Dissertation

submitted to the
Combined Faculties for the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg, Germany
for the degree of

Doctor of Natural Sciences

Put forward by
Katharina Kreplin
born in Nürnberg, Germany

Oral examination: 10.06.2015

A Novel Flavour Tagging Algorithm using Machine Learning Techniques and a Precision Measurement of the B^0–$\overline{B}^0$ Oscillation Frequency at the LHCb Experiment

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

Abstract   This thesis presents a novel flavour tagging algorithm using machine learning techniques and a precision measurement of the B^0–$\overline{B}^0$ oscillation frequency Δm_d using semileptonic B^0 decays. The LHC Run I data set is used which corresponds to $3\,\text{fb}^{-1}$ of data taken by the LHCb experiment at a center-of-mass energy of $7\,\text{TeV}$ and $8\,\text{TeV}$. The performance of flavour tagging algorithms, exploiting the $b\bar{b}$ pair production and the b quark hadronization, is relatively low at the LHC due to the large amount of soft QCD background in inelastic proton-proton collisions. The standard approach is a cut-based selection of particles, whose charges are correlated to the production flavour of the B meson. The novel tagging algorithm classifies the particles using an artificial neural network (ANN). It assigns higher weights to particles, which are likely to be correlated to the b flavour. A second ANN combines the particles with the highest weights to derive the tagging decision. An increase of the opposite side kaon tagging performance of 50% and 30% is achieved on $B^+ \to J/\psi\,K^+$ data. The second number corresponds to a readjustment of the algorithm to the B^0_s production topology. This algorithm is employed in the precision measurement of Δm_d. A data set of 3.2×10^6 semileptonic B^0 decays is analysed, where the B^0 decays into a D^- ($K^+\ \pi^-\ \pi^-$) or D^{*-} ($\pi^-\ \overline{D}^0$ ($K^+\ \pi^-$)) and a $\mu^+\nu_\mu$ pair. The ν_μ is not reconstructed, therefore, the B^0 momentum needs to be statistically corrected for the missing momentum of the neutrino to compute the correct B^0 decay time. A result of $\Delta m_d = 0.503 \pm 0.002\,(\text{stat}) \pm 0.001\,(\text{syst})\,\text{ps}^{-1}$ is obtained. This is the world's best measurement of this quantity.

Zusammenfassung   In dieser Arbeit wird ein neuartiger, auf maschinellem Lernen basierender Flavour Tagging Algorithmus sowie eine Präzisionsmessung der B^0–$\overline{B}^0$ Oszillationsfrequenz Δm_d mittels semileptonischer B^0 Zerfälle vorgestellt. Die Daten des ersten LHC Laufes werden genutzt; diese entsprechen $3\,\text{fb}^{-1}$, aufgenommen vom LHCb Experiment bei einer Schwerpunktsenergie von $7\,\text{TeV}$ und $8\,\text{TeV}$. Die Leistung von Flavour Tagging Algorithmen, welche die paarweise $b\bar{b}$ Produktion und die b Quark Hadronisierung ausnutzen, ist durch den großen Anteil an niederenergetischem QCD Untergrund in inelastischen Proton-Proton Kollisionen am LHC relativ niedrig. Der Standardansatz ist eine schnittbasierte Selektion von Teilchen, deren Ladungen mit dem Produktionsflavour des B Mesons korreliert sind. Der neuartige Tagging Algorithmus klassifiziert die Teilchen mit Hilfe eines künstlichen Neuronalen Netzes (ANN). Dieses weist umso höhere Gewichte den Teilchen zu, je wahrscheinlicher diese mit dem b Flavour korreliert sind. Ein zweites ANN kombiniert die höchstgewichteten Teilchen um die Taggingentscheidung zu treffen. Die Leistungssteigerung des Opposite Side Kaon Tagging Algorithmus, gemessen auf $B^+ \to J/\psi\,K^+$ Daten, beträgt 50% bzw. 30% nach Anpassung auf die B^0_s Produktiontopologie. Dieser Algorithmus wird in der Δm_d Präzisionsmessung genutzt. Ein Datensatz von 3.2×10^6 semileptonischen B^0 Zerfällen, wobei das B^0 in ein D^- ($K^+\ \pi^-\ \pi^-$) oder D^{*-} ($\pi^-\ \overline{D}^0$ ($K^+\ \pi^-$)) und ein $\mu^+\nu_\mu$ Paar zerfällt, wird analysiert. Das ν_μ wird nicht rekonstruiert, daher muss der B^0 Impuls statistisch auf den fehlenden Impuls des Neutrinos korrigiert werden um die korrekte B^0 Zerfallszeit zu berechnen. Als Resultat wird $\Delta m_d = 0.503 \pm 0.002\,(\text{stat}) \pm 0.001\,(\text{syst})\,\text{ps}^{-1}$ erzielt. Dies ist die weltbeste Messung dieser Größe.

Contents

Introduction

Fundamental particles and their interactions are described by the Standard Model (SM) of Particle Physics. Developed in the 1960's and 70's, it combines all our current knowledge these days. Decades of collider experiments have not disproved its precise predictions. The last missing predicted particle, the Higgs boson, has been discovered by ATLAS and CMS in 2012 at the Large Hadron Collider (LHC).

However, fundamental questions remain unanswered. Why is gravity that weak at the electroweak scale? What prevents gravity quantum corrections $\mathcal{O}(m_{\mathrm{Planck}})$ from modifying SM processes? When the SM describes all visible matter, what are the other 95% of the observed energy density of the universe? What drove the early evolution of the universe e.g. its inflationary phase? Where did all the anti-matter go?

The LHC was built to find answers to these fundamental questions. It produces collision energies and intensities never reached before on earth. It opens a window into conditions very close to the Big Bang. In this environment beyond the SM effects, so called New Physics (NP), might show up to bring the desired answers. NP can show up either directly in form of heavy new particles at the TeV scale, or indirectly via quantum loop corrections to the SM processes at the multiple TeV scale.

LHCb performs precision measurements in the beauty sector of the SM, which is sensitive to quantum loop corrections. Thus, it is sensitive to indirect contributions from possible NP. High $b\bar{b}$ and $c\bar{c}$ production rates at the LHC enable precise CP[1] asymmetry measurements. The SM mechanism of CP violation explains the dominance of particles over anti-particles, however, orders of magnitudes too small. To explain the matter access observed in the universe NP contributions are possible, e.g. in the semileptonic CP asymmetry in the B^0–$\bar{B}^0$ system (a_{sl}^d). A very small, $\mathcal{O}(10^{-4})$, CP asymmetry is precisely predicted by the SM, while the experimental precision on a_{sl}^d, $\mathcal{O}(10^{-3})$ [1], offers quite some room for NP contributions. The measurement of a_{sl}^d is time-dependent, thus, the measurement of a_{sl}^d also depends on the B^0–$\bar{B}^0$ oscillation frequency Δm_d. If LHCb wants to pin down a_{sl}^d, it has to demonstrate the ability to precisely measure Δm_d in semileptonic B^0 decays[2], which is presented in this thesis.

Semileptonic B^0 decays, where the B^0 decays into a charmed meson and a muon plus muon neutrino, have the advantage of high statistics. The muon leaves a signature which is easy to detect. However, the muon neutrino is not reconstructible. This poses a challenge to the mixing analysis because the decay reconstruction is kinematically under-constrained. Especially the reconstructed momentum and thus the decay time and mass resonance of the B^0 are underestimated by the missing momentum of the neutrino. Therefore, the B^0 decay time is statistically corrected using Monte Carlo (MC) simulation, which is one of the technical challenges of the analysis.

---

[1] The combination of charge conjugation (C) and parity transformation (P) stands for particle anti-particle conjugation.

[2] Charge conjugation of particles is implied unless explicitly stated.

Another challenge is the identification of the B production flavour. To determine the b quark flavour experimentally so-called flavour tagging algorithms are used. These algorithms exploit the $b\bar{b}$ pair production and the signal b hadronization process. However, B mesons produced in inelastic proton-proton interactions are accompanied by plenty of soft QCD background, which mainly consists of kaons and pions coming from the primary interaction. That is a challenge for the performance of flavour tagging algorithms. Therefore, the author has developed a novel flavour tagging algorithm using machine learning techniques. So far tagging particles were selected by cutting on kinematical and quality variables. This is replaced by a multi-objective optimizing algorithm based on machine learning.

Two artificial neural networks (ANN) have been trained. A first ANN classifies the reconstructed particles to find the correct tagging particles. It assigns higher weights to particles which are likely to be correlated to the b flavour and lower weights to those which are less likely to be correlated to the flavour of the b hadron. A second ANN combines the particles with the highest weights to derive the b flavour. Using these techniques a significant improvement in the performance of the opposite side kaon tagging algorithm is achieved. The new algorithm has been documented LHCb internally in [2] and is provided to the collaboration.

The measurement of Δm_d benefits from the new flavour tagging algorithm. The measurement, currently under review within the LHCb collaboration, will be published, soon. Details about the measurement are documented LHCb internally in [3].

The thesis is structured as follows. First, an introduction to the theoretical background, relevant for the measurement of Δm_d, is given in chapter 1. The LHCb experiment is outlined in chapter 2. Chapter 3 presents the work done on flavour tagging. To understand the challenges for flavour tagging at the LHC a study to evaluate the differences in tagging performance in MC compared to data is performed. Then, the novel flavour tagging algorithm is introduced, followed by a comparison to the existing LHCb flavour tagging algorithms. Systematical effects that influence the calibration and application of the new algorithm are determined. Chapter 4 presents the precision measurement of Δm_d. After demonstrating the analysis strategy, the determination of the relevant measurement inputs is revealed. A systematic uncertainty on Δm_d is determined and the final result is given. At the end, a conclusion of the work presented in this thesis is given. In this thesis natural units are used, i.e. $\hbar = c = 1$.

1. Theory

This chapter gives a brief introduction to the Standard Model of particle physics. It highlights the part of the theory which is necessary for the measurements presented in this thesis. A detailed introduction is given for example in [4, 5, 6, 7, 8].

1.1. The Standard Model of Particle Physics

The Standard Model (SM) of particle physics is a theory to describe fundamental particles and their interactions. Fundamental particles are quarks and leptons, depicted in figure 1.1, building up all visible matter. Interactions between these particles, described by the SM, are the electro magnetic force, binding electrons to form atoms, the strong force, binding protons and neutrons to form the nucleus of an atom, and the weak force, responsible for the decay of quarks and leptons. The fourth fundamental force, gravity, is not included in the SM.
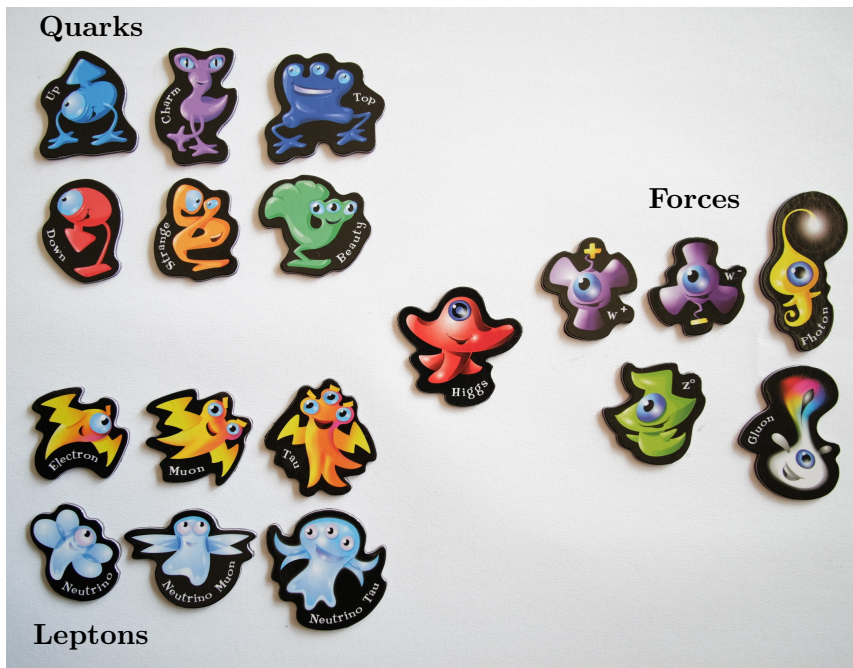

Figure 1.1.: Fundamental particles and force carriers of the Standard Model; stickers available at CERN gift shop.

1. Theory

Mathematically, the SM is a renormalizable Quantum Field Theory, exploiting the principle of local gauge invariance. The choice of gauge transformations is motivated by symmetries observed in nature, connected to the conservation of quantum numbers. The SM Lagrangian $\mathcal{L}$ is invariant under local transformations of the

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{1.1}$$

symmetry group. Quarks and leptons are realized as matter fields with spin $\frac{1}{2}$. There are 6 different quarks, grouped into three generations, each of them building a doublet of an up-type and a down-type quark. The up-type quarks are called up, charm and top. They carry an electrical charge of $+2/3$, while the down-type quarks, called down, strange and bottom, carry a charge of $-1/3$ in units of the elementary charge e. There are three charged leptons, the electron e, the muon μ and the tau τ, each belonging to another generation. The leptons are accompanied by electrically neutral, neutrinos, the electron neutrino ν_e, the muon neutrino ν_μ and the tau neutrino ν_τ, see figure 1.1. Forces are mediated by bosons, spin 1 particles. They are the generators of the corresponding symmetry group.

The strong force is described by Quantum Chromodynamics with the $SU(3)_C$ symmetry group. Its generators are eight gluons. The conserved quantum numbers are 3 colors, red, blue and green, carried by quarks, while anti-quarks carry anti-colors. The gluons are massless and electrically neutral, but carry color charge. Hence, they couple to color. Two characteristics of the strong force are the principles of confinement and asymptotic freedom. Confinement allows color neutral objects, hadrons, only. Hadrons are either composed of a quark anti-quark ($q\bar{q}$) pair to form a meson or of three quarks, forming a baryon. Hence, quarks are not free and they are produced in $q\bar{q}$ pairs, only, given the strong interaction as production mechanism. However, due to the running strong coupling constant, which is very large at low energies, but small at high energies they are quasi-free at small distances, called asymptotic freedom.

The electromagnetic and weak forces, unified to the electroweak force [9, 10, 11], are given by the $SU(2)_L \otimes U(1)_Y$ symmetry group. Its generators are the massless photon γ, and the massive gauge bosons W^+ and W^-, carrying the electric charge $\pm 1e$, and Z, which is electrically neutral. Their masses are generated by spontaneous broken symmetry, described by the Higgs mechanism [12, 13]. An additional scalar (spin 0) field is introduced, that has a non-zero vacuum expectation value. An excitation of this field results in a Higgs boson, discovered by the ATLAS and CMS experiments at CERN in July 2012 [14, 15]. Fermion masses are generated by Yukawa interactions between the Higgs and the fermion fields. The massless γ couples to the electric charge of fermions. The Z boson couples with different strength to left- and right-handed fermions, denoted as chiral property. Whereas the charged $W^\pm$ bosons couple only to left-handed fermions or right-handed anti-fermions. Corresponding to the electric charge of the bosons, these processes are called neutral and charged currents, respectively. The charged currents cause transitions of up- to down-type quarks and charged to neutral leptons, or vice versa. The conserved quantum numbers are the weak isospin T, its third component T_3 and the weak hypercharge $Y = Q - T_3$, with Q being the electric charge in units e.

1.2. Quark Sector of the Standard Model

Quark masses are generated by the Yukawa interaction terms in the SM $\mathcal{L}$. Their mass eigenstates q are not identical to the weak eigenstates q' [16]. They are connected by an unitary transformation matrix $\boldsymbol{V}_{\text{CKM}}$:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \boldsymbol{V}_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{1.2}$$

which rotates the down-type quarks, by convention. In the weak basis the charged current induces transitions within the doublet of each generation, only. However, transitions between different generations of quarks also occur, because $\boldsymbol{V}_{\text{CKM}}$ is not diagonal. $\boldsymbol{V}_{\text{CKM}}$ is the Cabibbo-Kobayashi-Maskawa (CKM) matrix

$$\boldsymbol{V}_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.3}$$

The matrix V_{CKM} has 4 free parameters, three amplitudes and one phase, that causes *CP* symmetry violation within the SM. The particle anti-particle conjugation *CP* is a combination of the parity transformation P, that inverts the spatial coordinates of a state, $P\vec{x} = -\vec{x}$, and the charge conjugation C, that inverts the electric charge of a particle $C|e^-\rangle = |e^+\rangle$. The single symmetries C and P are maximally violated by the weak interaction. Therefore, charged currents couple to left-handed particles or right-handed anti-particles, only. However, the combined *CP* transformation is conserved up to a small *CP* violating phase.

The Wolfenstein parameterization illustrates the hierarchy of the CKM matrix elements [17]. Using the expansion parameter $\lambda \sim 0.23$ [18], the CKM matrix can be written as

$$\boldsymbol{V}_{\text{CKM}} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3 (\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3 (1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{1.4}$$

where the Wolfenstein parameters are defined as

$$\begin{aligned} \lambda &= \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \\ A\lambda^2 &= \lambda \frac{|V_{cb}|}{|V_{us}|}, \\ A\lambda^3 (\rho + i\eta) &= V_{ub}^*. \end{aligned} \tag{1.5}$$

The diagonal elements are close to 1, the probability of transitions within the generation is most likely. Transitions between the first and second, or second and third generations are suppressed by λ and λ^2. Transitions between the first and third generation, suppressed by λ^3, involve the *CP* violating phase. The CKM matrix elements are fundamental parameters of the SM. They are determined experimentally. Using the unitarity condition of the CKM matrix, one can construct six unitarity triangles of the form [18]

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{1.6}$$

The corresponding angles are defined as

$$\alpha \equiv \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta \equiv \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma \equiv \left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{1.7}$$

The goal of quark flavour physics is to over-constrain these triangles by the combination of different measurements, see figure 1.2. Up to now, all measurements have confirmed the SM.
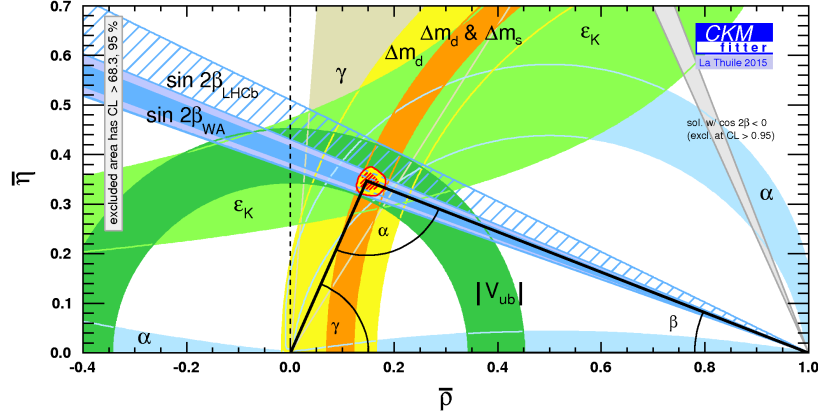
Figure 1.2.: Unitarity triangle, see equation 1.6, in the $\bar{\rho}-\bar{\eta}$ plane. Including the new LHCb $\sin 2\beta$ measurement (blue dashed band), presented at the La Thuile 2015 conference [19], into the World Average of $\sin 2\beta$ (blue band) leaves not much space for NP to break the triangle condition. Taken from [20]

The power of flavour physics measurements is, that they are sensitive to higher energy scales, not directly accessible by the center-of-mass energy at the collider, where New Physics (NP) might occur. NP can enter via higher order corrections to the processes studied. Especially rare processes, dominated by loop corrections, are sensitive to NP and can be used to extract CKM matrix elements, only if the SM is assumed. Neutral meson mixing is an example of second order corrections, discussed in the next section.

1.3. B^0–$\overline{B}^0$ Mixing Phenomenology

The weak interaction causes flavour oscillations of neutral mesons, B_q^0 oscillate into $\overline{B}_q^0$ mesons and vice versa, see figure 1.3. Here, q stands for down or strange quarks. These mixing processes belong to the class of flavour-changing neutral current processes, involving different flavours with the same electric charge. In the SM, they are forbidden at tree level. Possible NP heavy particles can appear as virtual particles in the loop diagram.

Flavour eigenstates are linear superpositions of mass eigenstates

$$|B_q^0\rangle \quad = \frac{1}{2p}(|B_{q,L}^0\rangle + |B_{q,H}^0\rangle) \tag{1.8}$$

$$|\overline{B}_q^0\rangle \quad = \frac{1}{2q}(|B_{q,L}^0\rangle - |B_{q,H}^0\rangle). \tag{1.9}$$
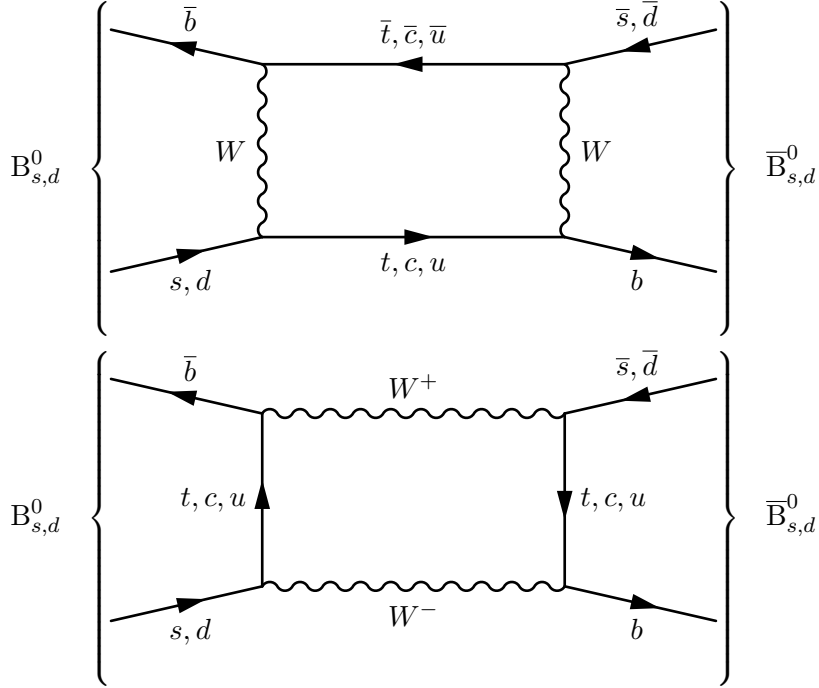
Figure 1.3.: Feynman diagrams of the leading order contributions to B^0–$\overline{B}^0$ mixing. B^0_q oscillate into $\overline{B}^0_q$ mesons via an internal loop of up-type quarks and $W^{\pm}$ bosons.

The indices L, H denote light and heavy in terms of masses, respectively. The parameters q and p fulfill the normalization condition $|p|^2 + |q|^2 = 1$. They are a measure of the decoupling of the flavour and mass eigenstates. The oscillation process is time-dependent, described by the phenomenological Schrödinger equation

$$i\frac{d}{dt}\begin{pmatrix}|B^0_q(t)\rangle \\ |\overline{B}^0_q(t)\rangle\end{pmatrix} = (\mathcal{M} - i\frac{\Gamma}{2})\begin{pmatrix}|B^0_q(t)\rangle \\ |\overline{B}^0_q(t)\rangle\end{pmatrix} \tag{1.10}$$

$\mathcal{M}$ and Γ are hermitian 2×2 mass and decay matrices. *CPT* symmetry enforces the diagonal elements to be equal

$$M_{11} = M_{22}, \qquad \Gamma_{11} = \Gamma_{22}. \tag{1.11}$$

The off-diagonal mass matrix element M_{12} stems from the dispersive (real) part of the box diagram in figure 1.3, given by the internal top loop. While the off-diagonal decay matrix element Γ_{12} stems from the absorptive (imaginary) part of the box diagram, given by the internal charm and up quark loop [21]. The eigenvectors that diagonalize the Hamiltonian $\mathcal{M} - \frac{i}{2}\Gamma$, are given by the time-dependent mass eigenstates

$$\begin{aligned}|B_H(t)\rangle &= e^{-(iM_H + \Gamma_H/2)t}|B_H\rangle, \tag{1.12}\\ |B_L(t)\rangle &= e^{-(iM_L + \Gamma_L/2)t}|B_L\rangle, \tag{1.13}\end{aligned}$$

where $M_{L,H}$ are the masses of the light and heavy mass eigenstates and $\Gamma_{L,H}$ their decay widths. The actual observable quantities are the average mass and decay width and the difference in masses and decay widths

$$m = \frac{M_H + M_L}{2} = M_{11} = M_{22}, \qquad\qquad \Delta m = M_H - M_L, \qquad (1.14)$$

$$\Gamma = \frac{\Gamma_L + \Gamma_H}{2} = \Gamma_{11} = \Gamma_{22}, \qquad\qquad \Delta\Gamma = \Gamma_L - \Gamma_H. \qquad (1.15)$$

The mass difference Δm_q, where $q = b, s$, is identified with the mixing frequency of the neutral B_q^0 meson oscillation. It is typically measured in ps^{-1}, setting $\hbar$ to one. The observable B_q^0–$\bar{B}_q^0$ meson oscillations can than be expressed as

$$|B_q^0(t)\rangle = g_+(t)|B_q^0\rangle + \frac{q}{p}g_-(t)|\bar{B}_q^0\rangle \qquad (1.16)$$

$$|\bar{B}_q^0(t)\rangle = \frac{p}{q}g_-(t)|B_q^0\rangle + g_+(t)|\bar{B}_q^0\rangle, \qquad (1.17)$$

where $g_+(0) = 1$ and $g_-(0) = 0$ and $g_\pm(t) \neq 0$ for $t > 0$ if $\Delta\Gamma \neq 0$. Hence, an initially produced B_q^0 will never turn into a pure $\bar{B}_q^0$ or back into a pure B_q^0. The time dependence of the mixing process is given by

$$g_+(t) = e^{-i\,m\,t}e^{-i\Gamma\,t/2}\left[\cosh\frac{\Delta\Gamma\,t}{4}\cos\frac{\Delta m\,t}{2} - i\sinh\frac{\Delta\Gamma\,t}{4}\sin\frac{\Delta m\,t}{2}\right], \qquad (1.18)$$

$$g_-(t) = e^{-i\,m\,t}e^{-i\Gamma\,t/2}\left[-\sinh\frac{\Delta\Gamma\,t}{4}\cos\frac{\Delta m\,t}{2} + i\cosh\frac{\Delta\Gamma\,t}{4}\sin\frac{\Delta m\,t}{2}\right]. \qquad (1.19)$$

Measurements of the mixing frequencies Δm_d and Δm_s constrain the CKM matrix elements, because

$$\Delta m_d \propto |V_{tb}V_{td}^*|, \qquad \Delta m_s \propto |V_{tb}V_{ts}^*| \qquad (1.20)$$

Compare to figure 1.2, where the measurements of Δm_d and Δm_s, indicated by yellow and orange circular bands, give constraints on the length of one side of the triangle $|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}|$.

1.3.1. Time-Dependent Mixing Asymmetry

This thesis presents a precision measurement of Δm_d using semileptonic B^0 decays. Semileptonic decays are flavour specific. That means, the final state f can only be reached by B decays not by $\bar{B}$ decays. However, if the $\bar{B}$ meson mixes into a B meson first, then it can also decay into f. The time-dependent mixing asymmetry $\mathcal{A}$, typically used to measure the oscillation frequency Δm, is defined as [6]

$$\mathcal{A}(t) \equiv \frac{\Gamma(B(t) \to f) - \Gamma(B(t) \to \bar{f})}{\Gamma(B(t) \to f) + \Gamma(B(t) \to \bar{f})} + c.c. \qquad (1.21)$$

Charge conjugated decays are included. Hence, the mixing asymmetry is given by the difference of the number of B mesons which do not change flavour between production and decay (unmixed)

and the number of B mesons which change flavour between production and decay (mixed), divided by the sum

$$\mathcal{A}(t) = \frac{N(t)_{\text{unmixed}} - N(t)_{\text{mixed}}}{N(t)_{\text{unmixed}} + N(t)_{\text{mixed}}} \tag{1.22}$$

$$= \frac{\cos(\Delta m\, t)}{\cosh(\Delta\Gamma\, t/2)} + \frac{a}{2}\left[1 - \frac{\cos^2(\Delta m\, t)}{\cosh^2(\Delta\Gamma\, t/2)}\right], \tag{1.23}$$

where a is a measure of the amount of *CP* violation in mixing, defined as

$$a = 1 - \left|\frac{q}{p}\right|^2. \tag{1.24}$$

In the B meson systems a is $\mathcal{O}(10^{-3})$ [22]. In order to measure Δm, the B production flavour has to be determined, typically done using flavour tagging algorithms. In the B^0–$\overline{B}^0$ system $\Delta\Gamma_d$ is very small, so that $\cosh(\Delta\Gamma_d) \sim 1$ and

$$\mathcal{A}(t) \simeq \cos(\Delta m_d\, t). \tag{1.25}$$

B^0–$\overline{B}^0$ mixing was discovered by the ARGUS collaboration at DESY in 1987 [23]. At the DORIS II storage ring, which was a $e^+\, e^-$ collider running at the $\Upsilon(4S)$ resonance to produce predominantly $B^0\, \overline{B}^0$ pairs, like-sign dilepton events were observed in semileptonic B^0, $\overline{B}^0$ decays. The $B^0\, \overline{B}^0$ pair is in an quantum-mechanical entangled state until the decay of one of the mesons is observed. Semileptonic events with two equally charged leptons can only appear, if one of the B mesons has undergone mixing. By counting the number of like-sign versus opposite-sign events, B^0-$\overline{B}^0$ mixing has been observed without measuring the decay time dependence of the mixing asymmetry $\mathcal{A}$.

