1+\lambda^2}, \quad D_{\bar{f}} = \frac{2\lambda\cos\left(\Delta + (\gamma+\phi_s)\right)}{1+\lambda^2} \tag{92}$$

$$S_f = \frac{2\lambda\sin\left(\Delta - (\gamma+\phi_s)\right)}{1+\lambda^2}, \quad S_{\bar{f}} = \frac{2\lambda\sin\left(\Delta + (\gamma+\phi_s)\right)}{1+\lambda^2}, \tag{93}$$

which depend on the strong phase Δ, the CKM-angles γ and the weak phase ϕ_s.

B DESCRIPTION OF THE FUNCTION TO LIMIT THE MISTAG PROBABILITY IN THE PDF

The calibration parameters p_0 and p_1 of the predicted mistag probability are floating in the fit of the B_s^0 decay time distribution. To avoid that the calibrated mistag probability is taking on unphysical values, it is limited in the fit to the range [0, 0.5] by a smooth transition function which is shown in Figure 85 and is given by

$$
\begin{aligned}
R &= 0.1 \\
a_{Hi} &= 1 + R/\tan(3 \times \pi/8.0) \\
a_{Lo} &= a_{Hi} - R/\sqrt{2} \\
f &= \begin{cases}
0.5, & \text{if } 4x - 1 > a_{Hi} \\
(\sqrt{(R^2 - (x - a_{Hi})^2 + 1 - R}) + 1)/4, & \text{if } a_{Hi} > 4x - 1 > a_{Lo} \\
(x + 1)/4, & \text{if } a_{Lo} > x > -a_{Lo} \\
(-\sqrt{(R^2 - (x + a_{Hi})^2 - 1 + R}) + 1 + 1)/4, & \text{if } -a_{Hi} < 4x - 1 < -a_{Lo} \\
0, & \text{if } 4x - 1 < -a_{Hi}
\end{cases}
\end{aligned}
\tag{94}
$$

This function is chosen because it can be differentiated analytically which avoids problems in the fitting procedure, *cf.* [44].

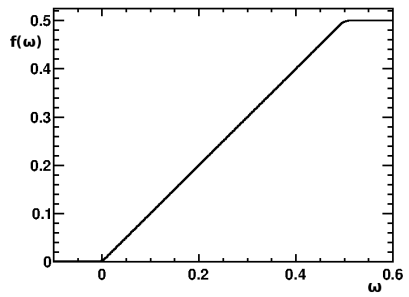


Figure 85: Function used to limit the mistag probability to the physical range of [0,0.5]. Figure taken from [44]

C TAGGING PERFORMANCE USING AN AVERAGE MISTAG FRACTION

In addition to the tagging performance using per-event mistag probabilities, also the performance of the new same side kaon tagging algorithm using an average mistag fraction is evaluated. This scenario is potentially beneficial in some cases, where the use of an per event mistag probability is not possible or not desirable.

In this case only the tagging decision is used in the fit. The tagging performance is depending on both the tagging efficiency and the mistag fraction. To find an optimal balance between the two, a cut is placed on the maximum predicted mistag probability and only events with a smaller predicted mistag probability are considered as tagged. For this, the predicted mistag probability is calibrated using the calibration parameters which have been determined in this chapter. Because the p_1 value found is smaller than 1 and the p_0 value smaller than $\langle\eta\rangle$, see Table 22, the largest predicted mistag probability after the calibration is 0.4680 per construction. Consequently, a value of 0.5 represents the default case where no cut is placed. The tagging performance using an average mistag fraction and a per event mistag probability for several cuts on the maximum predicted mistag probability are evaluated and listed in Table 45. A cut requiring the predicted mistag probability of the events to be smaller than 0.44 performs best when an average mistag fraction is used. The reached effective tagging power is $\varepsilon_{eff} = \varepsilon_{tag}(1-2\omega)^2 = 1.87 \pm 0.36\%$ at a tagging efficiency of $\varepsilon_{tag} = 32.01 \pm 0.36\%$. The additional tagging power when using per event mistag probabilities originates from the capability of the neural network to precisely predict the mistag probability of the single events.

cut on predicted ω	$\epsilon(\%)$	$\omega(\%)$	$\epsilon D^2(\%)$
average mistag fraction			
0.5	49.51±0.38	40.7±0.9	1.70±0.33
0.46	42.28±0.38	40.0±1.0	1.69±0.33
0.45	35.76±0.37	38.9±1.1	1.75±0.34
0.44	31.37±0.35	38.0±1.1	1.82±0.35
0.43	26.72±0.34	37.6±1.2	1.63±0.33

Table 45: Tagging performance for different cuts on the calibrated predicted mistag probability ω.