1.3.2. *CP* violation

There are three ways to violate *CP* symmetry. The decay amplitude of a B ($\overline{B}$) meson decaying into a final state f ($\bar{f}$) is given by

$$A_f = \langle f|\mathcal{H}|B\rangle, \qquad \bar{A}_{\bar{f}} = \langle \bar{f}|\mathcal{H}|\overline{B}\rangle. \tag{1.26}$$

where $\mathcal{H}$ is the Hamiltonian of the weak interaction. Direct *CP* violation (*CPV*) occurs, if

$$A_f \neq \bar{A}_{\bar{f}}. \tag{1.27}$$

This kind of *CPV* is also called *CPV* in decay. It is the sole source of *CPV* for all charged mesons and all baryons. In contrast, neutral mesons also undergo flavour oscillations. That leads to non-vanishing decay amplitudes of

$$\bar{A}_f = \langle f|\mathcal{H}|\overline{B}\rangle, \qquad A_{\bar{f}} = \langle \bar{f}|\mathcal{H}|B\rangle. \tag{1.28}$$

CPV in mixing occurs, if

$$\left|\frac{q}{p}\right| \neq 1. \tag{1.29}$$

A *CP* eigenstate fulfills the condition:

$$CP|f_{CP}\rangle = \eta_{CP}|f_{CP}\rangle \quad \text{with} \quad \eta_{CP} = \pm 1. \tag{1.30}$$

If the final state is common for B and $\bar{B}$ decays, which is always fulfilled for *CP* eigenstates, *CPV* can occur in the interference between mixing and decay. The related quantity is given as

$$\lambda_f = \frac{q}{p}\frac{\bar{A}_f}{A_f} \quad \text{with} \quad \Im(\lambda_f) \neq 0. \tag{1.31}$$

The flavour specific *CP* asymmetry is defined as [6]

$$a_{sl} \equiv \frac{\Gamma(\bar{B}(t) \to f) - \Gamma(B(t) \to \bar{f})}{\Gamma(\bar{B}(t) \to f) + \Gamma(B(t) \to \bar{f})}, \tag{1.32}$$

where the subscript *sl* denotes semileptonic decays. In semileptonic decays *CP* symmetry can be violated either directly or in mixing, not in interference between mixing and decay because the final state f is reached by B decays, only, while $\bar{f}$ is reached by $\bar{B}$ decays, only. Note, that a_{sl} can be measured without the knowledge of the production flavour of the B mesons, using the untagged decay rate

$$\Gamma[f,t] = \Gamma(\bar{B}(t) \to f) + \Gamma(B(t) \to f) \tag{1.33}$$

to obtain the *CP* asymmetry

$$\mathcal{A}_{CP}(t) = \frac{\Gamma[f,t] - \Gamma[\bar{f},t]}{\Gamma[f,t] + \Gamma[\bar{f},t]} = \frac{a_{sl}}{2} - \frac{a_{sl}}{2}\frac{\cos(\Delta mt)}{\cosh(\Delta\Gamma t/2)}. \tag{1.34}$$

Thus, measuring $\mathcal{A}_{CP}$ does not require the knowledge of the initial B flavour. It gives a handle to $\frac{a_{sl}}{2}$, thus, $\mathcal{A}_{CP}$ is damped by a factor 2 compared to equation 1.32. However, compared to the mixing asymmetry $\mathcal{A}$, $\mathcal{A}_{CP}$ is damped by a factor of 2, only, while the measurement of $\mathcal{A}$ requires the knowledge of the production flavour, which damps it by the effective tagging efficiency of $\mathcal{O}(5\%)$ at a hadron collider.

1.4. Theoretical Prediction for Δm_d

In B_q^0–$\bar{B}_q^0$ mixing three physical quantities are involved [24]: $|M_{12}^q|$, $|\Gamma_{12}^q|$ and the *CP* phase

$$\phi_q = \arg(M_{12}^q/\Gamma_{12}^q). \tag{1.35}$$

The mass and width differences between the heavy and the light mass eigenstates are related to these physical quantities as

$$\Delta m_q = M_H^q - M_L^q \simeq 2|M_{12}^q| \tag{1.36}$$

$$\Delta\Gamma_q = \Gamma_L^q - \Gamma_H^q \simeq 2|\Gamma_{12}^q|\cos\phi_q, \tag{1.37}$$

up to corrections of $\mathcal{O}(m_b^2/M_W^2)$ [24]. B_q^0 mixing processes are so-called $|\Delta B| = 2$ transitions, because they change the bottom quantum number by two units. The derivation of Δm_q within

the SM is sketched in the following. The argumentation is based on [24]. The dispersive term M_{12}^q is related to the effective $|\Delta B| = 2$ Hamiltonian $H_q^{|\Delta B|=2}$

$$|M_{12}^q| = \frac{\langle B_q^0 | H_q^{|\Delta B|=2} | \bar{B}_q^0 \rangle}{2 M_{B_q^0}}, \tag{1.38}$$

which is given in the SM as

$$H_q^{|\Delta B|=2} = (V_{tq}^* V_{tb})^2 C Q + h.c., \tag{1.39}$$

where Q is a four-quark operator and C is the Wilson coefficient, which depends on the heavy mass scales of the theory. The involved hadronic matrix element can be parametrized as

$$\langle B_q^0 | Q(\mu_B) | \bar{B}_q^0 \rangle = \frac{2}{3} M_{B_q^0}^2 f_{B_q^0}^2 \mathcal{B}_{B_q^0}(\mu_B), \tag{1.40}$$

with the decay constant $f_{B_q^0}$ and the "bag" factor $\mathcal{B}_{B_q^0}$, which has to be calculated on the lattice. That leads to large theoretical uncertainties on the predictions of Δm_d and Δm_s within the SM [25]:

$$\Delta m_d|^{SM} = (0.543 \pm 0.091)\,\text{ps}^{-1} \tag{1.41}$$

$$\Delta m_s|^{SM} = (17.30 \pm 2.6)\,\text{ps}^{-1}. \tag{1.42}$$

The relative uncertainty on the theoretical prediction is 17% for the B^0 and 15% for B_s^0 oscillation frequency. Therefore, deviations from the SM predictions are searched for using the semileptonic CP asymmetry

$$a_{sl}^q = \left| \frac{\Gamma_{12}^q}{M_{12}^q} \right| \sin \phi_q, \tag{1.43}$$

which is sensitive to the CP violating phase ϕ_q and possibly to additional NP phases. The SM predicts a_{sl}^d and a_{sl}^s to be close to zero [26]:

$$a_{sl}^d|^{SM} = -(4.1 \pm 0.6) \cdot 10^{-4}, \quad \text{with} \quad \phi_d = -4.3° \pm 1.4° \tag{1.44}$$

$$a_{sl}^s|^{SM} = (1.9 \pm 0.3) \cdot 10^{-5}, \quad \text{with} \quad \phi_s = 0.22° \pm 0.06°. \tag{1.45}$$

Hence, NP effects can be detected, if experimental determinations of a_{sl}^d and a_{sl}^s differ significantly from zero.

Mainly driven by the 3.9σ deviation of $A_{sl} = 0.6 a_{sl}^d + 0.4 a_{sl}^s$ from the SM prediction measured by the D0 collaboration in June 2011 [27], NP contributions to the B–$\bar{B}$ mixing amplitudes were motivated. A common scenario to introduce NP in neutral meson mixing is shown in figure 1.4 for the B^0–$\bar{B}^0$ system. In this NP scenario the complex parameter Δ_q, where $q = d, s$, is introduced independently for B^0, B_s^0 and K^0 [24]:

$$M_{12}^q = M_{12}^{SM,q} \Delta_q. \tag{1.46}$$

A deviation from the SM point $\Delta_d = 1(\Re(\Delta_d) = 1, \Im(\Delta_d) = 0)$ of 0.9σ is present, if the D0 A_{sl} measurement is excluded from the fit. If the D0 A_{sl} measurement is included, a deviation from the SM point of 1.2σ is present.
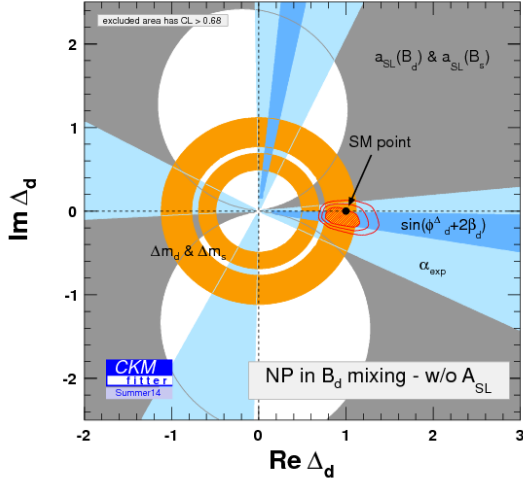
(a) NP in B^0–$\overline{B}^0$ mixing w/o A_{sl} from D0
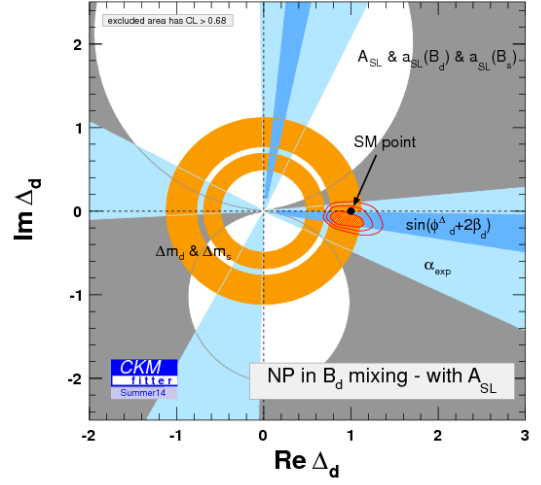
(b) NP in B^0–$\overline{B}^0$ mixing with A_{sl} from D0

Figure 1.4.: New Physics scenario: $M_{12}^q = M_{12}^{SM,q} \Delta_q$. The two-dimensional SM hypothesis in the B^0–$\overline{B}^0$ system is $\Delta_d = 1 (\Re(\Delta_d) = 1, \Im(\Delta_d) = 0)$. Combining all measurements but without the D0 A_{sl} measurement a 0.9σ deviation from the SM is obtained (a), with the D0 A_{sl} measurement a 1.2σ deviation from the SM is obtained (b). Taken from [20]

However, in [28], which is a reconsideration of SM processes contributing to A_{sl}, the 3.9σ deviation from the SM shrinks to 3σ. Therefore, precise and independent measurements of a_{sl}^s and a_{sl}^d are needed [29]. Within the current experimental precision a_{sl}^d $(2 \cdot 10^{-3})$ and a_{sl}^s $(3 \cdot 10^{-3})$ [22] are SM compatible.

Measurements of the oscillation frequencies Δm_d and Δm_s, however, serve as indirect constraints to NP scenarios. Given the large theoretical uncertainties on their SM predictions the measured values are used as inputs to global SM fits, rather than to exclude NP scenarios [29].

2. LHCb Experiment

2.1. The Large Hadron Collider

The Large Hadron Collider (LHC), situated at CERN[1] near Geneva, Switzerland, has finished its first run period (Run I) in the beginning of 2013. Two years of data taking are accomplished: in the year 2011 proton-proton collisions took place at a center-of-mass energy $\sqrt{s} = 7\,\mathrm{TeV}$ and in 2012 at $\sqrt{s} = 8\,\mathrm{TeV}$. This year Run II will begin, first at $\sqrt{s} = 13\,\mathrm{TeV}$, later at $\sqrt{s} = 14\,\mathrm{TeV}$, which was the initial design energy of the LHC.

The LHC is a proton-proton collider with a circumference of $27\,\mathrm{km}$, roughly $100\,\mathrm{m}$ underground. Its design instantaneous luminosity is $1 \cdot 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, having two proton beams with 2808 proton bunches, where each bunch consists of $1.15 \cdot 10^{11}$ protons. At $40\,\mathrm{MHz}$ bunch crossing rates, the spacing between the colliding bunches is $25\,\mathrm{ns}$.

To keep the proton beams on the circular orbit superconducting magnets with field strengths of up to $8.33\,\mathrm{T}$ are necessary, operated at $1.9\,\mathrm{K}$. With a beam current of $0.584\,\mathrm{A}$, $362\,\mathrm{MJ}$ are stored in the beams. Together with the energy stored in the magnets of $\approx 600\,\mathrm{MJ}$, the LHC stores $1\,\mathrm{GJ}$. A detailed description of the LHC is given in [30].

Four large experiments were built at the LHC, namely ATLAS, CMS, ALICE and LHCb, located at four different interaction points, see figure 2.1. The high luminosity experiments ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) have built large, general purpose particle detectors. The discovery of the Higgs boson on July 4th 2012 [14, 15] was a big success for the LHC. The ALICE (A Large Ion Collider Experiment) experiment is specialized on lead ion collisions to study QCD and quark-gluon plasma. The LHC collides lead ions during dedicated time periods. The LHCb (Large Hadron Collider beauty) experiment is designed for precision measurements in the B-system. Thus, a broad physics spectrum is covered by the experiments at the LHC.

After the magnet quench incident in 2008, the LHC started operation at the end of 2009. Shortly afterwards the LHC became with $1.18\,\mathrm{TeV}$ per beam the world's highest-energy particle accelerator. Operations in 2010, at $\sqrt{s} = 7\,\mathrm{TeV}$, were devoted to commissioning and establishing confidence in the critical machine protection system. In 2011, at $\sqrt{s} = 7\,\mathrm{TeV}$, the instantaneous luminosity was pushed to explore the performance limits of the machine. Then, 2012 became a year of efficient collision data accumulation at $\sqrt{s} = 8\,\mathrm{TeV}$ [32].

To discover the Higgs boson a maximum number of proton-proton collisions was crucial. In 2012 the LHC reached a peak instantaneous luminosity of $7.7 \cdot 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. To achieve this performance[2] at a bunch spacing of $50\,\mathrm{ns}$, the average number of protons per bunch was increased

---

[1] European Organization for Nuclear Research

[2] The injector complex of the LHC delivered excellent beam quality. The beam emittance, which is a measure for the average spread of particle coordinates in position and momentum phase space, was significantly lower
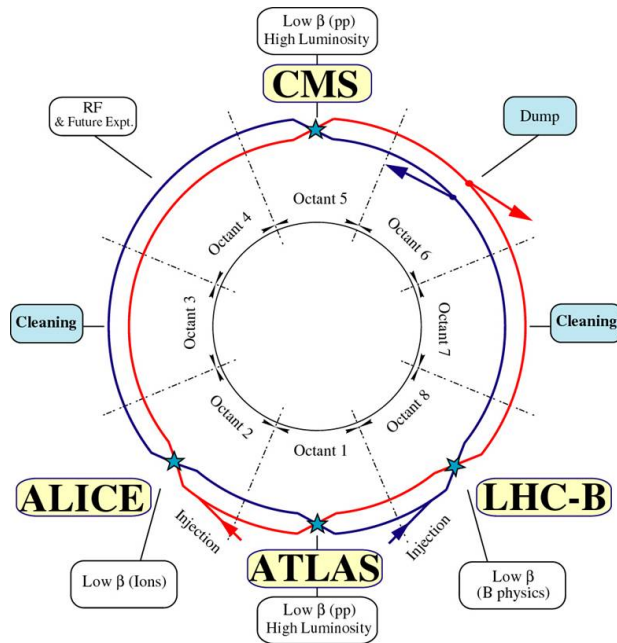
Figure 2.1.: The LHC is split into eight octants. The four big experiments sit at different interaction points. The high luminosity experiments ATLAS and CMS are placed at opposite interaction points, point 1 and point 5, to guarantee similar instantaneous luminosity conditions. LHCb and ALICE are located at point 8 and 2, respectively. The remaining octants are used for technical machine operations. The proton beams are collimated at the cleaning segments, the radio frequency (RF) cavities accelerate the proton beams and the beam dump facility exits the proton beams. Taken from [31]

beyond the design value to $1.6 - 1.7 \cdot 10^{11}$. Figure 2.2 shows characteristic plots provided by the ATLAS experiment to illustrate the luminosity increase during 2011 and the effect on the number of interactions per bunch crossing. With brighter proton beams the mean number of interactions per bunch crossing increased to values of 20 to 25 for ATLAS and CMS.

However, to perform precision measurements in the B-system a constant moderate instantaneous luminosity is required. Therefore, the proton beams are vertically displaced at the LHCb interaction point. This technique is called luminosity leveling. The effective collision area of the beams is readjusted during a fill to achieve a constant instantaneous luminosity. A fill typically lasts for 8 to 12 hours, sometimes up to 20 hours. It is counted from the injection of the proton beams into the LHC from the pre-accelerators at CERN, at 450 GeV, to the beam dump. The acceleration of the beams to collision energy takes 20 minutes, using the superconducting radio frequency cavities, that increase the beam energy with 0.5 MeV per turn at 400 MHz.

During Run I LHCb recorded data efficiently, see figure 2.3. About 1300 bunches were collided

---

(1.8 μm) for beams of 50 ns bunch spacing with $1.7 \cdot 10^{11}$ protons per bunch, than for beams with 25 ns bunch spacing with $1.2 \cdot 10^{11}$ protons per bunch (2.7 μm) [32].

(a) Colliding Bunches vs Year

(b) Instantaneous Luminosity vs Year

(c) Interactions per Crossing vs Year
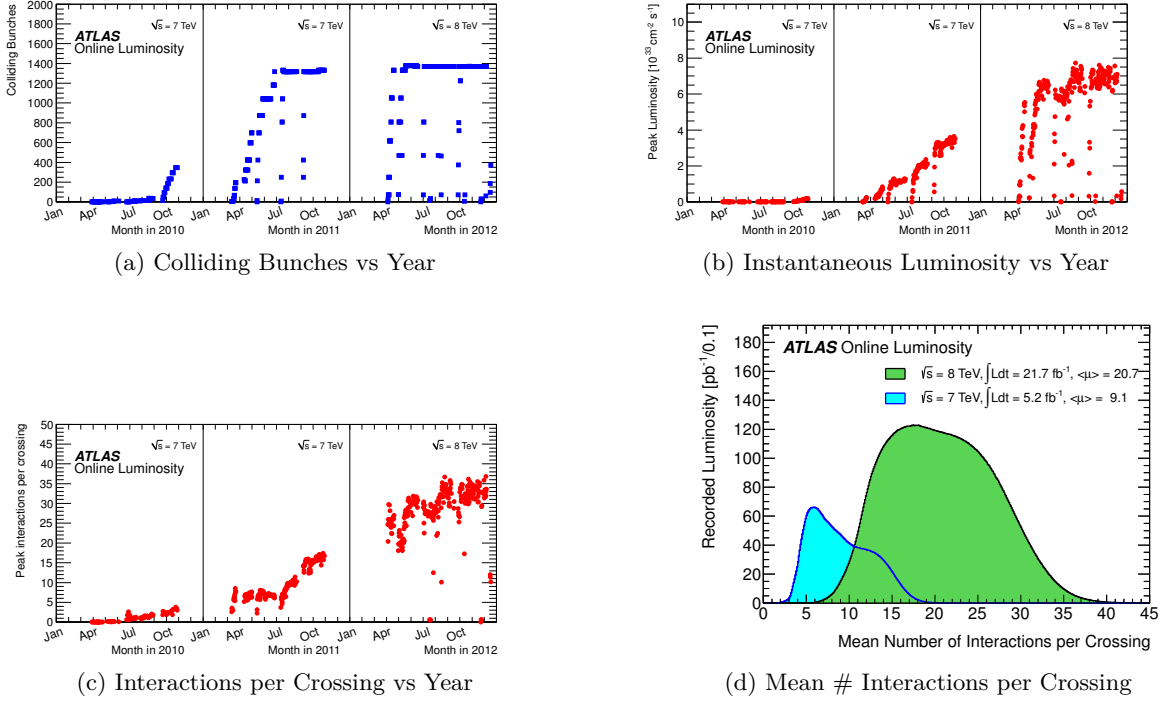

(d) Mean # Interactions per Crossing

Figure 2.2.: Characteristic plots showing high luminosity data taking conditions of the LHC Run I, provided by the ATLAS experiment. At a bunch spacing of 50 ns the number of colliding bunches was increased during 2011 to 1374 colliding bunches, but kept stable in 2012 (a). The instantaneous luminosity was increased throughout 2011 (b), to collect data efficiently in 2012. By simultaneously increasing the number of protons per bunch beyond design value, the interactions per bunch crossing increased drastically over the data taking period (c). The mean number of interactions per crossing was 20 to 25 in 2012 (d) in CMS and ATLAS. Taken from [33]

at a spacing of 50 ns. In 2011, the majority of data was taken at an instantaneous luminosity of $3.5 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which is a factor 1.75 larger than the design luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at LHCb. In 2012, data was taken at an instantaneous luminosity of $4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which is twice the design luminosity at LHCb. The pile-up, defined as the average number of visible interactions per beam-beam crossing, is a critical parameter for the performance of the LHCb detector. The detector was designed with a pile-up of 0.7, however it was successfully operated with an average pile-up of 1.7. The data set, corresponding to an integrated luminosity of $3\,\mathrm{fb}^{-1}$, collected in 2011 at a center-of-mass energy $\sqrt{s} = 7\,\mathrm{TeV}$ and 2012 at $\sqrt{s} = 8\,\mathrm{TeV}$ is used for the precision measurement of Δm_d, presented in this thesis.
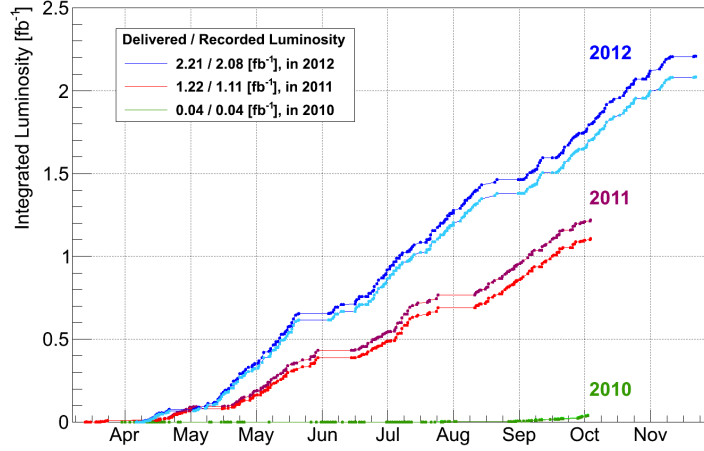
Figure 2.3.: Delivered and recorded luminosity at LHCb during Run I. Taken from [34]

2.2. Heavy Quark Production at the LHC

Protons are built of three light valence quarks, two up quarks and one down quark, seaquarks and gluons interacting via the strong interaction. The strong interaction conserves flavour, hence, quarks are produced as quark anti-quark pairs in inelastic proton-proton collisions. Compared to up, down and strange quarks, charm and beauty quarks are heavy[3]. The production of heavy quarks is described by perturbative QCD [35]. The $\mathcal{O}(\alpha_s^2)$ leading order Feynman diagrams to produce $Q\bar{Q}$ pairs are $q\bar{q} \to Q\bar{Q}$ and $gg \to Q\bar{Q}$ fusion processes, shown in figure 2.4. Additional contributions, $\mathcal{O}(\alpha_s^3)$, are pair creation with gluon emission $q\bar{q} \to Q\bar{Q}g$, flavour excitation $qg \to Q\bar{Q}q$, gluon splitting $gg \to Q\bar{Q}g$ and gluon splitting with flavour excitation, see figure 2.4. At LHC energies ($\sqrt{s} > 1\,\text{TeV}$), the flavour excitation process is dominant to produce $b\bar{b}$ pairs [35].

Figure 2.5 shows the predicted production cross-sections for different processes as a function of the center of mass energy $\sqrt{s}$. The cross section of $b\bar{b}$ pairs of $\sim 300\,\mu\text{b}$, is relatively high compared to Higgs production at the LHC.

At the LHC $b\bar{b}$ pairs are boosted. The invariant mass of the $b\bar{b}$ pair, $\mathcal{O}(10\,\text{GeV}/c^2)$, is small compared to the center-of-mass energy, $\mathcal{O}(10\,\text{TeV})$. Hence, partons with low and different proton momentum fractions enter the $b\bar{b}$ production process. It is very likely, that the $b\bar{b}$ pair is not produced at rest but boosted in the direction of one of the colliding protons, see figure 2.6. Therefore, the LHCb detector was designed as a single arm forward spectrometer. The detector acceptance covers $\approx 32\%$ of the produced $b\bar{b}$ pairs.