D USE OF PROMPT D_s^+ DECAYS FOR THE TRAINING OF THE NEURAL NETWORK

The fact that the neural network based same side kaon tagging algorithm also works for prompt D_s^+ decays, as discussed in Chapter 8.3, opens up another possibility: If the neural network trained on simulated B_s^0 works correctly for D_s^+ decays in the data, these should also be usable for training. Provided that sufficient statistics are available so that the full phase space of the B_s^0 is covered, the same side kaon tagging algorithm could be optimized and calibrated on data without resolving the B_s^0 oscillation, using the same techniques as for the opposite side tagging algorithms in the $B^+ \to J/\psi K^+$ decay channel.

To test this hypothesis, the second neural network used in the same side kaon tagging algorithm is trained on the sample of sideband subtracted prompt D_s^+ decays used in the previous chapter for the determination of the systematic uncertainties on the calibration parameters. The same network configuration is used as with the training on simulated events. As the cut on the first neural network has an efficiency of about 50%, 50.000 prompt D_s^+ are usable for the training which corresponds to one third of the statistics available for simulated B_s^0 decays. Figure 86 shows the response of the trained neural network for D_s^+ and D_s^- respectively. A good separation is visible between the two flavours.

The neural network based same side kaon tagging algorithm trained on the prompt D_s^+ sample is applied to the data and the calibration parameters and effective tagging power is determined. The calibration parameters are $p_0 = 0.430 \pm 0.008$, $p_1 = 0.991 \pm 0.104$ and $\langle \eta \rangle = 0.4234$. The calibration parameters are consistent with the expected values of $p_0 - \langle \eta \rangle = 0$ and $p_1 = 1$ within one standard deviation of their respective systematic uncertainty, which confirms that the neural network trained on the D_s^+ sample has predictive power in the B_s^0 sample. The effective tagging power is $\varepsilon_{\text{eff}} = 2.05 \pm 0.31\%$ at a tagging efficiency of $\varepsilon_{\text{tag}} = 49.5 \pm 0.4\%$. The effective tagging efficiency is about $\Delta\varepsilon_{\text{eff}} = 0.37\%$ lower compared to the default neural network based same side kaon tagging algorithm, where the second neural network is trained on simulated events. Possible reasons for the smaller performance can be either, that the prompt D_s^+ sample is smaller in statistics than the simulated B_s^0 sample or that the different phase space occupation of the D_s^+ with respect to the B_s^0 does not provide the best predicted mistag probability for all B_s^0 kinematics. If the same statistics as for the prompt D_s^+ sample is used in the simulated B_s^0 sample for the training of the second neural network, a tagging power of $\varepsilon_{\text{eff}} = 2.14 \pm 0.36\%$ is achieved, see Figure 80. This result can be considered compatible with the result from the prompt D_s^+ training.

With the additional statistics available from the other D_s^+ decay modes, $D_s^+ \to K^*K^+$ and $D_s^+ \to K^+K^-\pi$ non resonant decays, and the additional dataset taken in the 2012 run of the LHC, the statistics of the prompt D_s^+ sample can surpass the statistics available in the simulated sample. From the results that are discussed in this section and the previous chapter, it can be concluded that decays of prompt D_s^+ mesons can be used for a training

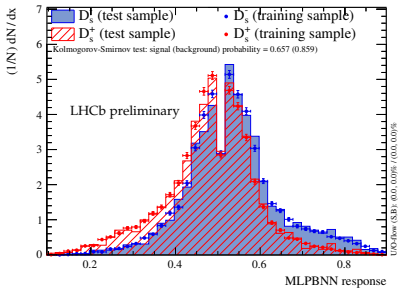


Figure 86: Distribution of the neural network response for a training of the second neural network using prompt D_s^+ and D_s^- decays extracted from the data.

and the calibration of the neural network same side kaon tagging algorithm and can supplement or replace the use of $B_s^0 \to D_s^- \pi^+$ decays. Further studies are necessary to exclude possible biases by the different kinematics of the D_s^+ and the B_s^0 fragmentation process, differences in the phase space of the two decays and influences of backgrounds from non prompt D_s^+ or other sources.

BIBLIOGRAPHY

[1] Ian C. Brock and Thomas Schorner-Sadenius. "Setting the Scene." In: *Physics at the Terascale*. Wiley, 2011.