In 2011, at $\sqrt{s} = 7\,\text{TeV}$, the production cross-sections of B^+, B^0 and B_s^0 mesons were measured at LHCb, within the rapidity[4] range of $2.0 < y < 4.5$ and a transverse momentum range of

---

[3]$m_{c,b} > \Lambda_{QCD} = 217\,\text{MeV}$

[4]Using the rapidity definition $y \equiv \frac{1}{2}\ln(\frac{E+p_z c}{E-p_z c})$. The z axis is along the beam line.
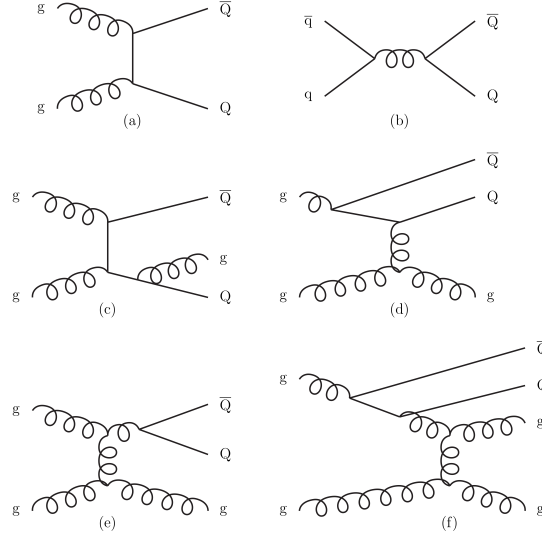
Figure 2.4.: Charm and beauty quark production processes at the LHC. Heavy quarks are produced as $Q\bar{Q}$ pairs, at $\mathcal{O}(\alpha_s^2)$ by gluon gluon fusion (a) and $q\bar{q}$ fusion (b). Additional contributions ($\mathcal{O}(\alpha_s^3)$) come from pair creation with gluon emission (c), flavour excitation (d), gluon splitting (e) and gluon splitting with flavour excitation (f). Taken from [35]

$0 < p_T < 40\,\text{GeV}/c$ [36]

$$\sigma(pp \to B^+ + X) = 38.9 \pm 0.3\,(\text{stat}) \pm 2.5\,(\text{syst}) \pm 1.3(\text{norm.})\text{µb}, \tag{2.1}$$

$$\sigma(pp \to B^0 + X) = 38.1 \pm 0.6\,(\text{stat}) \pm 3.7\,(\text{syst}) \pm 4.7(\text{norm.})\text{µb}, \tag{2.2}$$

$$\sigma(pp \to B_s^0 + X) = 10.5 \pm 0.2\,(\text{stat}) \pm 0.8\,(\text{syst}) \pm 1.0(\text{norm.})\text{µb}. \tag{2.3}$$

Charge conjugate states are included. To form a B meson or baryon, the b quark hardonizes into a confined state by picking up light up, down or strange quarks, coming from proton remnants or the fragmentation process of the initial interaction. The ratio between B^+, B^0 to B_s^0 to b-baryons is roughly 41:41:11:8 % [22].
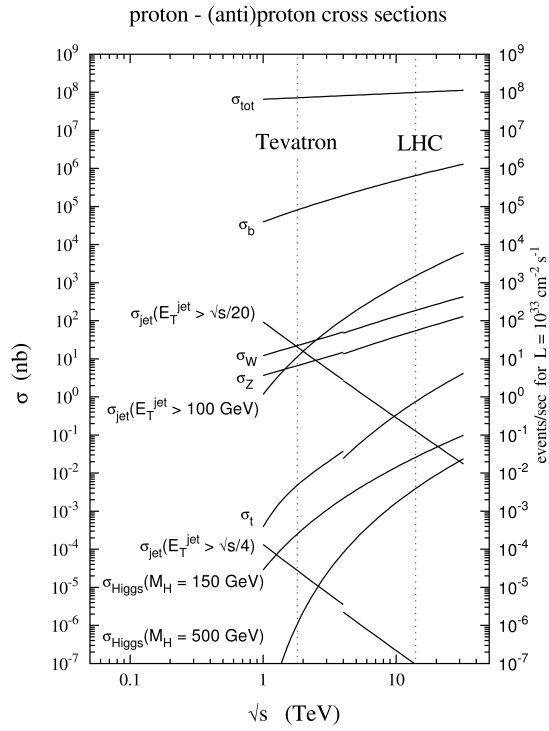
Figure 2.5.: Predicted cross-sections of different processes as a function of the center-of-mass energy $\sqrt{s}$. The cross section of $b\bar{b}$ pairs is by far the highest with $\sim 300\,\mu$b at $\sqrt{s} = 7\,\mathrm{TeV}$ to $8\,\mathrm{TeV}$, corresponding to the 2011 and 2012 running conditions of the LHC. Taken from [30]
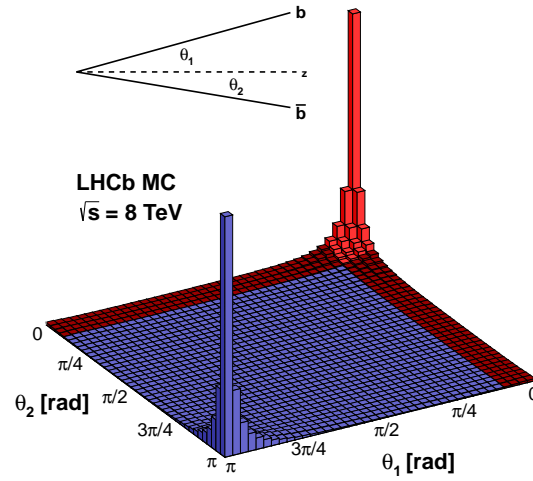
Figure 2.6.: At the LHC $b\bar{b}$ pairs are predominantly produced in the same forward or backward direction. Taken from [34]

2.3. The LHCb Experiment

The LHCb detector is a single-arm forward spectrometer, shown in figure 2.7 in the *y-z* plane. The coordinate system originates at the interaction point, the *z* axis follows the beam line into the detector, the *y* axis goes vertically from the experiment to the surface and the *x* axis points horizontally from the experiment to the outside of the LHC ring.
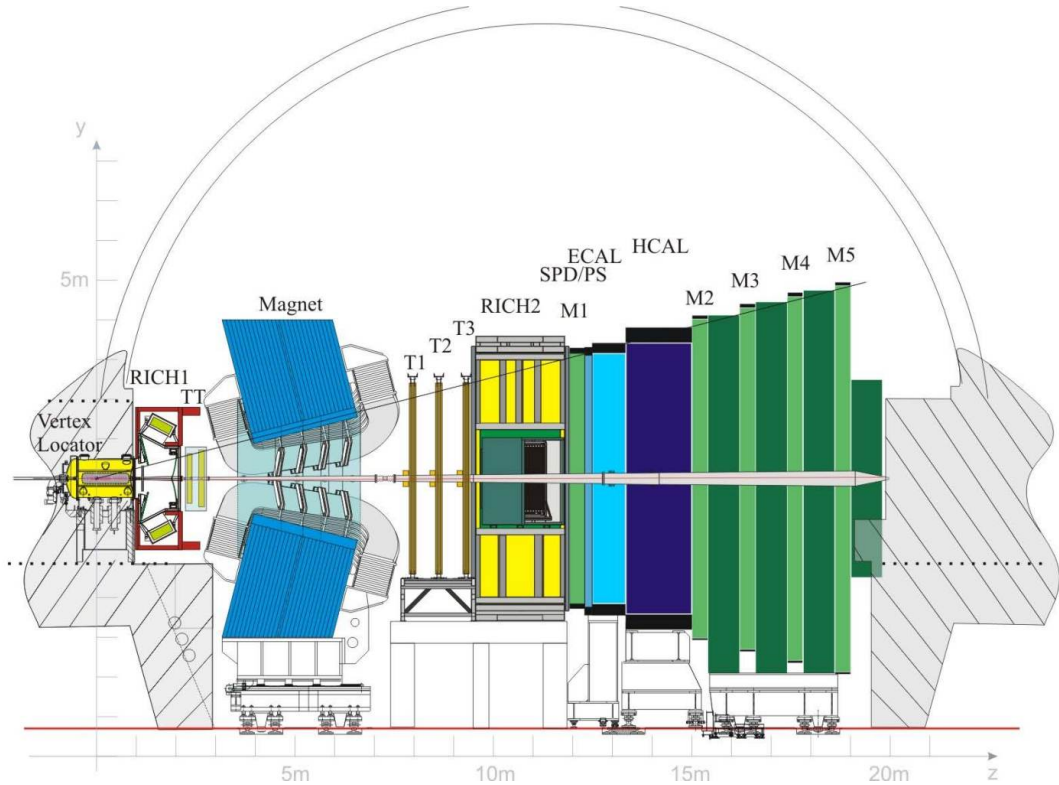
Figure 2.7.: The LHCb detector in the $z - y$ plane. The proton-proton collisions take place on the left hand side of the figure, inside the Vertex Locator (VELO). From left to right, the VELO is followed by the first Ring Imaging Cherenkov detector (RICH1), the Trigger Tracker (TT), the dipole magnet, bending tracks of charged particles in the $x - z$ plane, the three main tracking stations (T1-T3), a second RICH (RICH2), the first of the five Muon stations (M1), the Scintillating Pad Detector (SPD) and Pre-Shower Detector (PS), the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL), the four of the five Muon stations (M2-M5). Taken from [34]

The detector is designed to perform precision *B* physics[5] measurements. Due to the boost, *B* hadrons fly about 1 cm in the laboratory system. A precise determination of the *B* flight distance, which is necessary for many *B* physics analyses, requires a good vertex resolution of both vertices, the primary vertex, where the proton-proton interaction takes place, and the *B* decay vertex.

---

[5]Given the large charm production cross-section at the LHC, LHCb has also a charm physics program.

At a hadron collider plenty of low QCD background, mainly kaons and pions, is produced in the proton-proton collision. A precise vertex reconstruction allows to suppress this background. Additionally, the detector design is optimized to minimize material budget to precisely reconstruct single tracks, assigning each a particle hypothesis. The tracks' momentum resolution is enhanced by a long lever arm, given there is no tracker inside the dipole magnet.

As shown in figure 2.7, the proton collision point is surrounded by a vertex detector, the Vertex Locator (VELO). To achieve an optimal vertex resolution, the cylindrical shaped VELO is built in two halves, that are closed around the interaction point at the beginning of each fill, once the beams are stable. B mesons decay inside the VELO, hence the primary and secondary vertex positions are determined by the VELO. The VELO is followed by a Ring Imaging Cherenkov Detector (RICH1) that separates mainly pion from kaon tracks. After RICH1 the Trigger Tracker (TT) is installed to reconstruct charged particles from decays of particles outside the VELO, e.g. K_S^0. The TT is followed by a dipole magnet with a maximum magnetic field strength of 1 T bending the trajectories of charged particles in the $x-z$ plane. Inside the magnet, no sub-detector is installed. The main tracker, consisting of three tracking stations (T1-T3) is installed behind the magnet. A second RICH (RICH2) provides particle identification information covering a different momentum range than RICH1. Before the calorimeters the first Muon station (M1) is installed, to serve the hardware trigger improving the momentum estimate. The electromagnetic (ECAL) and hadronic calorimeters (HCAL) absorb all particles except muons, measuring the energy deposit of the stopped particle. After the calorimeters the four of the five Muon stations (M2-M5) are installed, measuring the trajectories of muons. The total LHCb detector acceptance for B events, including all B daughters, is very large, about 30%. In the bending plane the acceptance covers the angular range from 10 m rad to 300 m rad and in the non-bending plane from 10 m rad to 250 m rad. The sub-detectors are explained in more detail in the following. The LHCb detector design is documented in [37]. The LHCb detector performance of Run I is given in [38].

2.3.1. Tracking System

The tracking system is designed to deliver precise vertex and momentum resolution. It consists of the dipole magnet, the VELO and tracking stations, the TT in front of the magnet and the T stations behind the magnet. The main component of the magnetic field B_y is shown in figure 2.8 as a function of the z direction.

The magnetic field deflects charged particles depending on the momentum and the charge. The polarity of the field is reversed frequently to minimize detection asymmetries. Particles traversing the entire tracking system are exposed to an integrated magnetic field of $\int B dl \approx 4 T m$. All tracking detectors are installed outside the magnet. This results in a long lever arm in the track extrapolation between the tracking detectors giving a precise momentum resolution of $\Delta p/p = 0.4\%$ at $2\,\text{GeV}/c$ and $\Delta p/p = 0.6\%$ at $100\,\text{GeV}/c$. However, this also enhances the mis-matching of track parts, resulting in a significant amount of so-called ghost tracks $\mathcal{O}(10\%)$ [39].

Vertex Locator

The VELO is a silicon strip vertex detector built in R-ϕ geometry around the interaction point, see figure 2.9. The LHC beam pipe was removed inside the VELO in order to achieve a very
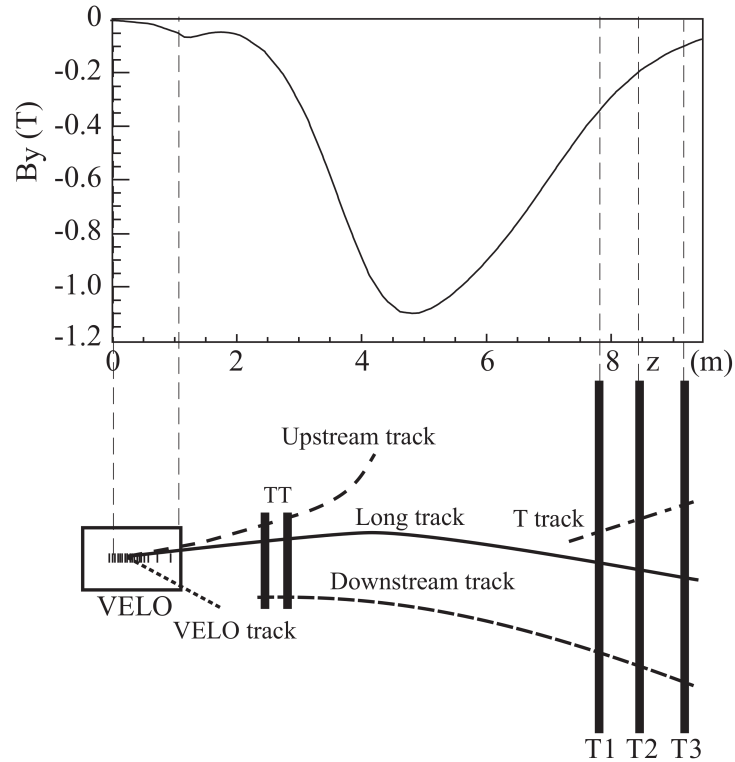
Figure 2.8.: Top: The main magnetic field component B_y of the LHCb dipole magnet as a function of the z direction. Bottom: The tracking detectors VELO, TT and T stations. Particles, that are detected before and after the magnet are reconstructed as long tracks. Taken from [37]

precise vertex resolution. It is replaced by a thin corrugated (RF) foil, 5 mm from the beam, to shield the vacuum of the beam and possible interference of the beam current with the electronics. The VELO is built in two halves. During beam injection and calibration the VELO is open, the two halves are moved in opposite direction along the x axis. For data taking the VELO is closed around the interaction point. In this position the distance between the beam and the VELO sensors is 7 mm. For comparison, the inner radii of the vertex detectors of ATLAS or CMS are 50.5 mm and 44 mm, with an inner radius of the beam pipe of 29 mm. In the transverse plane $(x - y)$ the primary vertex resolution is $13\,\mu\mathrm{m}$ and $71\,\mu\mathrm{m}$ along the beam axis, measured for vertices with at least 25 tracks [41].

Trigger Tracker

The TT is a large area silicon strip detector, installed in front of the magnet, covering the full acceptance of the detector. The two stations consist of two layers each, see figure 2.10. The strips of one layer are oriented in y direction, the strips of the second layer are tilted $\pm 5°$ with respect to the first layer. In this way, a two dimensional measurement of the particles position is achieved.
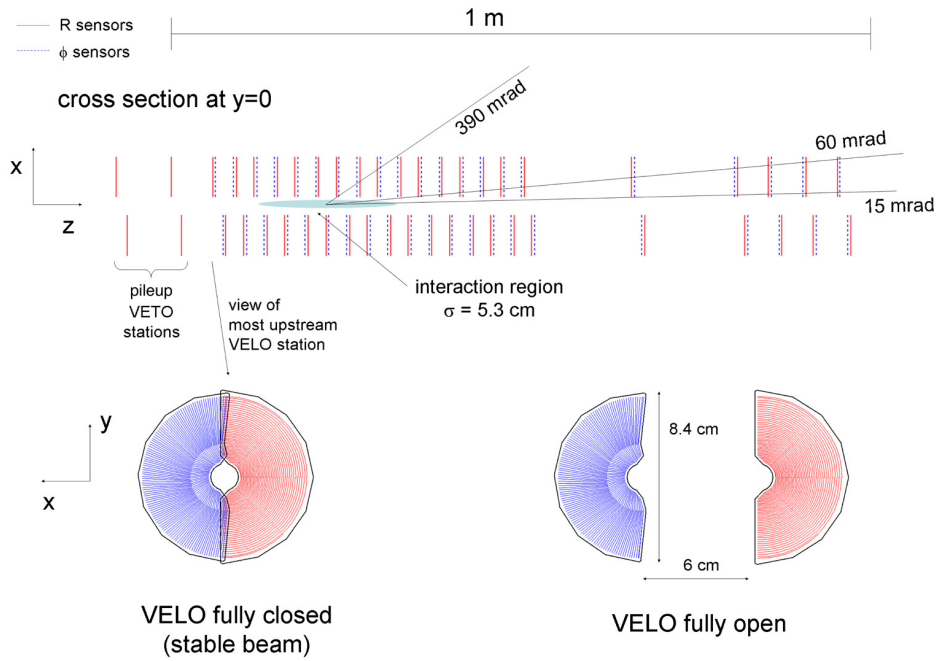
Figure 2.9.: The silicon strip Vertex Locator (VELO) consists of two halves. During beam injection the VELO is open. For data taking it is closed around the interaction point. Taken from [40]
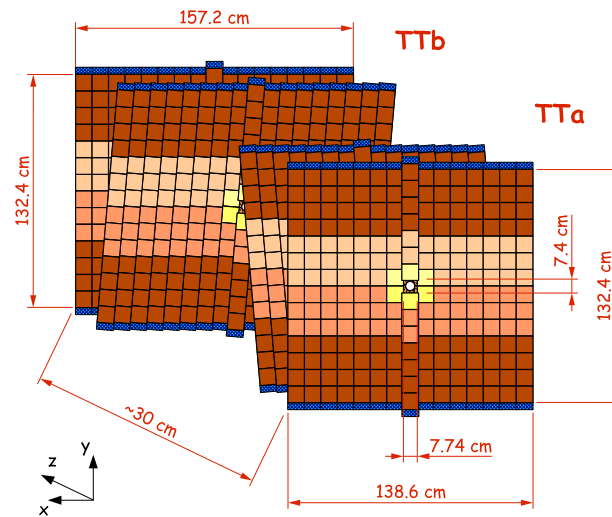
Figure 2.10.: The Trigger Tracker (TT) consisting of two stations. The silicon strips of one layer are oriented in y direction, the ones of the other layer are tilted $\pm 5°$. Taken from [42]

The main purpose of the TT is to reconstruct low momentum particles, bent out of the detector acceptance, and long-lived particles, decaying after the VELO [42].

Tracking Stations

The three tracking stations (T stations) are located behind the magnet with a size of $6\,\text{m} \times 5\,\text{m}$. Each station holds two detectors, the Inner Tracker (IT) and the Outer Tracker (OT), see figure 2.11.
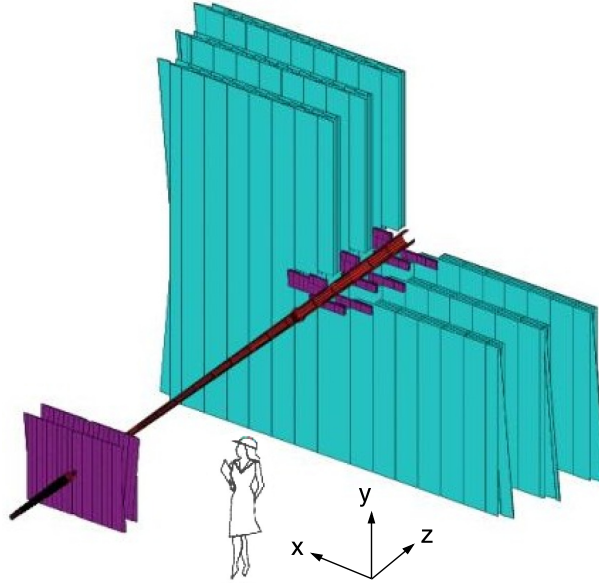
Figure 2.11.: The three tracking stations (T stations) (right) and the TT (left) surrounding the beam pipe. The silicon strip detectors TT and IT are shown in purple, the straw tube drift-time OT detector is shown in turquoise. Taken from [37]

The IT is a silicon strip detector positioned in the inner region with high particle flux, see figure 2.12. There are four active layers per station. The strips of the first and the last layers
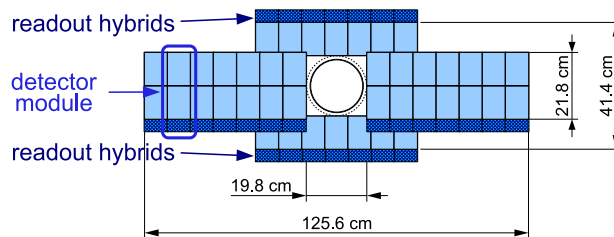
Figure 2.12.: Design of an IT layer with silicon strips oriented in y direction. Taken from [40]

are oriented in y direction, while the strips of the inner two layers are tilted $\pm 5°$, similar to the TT. Close to the beam pipe the occupancy of the IT is 2.5%. A hit resolution of $\sim 50\,\mu\text{m}$ is achieved [43].

The OT is a straw tube drift-time detector, covering the large region around the IT at low particle flux. Analog to the IT, it consists of four layers per station, but each layer is made of two layers of straw tubes, shifted half a straw diameter for optimal spatial resolution, see figure 2.13. The straw tubes are filled with gas, serving as drift cells. A position resolution of $200\mu m$ is
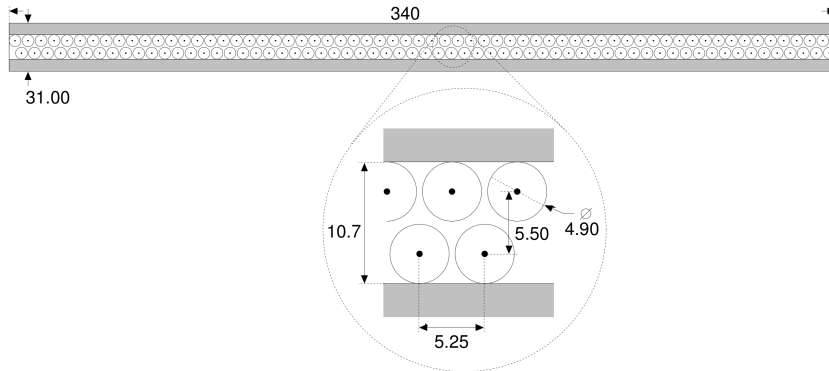
Figure 2.13.: Design of an OT layer with two layers of straw tubes. Taken from [40]

achieved at a hit efficiency of 99.2% [44, 45].

2.3.2. Particle Identification System

Beside a good tracking system an accurate particle identification (PID) is essential for precision B physics measurements at LHCb. Kaons, pions, protons, electrons and muons are mainly reconstructed at LHCb as final states of B and D decays. A particle hypothesis is assigned to the reconstructed tracks, which is based on a probability provided by the particle identification system. As many B decay channels contain hadronic final state particles, a good separation between kaons and pions is essential. The particle identification system consists of RICH1 and RICH2, located before and after the magnet, the ECal and the HCal and the five muon chambers, M1 – M5.

In analyses a special variable, called DLL, can be used to distinguish final state particles. Based on the particle hypothesis probability a likelihood $\mathcal{L}$ is calculated. The difference in the logarithmic likelihood of a given kaon hypothesis with respect to a pion hypothesis is computed

$$DLL_{K\pi} = \ln \mathcal{L}(K) - \ln \mathcal{L}(\pi). \tag{2.4}$$

If $DLL_{K\pi} > 0$ the particle is more likely to be a kaon, otherwise a pion. A second variable to distinguish final state particles is the ANN-based PID variable [46], which combines the available the particle hypothesis information using an artificial neural network (ANN).

Ring Imaging Cherenkov Detector

The RICH1 and RICH2 are solid state and gas radiator detectors utilizing the Cherenkov effect. Relativistic charged particles traversing the radiator gas emit Cherenkov radiation in form of a light cone, because their velocity is greater than c, the speed of light, in the gas. The light cone is

radiated under the Cherenkov angle Θ_C relative to the flight direction depending on the velocity of the particle v, and the refractive index of the medium n

$$\cos\Theta_C = \frac{c}{nv}\,. \tag{2.5}$$

Using the momentum estimate, provided by the tracking system, the Cherenkov angle is related to the mass of the particle. In RICH detectors the radiated photons are collected by spherical mirrors and imaged onto Pixel Hybrid Photon Detectors, located outside of the detector acceptance. Θ_C is related to the radius of the imaged ring.

RICH1 and RICH2 operate at different momentum ranges. RICH1, installed before the magnet, contains Aerogel and a Fluorobutane (C_4F_{10}) gas radiator to cover a momentum range of $2-40\,\mathrm{GeV/c}$, see figure 2.14. Aerogel, with a refractive index of $n = 1.03$, provides separation power for particles with momenta below 9.2 GeV/c, while C_4F_{10} gas separates particles up to $40\,\mathrm{GeV}/c$, see figure 2.15. RICH2, installed after the magnet, contains a CF_4 gas radiator to cover a momentum range of $15-100\,\mathrm{GeV/c}$ [47], also shown in figure 2.14. As RICH1 separates low momentum particles it covers the full angular acceptance of the tracking system of 25 - 300 mrad. RICH2, however, separating high momentum particles, preferentially produced in the forward region, covers 15 - 120 mrad [40].

Figure 2.15 shows the performance of RICH1. Pion, kaon and proton tracks are clearly separable up to a momentum of 10 GeV. The Cherenkov angle resolutions are in agreement with the expected design performance of the gas radiators [48].

Calorimeter System

The calorimeter system stops all particles except muons and neutrinos. The deceleration energy is transformed into particle showers, detected by scintillating material. The scintillator light is collected by optical fibers and guided to photomultipliers.

The calorimeter system is built of the Scintillating Pad Detector (SPD), the Preshower detector (PS), the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL), in positive z direction. Between the SPD and the PS a lead plate is installed. The combination of SPD, PS and ECAL delivers a longitudinal shower profile of electromagnetic showers. Charged particles start to shower in the SPD, while neutral particles start to shower in the lead plate between the SPD and the PS.

The SPD and the PS material corresponds to 4 electromagnetic interaction lengths and 0.2 hadronic interaction lengths. The ECAL, with 25 electromagnetic interaction lengths, is designed to stop all photons and electrons. It corresponds to 1.1 hadronic interaction lengths and is built in shashlik structure of alternating layers of lead and scintillating tiles. The HCAL, with 5.6 hadronic interaction lengths, stops neutral and charged hadrons. It is built as sampling calorimeter, iron blocks interspersed by scintillating tiles, see figure 2.16.

The calorimeters are segmented into square cells in the plane perpendicular to the beam line. The granularity of the cells is highest close to the beam, where the particle flux is high, and decreases for outer parts. The energy resolution of the ECAL is given as

$$\frac{\sigma(E)}{E} \leq \frac{9.5\%}{\sqrt{E/\mathrm{GeV}}} \oplus 0.8\%\,, \tag{2.6}$$
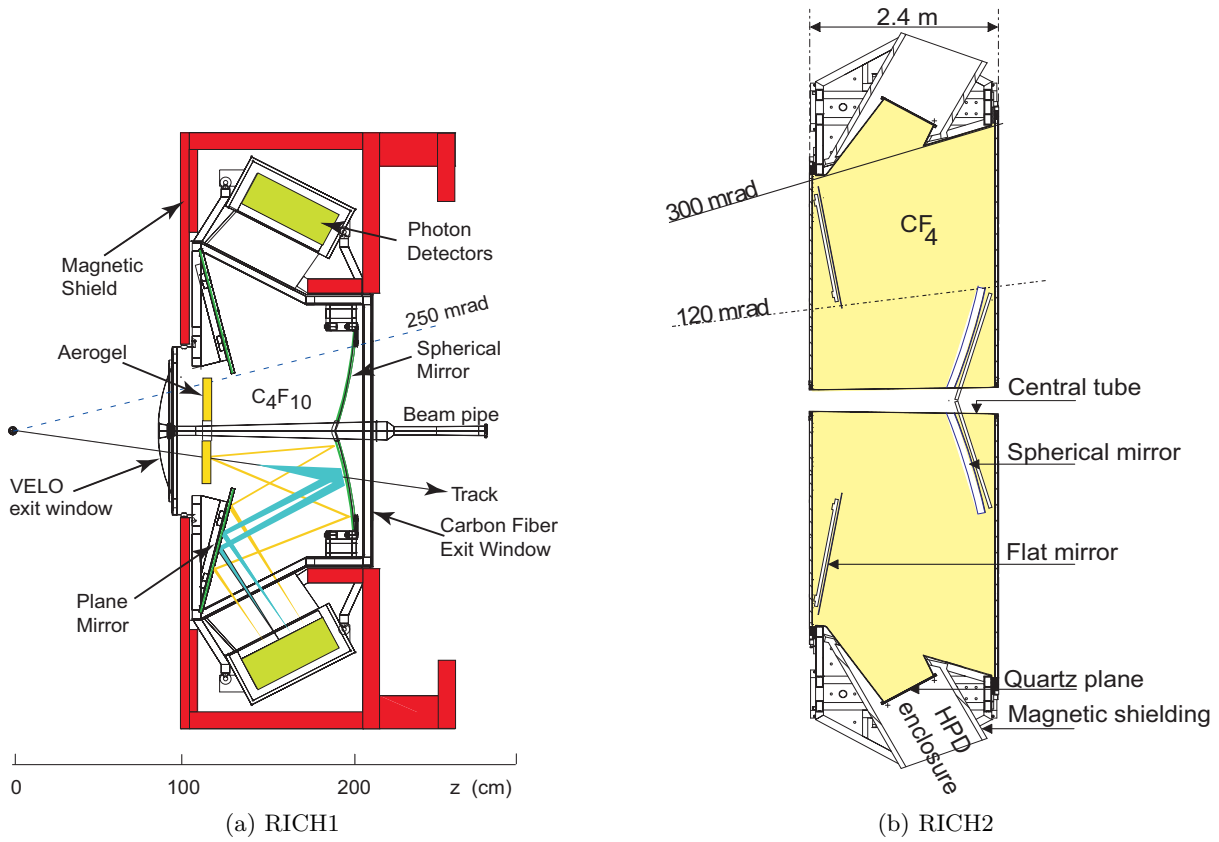
(a) RICH1

(b) RICH2

Figure 2.14.: Design of the RICH1 (a, side view) and the RICH2 (b, top view) detectors. Taken from [40]

(a) Designed separation

(b) Measured separation
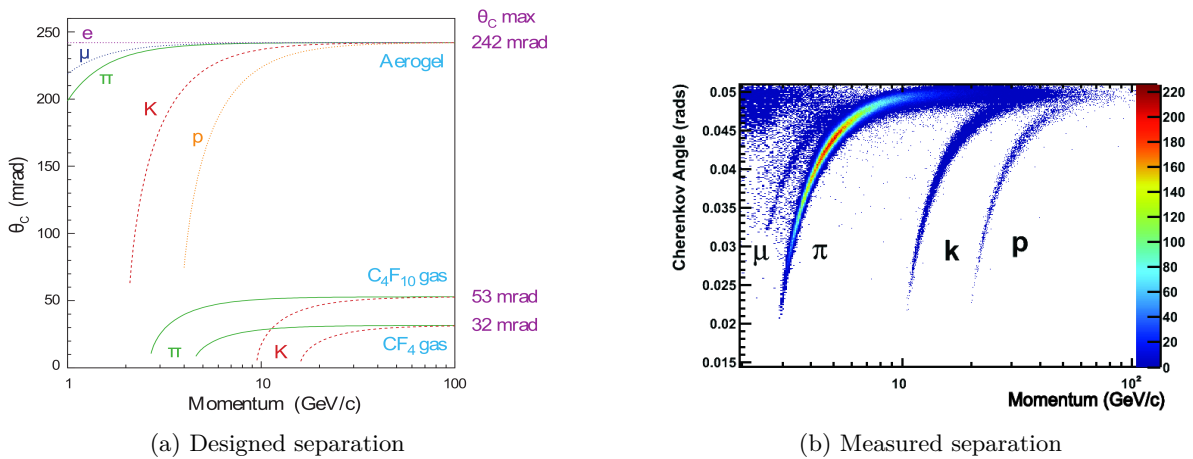
Figure 2.15.: Cherenkov angle as a function of the track momentum. Design performance of the different gas radiators (a), measured performance of RICH1 (b). Taken from [48]
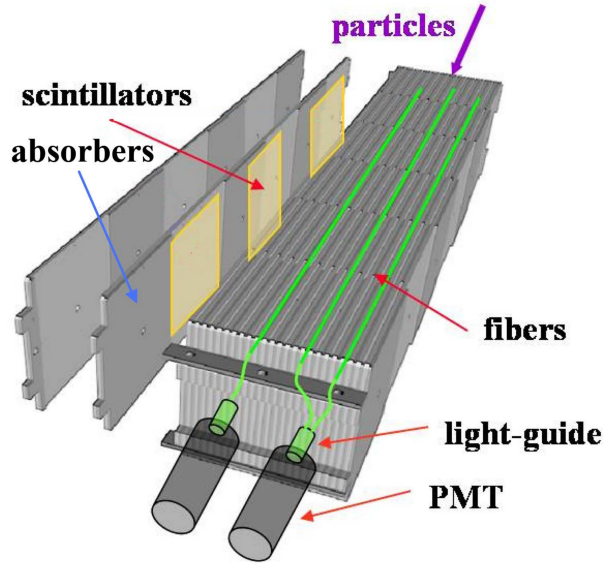
Figure 2.16.: Design of the HCAL as sampling calorimeter of iron blocks used as absorber material and scintillating tiles to collect the deposit energy. The scintillator light is collected by optical fibers and guided to the Photomultipliers (PMT). Hadronic particles come in lengthwise. Taken from [40]

where ⊕ denotes quadratic summation. The energy resolution of the HCAL is given as

$$\frac{\sigma(E)}{E} \leq \frac{69\%}{\sqrt{E/\text{GeV}}} \oplus 9\% \,. \tag{2.7}$$

The first term accounts for the uncertainty due to the stochastic behavior of showers, the second term accounts for residual mis-calibrations [49]. The calorimeter system identifies high transverse energy hadron, electron and photon candidates. The transverse energy of a 2×2 cells cluster is defined as

$$E_T = \sum_{i=1}^{4} E_i \sin \Theta_i, \tag{2.8}$$

where E_i is the energy deposited in cell i and Θ_i is the angle between the beam line and the assumed flight direction of the neutral particle. The calorimeter information is used in the trigger.

Muon system

The identification of muons plays an important role in the LHCb physics programme. Muons are present in many final states of decay channels used to perform precision *CP* measurements. Additionally, muons are used to tag the flavour of neutral *B* mesons. The muon system consists of five muon stations M1 – M5, see figure 2.17.

M1 is installed in front of the calorimeters. It is used to improve the transverse momentum measurement utilized by the hardware trigger. The other stations, M2 to M5, are located behind

(a) Muon System Sideview

(b) Schematic view of one Muon Station
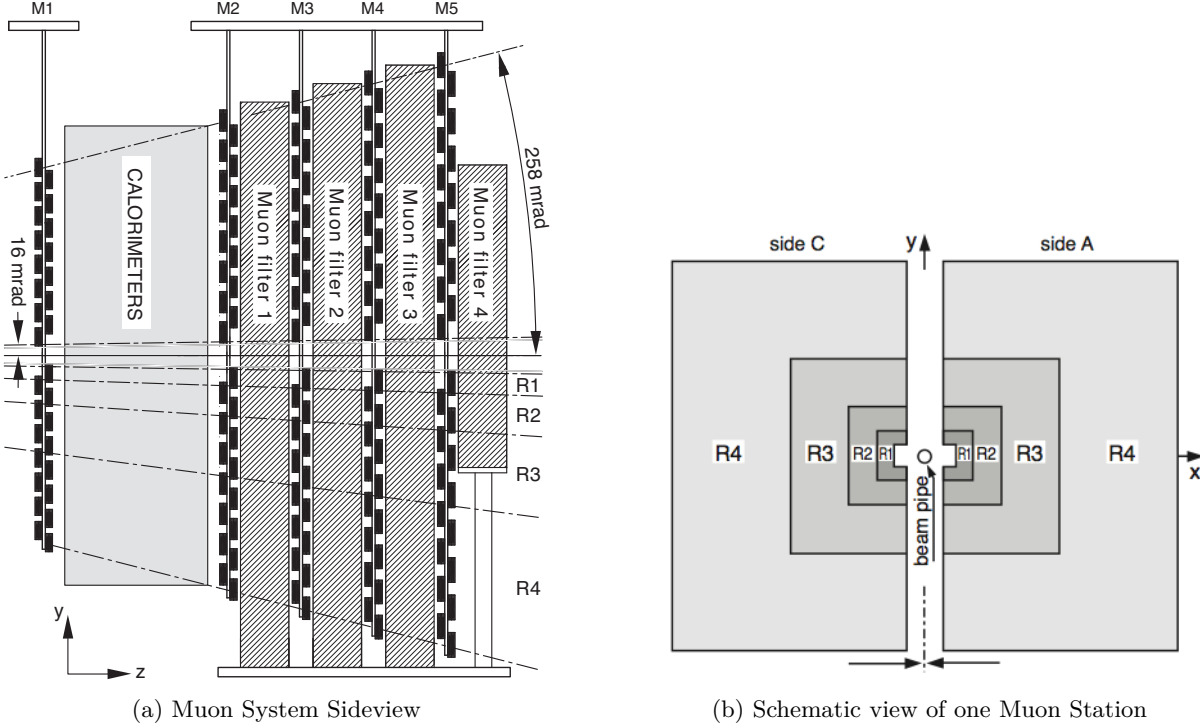
Figure 2.17.: Side view of the muon system (a). The first muon station is located in front of the calorimeters. The other four muon stations are located directly behind the calorimeters, separated by muon filters. Schematic of one Muon Station (b) to illustrate the granularity. Taken from [40, 50]

the calorimeter system, at the end of the LHCb spectrometer. They are interleaved with 80 cm thick iron absorbers. To pass all five muon stations a muon must have a minimum momentum of around 6 GeV. The inner part of M1 consists of Gas Electron Multipliers (GEM), the outer part of M1 and M2 – M5 is equipped with multi-wire proportional chambers (MWPC) [51].

The muon chambers provide a hermetic geometric acceptance to high momentum particles coming from the interaction point, see figure 2.17. They provide space point measurements on the particle trajectories, which are also used in the trigger [52]. The muon identification efficiency is 93% at a hadron mis-identification probability of below 0.6% [50].

2.4. LHCb Trigger

The production cross-section of B mesons in proton-proton collisions within the LHCb acceptance is $\sigma(pp \rightarrow BX) \simeq 90\,\mu$b at $\sqrt{s} = 7$ TeV, where B is B^+, B^0 or B_s^0, as shown in section 2.2. The instantaneous luminosity at LHCb was $\mathcal{L} \simeq 3.5 \cdot 10^{32} cm^{-2}s^{-1}$ in 2011. Hence, in 2011 the production rate of $pp \rightarrow BX$ at LHCb was

$$\mathcal{R}(pp \rightarrow BX) = \sigma \cdot \mathcal{L} = 90 \cdot 10^{-6} \cdot 10^{-24} cm^2 \cdot 3.5 \cdot 10^{32} cm^{-2}s^{-1} = 31.5\,\text{kHz}\,. \tag{2.9}$$

The main task of the LHCb trigger is to identify these heavy flavour events among the light quark background processes[6]. A trigger always has to face a trade off problem. It has to form decisions very fast, but correct and efficient.

To identify *B* decays within a hadronic interaction environment partial reconstruction of the decay is necessary, performed on-line by a software trigger. This takes up to 30 ms per bunch crossing, executed in parallel on 3×10^4 CPUs. Hence, the 40 MHz collision rate has to be reduced to 1 MHz beforehand, done by a fast hardware trigger.

LHCb has a trigger system, that consists of a level 0 hardware trigger (L0) and a two step software trigger, the High Level Triggers HLT1 and HLT2, that run on a computer farm of 3×10^4 CPUs. The trigger identifies heavy flavour processes by determining the impact parameter of tracks, by measuring the momentum of tracks with sufficiently high impact parameter or by identification of muon candidates.

2.4.1. Level 0 Hardware Trigger

L0 reduces the bunch crossing rate to below 1 MHz. It is synchronized to the LHC 40 MHz clock and operates at latencies below 4 µs [53], which corresponds to the maximum pipeline depth of the read out system. It uses information from the calorimeter system and from the muon system as input. The VELO pile-up veto, as reported in [40], is not used to trigger flavour physics events, since LHCb does not limit the average number of visible interactions per bunch crossing to $\mu \simeq 0.4$ (design), but runs at $\mu \simeq 1.4$ in 2011 [53] and $\mu \simeq 1.6$ in 2012 [54].

An event is triggered if a hadronic shower with transverse energy $E_T > 3.5$ GeV, an electromagnetic shower with $E_T > 2.5$ GeV or up to two high energetic muon tracks are detected. The minimum transverse momentum required for one muon track is $p_T > 1.48$ GeV and for two muon tracks it is $\sqrt{p_{T,1} \cdot p_{T,2}} > 1.296$ GeV [53]. A simplified momentum estimation with a relative momentum resolution of 20% is done.

In the 2012, the center of mass energy was increased to 8 TeV, the thresholds were increased to $E_T > 3.7$ GeV for hadronic showers, $E_T > 3$ GeV for electromagnetic showers and $p_T > 1.76$ GeV, $\sqrt{p_{T,1} \cdot p_{T,2}} > 1.6$ GeV for one and two muon tracks [54].

2.4.2. High Level Trigger

The HLT reduces the rate of accepted events to 3 kHz in 2011 and 5 kHz in 2012. A higher rate would be desirable since the $pp \to BX$ production rate in the LHCb acceptance was 31.5 kHz in 2011, however, the rate is limited by the implementation and execution of the software trigger. All of these events are written to storage. The HLT is written in C++, running on an Event Filter Farm (EFF), which is a farm of multiprocessor PCs with $\sim 3 \times 10^4$ CPUs. If an event is excepted by L0, it is transferred from the front end electronic boards to the EFF. The processing time per event in the HLT is up to 30 ms [53].

In 2011, the L0 rate of 870 kHz was reduced by the HLT1 to 43 kHz. HLT1 reconstructs tracks using a simplified reconstruction and performs muon identification. A combination of cuts on transverse momentum, p_T, and impact parameter is applied. At 43 kHz HLT2 performed an

---

[6]The charm production rate is a factor 20 larger than the Beauty production rate.

almost complete event reconstruction, using tracks above $p_T > 0.5\,\text{GeV}$, only. HLT2 selection criteria are lepton identification, decay time information and invariant mass. The HLT2 output rate of $3\,\text{kHz}$ consisted of 50% inclusive hadronic triggers[7] and 25% triggers on leptons. The remaining rate is covered by exclusive triggers, mainly on charmed hadrons [53].

In 2012 the new concept of deferred triggering was introduced [55]. During the 2012 data taking period the LHC machine delivered stable beams for about $\sim 30\%$ of the time on average, see figure 2.18.
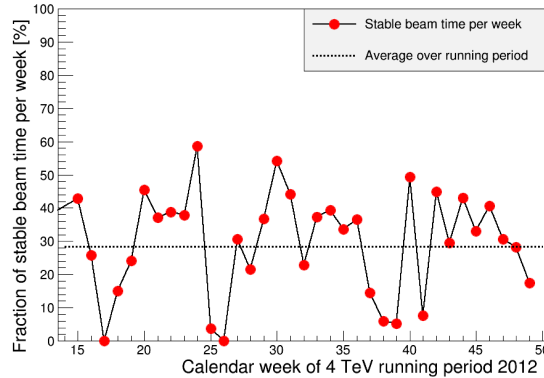
Figure 2.18.: Fraction of stable beam time per week versus the 2012 data taking year. Dips in the distribution occur at machine development phases, that typically last for more than a week. Taken from [55]

When the LHC delivers stable beams physics data is recorded. During that time, the HLT computing farm operates at full capacity, else the CPUs idle. The bottleneck of the LHCb trigger is the time consumption of the partial event reconstruction of the HLT.

Hence, a new approach was followed. 20% of the L0 accepted events were temporarily saved to be processed by the HLT during the inter-fill gaps. This increased the HLT output rate from $3\,\text{kHz}$ to $5\,\text{kHz}$ and allowed to improve the HLT2 track reconstruction. The minimum p_T requirement was lowered from $500\,\text{MeV}$ to $300\,\text{MeV}$ and long-lived particles like K_S^0, that decay outside of the VELO, were reconstructed.

In 2012, the L0 event rate of $1\,\text{MHz}$ is reduced by the HLT1 to $80\,\text{kHz}$, which is further reduced by the HLT2 to $5\,\text{kHz}$. The HLT2 output rate of $5\,\text{kHz}$ includes the deferred triggered events, see figure 2.19 [54].

The HLT2 includes generic inclusive beauty trigger selections, based on multivariate selections to select N-body beauty decays, where N is $2, 3, 4$. These so called topological triggers select beauty decays with charged final state tracks. They do not require all final state particles to be reconstructed, performing at high efficiency. They are used by the precision measurement of Δm_d, presented in this thesis.

---

[7]An inclusive trigger does not require all final state particles to be reconstructed in contrast to exclusive triggers.
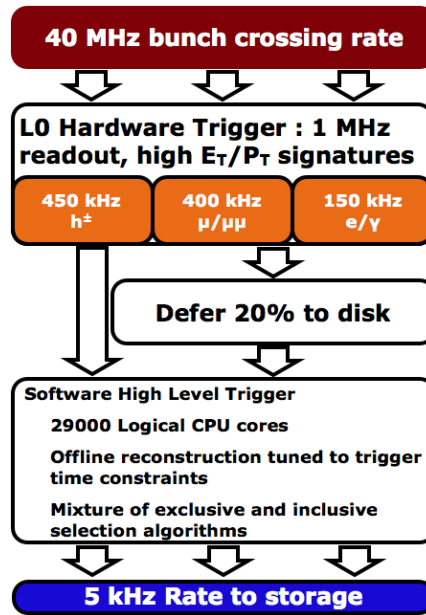
Figure 2.19.: Illustration of the 2012 Trigger System at LHCb. The 40 MHz bunch crossing rate is reduced by the L0 hardware trigger to 1 MHz. 20% of the L0 triggered events are written to temporary storage, to be processed during the inter-fill gaps of the LHC. The other 80% of the L0 triggered events are processed by the software HLT on-line. Taken from [34]

2.5. LHCb Software Environment

LHCb software applications are based on the GAUDI framework [56, 57], see figure 2.20. The framework is structured in an object oriented paradigm and written in C++. Originally developed by and for the LHCb experiment, GAUDI has been adopted by ATLAS and other experiments [58]. The phases of data processing are encapsulated in LHCb applications, as depicted in figure 2.20. All applications share the LHCb event model and make use of the LHCb detector description. This ensures consistency.

GAUSS software [59, 60] generates events using Monte Carlo (MC) simulation methods. Proton-proton collisions are generated using PYTHIA [61, 62] with a LHCb specific configuration [63]. Decays of hadronic particles are described by EVTGEN [64], where final state radiation is generated by PHOTOS [65]. Electromagnetic and hadronic interactions of particles with the detector material are described by the GEANT toolkit [66, 67]. MC simulation is an important tool to understand the detector response and the physics processes. The originally generated physics process is accessible from each particle by a link to MC truth information.

BOOLE software [68] generates detector response for MC simulation events. It digitizes the simulated physics processes into the detector read out format. Afterwards the format of MC simulation data is identical to real data. Hence, it is processed by the same reconstruction software packages in the following.
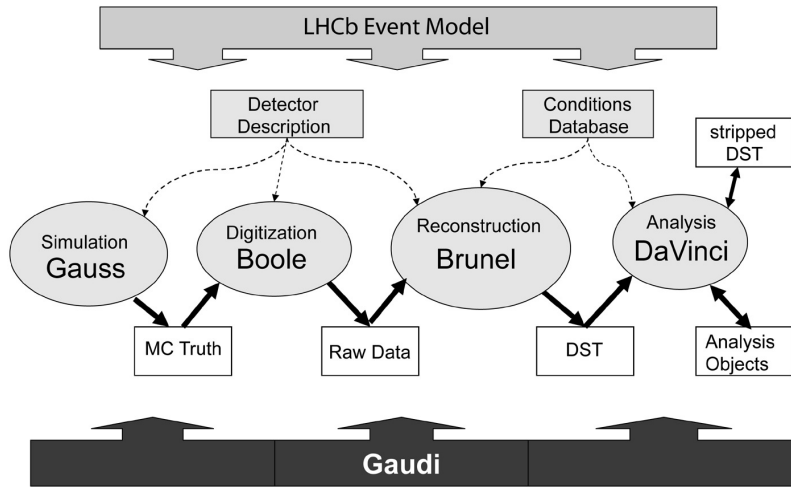
Figure 2.20.: Different LHCb software applications, based on the GAUDI framework. Simulated data is generated by GAUSS software. It is digitised by BOOLE software transforming simulation into detector response as real data. After this stage, simulated and real data events are reconstructed by BRUNEL software. Physics decays are reconstructed by DAVINCI software. Taken from [56]

BRUNEL software [69] performs off-line reconstruction. From measurements in the tracking and calorimeter detectors tracks are reconstructed. Single tracks are transformed into particle candidates using the information of the particle identification system. This reconstruction process is very time consuming. Therefore, it is performed once for the full data set by the LHCb computing group.

DAVINCI software [70] reconstructs the physics processes of interest. *B* decays are reconstructed. By adding four-momenta of final state particles the decay chain is reconstructed from back to front. Vertices are fitted and selection requirements are applied. This final reconstruction process is also very time consuming. Analogue to the BRUNEL reconstruction the DAVINCI reconstruction is done once on the full data set, but using a variety of decay channel and selection requirements, predefined by the different analysis groups. Events, that do not fulfill the requirements are stripped off, hence, this process is called "stripping". Typically, the data analyst accesses fully reconstructed and stripped data sets.

The HLT is written in MOORE software [71]. It reconstructs tracks on-line to reject background events, that do not fulfill momentum, impact parameter or invariant mass requirements. Compared to the off-line reconstruction the on-line HLT processing time per event of 30 ms is very short. The reconstruction and selection algorithms used by the HLT are shared with the BRUNEL and DAVINCI software, but optimized on timing performance rather than on accuracy.

3. Flavour Tagging

To perform a mixing asymmetry measurement in the B^0–$\overline{B}^0$ or B_s^0–$\overline{B}_s^0$ system the production flavour of the B needs to be determined. This is done by flavour tagging algorithms.

Flavour tagging algorithms use particles which are correlated to the production of the B meson. Exploiting the fact that b quarks are predominantly produced in $b\bar{b}$ pairs the B event is naturally partitioned in two sides: the signal side, which are all particles related to the hadronization of the b quark that forms the signal B meson, and the opposite side, which are all particles related to the hadronization and decay of the associated opposite side $\bar{b}$ quark.

Figure 3.1 sketches the structure of a typical B event using a $B_s^0 \to J/\psi\,\phi$ decay. The $b\bar{b}$ quark pair is produced at the primary vertex. The $\bar{b}$ quark hadronizes into the signal B_s^0 meson. The associated b undergoes either $b \to c \to s$ transitions and / or a semileptonic $b \to Xl^-$ decay.

The different types of flavour tagging algorithms, also called taggers, available at LHCb are explained in the following. On the signal side, where the $\bar{b}$ hadronizes together with an s quark into the signal B_s^0 meson, the associated $\bar{s}$ quark from the $s\bar{s}$ pair can hadronize together with an u quark into a charged kaon. This charged kaon is utilized by the Signal Side Kaon tagger (SSK). On the opposite side the $b \to c \to s$ transition can result in a charged kaon, utilized by the Opposite Side Kaon tagger (OSK), while the semileptonic $b \to Xl^-$ decay, where the lepton is either a muon or electron, is exploited by the Opposite Side Muon (OSμ) or Opposite Side Electron (OSe) taggers. Additionally, the opposite b decay vertex can be reconstructed inclusively by the Opposite Side Vertex charge tagger (Q_{vtx}).

For the single particles taggers, the best kaon, electron or muon per B event is selected. This particle will be called tagging candidate in the following. Its charge determines the B production flavour. Depending on the B type several taggers can be applied.

The following sections explain the principles of flavour tagging in detail. First, the key quantities of flavour tagging are introduced and a method to combine multiple taggers is shown. Then, different taggers are presented, highlighting their advantages and challenges. A typical tagger development procedure is sketched. A derivation of the tagging power relevant for mixing asymmetry analyses is given.

3.1. Tagging Key Quantities

Flavour taggers are never perfect. There is a natural non zero probability of mistag as will be explained below. Additionally, in some B events none of the particles fulfill the requirements of the tagger. These events remain untagged. Therefore, the tagging efficiency ε_{tag} is defined as the number of tagged B events divided by all B events

$$\varepsilon_{\text{tag}} = \frac{\text{tagged events}}{\text{all events}}. \tag{3.1}$$
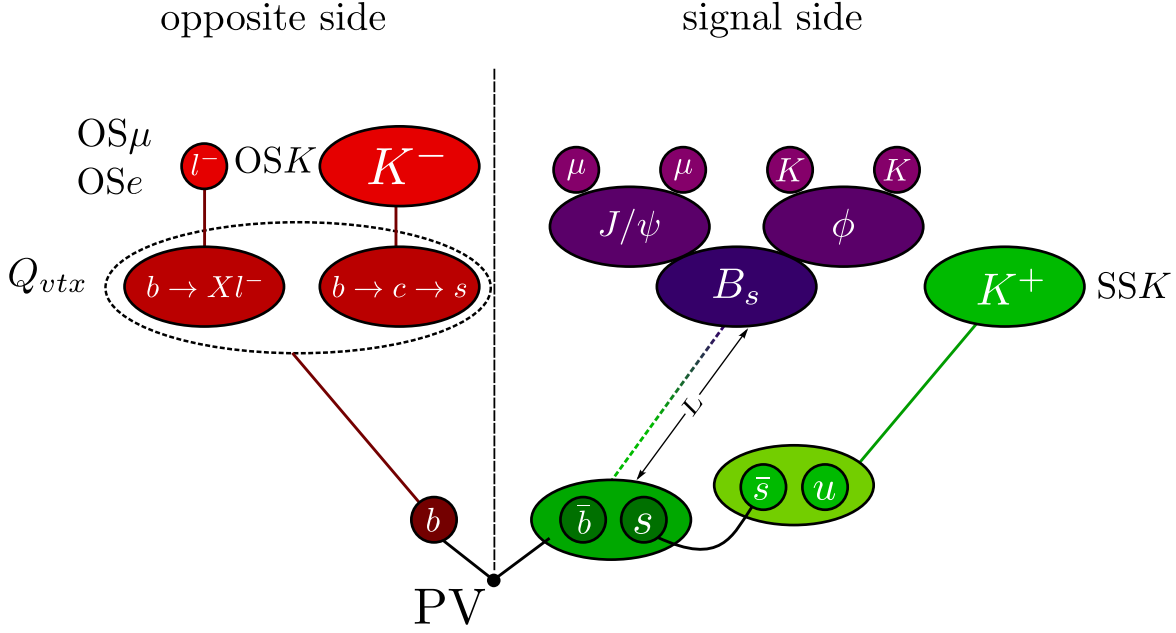
Figure 3.1.: Schematic to illustrate a typical B event. A $b\bar{b}$ quark pair is produced at the primary vertex (PV). The $\bar{b}$ quark hadronizes into the signal B meson, here a B_s^0, which further decays into $J/\psi \rightarrow \mu^+\mu^-$ and $\phi \rightarrow K^+K^-$ final states, classified as signal side of the event. On the opposite side the associated b undergoes either $b \rightarrow c \rightarrow s$ transitions and / or a semileptonic $b \rightarrow Xl^-$ decay. At LHCb different flavour taggers are applied to determine the production flavour of the B_s^0: the Signal Side Kaon tagger (SSK), the Opposite Side Kaon (OSK), Muon (OSμ) or Electron (OSe) taggers and the vertex charge tagger (Q_{vtx}).

The average mistag probability ω is defined as the number of wrongly tagged B events divided by all tagged B events

$$\omega = \frac{\text{wrong tagged events}}{\text{all tagged events}}. \tag{3.2}$$

A closely related quantity, the tagging dilution D, is defined as

$$D = (1 - 2\omega). \tag{3.3}$$

The figure of merit is the effective tagging efficiency ε_{eff}, also called tagging power, which is defined as

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}} D^2 = \varepsilon_{\text{tag}}(1 - 2\omega)^2. \tag{3.4}$$

These variables quantify the performance of flavour tagging algorithms. The tagging power ε_{eff} is directly connected to the statistical power of a tagged data set used for the mixing asymmetry measurement. The derivation of this connection is shown in section 3.4.2. A tagger optimization maximizes ε_{eff}.

3.2. Opposite Side Tagging Algorithm

Opposite side tagger exploit the decay of the associated b hadron. They partially reconstruct the opposite b hadron decay. Semileptonic decays of the opposite b hadron are exploited by the OSμ and OSe taggers using the charge of the lepton to tag the production flavour of the opposite b hadron and coincidentally the production flavour of the associated signal $\bar{b}$ hadron. The OSK tagger uses the charge of the kaon coming from $b \to c \to s$ transitions to tag the production flavour. The OS vertex charge tagger computes a weighted sum of the charges of the tracks that originate from the opposite b decay. It sums over all tracks used to reconstruct the vertex.

A hadron and coincidentally the production flavour of the associated signal is, that they are independent of the signal B decay. They can be developed on high statistics flavour specific channels (e.g. $B^+ \to J/\psi\, K^+$), that do not undergo flavour oscillations. Thus, the production flavour of the signal B meson is equal to the decay flavour. The number of wrongly tagged events is determined by comparing the tagging decision and the charge of the final state particle.