[2] Markus Diehl and Wolfgang Hollik. "The Standard Model: Our Picture of the Microcosm." In: *Physics at the Terascale*. Wiley, 2011.

[3] Gudrun Hiller and Ulrich Uwer. "Quark Flavour Physics." In: *Physics at the Terascale*. Wiley, 2011.

[4] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Physics Letters B* 716.1 (2012), pp. 30 –61. ISSN: 0370-2693. DOI: http://dx.doi.org/10.1016/j.physletb.2012.08.021. URL: http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[5] 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." In: *Physics Letters B* 716.1 (2012), pp. 1 –29. ISSN: 0370-2693. DOI: http://dx.doi.org/10.1016/j.physletb.2012.08.020. URL: http://www.sciencedirect.com/science/article/pii/S037026931200857X.

[6] J. Beringer et al. "Review of particle physics." In: *Phys. Rev.* D86 (2012), p. 010001. DOI: 10.1103/PhysRevD.86.010001.

[7] R Aaij et al. "Measurement of b-hadron branching fractions for two-body decays into charmless charged hadrons." In: *JHEP* 10 (2012), p. 37. DOI: 10.1007/JHEP10(2012)037. arXiv:1206.2794 [hep-ex].

[8] Suvayu Ali et al. "Measurement of time-dependent CP-violation observables in $B_s^0 \to D_s^{\mp} K^{\pm}$." In: LHCb-ANA-2012-068 (2012).

[9] J. Charles et al. "CKM fitter group." In: *Eur.Phys.J.* C41 (2005), pp. 1–131.

[10] H. Albrecht et al. "Observation of Bo-Bo mixing." In: *Physics Letters B* 192.1,Äì2 (1987), pp. 245 –252. ISSN: 0370-2693. DOI: http://dx.doi.org/10.1016/0370-2693(87)91177-4. URL: http://www.sciencedirect.com/science/article/pii/0370269387911774.

[11] A. Abulencia et al. "Observation of Bo(s) - anti-Bo(s) Oscillations." In: *Phys.Rev.Lett.* 97 (2006), p. 242003. DOI: 10.1103/PhysRevLett.97.242003. arXiv:hep-ex/0609040 [hep-ex].

[12] V.M. Abazov et al. "First direct two-sided bound on the B_s^0 oscillation frequency." In: *Phys.Rev.Lett.* 97 (2006), p. 021802. DOI: 10.1103/PhysRevLett.97.021802. arXiv:hep-ex/0603029 [hep-ex].

[13] R. Aaij et al. "Measurement of the B_s^0–$\overline{B}_s^0$ oscillation frequency Δm_s in the decay $B_s^0 \to D_s^+ \pi^-$." In: *New J. Phys.* 15 (2013), p. 053021. DOI: 10.1088/1367-2630/15/5/053021. arXiv:1304.4741 [hep-ex].

[14] R. Aaij et al. "Measurement of CP-violation and the B_s^0-meson decay width difference with $B_s^0 \to J/\psi K^+ K^-$ and $B_s^0 \to J/\psi \pi^+ \pi^-$ decays." In: (2013). to appear in Phys. Rev. D. arXiv:1304.2600 [hep-ex].

[15] P. Nason *et al.*. *Bottom Production*. arXiv:hep-ph/0003142v2, 2001.

[16] R. Aaij et al. "Measurement of J/ψ production in pp collisions at $\sqrt{s} = 2.76\,\text{TeV}$." In: *JHEP* 02 (2013), p. 41. DOI: 10.1007/JHEP02(2013)041. arXiv:1212.1045 [hep-ex].

[17] J. Nardulli. *Reconstruction of two-body B decays in LHCb*. CERN-THESIS-2007-063, 2007.

[18] G. Altarelli and M.L. Mangano (editors). *Standard model physics (and more) at the LHC*. CERN-2000-004, 2000.

[19] R. Hierck. *Optimisation of the LHCb detector*. CERN-THESIS-2003-025, 2003.

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

[21] R. Aaij et al. "The LHCb trigger and its performance in 2011." In: *JINST* 8 (2013), P04022. DOI: 10.1088/1748-0221/8/04/P04022. arXiv:1211.3055 [hep-ex].

[22] Christian Linn. "Measurement of the CP-violating phase Φ_s using $B_s^0 \to J/\psi\,\phi$ and $B_s^0 \to J/\psi\,\pi^+\pi^-$ decays with the LHCb experiment." PhD thesis. Heidelberg University, Heidelberg, 2013.