A challenge for opposite side taggers are B events with high track multiplicities. Of $\mathcal{O}(100)$ tracks per event are typically produced at the LHC, see figure 3.2, showing a graphical representation of a simulated $B_s^0 \to J/\psi\, \phi$ event in the LHCb detector.
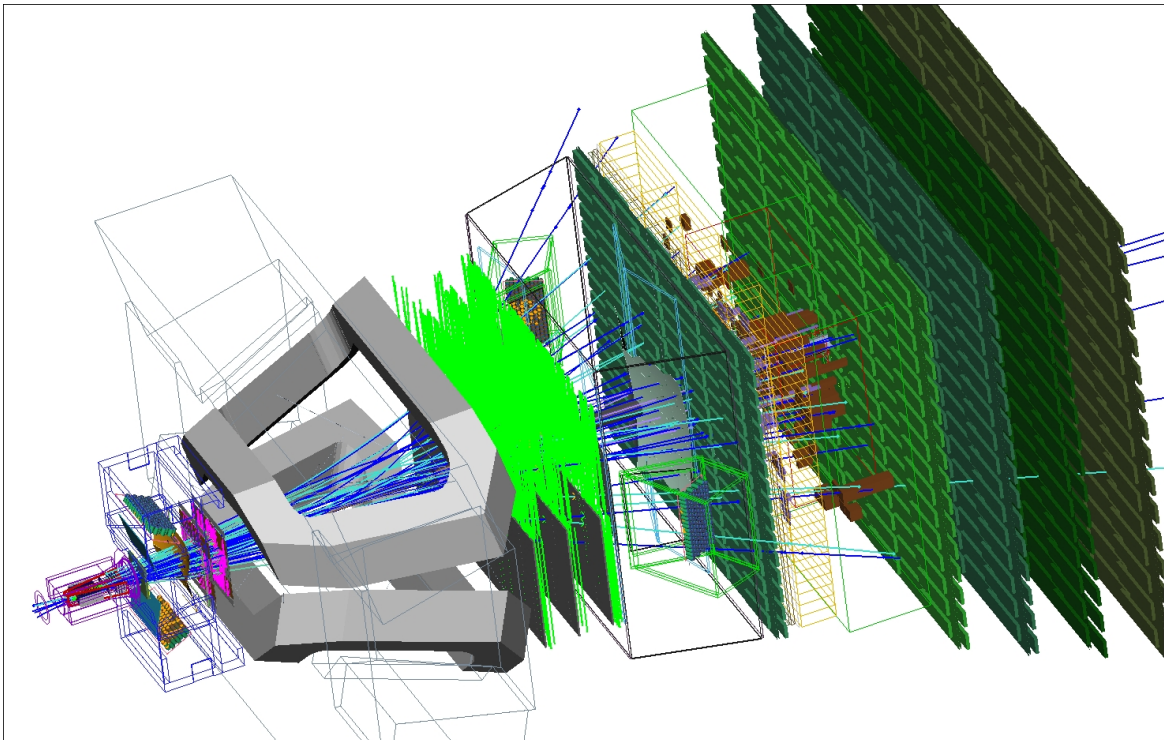

Figure 3.2.: Graphical representation of a simulated $B_s^0 \to J/\psi\, \phi$ event in the LHCb detector. The reconstructed track multiplicity is of $\mathcal{O}(100)$.

Charged soft QCD background particles from the primary vertex or particles from any other vertex in the event, not related to the opposite b decay, can be selected as tagging candidates. The

charge of these background particles gives a random tagging decision. If the opposite b hadron oscillates, in case it is a B^0 or B_s^0 meson, the tagging decision is no longer related to the signal B production.

The relative fractions of producing b flavoured hadrons at the LHC (B^+: B^0: B_s^0: b-baryon) is roughly given by 41:41:11:8 % [22]. Among these, B^0 and B_s^0 mesons oscillate. B_s^0 mesons oscillate very fast, the mean B_s^0 lifetime is $\tau = (1.512 \pm 0.007)\,\text{ps}$ with an oscillation frequency $\Delta m_s = (17.761 \pm 0.022)\,\text{ps}^{-1}$ [72]. A tagging decision based on the charge of an opposite B_s^0 decay final state track gives a random answer with a mistag probability of 50%. Whereas, B^0 mesons oscillate relatively slow, the mean B^0 lifetime is $\tau = (1.519 \pm 0.005)\,\text{ps}$ with an oscillation frequency $\Delta m_d = (0.510 \pm 0.003)\,\text{ps}^{-1}$ [72]. If the tagger uses the charge of a B^0 decay final state track, it is more likely to give the correct tagging decision, than the wrong tagging decision.

OS lepton taggers can tag the B event wrongly, if a lepton from a semileptonic $b \to c \to s\ell\nu$ decay is selected instead of a lepton from a semileptonic b hadron decay, see figure 3.3.

(a) Semileptonic b decay (b) Semileptonic c decay
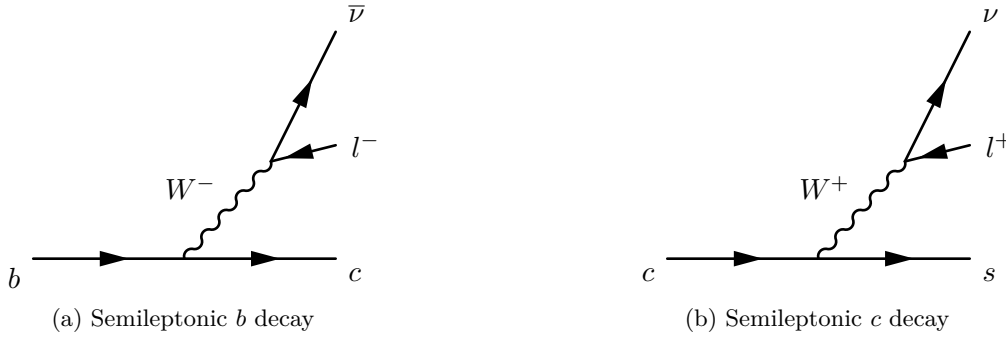
Figure 3.3.: The charge of the lepton from a semileptonic b hadron decay (a) is opposite to the charge of the lepton from a semileptonic c hadron decay (b). If the OS lepton tagger selects a lepton from a semileptonic $b \to c$-hadron decay instead of a semileptonic b hadron decay, the tagging decision is wrong.

The OS vertex charge tagger can derive the wrong tagging decision, if background particles are selected by the tagger, which did not originate from the vertex. On the other hand, it can happen that particles, that originate from the vertex, are not selected by the OS vertex tagger. The weighted sum over the charges of the particles can yield the wrong sign in both cases. All effects together result in a rather high mistag probability of $\sim 30 - 40\%$ for the OS taggers.

3.3. Signal Side Tagging Algorithm

Signal side taggers select charged particles produced during the fragmentation of the signal b hadronization. In the B_s^0 decay shown above, a $\bar{b}$ hadronizes together with a s quark into a B_s^0 meson. The s quark also comes from a $s\bar{s}$ quark pair. Hence, the associated $\bar{s}$ quark can hadronize into a charged kaon, which happens in about 50% of the cases [73], picked up by the Signal Side Kaon (SSK) tagger. When the signal B is a B^0 meson, the $s\bar{s}$ pair is replaced by a $d\bar{d}$ pair, so that the associated $\bar{d}$ quark can hadronize into a charged pion, utilized by the Signal Side Pion tagger (SSπ).

An advantage of signal side taggers is, they contribute significantly to the tagging power[1] achieved by opposite side taggers. SS and OS taggers are almost completely uncorrelated.

SS taggers depend on the signal B type signal. They are applicable to B^0 and B_s^0 decays, only. Thus, a challenge for the algorithm development and calibration is the oscillation of the neutral B, which has to be resolved. Another challenge for SS taggers is the high track multiplicity at the LHC. The mistag probability is 50% if the SS tagger selects randomly a particle from the PV, which is not involved in the signal b hadronization process. In this case, the particle stems from the so-called underlying event (UE). On the other hand, the tagging efficiency is additionally reduced by a factor of 2, because only in 50% of the cases a charged kaon or pion is produced during fragmentation. This results in a rather high mistag probability of $\sim 35\%$ and a reduced tagging efficiency of $\sim 16\%$.

3.4. Tagging Algorithms used in LHCb

The standard LHCb taggers are documented in [75] and [76]. Taggers are applied to events, that contain signal B decays. To execute a tagger all tracks in such events have to be reconstructed. Tracks are the trajectories of charged particles reconstructed in the tracking detectors. Taggers apply selection requirements to these particles with the goal, to find the correct tagging candidate, which is the particle, whose charge is correlated to the production flavour of the signal b quark. Different taggers apply different selection requirements, accordingly.

A common pre-selection is applied to all particles in the event prior to the execution of the taggers, see table 3.1. The purpose of the pre-selection is to suppress a large amount of soft QCD background. This reduces the execution time of the taggers. Additionally, the pre-selection ensures that particles from the signal B decay are not selected by the taggers. This is important, because a decay product of the B decay would tag the decay flavour of the B, not the production flavour.

A geometric requirement, which is determined as the sum of the minimum distances between the B decay daughters and the tagging particle, ensures that the particle is not part of the signal B decay chain. Additionally, the azimuthal angle between any B daughter and the tagging particle has to be greater than 5 mrad. The momentum, p, of the particle has to be above $2\,\text{GeV}/c$ to reduce the soft QCD background. An upper limit on p and transverse momentum, p_T, of the particle enhances the probability to not pick up a composed particle rather than a LHCb "stable" kaon, pion, electron or muon[2].

In order to suppress poorly reconstructed tracks, the track fit quality of the tagging particle has to fulfill a minimum requirement. The track needs to have hits in the vertex detector and in the tracking stations, either before the dipole magnet (upstream tracks) or before and behind the magnet (long tracks), see figure 2.8. Particles that traverse the beam pipe suffer typically from multiple scattering. Hence, they are poorly reconstructed. To avoid picking up those tracks a minimum requirement is applied to the polar angle, which is the angle between the particle's momentum direction and the beam line.

---

[1]Compared to lepton colliders the tagging power at hadron colliders is roughly a factor 10 worse, $\varepsilon_{\text{eff}} \simeq 3\%$ instead of $\varepsilon_{\text{eff}} \simeq 30\%$ [74].

[2]Kaons, pions, electrons and muons leave hits in all LHCb tracking detectors.

Quantity	Cut Value
geometric	$\notin$ signal decay chain
$|\Delta\phi|$	$> 5\,\mathrm{mrad}$
p	$> 2\,\mathrm{GeV}/c$
p	$< 200\,\mathrm{GeV}/c$
p_t	$< 10\,\mathrm{GeV}/c$
track type	*long* or *upstream*
charge	$\pm$
θ	$> 12\,\mathrm{mrad}$
PU IP significance	> 3

Table 3.1.: LHCb tagging particle pre-selection. This set of selection requirements is applied to all reconstructed particles in the signal B event prior to the execution of the taggers.

At typical LHCb running conditions up to five primary vertices are reconstructed per event, called pile up (PU). In order to not select a particle, which is produced in a PU decay, a minimum requirement is applied to the PU impact parameter significance. The impact parameter (IP) with respect to a vertex, is the shortest distance between the particle's track and the vertex. The IP significance takes the uncertainty on the IP determination into account. It is computed as the quotient of the IP divided by its uncertainty.

Then, each tagger applies an individual set of selection requirements to the particles passing the pre-selection. These selections are optimized for each tagger separately. If more than one particles passes the tagger selection, the one with the highest p_T is used as tagging candidate. Its charge determines the tagging decision.

Finally, an artificial Neural Net (ANN) predicts the mistag probability of the tagging decision. The ANN output η_i is a prediction for the mistag probability ω_i of the event i. The ANN is trained on MC simulation. To ensure a correct mistag probability prediction on data, the predicted mistag probability is calibrated to the measured mistag probability $\omega_i(\eta_i)$ on data using a first order polynomial

$$\omega_i = p_0 + p_1 \cdot (\eta_i - <\eta>), \tag{3.5}$$

where $<\eta>$ is the mean value of the predicted mistag distribution on the tagged data set. The calibrated predicted mistag probability ω_i per event i can than be used to sum up an event-by-event dilution, defined as

$$D_{event-by-event} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(1 - 2\omega_i)^2}, \tag{3.6}$$

where N is the number of tagged B mesons. In an event-by-event approach, the tagged events can be weighted according to their predicted mistag probability. Compared to the average dilution,

introduced in equation 3.3, this improves the tagging power, $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}D^2$, because

$$D_{event-by-event} > D_{average} = \sqrt{(1 - 2\frac{1}{N}\sum_{i=1}^{N}\omega_i)^2}. \tag{3.7}$$

The single flavour tagging algorithms have different working points. The OS lepton taggers, for example, deliver more often a correct tagging decision, thus their mistag probability is relatively low. However, they are also less efficient, thus have a low tagging efficiency. In contrast, the OS and SS kaon taggers are relatively efficient, but their mistag probability is also higher, because the selection of a hadronic tagging particle is more difficult due to the high amount of soft QCD background than the selection of a leptonic particle.

To get a maximum tagging performance on a given data set the combination of multiple taggers is used. Each tagger is calibrated individually beforehand. The responses of the different taggers are combined to derive a single tagging decision and mistag probability prediction per signal B meson. The combination method is shown in the next subsection.

3.4.1. Combination of Tagging Algorithms

Each tagging algorithm provides a tagging decision d_i and a predicted mistag probability η_i per signal B meson. Here, i stands for the ith tagger, applied to the same B meson. In case, the B is not tagged by the tagger i, $d_i = 0$ and $\eta_i = 0.5$. The combined probability $P(b)$ / $P(\bar{b})$ that the B meson contains a b/ $\bar{b}$ quark is calculated as

$$P(b) = \frac{p(b)}{p(b) + p(\bar{b})}, \qquad P(\bar{b}) = 1 - P(b), \tag{3.8}$$

where $p(b)$ is the product of in the current event available tagging decisions d_i, weighted by the corresponding predicted mistag probability η_i

$$p(b) = \prod_i(\frac{1 + d_i}{2} - d_i(1 - \eta_i)), \qquad p(\bar{b}) = \prod_i(\frac{1 - d_i}{2} + d_i(1 - \eta_i)). \tag{3.9}$$

The index i loops over the available taggers in the given signal B event. The convention is, that the combined tagging decision $d_i = 1(-1)$ if the signal B contains a $\bar{b}$ (b). Accordingly,

- if $P(b) > P(\bar{b})$, than $d = -1$ and $\eta = 1 - P(b)$

- if $P(b) < P(\bar{b})$, than $d = +1$ and $\eta = 1 - P(\bar{b})$.

A basic assumption entering this combination technique is that the single taggers are uncorrelated. However, to account for possible correlations the combined predicted mistag probability is also calibrated on data against the measured mistag probability using the same calibration function as for the single taggers, see equation 3.5.

3.4.2. Derivation of the Tagging Power ε_{eff}

The mixing asymmetry $\mathcal{A}$ is defined as the difference of the number of B mesons, that decay with the same flavour compared to the production flavour (unmixed), minus the number of B mesons, that decay with the opposite flavour compared to the production flavour (mixed), divided by the sum

$$\mathcal{A} \equiv \frac{N_{\text{unmixed}} - N_{\text{mixed}}}{N_{\text{unmixed}} + N_{\text{mixed}}} \equiv \frac{N - \overline{N}}{N + \overline{N}}. \tag{3.10}$$

The second definition is done for better readability. However, the production and decay flavours are determined experimentally. They are known up to a certain mistag probability ω. Hence, the measured number of unmixed (mixed) B mesons N_m ($\overline{N}_m$) is a superposition of the true number of unmixed (mixed) B mesons N ($\overline{N}$) and mixed (unmixed) B mesons $\overline{N}$ (N):

$$N_m = (1 - \omega)N + \omega\overline{N} \tag{3.11}$$

$$\overline{N}_m = (1 - \omega)\overline{N} + \omega N. \tag{3.12}$$

Analogue, the measured mixing asymmetry $\mathcal{A}_m$ is:

$$\mathcal{A}_m = \frac{N_m - \bar{N}_m}{N_m + \bar{N}_m} = (1 - 2\omega)\mathcal{A} = D\mathcal{A}. \tag{3.13}$$

In a perfect world, the mistag probability ω would be 0, resulting in a dilution $D = 1$ and $\mathcal{A}_m = \mathcal{A}$. In case of random tagging, the mistag probability is 50%, resulting in a dilution $D = 0$ and $\mathcal{A}_m = 0$. The statistical uncertainty on the measured asymmetry is given by

$$\sigma_{\mathcal{A}_m} = D\sigma_{\mathcal{A}}. \tag{3.14}$$

The measured numbers of B decays carry the corresponding uncertainties of $\mathcal{A}$. They are propagated into equation 3.14 using Gaussian error propagation:

$$\sigma_{\mathcal{A}_m}^2 = (\frac{\partial \mathcal{A}_m}{\partial N_m})^2 \sigma_{N_m}^2 + (\frac{\partial \mathcal{A}_m}{\partial \bar{N}_m})^2 \sigma_{\bar{N}_m}^2 = \frac{4\bar{N}_m^2 \sigma_{N_m}^2 + 4N_m^2 \sigma_{\bar{N}_m}^2}{(N_m + \bar{N}_m)^4}. \tag{3.15}$$

With the assumption of Poisson distributed uncertainties

$$\sigma_{N_m} = \sqrt{N_m}, \quad \sigma_{\bar{N}_m} = \sqrt{\bar{N}_m} \tag{3.16}$$

and the help of the following identity

$$1 - \mathcal{A}_m^2 = \frac{4N_m \bar{N}_m}{(N_m + \bar{N}_m)^2}, \tag{3.17}$$

the statistical uncertainty on the measured asymmetry becomes

$$\sigma_{\mathcal{A}_m}^2 = \frac{4N_m \bar{N}_m}{(N_m + \bar{N}_m)^3} = \frac{1 - \mathcal{A}_m^2}{N_m + \bar{N}_m} = \frac{1 - \mathcal{A}_m^2}{\varepsilon_{\text{tag}}\mathcal{N}} = \frac{1 - D^2\mathcal{A}^2}{\varepsilon_{\text{tag}}\mathcal{N}}. \tag{3.18}$$

In the second last step the relation, that the measured number of unmixed plus mixed B mesons is equal to the number of all B mesons in the sample $\mathcal{N}$ times the tagging efficiency ε_{tag}, is used. In the last step equation 3.13 is used. Using equation 3.14, the statistical uncertainty on the mixing asymmetry becomes

$$\sigma_{\mathcal{A}}^2 = \frac{1 - D^2 \mathcal{A}^2}{\varepsilon_{\text{tag}} \mathcal{N} D^2} = \frac{1}{\varepsilon_{\text{tag}} \mathcal{N}} \left(\frac{1}{D^2} - \mathcal{A}^2 \right). \tag{3.19}$$

At LHCb typical values of the mistag probability are $\omega \sim 40\%$. That yields a dilution of $D \sim 0.2$, while mixing asymmetries $\mathcal{A}$ are in $[-1, +1]$. Hence, the approximation

$$\frac{1}{D^2} \gg \mathcal{A}^2 \tag{3.20}$$

is valid, which gives a statistical uncertainty on the mixing asymmetry of

$$\sigma_{\mathcal{A}} = \sqrt{\frac{1}{\mathcal{N} \varepsilon_{\text{tag}} D^2}} = \sqrt{\frac{1}{\mathcal{N} \varepsilon_{\text{eff}}}} \tag{3.21}$$

The statistical power of a tagged data set is at maximum, if the uncertainty $\sigma_{\mathcal{A}}$ is at minimum. To minimize $\sigma_{\mathcal{A}}$ the data set needs to be large, large $\mathcal{N}$, and the tagging power ε_{eff} needs to be maximal. Hence, ε_{eff} is the figure of merit for flavour tagging optimisation.

3.5. Data MC Comparisons for the cut-based OS Kaon Tagging Algorithm

This section presents a preparatory study for the development of an improved OS kaon tagger. The existing LHCb OS kaon tagger is studied in detail, especially the reason for the better performance in MC simulation than in real data is evaluated. Five topics are investigated: the possible data MC differences in particle identification, in the number of primary vertices per event, in track multiplicities per event, in the impact parameter resolution and in the quality of the track fit. Effective corrections are successively applied to MC simulation to obtain similar tagging performance on MC compared to data. The goal of this study is not to reproduce the data tagging performance found in MC simulation, but rather to understand the main causes for the different performances. At the end of the section a short summary highlights the most important points.

The performance of the flavour tagging algorithms is significantly better in MC simulation than in real data. Table 3.2 gives the tagging performance of the single OS taggers and their combination in MC simulation and data. The performance is measured on $B^+ \to J/\psi K^+$ decays in $1\,\text{fb}^{-1}$ of data, collected at $\sqrt{s} = 7$ TeV in 2011 by the LHCb experiment. The signal selection is taken from the LHCb measurement of the B^+, B^0, B_s^0 meson and Λ_b^0 baryon lifetimes [77].

The reconstructed B^+ mass distribution is fitted in data by a double Gaussian distribution to model the signal and an exponential distribution to model the combinatoric background, see figure 3.4. Due to the clean trigger signature of the two muons and the minimum requirement on the B^+ decay time the amount of background is low. A signal yield of 2.17×10^5 $B^+ \to J/\psi K^+$

sample	OS tagger	$\epsilon_{tag}(\%)$	$\omega(\%)$	$\epsilon_{eff}(\%)$
MC	muon	4.60 ± 0.02	30.62 ± 0.23	0.69 ± 0.02
MC	electron	2.67 ± 0.02	33.63 ± 0.31	0.29 ± 0.01
MC	kaon	15.91 ± 0.06	35.13 ± 0.18	1.41 ± 0.03
MC	vertex	17.86 ± 0.04	37.00 ± 0.12	1.21 ± 0.02
MC	combination	31.10 ± 0.05	36.29 ± 0.09	2.34 ± 0.03
data	muon	5.16 ± 0.05	30.20 ± 0.43	0.81 ± 0.03
data	electron	2.44 ± 0.03	31.77 ± 0.64	0.32 ± 0.02
data	kaon	17.57 ± 0.08	39.33 ± 0.25	0.80 ± 0.04
data	vertex	18.29 ± 0.08	40.18 ± 0.24	0.71 ± 0.03
data	combination	33.47 ± 0.10	39.28 ± 0.18	1.54 ± 0.05

Table 3.2.: Tagging performance of the single OS taggers and of the combined OS tagger. The large difference in the tagging power ε_{eff} of the OS combination between MC and data is caused by the worse performance of the OS kaon and vertex charge taggers in data.
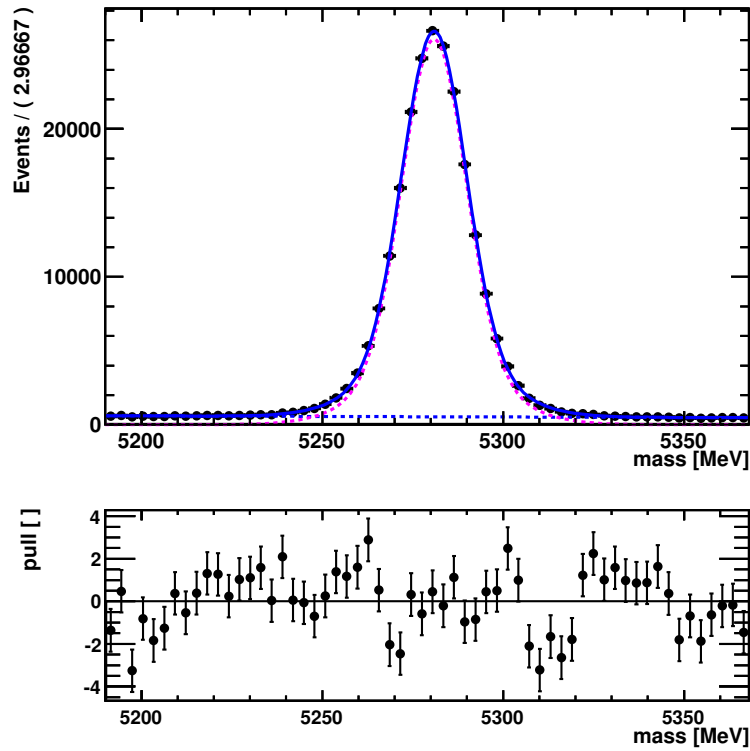

Figure 3.4.: Fit to the reconstructed B^+ mass distribution.

Fit Parameter	Fitted Value
# signal	217429 ± 527
# background	32508 ± 309
μ	5280.920 ± 0.024
σ_1	8.706 ± 0.071
σ_2	16.92 ± 0.38
fraction	0.756 ± 0.015
slope	-0.0013 ± 0.0001

Table 3.3.: Parameters obtained from a fit to reconstructed B^+ mass distribution.

decays is extracted, see table 3.3. For the computation of the tagging performance, the background events are statistically removed using the sWeights technique [78, 79].

Since the B^+ meson does not undergo flavour oscillations, the number of wrong tagged events is simply determined by comparing the charge of the final state signal decay kaon with the tagging decision. Then, the mistag probability on the data set is computed by the quotient of wrong tagged events, divided by all events. The average mistag probability fulfills the aim of this study. The predicted mistag probability, used to compute an event-by-event weighted tagging power, is not considered here, however, the trend of the difference in performance between MC and data is the same. The tagging performance numbers, shown in table 3.2, are the starting point of the study presented in the following.

The combined OS tagging power is a factor of 1.52 better in MC simulation than in data. It is computed as a combination of the single OS taggers following the combination technique given in subsection 3.4.1. The OS lepton tagger performance does not cause this huge difference. The mistag probability is compatible in data and MC for these taggers, while for the OSμ tagger the tagging efficiency is even slightly better in data than in MC.

However, the mistag probability of the OS kaon and vertex charge tagger is significantly underestimated by the MC simulation. In case of the OSK tagger this is 12% in simulation. The tagging efficiency of both taggers is also higher in data. That indicates, that the taggers pick up more tracks in data in total, but these tracks are more likely to give a random tagging decision. Hence, they are probably not related to the signal B decay. Therefore, a dedicated study on the OSK tagger has been performed to determine the reason for that large difference in performance between data and MC shown in the following. The study has been made available to the LHCb collaboration as well as the improved OSK tagger, documented LHCb-internally in [2].

Step by step different effects, that are likely to cause the difference in tagging power, are investigated. The order of the effects is important. When a significant change on the tagging power is found, the effect is corrected in MC simulation. All correction methods have been developed in the course of this thesis.

3.5.1. The cut-based OS Kaon Tagging Algorithm

The OS kaon tagger tries to find the charged kaon from opposite side $b \rightarrow c \rightarrow s$ transition. It applies a selection of cuts, listed in table 3.4, to the particles that fulfill the requirements of the tagging pre-selection.

Quantity	Cut Value
p	$> 5\,\text{GeV}$
p_t	$> 0.7\,\text{GeV}$
track $\frac{\chi^2}{ndf}$	< 2.45
$\text{DLL}_{K-\pi}$	> 0.75
DLL_{K-p}	> -3.0
IP significance	> 4.3
IP	$< 1.45\,\text{mm}$
PU IP significance	> 7.5

Table 3.4.: Tagging particle selection requirements of the OS kaon tagger. The requirements of the tagging pre-selection, listed in table 3.1, are applied prior to the execution of the OS kaon tagger.

The right tagging particle comes from a b hadron decay. Thus, it has on average a higher momentum than background particles from the underlying event. In order to suppress mis-reconstructed tracks, the particle's track has to be of good track fit quality. Requirements on the particle identification likelihood ($\text{DLL}_{K-\pi}$, DLL_{K-p}) ensure to select kaons not pions or protons.

Opposite b hadron decays are displaced from the PV. This property is used by the tagger, when placing a minimum requirement on the IP significance of the particle with respect to the primary vertex. However, due to the pile up (PU) of multiple primary vertices in the event, a maximum requirement on the IP and a minimum requirement on the PU IP significance ensure that the selected particle does not come from a PU vertex.

This rather tight selection results in a tagging efficiency of $\sim 18\%$ on data. On average 1.2 particles are selected per signal B event by the tagger. The charge of the particle with the highest transverse momentum is used to give the tagging decision.

3.5.2. Effect of the Particle Identification

This subsection determines the influence of possible differences in the particle identification (PID) in MC simulation compared to data. In data the PID cut efficiencies are measured on clean samples of kaons, pions and protons, provided by the LHCb PID calibration group [80]. In MC simulation the MC truth information is used to test the particle hypothesis and to measure the PID selection efficiency. The particles of the calibration samples cover a different phase space compared to the tagging particles. Thus, the calibration samples are re-weighted to match the tagging particle phase space.

A weighting in the track multiplicity and in the two dimensional momentum and pseudo rapidity distribution is applied. The PID distributions before and after reweighting are shown in figure 3.5 for the kaons calibration sample. The distributions for the pions and protons samples look similar, given in appendix A.1. The distributions of the PID calibration data samples are shown on top of the tagging particle MC simulation sample. The phase space reweighting reduces the statistical power of the calibration sample significantly.
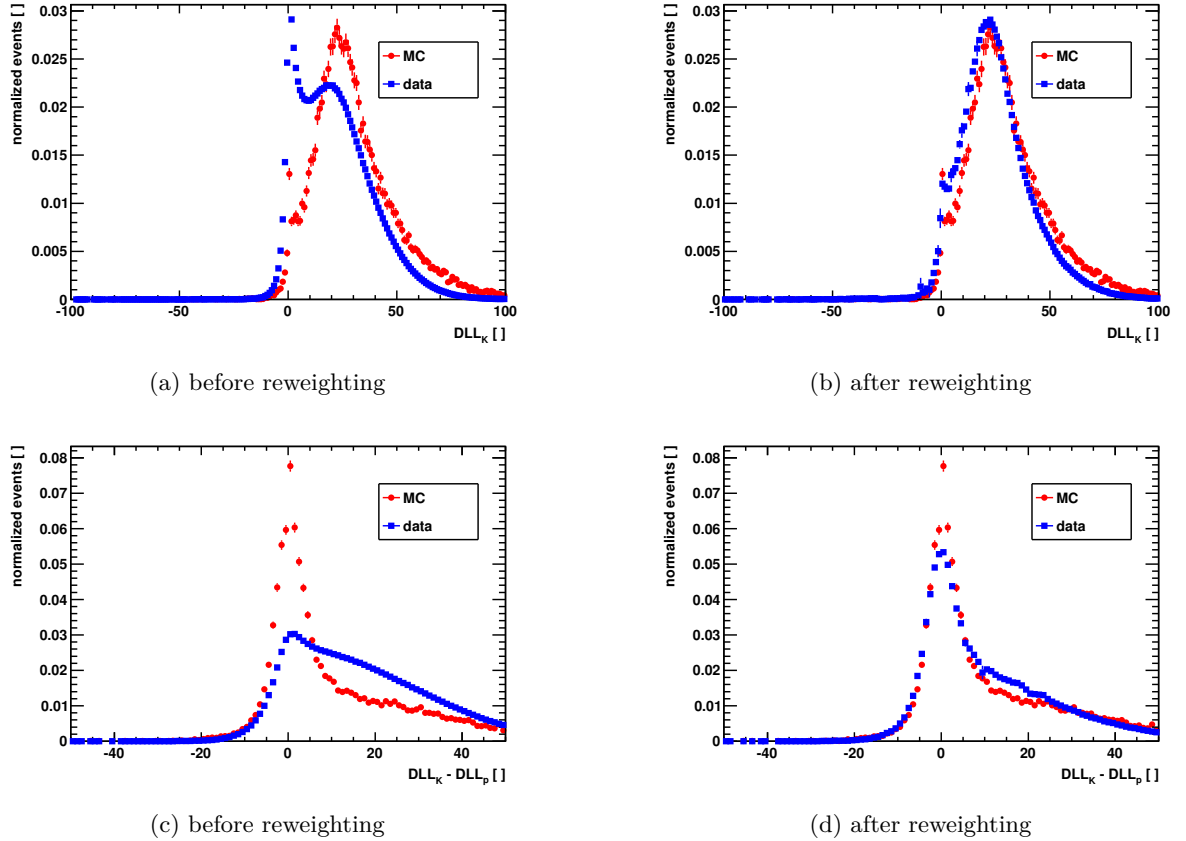

(a) before reweighting

(b) after reweighting

(c) before reweighting

(d) after reweighting

Figure 3.5.: $DLL_{K-\pi}$ distributions of true kaons in MC and in calibration data (a) before reweighting, (b) after reweighting. DLL_{K-p} distributions of true kaons in MC and in calibration data (c) before reweighting, (d) after reweighting.

Relevant for this study is cut efficiency of the PID cuts applied by the tagger. The PID cut efficiencies, measured on MC and on PID calibration data separately, are given in table 3.5.

The PID cuts are more efficient on MC simulation. Therefore, the effect of this difference on the tagging performance in MC is evaluated.

A worse PID cut efficiency is simulated in MC during the execution of the tagger. The effect on the tagging performance is determined, shown in table 3.6. The absolute and relative changes in the three tagging performance quantities: ε_{tag}, ω and ε_{eff} are given.

PID efficiency	$\epsilon(K-\pi)(\%)$	$\epsilon(K-p)(\%)$	combined $\epsilon(\%)$
MC kaon	97.09 ± 0.09	89.22 ± 0.17	86.86 ± 0.19
data kaon	96.23 ± 0.25	87.74 ± 0.12	84.95 ± 0.23
MC pion	15.89 ± 0.11	53.19 ± 0.15	9.73 ± 0.09
data pion	11.49 ± 0.04	51.68 ± 0.04	7.89 ± 0.04
MC proton	88.58 ± 0.27	27.65 ± 0.38	25.34 ± 0.37
data proton	85.75 ± 1.15	23.78 ± 0.99	21.95 ± 0.82

Table 3.5.: Kaon, pion and proton PID efficiencies and misidentifications, respectively, obtained with cuts at $\mathrm{DLL}_{K-\pi} > 0.75$ and $\mathrm{DLL}_{K-p} > -3.0$. The numbers for MC are determined with the help of MC truth match information, for data, clean PID calibration data samples are used.

change	$\Delta\varepsilon_{\mathrm{tag}}(\%)$	$\Delta\omega(\%)$	$\Delta\varepsilon_{\mathrm{eff}}(\%)$
absolute	-0.35 ± 0.01	-0.21 ± 0.01	$+0.01 \pm 0.03$
relative	-2.18 ± 0.07	-0.61 ± 0.19	$+0.65 \pm 0.06$

Table 3.6.: Absolute and relative change in the tagging performance due to the differences in the PID cut efficiency estimated on MC compared to the nominal tagging performance of the OSK tagger on MC. The Variation accounts for the PID efficiency and misidentification measured in data using reweighted calibration channels. This effect is negligible.

Modeling a worse PID cut efficiency in MC simulation lowers $\varepsilon_{\mathrm{tag}}$. However, the change in $\varepsilon_{\mathrm{eff}}$ is negligible. Hence, the difference in PID performance does not cause the different tagging performance in MC simulation and data.

3.5.3. Effect of the Number of Primary Vertices

The pile up of multiple primary interactions per event is a problem for flavour tagging. The tagging algorithms perform significantly better in events with less primary vertices, both in data and in MC simulation. Table 3.7 shows the tagging performance in data and in MC using subsamples separated by the number of PVs per event. In data the tagging power is 60% worse in events with 5 or more PVs compared to events with 1 PV, only. While in MC the tagging power is 30% worse in events with 5 or more PVs compared to events with 1 PV, only.

Figure 3.6 shows the number of PVs distribution in data and MC. The distributions are determined on the sample of tagging candidates using $B^+ \to J/\psi K^+$ decays, having applied the tagging preselection and the PID cuts. The histograms are normalized to an area of one. The distributions are very similar, in data there are 2.01 PVs per event on average, while in MC there are 1.99 PVs per event on average. However, a correction to MC is applied, in order to match

# PVs/event	$\epsilon_{tag}(\%)$	$\omega(\%)$	$\epsilon_{eff}(\%)$
1	17.66 ± 0.14	38.91 ± 0.42	0.87 ± 0.06
2	17.66 ± 0.13	39.65 ± 0.41	0.76 ± 0.06
3	17.16 ± 0.18	38.58 ± 0.58	0.84 ± 0.08
4	16.98 ± 0.33	40.29 ± 1.04	0.64 ± 0.13
≥ 5	17.09 ± 0.65	42.81 ± 2.06	0.35 ± 0.19
1	15.84 ± 0.09	33.74 ± 0.30	1.67 ± 0.06
2	15.83 ± 0.09	35.23 ± 0.30	1.38 ± 0.05
3	16.04 ± 0.13	36.86 ± 0.42	1.11 ± 0.07
4	16.41 ± 0.23	36.83 ± 0.75	1.14 ± 0.12
≥ 5	16.08 ± 0.46	36.74 ± 1.51	1.13 ± 0.25

Table 3.7.: Tagging performance of the OSK tagger split by the number of reconstructed PVs per event determined in 1 fb^{-1} data (top), collected in 2011, and in MC (MC11a) (bottom) using the FlavourTagging package v12r7.

exactly the data distribution, by throwing away randomly events in MC until the data distribution is reproduced, see subfigure (b) in figure 3.6.
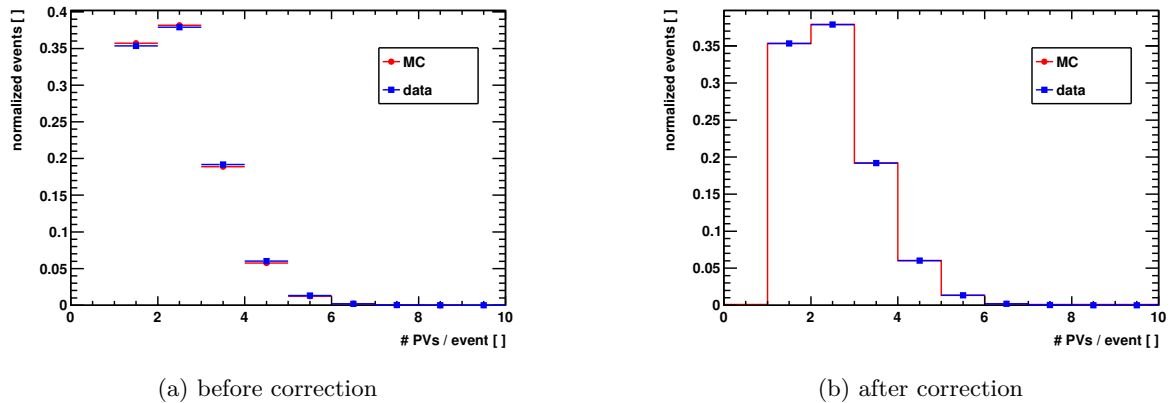

(a) before correction

(b) after correction

Figure 3.6.: Number of PVs per event in data and MC (a) before correction, (b) after correction.

The change in tagging performance on MC is given in table 3.8. The mistag probability is increased by the correction, which in turn lowers the tagging power by relatively 1.4%. This effect is small. Further studies are determined on the PV corrected MC sample, which is split up into five subsamples according to the number of PVs per event.

change	$\Delta\varepsilon_{\text{tag}}(\%)$	$\Delta\omega(\%)$	$\Delta\varepsilon_{\text{eff}}(\%)$
absolute	-0.02 ± 0.01	$+0.06 \pm 0.01$	-0.02 ± 0.04
relative	-0.13 ± 0.07	$+0.17 \pm 0.19$	-1.44 ± 0.07

Table 3.8.: Absolute and relative change in the tagging performance estimated on MC for a data driven correction applied to the number of primary vertices per event. For increasing number of PVs per event, the mistag fraction increases, while the tagging power decreases. This effect is small.

3.5.4. Effect of the Track Multiplicity

Track multiplicity per event is underestimated by the MC simulation, see figure 3.7. The correction to MC to match the number of reconstructed PVs in data is already applied. The simulation of hard processes in B decays is very well understood and modeled. Hence, a possible explanation is that the amount of low QCD background is not modeled correctly. This so-called "underlying event" (UE) is underestimated by simulation. If the tagger picks up a particle from UE the mistag probability is 50%, because the particle does not carry any information about the B production. This class of tagging background is most likely responsible for the worse tagging performance in data.

(a) Track Multiplicity before Correction

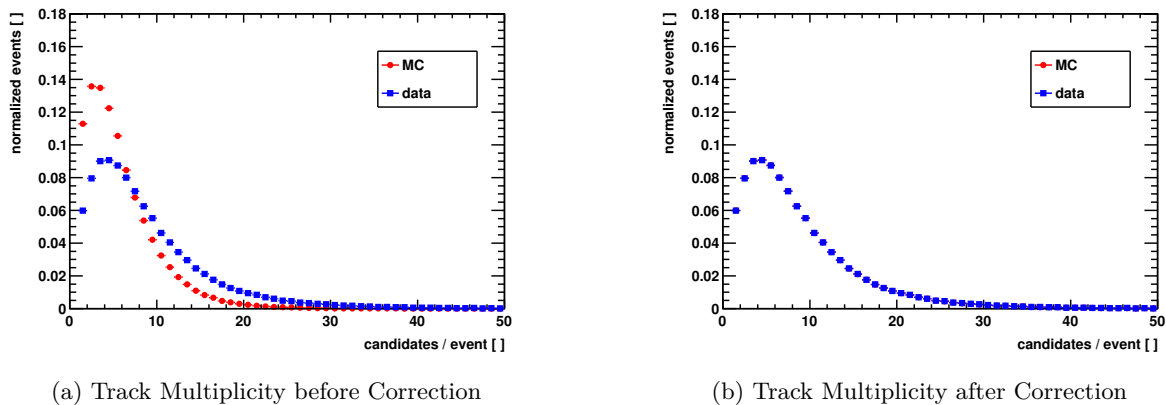(b) Track Multiplicity after Correction

Figure 3.7.: Track multiplicity distribution (a) before and (b) after correction.

The OS kaon tagger uses particles reconstructed from two types of tracks, labeled in LHCb with long tracks and upstream tracks, as introduced in chapter 2. Using additionally the class of upstream tracks does not improve the tagging power, see table 3.9. The tagging performance is given for data and MC, separately determined on a sample of tagging particles reconstructed as long and upstream tracks and as long tracks, only. Within the statistical uncertainty, the tagging power is the same within the two samples of track types in data and MC. In the following, particles reconstructed as upstream tracks are removed from the sample of tagging candidates.

sample	$\epsilon_{tag}(\%)$	$\omega(\%)$	$\epsilon_{eff}(\%)$
data (long + upstream)	17.57 ± 0.08	39.33 ± 0.25	0.80 ± 0.04
data (long)	16.49 ± 0.08	39.07 ± 0.26	0.79 ± 0.04
MC (long + upstream)	15.91 ± 0.06	35.13 ± 0.18	1.41 ± 0.03
MC (long)	13.93 ± 0.05	33.93 ± 0.19	1.44 ± 0.03

Table 3.9.: Average tagging performance of the OS kaon tagger in data (upper part) and in MC (lower part). Vetoing the small fraction of particles reconstructed from upstream tracks has no significant impact on the tagging performance.

To evaluate the effect of the higher track multiplicities due to the UE on the tagging performance in MC, a correction method is applied to MC. This correction method adds UE particles per event to MC until the track multiplicity distribution found on data is reproduced. The UE particles are taken from different events in MC than the events they are added to. A mean fraction of particles to be added, f, is determined by the difference of the mean number of particles in data minus the mean number of particles in MC divided by the mean number of particles in MC

$$f = \frac{\mu_{data} - \mu_{MC}}{\mu_{MC}}. \tag{3.22}$$

Typical values for f are 0.36, 0.53, 0.67, 0.77 and 0.77 corresponding to the 5 categories, respectively. Note, that f increases for increasing number of PVs per event. The number of particles to be added is calculated event by event by multiplying the fraction f by the actual number of particles in the event

$$n_{add} = f \times n_{\text{particles}}. \tag{3.23}$$

A Gaussian random number is drawn with $\mu = n_{add}$ and $\sigma = n_{add}/3$. The result, rounded to integer values, is used as the number of particles to be added per event. A histogram bookkeeping method guarantees, that all bins of track multiplicities are represented. This event by event correction is applied to MC, on the 5 subsamples corresponding to the number of PVs per events, separately. After correction the data track multiplicity distribution is reproduced by MC, see subfigure (b) in figure 3.7. Here, the total sample, containing all number of PV subsamples, is shown. The corrected track multiplicity distributions per PV subsample are shown in appendix A.2.

The UE particles are taken from events in MC that have the same number of PVs as the event they are added to. Additionally, the added particles have to fulfill kinematical requirements, represented by the transverse momentum p_T and the pseudo rapidity η. The kinematical requirements are determined two dimensionally in p_T and η, dividing the data distribution by the uncorrected MC distribution in these quantities.However, the uncorrected MC distribution is especially in the subsample with 1 PV per event not a subdistribution of the data $p_T - \eta$ distribution. Hence, the kinematical correction will not succeed perfectly, shown in figure 3.8. Here, all tagging candidates per event before and after correction are shown.

Figure 3.9 shows the same distributions before and after correction but filling only the highest p_T tagging candidate per event into the histograms. The corrected histograms show further

(a) before correction

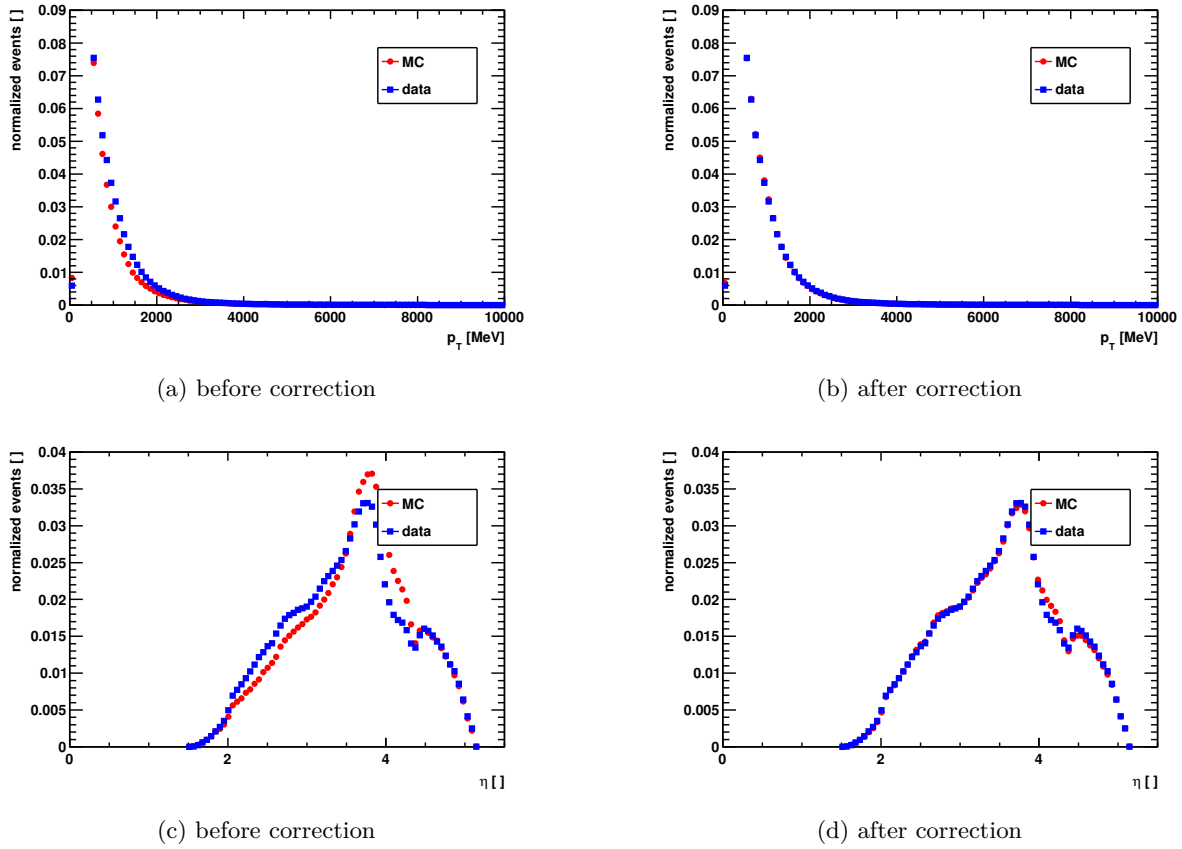(b) after correction

(c) before correction

(d) after correction

Figure 3.8.: Uncorrected and corrected distributions of the transverse momentum (a) and (b) and of the pseudo rapidity (c) and (d) of all tagging candidates per event.

non-perfect matching of the MC distributions relative to the data distributions, because the correction method corrects for average not single particle properties per event. The remaining differences are caused by momentum correlations of the particles within the event in data.

The effect of the track multiplicity correction on the tagging performance in MC is given in table 3.10. The tagging power is lowered by 23% percent, relatively, due to the increased mistag probability by 10%, relatively. This is a big effect. Therefore, all further MC data differences are studies using the track multiplicity corrected MC sample.

3.5.5. Effect of the Impact Parameter Resolution

The OS kaon tagger is tuned to select kaons from opposite b hadron decays. The b hadron decay vertex is significantly displaced from the PV. Thus, the most powerful variable to achieve that is the impact parameter (IP) of the particle's track with respect to the PV. If the IP resolution is better in MC compared to data, the tagging efficiency is higher in MC. However, a worse IP

(a) before correction

(b) after correction
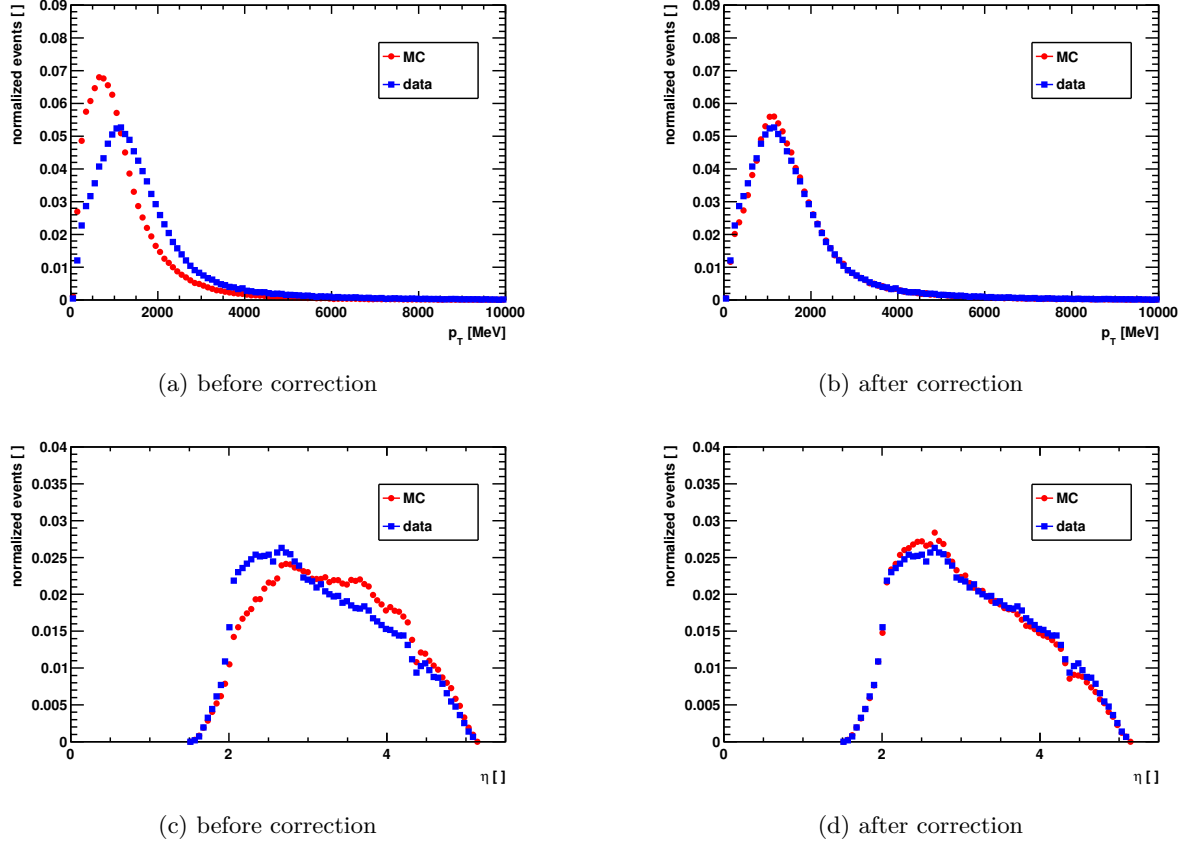
(c) before correction

(d) after correction

Figure 3.9.: Uncorrected and corrected distributions of the transverse momentum (a) and (b) and of the pseudo rapidity (c) and (d) of the highest p_T tagging candidate per event.

change	$\Delta\varepsilon_{\mathrm{tag}}(\%)$	$\Delta\omega(\%)$	$\Delta\varepsilon_{\mathrm{eff}}(\%)$
absolute	$+3.55 \pm 0.01$	$+3.47 \pm 0.01$	-0.33 ± 0.04
relative	$+25.50 \pm 0.07$	$+10.21 \pm 0.19$	-23.08 ± 0.07

Table 3.10.: Absolute and relative change in the tagging performance estimated on MC for a data driven correction applied to the number of tagging candidates per event.

resolution in data mimics UE particles to not originate from the PV, which increases the mistag probability in data. Therefore, the IP distributions in data and MC are carefully studied.

The IP resolution depends on the kinematical properties of the track. Therefore, figure 3.10 shows the widths of the x and y components of the IP of tagging candidates in bins of the inverse of the transverse momentum. Here, an effective IP resolution is studied, because the sample contains UE event particles and particles from b hadron decays. The true IP resolution is determined from a sample of prompt particles, that have a true IP of zero because all particles originate from the PV, but due to resolution effects, the particles are reconstructed with a non-zero IP.
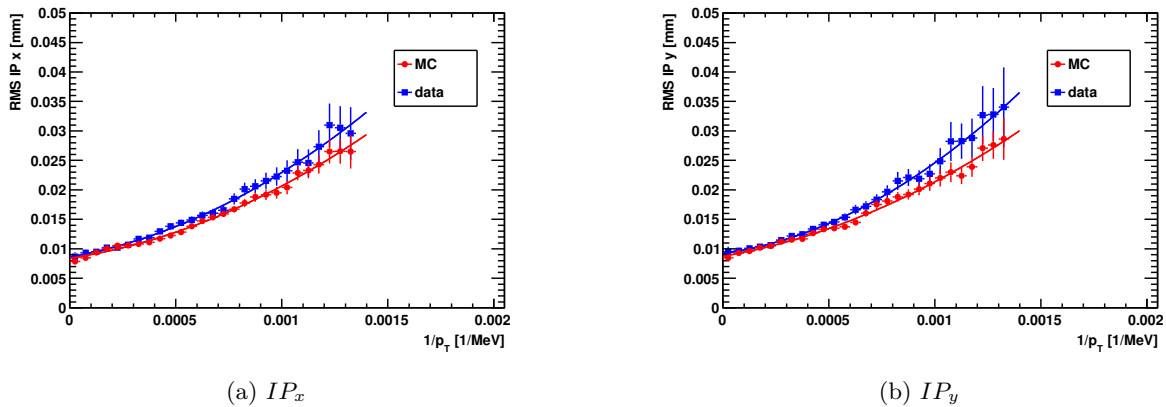

(a) IP_x (b) IP_y

Figure 3.10.: Impact parameter resolution distributions (a) in x direction and (b) in y direction.

Low p_T particles, which are most likely particles not coming from opposite b hadron decays, have a significantly worse IP resolution in data compared to MC. The MC sample is corrected to match the PV and track multiplicities per event in data. Due to the fact, shown in the previous section, that to events with more PVs more low p_T UE particles have to be added, the IP resolution is studied and corrected in subsamples corresponding to the number of PVs per event, separately. The corresponding distributions per PV subsample are shown n appendix A.3.

A correction method applied to MC to correct for the worse IP resolution seen in data is developed. A two dimensional distribution of the x and y components of the IP in bins of $1/p_T$ is fitted by Gaussian distributions per $1/p_T$ bin. Only the core part of the $IP_{x/y}$ distributions is fitted. The widths of the fitted Gaussian distributions is shown in figure 3.10. Then, the quadratic difference between the x and y components of the IP resolution in data and MC is determined

$$\Delta\sigma_{x,y}(p_T) = \sqrt{IP^2_{x,y(data)}(p_T) - IP^2_{x,y(MC)}(p_T)}. \tag{3.24}$$

A Breit-Wigner shaped random number is drawn with a mean value of 0 and a width of the difference between data and MC in the IP resolution $\Delta\sigma_{x,y}(p_T)$. This random number is added in quadrature to the x and y components of the IP resolution

$$IP_{\mathrm{smeared}}(p_T) = \sqrt{(IP_x(p_T) + \Delta\sigma_{x,random}(p_T))^2 + (IP_y(p_T) + \Delta\sigma_{y,random}(p_T))^2}. \tag{3.25}$$

The result of this IP resolution correction is shown in figure 3.11. The corresponding corrected distributions per PV subsample are shown in appendix A.3. Within the uncertainty of the single

$IP_{x/y}$ widths fits, the IP resolution in MC now matches the one in data. This correction method increases the IP resolution in MC for both particle types: the kaons coming from opposite b hadron decays and the UE event particles. On average the correction should be correct.
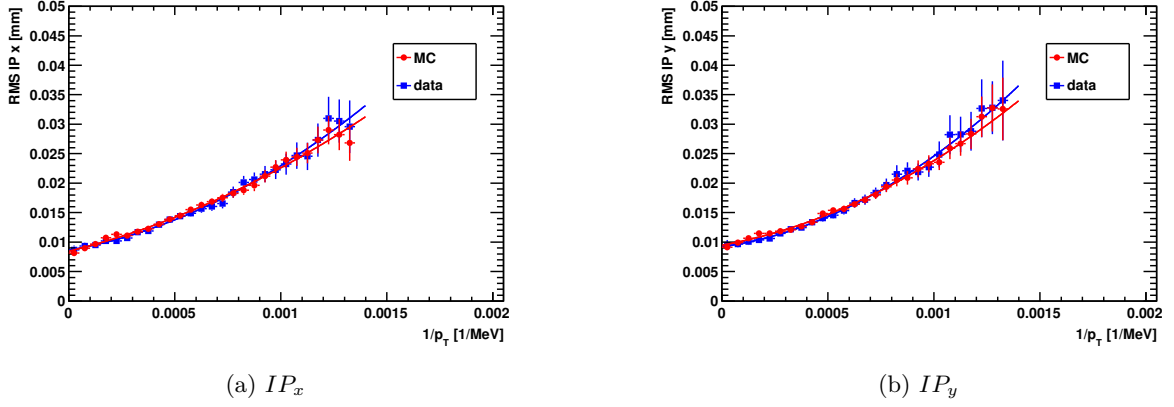

(a) IP_x (b) IP_y

Figure 3.11.: Smeared impact parameter resolution distributions (a) in x direction and (b) in y direction.