[23] Torbjörn Sjöstrand, Stephen Mrenna, and Peter" Skands. "PYTHIA 6.4 physics and manual." In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv:hep-ph/0603175 [hep-ph].

[24] I. Belyaev et al. "Handling of the generation of primary events in Gauss, the LHCb simulation framework." In: *Nuclear Science Symposium Conference Record (NSS/MIC)* IEEE (2010), p. 1155. DOI: 10.1109/NSSMIC.2010.5873949.

[25] D. J. Lange. "The EvtGen particle decay simulation package." In: *Nucl. Instrum. Meth.* A462 (2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[26] S. Agostinelli et al. "Geant4: a simulation toolkit." In: *Nucl. Instrum. Meth.* A506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[27] T Brambach et al. "Optimization and Calibration of the Flavour Tagging performance using 2010 data." In: LHCb-ANA-2011-003 (2011).

[28] R. Aaij et al. "Opposite-side flavour tagging of B mesons at the LHCb experiment." In: *Eur.Phys.J.* C72 (2012), p. 2022. DOI: 10.1140/epjc/s10052-012-2022-1. arXiv:1202.4979 [hep-ex].

[29] "Performance of flavor tagging algorithms optimised for the analysis of $B_s^0 \to J/\psi\phi$." In: LHCb-CONF-2012-026 (2012).

[30] X. Artru and G. Mennessier. "String model and multiproduction." In: *Nucl.Phys.* B70 (1974), pp. 93–115. DOI: 10.1016/0550-3213(74)90360-5.

[31] M.G. Bowler. "e+ e- Production of Heavy Quarks in the String Model." In: *Z.Phys.* C11 (1981), p. 169. DOI: 10.1007/BF01574001.

[32] Bo Andersson, G. Gustafson, and B. Soderberg. "A General Model for Jet Fragmentation." In: *Z.Phys.* C20 (1983), p. 317. DOI: 10.1007/BF01407824.

[33] Bo Andersson, G. Gustafson, and B. Soderberg. "A PROBABILITY MEASURE ON PARTON AND STRING STATES." In: *Nucl.Phys.* B264 (1986), p. 29. DOI: 10.1016/0550-3213(86)90471-2.

[34] J. Abdallah et al. "A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole." In: *Eur.Phys.J.* C71 (2011), p. 1557. DOI: 10.1140/epjc/s10052-011-1557-x. arXiv:1102.4748 [hep-ex].

[35] Stefan Gieseke and Zoltán Nagy. "Monte Carlo Generators and Fixed-order Calculations: Predicting the (Un)Expected." In: *Physics at the terascale*. Wiley, 2011.

[36] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "PYTHIA 6.4 Physics and Manual." In: *JHEP* 0605 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv:hep-ph/0603175 [hep-ph].

[37] Thomas Nikodem. "Study of the Performance of a Same-Side-Kaon-Tagging Algorithm for the LHCb Experiment using $D_s^+ \to \phi\pi^+$ and $B_s^0 \to D_s^-\pi^+$ decays." diploma thesis. Heidelberg University, Heidelberg, 2011.

[38] M Meissner. "Measurement of Charged Particle Multiplicities in pp-collisions with Forward Tracks." In: LHCb-ANA-2011-084 (2011).

[39] Georg Krocker et al. "Optimization and calibration of the Same Side Kaon tagging algorithm using hadronic B_s^0 decays in 2011 data." In: LHCb-ANA-2011-043 (2011).

[40] LHCb collaboration. "Optimization and calibration of the same-side kaon tagging algorithm using hadronic B_s^0 decays in 2011 data." In: LHCb-CONF-2012-033 (2012).

[41] Sebastian Wandernoth. "Measurement of the B_s^0–$\bar{B}_s^0$ Mixing Frequency and Calibration of the Same Side Tagger at LHCb." diploma thesis. Heidelberg University, Heidelberg, 2009.

[42] K Akiba et al. "Measurement of Δm_s in the decay $B_s \to D_s(K^+K^-\pi^-)\pi^+$ and $B_s \to D_s(K^+K^-\pi^-)\pi^+\pi^+\pi^-$." In: LHCb-ANA-2011-005 (2011).

[43] Stephanie Hansmann-Menzemer, Georg Krocker, and Sebastian Wandernoth. "Measurement of Δm_s in the decay $B_s \to D_s(K+K-\pi-)\pi+$ and $B_s \to D_s(K+K-\pi-)\pi+\pi+\pi-$ in 2011 data." In: LHCb-ANA-2011-043 (2011).