Figure 3.12 shows the uncorrected and corrected distributions of IP_x and IP_y in data and MC. After the IP resolution correction the distributions in MC match the data distributions.

The OS kaon tagger requires a minimum value of 4.3 for the IP significance and a maximum value of 1.45 mm for the IP. Figure 3.13 shows the uncorrected and corrected distributions of the IP and IP significance for data and MC. After the IP resolution correction the MC distributions match the ones in data better, however other reconstruction effects play a role here. Especially visible in the IP significance distribution, there are steps in the data distribution, which is likely due to PV multiplicity related reconstruction effects on the IP fit uncertainty. Further investigations on such reconstruction effects are beyond the scope of this study.

The effects of the IP resolution correction on the tagging performance in MC are given in table 3.11. The relative increase in the tagging efficiency of almost 38% is remarkable. Hence, the broadening of the IP resolution in MC has increased the number of particles that fulfill the IP significance lower cut of the OS kaon tagger, significantly. The mistag probability is also increased by 9%, which means that the majority of the particles in MC that now fulfill the IP significance requirement are not opposite b hadron decay products. In total this lowers the tagging power by relatively 35% in MC. This is a huge effect. Therefore, all further MC data differences are studies using the track multiplicity and IP resolution corrected MC sample.

3.5.6. Effect of the Track χ^2/ndf

The quality of the track reconstruction is quantified in the reduced χ^2 of the track fit, χ^2/ndf. The OS kaon tagger applies an upper cut on the track χ^2/ndf of 2.45 to suppress particles with badly reconstructed tracks. However, the χ^2/ndf distribution in data has a large tail to higher values compared to the distribution in MC, see figure 3.14. All OS kaon tagger selection requirements are

(a) IP_x uncorrected

(b) IP_y uncorrected

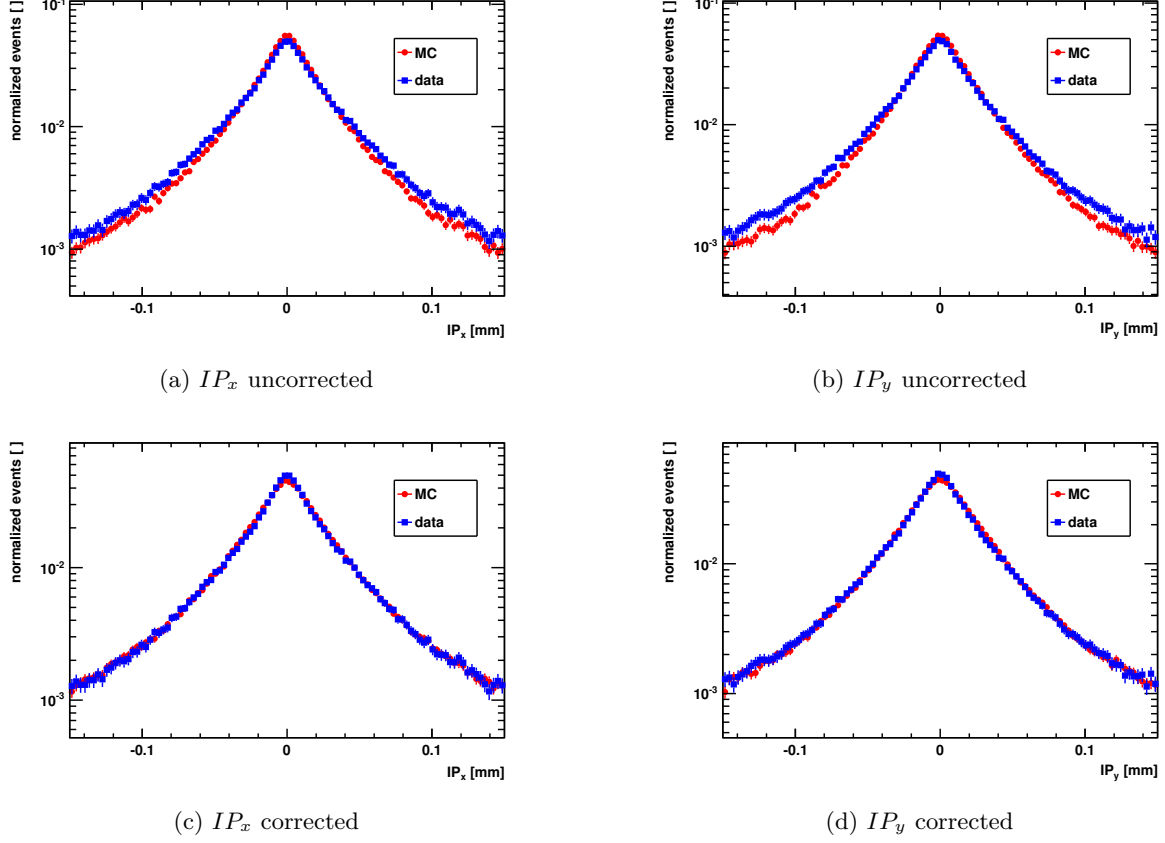(c) IP_x corrected

(d) IP_y corrected

Figure 3.12.: Impact parameter resolution distributions in x and y direction before the IP smearing correction (a) and (b) and after the correction (c) and (d).

change	$\Delta\varepsilon_{\text{tag}}(\%)$	$\Delta\omega(\%)$	$\Delta\varepsilon_{\text{eff}}(\%)$
absolute	$+6.57 \pm 0.00$	$+3.30 \pm 0.01$	-0.28 ± 0.05
relative	$+37.61 \pm 0.06$	$+8.81 \pm 0.20$	-34.55 ± 0.08

Table 3.11.: Absolute and relative change in the tagging performance estimated on PV, track multiplicity and IP resolution corrected MC compared to the MC performance without the IP correction.

(a) *IP* uncorrected

(b) *IP* corrected
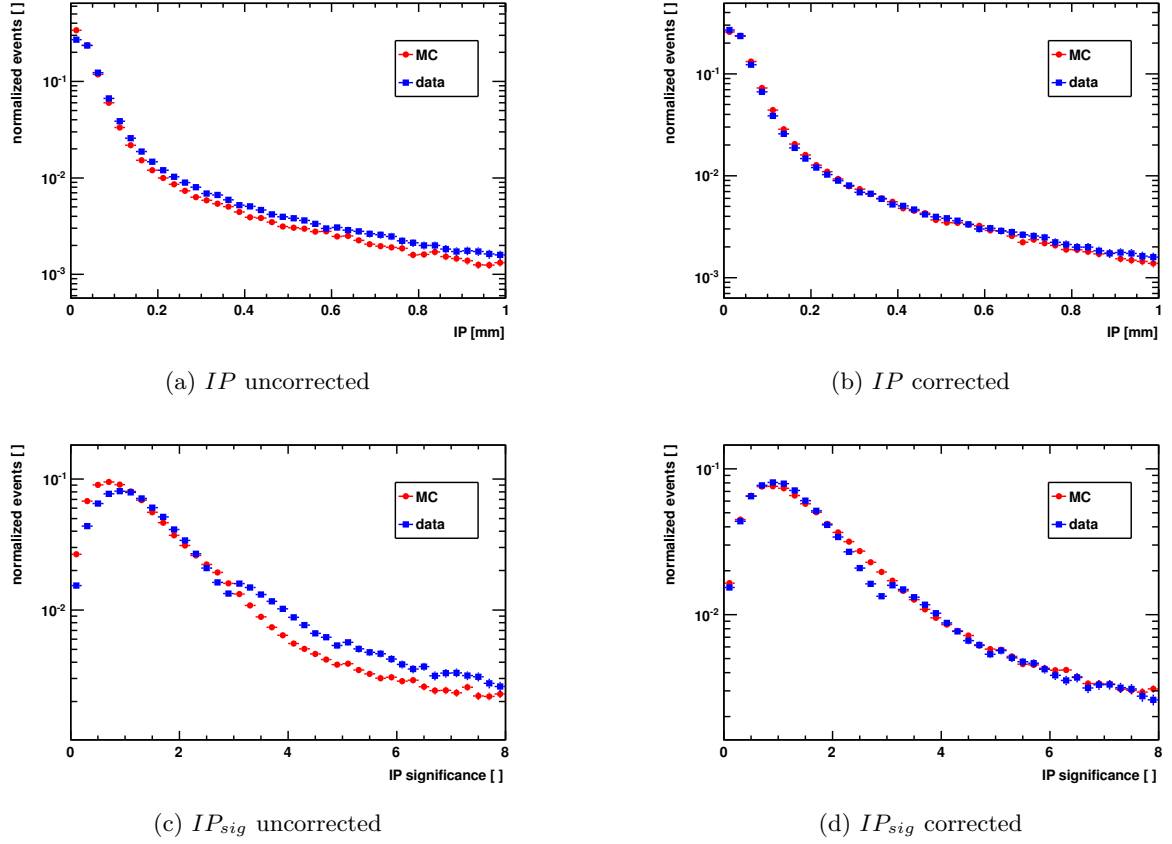
(c) IP_{sig} uncorrected

(d) IP_{sig} corrected

Figure 3.13.: Impact parameter and IP significance distributions uncorrected: (a) and (c), corrected: (b) and (d).

applied, except the cut on the track χ^2/ndf. In MC the PV, track multiplicity and IP resolution corrections are applied.

In order to achieve a similar tagging performance in data and MC, the track χ^2/ndf cut efficiency in data has to be translated to MC by determining a different cut value. High momentum tracks have typically a better track fit quality. Hence, the sample of tagging candidates is binned in 5 momentum bins of increasing momentum, see figure 3.15. The χ^2/ndf distribution of wrong (W) tagged events is subtracted from the one of right (R) tagged events. In this way, the effect of the χ^2/ndf cut on the tagging performance becomes visible. Particles with low quality track fits carry no information relevant for flavour tagging.

In figure 3.15 the shape of the MC distributions does not change significantly with the momentum, while, in data the width of the distributions decreases with increasing momentum, significantly. Hence, the χ^2/ndf cut acts differently on MC compared to data. Therefore, five different effective cut values are determined on MC, to mimic the same χ^2/ndf cut efficiency as in data.

Table 3.12 shows the cut efficiencies in data of a χ^2/ndf < 2.45 cut and the corresponding cut values, determined per momentum bin, for MC to yield an equivalent effective cut efficiency.
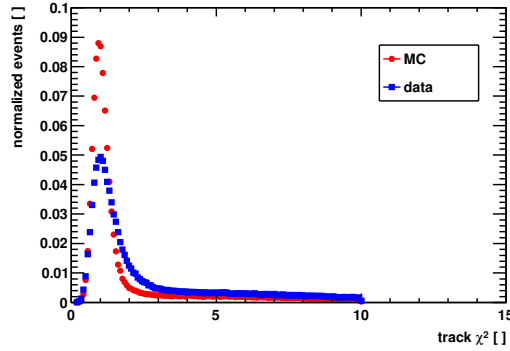
Figure 3.14.: Track χ^2/ndf distribution of tagging particles for data and MC. The full OS kaon tagger selection is applied, except the cut on this quantity. In MC the PV, track multiplicity and IP resolution corrections are applied.

The determination of the effective cut values on MC is limited by the bin widths of the χ^2/ndf distribution histogram. Hence, a range for the effective MC cut is given. However, within this range, the tagging performance on MC does not change significantly.

| momentum bin [GeV]	data efficiency [%]	MC cut
5 − 10	0.81 ± 0.01	1.35
10 − 14	0.81 ± 0.01	1.35 − 1.40
14 − 22	0.87 ± 0.01	1.40 − 1.45
22 − 40	0.94 ± 0.01	1.50 − 1.65
40 − 100	1.02 ± 0.01	1.55

Table 3.12.: Evaluation of the track χ^2/ndf cut efficiency in data (< 2.45) and the corresponding cut in MC for different momentum bins. All tagging selection cuts are applied except the cut on this quantity. The highest p_T particle per event is taken. Wrong tagged events are subtracted from right tagged events.

The change in the MC tagging performance using the effective track χ^2/ndf cuts per momentum bin are given in table 3.13. The tagging efficiency is decreased by 16%, as expected, but also the mistag probability is decreased. The relative change in the tagging power is small.
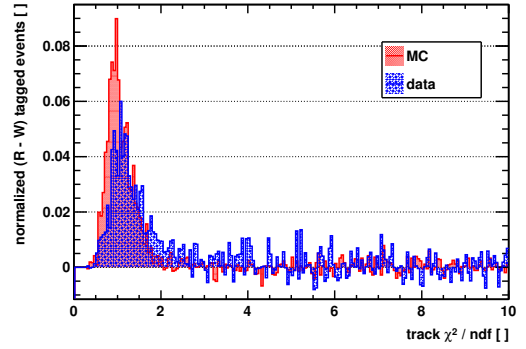
3.5.7. Summary

This section has evaluated all important differences between data and MC that caused the different tagging performance. Table 3.14 shows the tagging performance in data and in MC, split up in the different correction steps applied to MC. Each correction step is applied on top of the previous
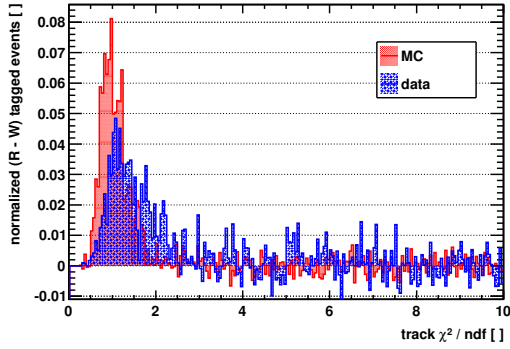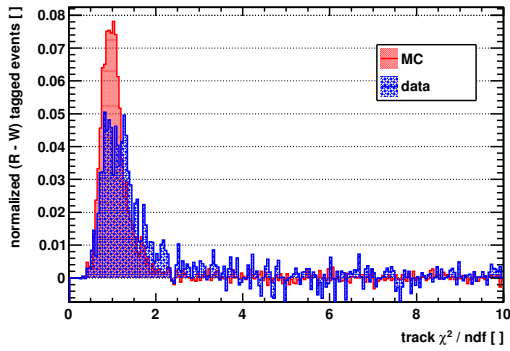
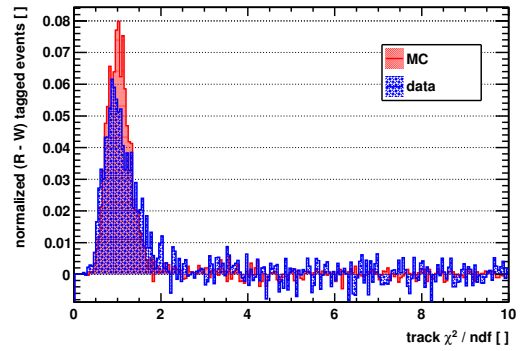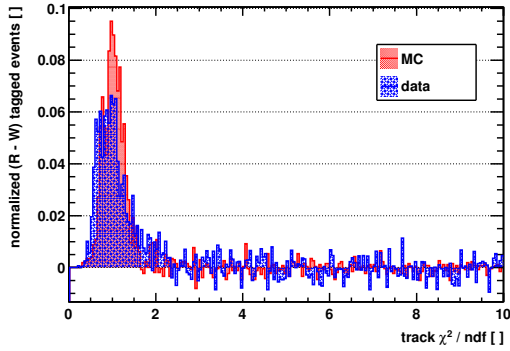(a) $5\,\text{GeV} - 10\,\text{GeV}$: R - W tagged events

(b) $10\,\text{GeV} - 14\,\text{GeV}$: R - W tagged events

(c) $14\,\text{GeV} - 22\,\text{GeV}$: R - W tagged events

(d) $22\,\text{GeV} - 40\,\text{GeV}$: R - W tagged events

(e) $40\,\text{GeV} - 100\,\text{GeV}$: R - W tagged events

Figure 3.15.: Track χ^2/ndf distribution of right (R) minus wrong (W) tagging candidates with momenta in (a) $[5, 10]\,\text{GeV}$, (b) $[10, 14]\,\text{GeV}$, (c) $[14, 22]\,\text{GeV}$, (d) $[22, 40]\,\text{GeV}$ and (e) $[40, 100]\,\text{GeV}$.

change	$\Delta\varepsilon_{\text{tag}}(\%)$	$\Delta\omega(\%)$	$\Delta\varepsilon_{\text{eff}}(\%)$
absolute	-3.94 ± 0.00	-0.77 ± 0.01	-0.01 ± 0.05
relative	-16.39 ± 0.07	-1.89 ± 0.17	-1.22 ± 0.08

Table 3.13.: Absolute and relative change in the tagging performance estimated on PV, track multiplicity, IP resolution and track χ^2/ndf cut efficiency corrected MC compared to the MC performance without the last correction.

corrections. Within the statistical uncertainty, the tagging power in MC matches the one in data after the IP resolution correction or after the effective χ^2/ndf cut determination.

sample	$\epsilon_{tag}(\%)$	$\omega(\%)$	$\epsilon_{eff}(\%)$
data	16.49 ± 0.08	39.07 ± 0.26	0.79 ± 0.04
MC initial	13.93 ± 0.05	33.93 ± 0.19	1.44 ± 0.03
MC PV	13.92 ± 0.06	33.99 ± 0.21	1.43 ± 0.04
MC PV + ntracks	17.47 ± 0.06	37.46 ± 0.19	1.10 ± 0.03
MC PV + ntracks + IP	24.04 ± 0.07	40.76 ± 0.16	0.82 ± 0.03
MC PV + ntracks + IP + χ^2	20.10 ± 0.07	39.99 ± 0.18	0.81 ± 0.03

Table 3.14.: Tagging performance of the OS kaon tagger showing the effect of each significant correction step using average mistag probabilities. The PID cut efficiency effect on the tagging performance is negligible.

The corrections of PVs per event and adapted track χ^2/ndf effective cut efficiencies do not change the tagging performance significantly. Whereas the impact of the track multiplicity and IP resolution corrections on the tagging performance is large. However, the tagging efficiency is now over-corrected. The mistag probability is now also slightly too high. The reason is probably, that all correction methods were determined on average. Possible correlations were not taken into account. Nevertheless, the most important effects, that cause different tagging performance in MC compared to data, are understood.

3.6. Novel Artificial Neural Network Tagging Algorithm

The main task of a tagging algorithm is to find the correct tagging particle in the B event. In case of the Opposite Side (OS) kaon tagger, this task is not trivial. In proton-proton collisions it is very difficult to distinguish a hadronic tagging particle from the large amount of soft QCD background, which is produced in the primary interaction as underlying event (UE). Up to now, a cut-based selection was used to find the tagging particle.

A novel approach is presented here. The task is reformulated as a multi-dimensional classification problem, i.e. not a separate optimization of one variable after the other. Thus, it can be efficiently solved using machine learning techniques.

Machine learning employs three techniques: supervised, unsupervised and reinforcement learning. In the given tagging classification task, the answer is binary: yes (it is a tagging particle) or no (it is not). Using MC simulation, this answer is given for every particle in the event before the classification is done. Hence, supervised learning can be used.

Unsupervised learning is used for example in clustering problems. An algorithm learns unsupervised from a set of inputs how to cluster it into subgroups. An application of such algorithms is a search engine which bundles information about different topics or even gathers information from object recognition in images[3].

Reinforcement learning is inspired by behaviourist psychology. An algorithm learns to interact with a dynamic environment. The focus is laid on online performance, applied for example in automated driving vehicles. Automated driving utilizes deep[4] Artificial Neural Networks (ANN) for image recognition, where several deep ANNs are connected, each processing a piece of the environment, but combined they deliver a high-abstraction recognition, such as separating between a truck and an SUV, that accelerates or breaks[5].

In general, machine learning techniques are used to solve complex problems, where a conventionally programmed solution would be too complicated or too time consuming in terms of programmer's working hours. The given tagging classification problem is an ideal application for an ANN, which learns supervised to distinguish tagging particles from particles from the UE. The ANN response is interpreted as a probability, how likely the particle fulfills the requirements to be a tagging particle, i.e. the particle is correlated to the production flavour of the signal B meson. This probability is evaluated for every particle in the B event. Hence, all particles can be weighted according to their assigned probability.

Another novelty is, that the highest-weighted particles are combined to derive the tagging decision. In the old implementation, only the charge of the highest p_T particle was used as tagging decision. Now, a second ANN is used to derive a tagging decision from multiple high-ranked particles. The advantage of this implementation is, that it covers cases, where particles of similar quality are present.

---

[3]See for example `http://googleresearch.blogspot.de/2014/09/building-deeper-understanding-of-images.html` and `http://googleresearch.blogspot.de/2014/11/a-picture-is-worth-thousand-coherent.html`

[4]The ANN has multiple hidden layers between the input and output layers.

[5]See for example Nvidia: deep neural network for auto-piloted driving `https://www.youtube.com/watch?v=zsVsUvx8ieo`

Hence, the working principle of the novel ANN-based OS kaon tagger is:

1. Distinguish between tagging particles, which are kaons from opposite b hadron decays, and background particles using an ANN. The ANN response is evaluated for each particle and assigned to it as weight.

2. Combine high-ranked particles from step 1 to derive the tag decision using a second ANN. The second ANN response is evaluated for each B event and assigned to it as mistag probability.

In the following the strategy of the ANN-based OS kaon tagger is explained in detail. First an introduction to the used ANN is given, then the development process of the new tagging algorithm is presented. The improvement in tagging performance achieved by the new tagger is evaluated. The new tagger is embedded into the combination with the LHCb flavour tagging algorithms, tagger correlations and systematic uncertainties on the calibration of the new tagger are determined. At the end of the section a short summary highlights the improvements.

3.6.1. The Artificial Neural Network (ANN)

Mathematically, an ANN is given by a nonlinear function from a set of input variables x_i $(i = 1, .., n)$ to a set of output variables y_k $(k \in \mathbb{N})$ controlled by a vector of adjustable parameters $\mathbf{w}$, the network weights [81]. The ANN is structured in layers. The basic structure is given by three layers, an input layer, a hidden layer and an output layer. Each layer consists of neurons, where a neuron is a linear or non-linear function, that maps the neuron input onto the neuron output.

The input variables x_i are connected to the neurons of the input layer $f^{(1)}(x_i)$ and linearly mapped onto the outputs of the first layer $y_i^{(1)}$

$$y_i^{(1)} = f^{(1)}(x_i) \quad \text{with} \quad f^{(1)}(x_i) = \alpha x_i + \beta. \tag{3.26}$$

From the outputs of the first layer, m linear combinations are built to form activations a_j $(j = 1, ..., m)$, that serve as inputs to the second layer

$$a_j = \sum_{i=1}^{n} w_{ji}^{(1)} y_i^{(1)} + w_{j0}^{(1)}. \tag{3.27}$$

$w_{ji}^{(1)}$ are weights, which are adjusted during the network training process. $w_{j0}^{(1)}$ are bias weights, used to stabilize the training process. Typically, $m > n$ is set in order not to loose information. In the second layer, the activations a_j are transformed using a differentiable, non-linear activation function $f^{(2)}(a_j)$ to yield the outputs $y_j^{(2)}$ of this layer

$$y_j^{(2)} = f^{(2)}(a_j) \quad \text{with} \quad f^{(2)}(a_j) = \frac{1}{1 + e^{-a_j}}. \tag{3.28}$$

The given flavour tagging task is a classification problem. As a direct consequence of the Bayesian probability theorem applied to a classification problem the neuron activation function $f^{(2)}(a_j)$ is given by a sigmoidal function, either a logistic sigmoid function, ranging from 0 to 1, as used

here, or a $\tanh(h)$ function ranging from -1 to 1. To obtain the inputs of the third layer, k linear combinations are built of the outputs of the second layer to form activations a_k of the third layer

$$a_k = \sum_{j=1}^{m} w_{kj}^{(2)} y_j^{(2)} + w_{k0}^{(2)}. \tag{3.29}$$

In the classification problem $k = 1$ is set[6]. Analog to the hidden layer, the activations are transformed using $f^{(2)}(a_k)$ to yield the output $y_1^{(3)}$ of the third layer, which corresponds to the ANN response y_{ANN}

$$y_{ANN} = y_1^{(3)} = f^{(2)}(a_k). \tag{3.30}$$

Hence, the combined formula which describes the three-layered ANN is

$$y_1^{(3)}(\mathbf{x}, \mathbf{w}) = f^{(2)} \left(\sum_{j=1}^{m} w_{kj}^{(2)} f^{(2)} \left(\sum_{i=1}^{n} w_{ji}^{(1)} f^{(1)}(x_i) + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right). \tag{3.31}$$

Figure 3.16 shows a visualization of a three-layered ANN. This network topology is also referred to as two-layered ANN, counting the layers that carry weights, only. The evaluation process can be interpreted as a forward propagation of information through the network. Hence, this structure is called feed-forward in contrast to recurrent neural networks, where connections form directed cycles.

Each of the two processing stages of the three-layered ANN can be identified with the perceptron model. Hence, the ANN is also called multilayer perceptron. However, the ANN uses continuous sigmoidal nonlinear functions in the hidden units, while a perceptron uses step-function nonlinearities.

ANNs are said to be universal approximators [81]. Given sufficient training statistics, the ANN can distinguish any data that is not linearly separable [82].

The determination of the event weights happens during the learning process. An Error function E is minimized by adjusting the vector of weights $\mathbf{w}$. E is the sum of N Error functions E_a, where $a = 1, ..., N$ and N being the number of training events. The Error function of the ath training event is defined as

$$E_a(\mathbf{x}_a|\mathbf{w}) = \frac{1}{2} (y_{ANN,a}(\mathbf{x}_a, \mathbf{w}) - \hat{y}_a)^2, \tag{3.32}$$

where $y_{ANN,a}(\mathbf{x}_a, \mathbf{w})$ is the network response computed from the vector of input variables $\mathbf{x}_a = (x_1, ..., x_n)_a$ and the vector of weights $\mathbf{w}$. $\hat{y}_a$ is the desired output, which is either 1 for signal or 0 for background. Then, E is given as

$$E(\mathbf{x}_1, ..., \mathbf{x}_N|\mathbf{w}) = \sum_{a=1}^{N} E_a(\mathbf{x}_a|\mathbf{w}) = \sum_{a=1}^{N} \frac{1}{2} (y_{ANN,a}(\mathbf{x}_a, \mathbf{w}) - \hat{y}_a)^2. \tag{3.33}$$

Depending on the chosen learning mechanism the optimal set of weights is found. Back propagation optimizes the weights iteratively via the first derivative of the Error function. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [83], [84], [85] and [86] uses the second derivatives of the Error function. The BFGS method is used in this thesis.

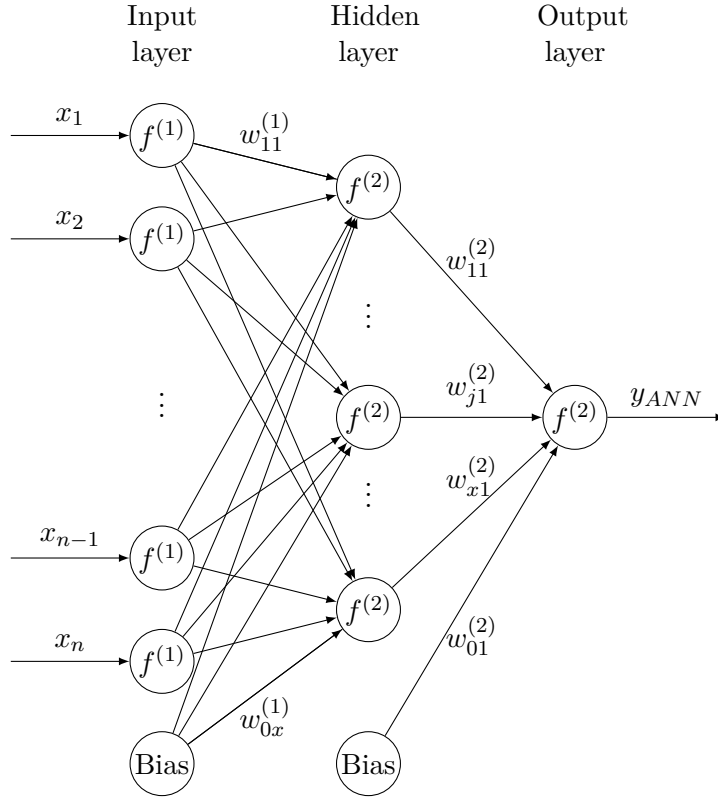[6]In regression problems $k \in \mathbb{N}$ and $k > 1$ is set.