[44] Stephanie Hansmann-Menzemer, Georg Krocker, and Sebastian Wandernoth. "Measurement of Δm_s in the decay $B_S^0 \to D_s^-\pi^+$ using $1fb^{-1}$." In: LHCb-ANA-2011-053 (2012).

[45] LHCb collaboration. "Measurement of Δm_s in the decay $B_s^0 \to D_s^-(K^+K^-\pi^-)(3)\pi$." In: LHCb-CONF-2011-005 (2011).

[46] LHCb collaboration. "Measurement of Δm_s in the decay $B_s^0 \to D_s^-(K^+K^-\pi^-)\pi^+$ using opposite-side and same-side flavour tagging algorithms." In: LHCb-CONF-2011-050 (2011).

[47] Vladimir V Gligorov, Christopher Thomas, and Michael Williams. *The HLT inclusive B triggers*. Tech. rep. LHCb-PUB-2011-016. CERN-LHCb-PUB-2011-016. LHCb-INT-2011-030. LHCb-INT-2011-030. Geneva: CERN, 2011.

[48] B. P. Roe et al. "Boosted decision trees as an alternative to artificial neural networks for particle identification." In: *Nucl.Instrum.Meth.* A543 (2005), pp. 577–584. DOI: 10.1016/j.nima.2004.12.018. arXiv:physics/0408124 [physics].

[49] L. Breiman et al. *Classification and regression trees*. Belmont, California, USA: Wadsworth international group, 1984.

[50] Robert E. Schapire and Yoav Freund. "A decision-theoretic generalization of on-line learning and an application to boosting." In: *Jour. Comp. and Syst. Sc.* 55 (1997), p. 119. DOI: 10.1006/jcss.1997.1504.

[51] M Williams et al. *The HLT2 Topological Lines*. Tech. rep. LHCb-PUB-2011-002. CERN-LHCb-PUB-2011-002. Geneva: CERN, 2011.

[52] Muriel Pivk and Francois R. Le Diberder. "sPlot: a statistical tool to unfold data distributions." In: *Nucl.Instrum.Meth.* A555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv:physics/0402083 [physics.data-an].

[53] G. Punzi. "Comments on Likelihood fits with variable resolution." In: *ArXiv Physics e-prints* (Jan. 2004). eprint: arXiv:physics/0401045.

[54] Stephanie Hansmann-Menzemer, Katharina Kreplin, and Georg Krocker. "The Opposite-side kaon tagger: Data MC performance comparison and optimization using neural networks." In: LHCb-INT-2013-014 (2013).

[55] R. Aaij et al. "Measurement of the $D^{\pm}$ production asymmetry in 7 TeV pp collisions." In: *Phys. Lett.* B718 (2013), 902,Äì909. DOI: 10.1016/j.physletb.2012.11.038. arXiv:1210.4112 [hep-ex].

[56] Andreas Hoecker et al. "TMVA 4, Toolkit for Multivariate Data Analysis with ROOT, Users Guide." In: 040 (2009). arXiv:physics/0703039.

[57] Jiahang Zhong, Run-Sheng Huang, and Shih-Chang Lee. "A program for the Bayesian Neural Network in the ROOT framework." In: *Computer Physics Communications* 182.12 (2011), pp. 2655–2660. arXiv:1103.2854.

[58] D F Shanno. "Conditioning of Quasi-Newton Methods for Function Minimization." In: *Mathematics of Computation* 24.111 (1970), pp. 647–656.

[59] D Goldfarb. "A Family of Variable-Metric Methods Derived by Variational Means." In: *Mathematics of Computation* 24.109 (1970), pp. 23–26.

[60] R Fletcher. "A new approach to variable metric algorithms." In: *The Computer Journal* 13.3 (1970), pp. 317–322.

[61] C. G. Broyden. "The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations." In: *IMA J Appl Math* 6.1 (1970), pp. 76–90.

[62] H Gordon et al. "A Measurement of the $K\pi$ Detection Asymmetry." In: LHCb-INT-2012-027 (2013).

[63] R. Aaij et al. "Measurement of the CP-violating phase ϕ_s in $\bar{B}^0_s \to J/\psi\pi^+\pi^-$ decays." In: *Phys.Lett.* B713 (2012), pp. 378–386. DOI: 10.1016/j.physletb.2012.06.032. arXiv:1204.5675 [hep-ex].