Figure 3.16.: Visualization of an ANN with one hidden layer. The neurons are connected in feed-forward structure.

In the given case of binary classification with logistic sigmoid activation function, where class $\mathcal{C}_1$ corresponds to $\hat{y} = 1$ and $\mathcal{C}_2$ to $\hat{y} = 0$, the ANN response $y_{ANN}(\mathbf{x}, \mathbf{w})$ can be interpreted as conditional probability $p(\mathcal{C}_1|\mathbf{x})$. Accordingly, $p(\mathcal{C}_2|\mathbf{x})$ is given by $1 - y_{ANN}(\mathbf{x}, \mathbf{w})$. The conditional probability of $\hat{y}$ to occur is given by a Bernoulli distribution

$$p(\hat{y}|\mathbf{x}, \mathbf{w}) = y_{ANN}(\mathbf{x}, \mathbf{w})^{\hat{y}} \left(1 - y_{ANN}(\mathbf{x}, \mathbf{w})\right)^{1-\hat{y}} \tag{3.34}$$

If the training set consists of independent observations, as it is here the case, the Error function, given by the negative log likelihood, is a cross-entropy Error function [81]

$$E(\mathbf{x}_1, ..., \mathbf{x}_N|\mathbf{w}) = -\sum_{a=1}^{N} \left[\hat{y}_a \ln(y_{ANN,a}(\mathbf{x}_a, \mathbf{w}) + (1 - \hat{y}_a) \ln(1 - y_{ANN,a}(\mathbf{x}_a, \mathbf{w})\right]. \tag{3.35}$$

The ANN implementation provided by the TMVA package is used here [87].

After the learning process an overtraining test of the network is done. The training sample is split in two halves beforehand, one half for training and one half for testing. The minimization of

the Error function is recalculated on the testing events and compared to the one for training. A possible sign for overtraining is, when the estimators differ significantly. If the performance on the training sample is significantly better, the network is sensitive to statistical fluctuations in the training sample. Normally, the training estimator distribution becomes constant after a sufficient amount of training cycles and performs slightly better than the testing estimator.

In order to determine the best ANN for a given problem, the importance I_i of each input variable x_i is determined using the weights between the input layer and the first hidden layer $w_{ij}^{(1)}$

$$I_i = \bar{x}_i^2 \sum_{j=1}^{n_h} \left(w_{ij}^{(1)} \right)^2 , \qquad (3.36)$$

where n_h is the number of neurons in the first hidden layer and $\bar{x}_i^2$ is the normalized sample mean of the ith input variable, i.e. prior to the determination of the sample mean the range of the variable is normalized to [-1,1] using the transformation

$$x_{i,norm} = \frac{2 \cdot (x_i - x_{i,min})}{x_{i,max} - x_{i,min}} - 1. \qquad (3.37)$$

The importance I_i of the ith input variable gives a measure of the connectivity between this variable and the ANN response. An input variable of good separation power will clearly influence the ANN response positively.

In the following, the development of the ANN-based OS kaon tagging algorithm is presented. First, the configuration and training of the first ANN (ANN1) is shown, which separates OS kaon tagging candidates from background. The ANN1 response is assigned as weight to each candidate. Then, the configuration and training of the second ANN (ANN2) is shown, which forms the tagging decision by combining the high-ranked candidates from ANN1.

The global figure of merit is the tagging power of the tagging algorithm. However, the optimization process consists of multiple steps depending on each other. The ANN1 is used as input for the ANN2. Hence, the tagging power can only be determined, once an ANN1 has been chosen and the ANN2 has been trained using the ANN1 response. The local figure of merit, used to select the ANN1 from several networks, is separation power and training stability.

3.6.2. First ANN

Training Sample

OS taggers can be developed independently of the signal B decay. The underlying assumption is, that the optimal working point of an OS tagger is independent of the signal B decay. Therefore, the flavour specific $B^+ \to J/\psi K^+$ decay is commonly used. The advantages are, that due to the clean trigger signature of the subsequent $J/\psi \to \mu^+ \mu^-$ decay, a high statistics sample of these decays is available in data. Additionally, charged B mesons do not oscillate. If the tagging decision differs from the charge of the final state kaon the B^+ meson is wrongly tagged. Thus, the predicted mistag probability of the tagger can be calibrated to the measured mistag probability in real data by counting the number of correctly and wrongly tagged events.

However, after the ANN-based OS kaon tagger was developed and optimized on $B^+ \to J/\psi K^+$ data, it turned out that the initial assumption was wrong. The novel tagging algorithm had been optimized for high tagging efficiencies. However, the fragmentation processes of B^+, B^0 and B_s^0 mesons are different. In the B_s^0 fragmentation, kaons are produced, that are correlated to the signal b production flavour, but carry the opposite charge compared to kaons from opposite side $b \to c \to s$ transitions. This type of background is not present in B^+ decays. Hence, it was not considered during the optimization of the ANN-based OS kaon tagger. Thus, the tagger had to be re-developed to find the new optimal working point applicable to oscillating B^0 and B_s^0 mesons.

A retraining of the first ANN (ANN1) was done using $B_s^0 \to J/\psi \phi$ decays instead of $B^+ \to J/\psi K^+$ decays. The $B_s^0 \to J/\psi \phi$ signal selection is taken from the measurement of the *CP*-violating phase ϕ_s at LHCb using $B_s^0 \to J/\psi \phi$ decays [88]. The ANN1 is trained on MC simulation to access the production information of the tagging candidates, while ANN2 is trained on $B^+ \to J/\psi K^+$ data. A cut on the response of ANN1 removes background particles including the fragmentation kaons relevant in B_s^0 decays. Thus, ANN2 can still be trained using B^+ decays. The full LHC Run I data set is used corresponding to $1\,\mathrm{fb}^{-1}$ of data taken in 2011 and $2\,\mathrm{fb}^{-1}$ of data taken in 2012.

Training Procedure

The training task of ANN1 is to distinguish between kaons from opposite b hadron decays (OS kaons) and background particles. In general, particles from b hadron decays are high energetic. Thus, OS kaons have a harder (transverse) momentum spectrum than soft QCD background from the underlying event (UE) in the primary proton-proton interaction. The b hadrons are relatively long-lived, hence, their decay vertices are displaced from the primary vertex (PV). Thus, OS kaons also have relatively large impact parameters (IP) with respect to the PV. The IP is measured as the perpendicular distance between the OS kaon track and the PV. For OS kaons also the IP significance is high, which is the IP divided by its uncertainty.

The above listed kinematic properties of OS kaons help to separate them from soft QCD background. An additional source of background are particles from mis-reconstructed tracks. The track fit quality serves as a useful variable for that task. The long extrapolation of tracks through the LHCb spectrometer provides an excellent momentum resolution. However, it also causes mis-matchings between the tracking detectors resulting in so-called ghost tracks. The ghost probability [39], a multivariate variable, is used to suppress ghost tracks.

Figure 3.17 shows the mentioned variables. Per variable two distributions are shown, OS kaons, referred to as signal, and background, which are all other particles in the event, mainly UE particles and particles from mis-reconstructed tracks. Two transformations are applied to the variables, before using them as inputs to ANN1. Variables with sharp peak structure, like momentum or IP distributions are transformed by a logarithmic function. To all variables, a linear decorrelation transformation is applied, provided by the TMVA framework[7].

The $b\bar{b}$ pair produced at the LHC is boosted. Hence, the flight directions of the opposite b hadron and the signal B meson are similar. This property can be used to select OS kaons. Thus, the difference in pseudo rapidity and azimuthal angle between the tagging candidate and the

---

[7]The vector of input variables $\mathbf{x}$ is multiplied with the inverse of the square-root matrix $C' \cdot \mathbf{x} = \mathbf{x}'$, which is defined as $C' = S\sqrt{D}S^T$, where D is the diagonal matrix of the symmetric covariance matrix S, with $D = S^T C S$.
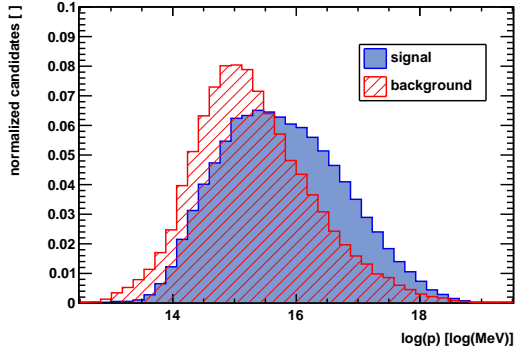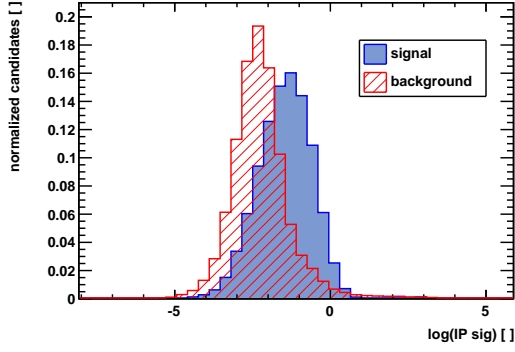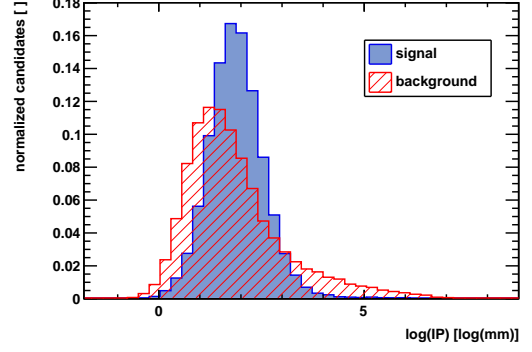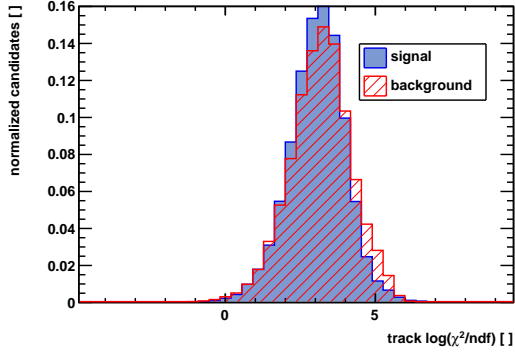
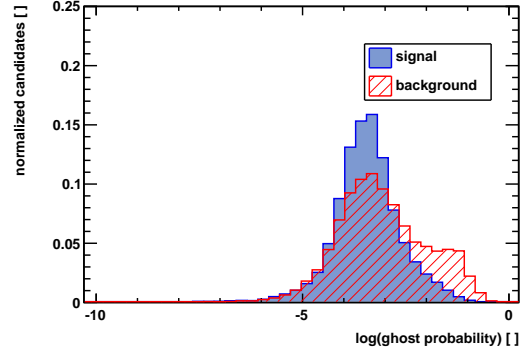(a) $\log(p)$
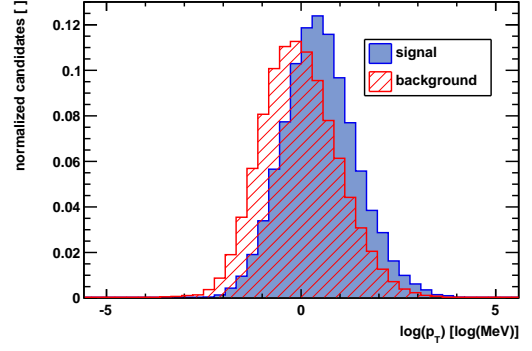
(b) $\log(p_T)$

(c) $\log(IP_{sig})$

(d) $\log(IP)$

(e) track $\log(\chi^2/ndf)$

(f) ghost probability

Figure 3.17.: ANN1 input variables separately drawn for signal, which are OS kaons (blue) and background particles, which are all other particles in the event (red): $\log(p)$, $\log(p_T)$, $\log(IP_{sig})$, $\log(IP)$, track $\log(\chi^2/ndf)$, $\log(\text{ghost probability})$. The linear decorrelation transformation is applied.

signal B is taken as inputs to ANN1, $\Delta\eta$ and $\Delta\phi$, respectively. The probability to tag signal B mesons with a large transverse momentum p_T is higher than the one for signal B mesons with small p_T due to the more distinct topology of the high energetic B decay. Therefore, the signal B p_T is used as input to ANN1. Generally, the probability to tag B mesons in events with less primary interactions and lower track multiplicities is higher. Thus, the number of primary vertices and the number of tagging candidates per event is also used.

These five quantities are shown in figure 3.18. The signal to background separation power seems to be much weaker than the one from the quantities in figure 3.17. However, they add information to the ANN, because they give valuable insights into the kinematics of the B event.

The tagging pre-selection, introduced in section 3.4, and the PID cuts as used in the cut-based OS kaon tagger, see section 3.5.1, are applied to all tagging candidates prior to the ANN1 training and application. The number of tagging candidates per event is determined after these cuts. The cuts are applied in order to formulate a distinct training task for the ANN1. The tagging pre-selection ensures, that particles from the signal B decay are rejected. Additionally, it removes particles from tracks of poor quality. The PID cuts suppress pions and protons. Hence, the ANN1 learns to distinguish between high energetic OS kaons and kaons from the UE.

During the development process of the ANN1 the above presented set of input parameters turned out to deliver the network with the best separation power and with the most robust learning rate. Other networks were trained with subsets of this set of input parameters. Those networks showed worse performance.

A ranking of the separation power of the input variables is shown in table 3.15. The separation power $\left\langle S^2 \right\rangle$ of the variable x between the signal $S(x)$ and background $B(x)$ distributions is defined as

$$\left\langle S^2 \right\rangle = \frac{1}{2} \int_{min}^{max} \frac{(S(x) - B(x))^2}{S(x) + B(x)} dx, \tag{3.38}$$

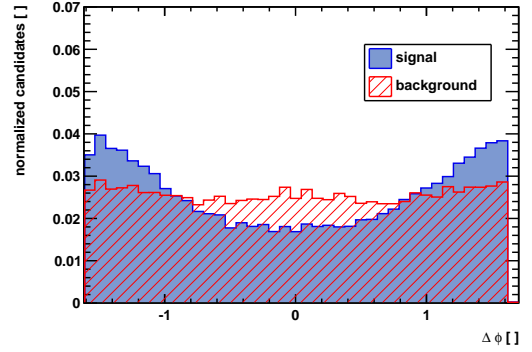where *min*, *max* are the lower and upper limits of the distribution of x.

Rank	Variable	Separation Power
1	$\log(|IP_{sig}|)(K)$	2.950×10^{-1}
2	$\log(|IP|)(K)$	2.100×10^{-1}
3	$\log(p_T)(K)$	1.897×10^{-1}
4	$\log(p)(K)$	1.251×10^{-1}
5	$\log(\text{ghost prob})(K)$	7.455×10^{-2}
6	$\log(\chi^2/\text{ndf})(K)$	4.230×10^{-2}
7	$log(\#\text{candidates})$	2.279×10^{-2}
8	$\Delta\eta(K, B)$	1.536×10^{-2}
9	$\Delta\phi(K, B)$	1.469×10^{-2}
10	$\#\text{vertices}$	3.390×10^{-3}
11	$\log(p_T)(B)$	9.498×10^{-4}

Table 3.15.: Separation Power of the input variables of ANN1.
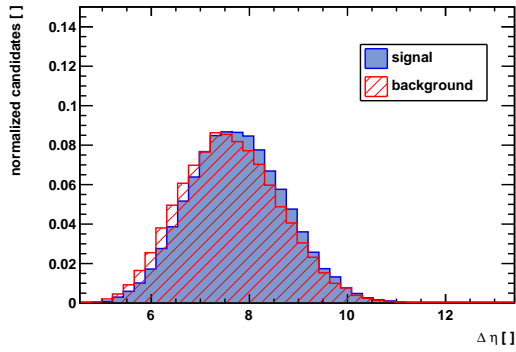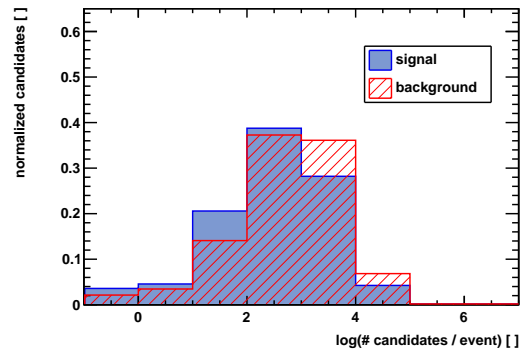
(a) $\Delta\eta$
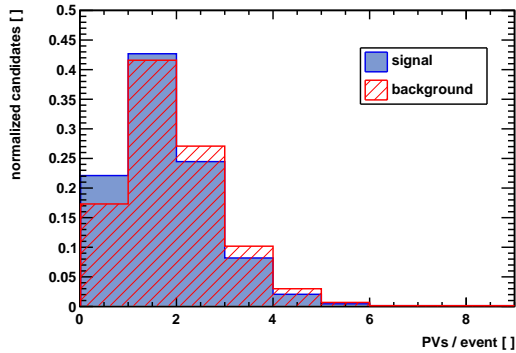
(b) $\Delta\phi$
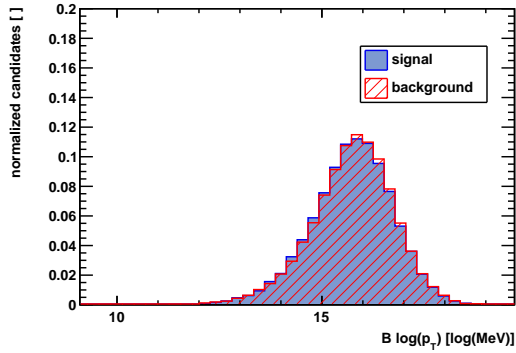
(c) $B \log(p_T)$

(d) $\log(\#candidates/event)$

(e) $\#PVs/event$

Figure 3.18.: ANN1 input variables separately drawn for signal, which are OS kaons (blue) and background particles, which are all other particles in the event (red): $\Delta\eta$, $\Delta\phi$, signal $B \log(p_T)$, $\log(\#candidates/event)$, $\#PVs/event$. The linear decorrelation transformation is applied.

The Bayesian ANN [89] implemented in the TMVA framework [87] is used. A normalization and decorrelation transformation provided by the TMVA framework is applied to the input variables. The estimator type cross-entropy is chosen, because the training set consists of independent observations. The neuron activation function is chosen to be a sigmoid logistic function, which results in an ANN response between 0 and 1. For training 5.0×10^4 signal and 5.0×10^4 background tagging candidates are used. The same amount of tagging candidates is used to test the estimator, allowing to compare the performance of the ANN on two independent but comparable samples. Several networks were trained with different numbers of hidden layers and neurons per layer. The network which showed the most reliable training success and gave the best performance was picked.

The topology of the final network consists of one hidden layer with $n+8$ neurons, where n is the number of input variables. The number of additional neurons in the hidden layer was set in different training scenarios to any number between 1 and 20. Eight turned out to yield the best training behaviour. The network topology of ANN1 is shown in figure 3.19.
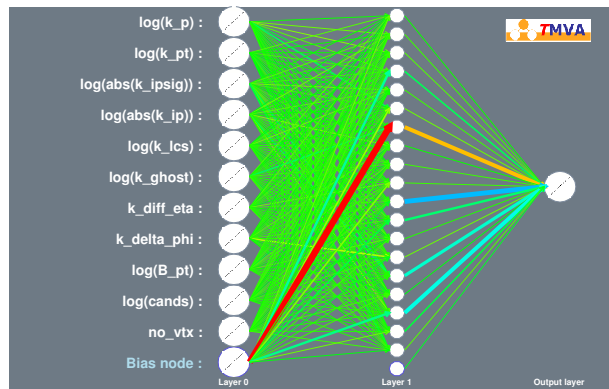

Figure 3.19.: Network Topology of ANN1.

The BFGS training method is chosen. These settings ensure that the ANN response can be interpreted as a normalized probability [89].

The response of ANN1 is shown in figure 3.20. It clearly peaks at 1 for signal (OS kaons) and at 0 for background (all other kaons). However, both distributions have long tails that leak into the other distribution's peak. The training convergence test is also shown in figure 3.20. After 400 training cycles (called epochs) the estimator curve becomes constant. The estimator of the training sample is slightly better than the one of the testing sample. The training and testing samples are statistically independent.

The correlation between the ANN1 response and each input variable is given in table 3.16. As expected, the IP significance is strongly correlated to the network response. The ranking of the correlations follows a similar order compared to the ranking of separation power of the variables, shown in table 3.15.

The separation significance σ_S of a network response is defined as

$$\sigma_S = \frac{(\bar{y}_{ANN1,S} - \bar{y}_{ANN1,B})^2}{RMS(y_{ANN1,S})^2 + RMS(y_{ANN1,B})^2}, \qquad (3.39)$$

(a) Response of ANN1
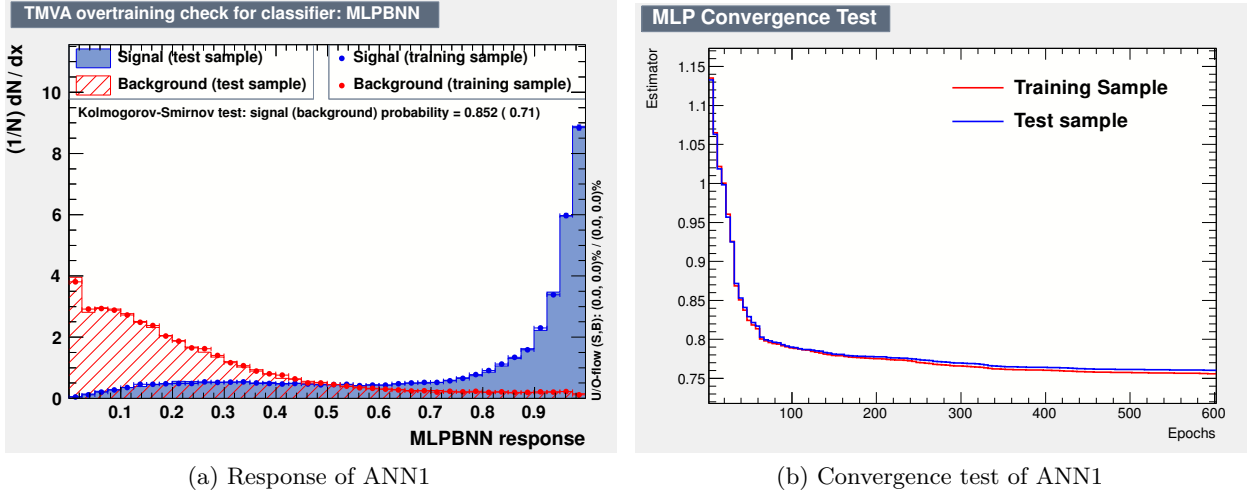
(b) Convergence test of ANN1

Figure 3.20.: (a) The response of the first ANN. (b) Convergence test of the training process of the first ANN. The estimator becomes constant over a large number of epochs. The estimator of the testing sample follows closely the estimator on the training sample.

$\bar{y}_{ANN1,S}$ and $\bar{y}_{ANN1,B}$ are the mean values of the response distributions for signal and background, respectively, and $RMS(y_{ANN1,S})$, $RMS(y_{ANN1,B})$ are the corresponding root-mean-square values of the distributions. The separation power and the separation significance were used in the iterative process to determine the optimal ANN1. Table 3.17 shows both quantities evaluated for networks, where the number of additional neurons in the hidden layer is varied. The configuration with 8 additional neurons was chosen.

Rank	Variable	Correlation to y_{ANN1}
1	$\log(|IP_{sig}|)(K)$	0.656
2	$\log(p_T)(K)$	0.481
3	$\log(|IP|)(K)$	0.421
4	$\log(p)(K)$	0.271
5	$\Delta\eta(K, B)$	0.153
6	$\log(\text{ghost prob})(K)$	0.143
7	$log(\#\text{candidates})$	0.101
8	$\log(\chi^2/\text{ndf})(K)$	0.090
9	$\#\text{vertices}$	0.024
10	$\Delta\phi(K, B)$	0.016
11	$\log(p_T)(B)$	0.009

Table 3.16.: Correlation of the ANN1 response to each input variable.

Additional Neurons	Separation Power	Separation Significance
3	0.522	1.472
4	0.521	1.470
5	0.520	1.467
6	0.521	1.470
7	0.520	1.469
8	0.522	1.472
9	0.520	1.469
10	0.522	1.472

Table 3.17.: Separation power and separation significance of networks, configured and trained as ANN1, but varying the number of additional neurons in the hidden layer.

Background rejection versus signal efficiency of ANN1 is shown in figure 3.21. At 90% signal efficiency 70% of background is rejected.
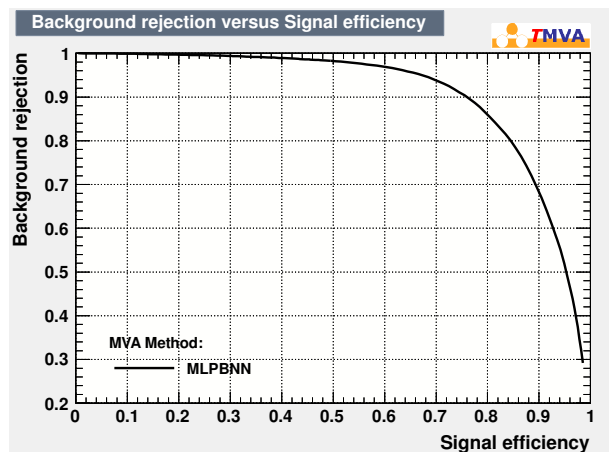

Figure 3.21.: Background rejection against signal efficiency of ANN1.

ANN1 is applied to three samples, $B_s^0 \to J/\psi\,\phi$ decays using simulation and $B^+ \to J/\psi\,K^+$ decays using simulation and data, shown in figure 3.22. A minimum cut is applied to the ANN1 response to reject kaons which are unlikely to come from an opposite b hadron decay. The cut value is optimized to achieve maximum tagging power on both types of B mesons: B^+ and B_s^0. Due to the different fragmentation processes of B^+ and B_s^0 mesons, a working point of the tagger has to be chosen that is suboptimal for B^+ decays.

When the tagger is optimized on $B^+ \to J/\psi\,K^+$ decays, only, a cut value of 0.3 is found. This rather loose cut is sufficient to reject background tagging particles in B^+ decays. The tagger works at a maximum tagging efficiency. However, when the tagger is also applied to $B_s^0 \to J/\psi\,\phi$ decays, the cut value needs to be tightened to 0.6 to reject fragmentation kaons. The distributions of the IP significance and of the number of tagging candidates per event, before and after applying the

cut on the ANN1 response, are also shown in figure 3.22. The cut on the ANN1 response removes tagging candidates with low IP significance. After the cut the number of tagging candidates per event is very similar in B_s^0 and B^+ decays, on average around 1.3.

The initial assumption, that OS taggers can be optimized on any B decay channel is withdrawn by this study. A multivariate tagger development, that optimizes not only in purity, low mistag probability, to enhance the tagging power, but also in tagging efficiency, is very sensitive to different background compositions between B types. The background is different in terms of the absolute amount of background particles per event and of charge correlation to the b flavour. Since the B decay time follows an exponential distribution, the tagger with the higher tagging efficiency predominantly tags additional opposite b hadrons with shorter decay times. Especially in these decays, where the separation power between the primary and secondary vertices is smaller, the tagger picks up tagging particles with lower IP values. However, a small IP is also the signature of fragmentation kaons from signal B_s^0 mesons. Since fragmentation kaons are oppositely charged than OS kaons, the tagging power of the efficient ANN-based OS kaon tagger is decreased. Hence, a different working point is chosen, which is less efficient, because it rejects more tagging candidates, but the tagging performance becomes comparable on B^+ and B_s^0 decays.

3.6.3. Second ANN

Training Sample

The second ANN (ANN2) is trained on $B^+ \to J/\psi\,K^+$ decays in data. Since ANN1 was re-trained to remove the additional fragmentation kaons background present in the B_s^0 production, B^+ decays are suitable here. Thus, the advantages of flavour specific B^+ decays, see section 3.6.2, can be exploited to optimize the ANN2.

The 2011 and 2012 $B^+ \to J/\psi\,K^+$ data sets are used. The same procedure to evaluate the yields is applied as in section 3.5. The fits to the reconstructed B^+ mass resonance are shown in figure 3.23, separately for 2011 and 2012 data.

The corresponding fit parameters are listed in table 3.18. In total, 770×10^3 B^+ candidates are obtained, 250×10^3 in 2011 data and 520×10^3 in 2012 data.

Fit Parameter	Fit Value in 2011 Data	Fit Value in 2012 Data
# signal	251370 ± 584	519360 ± 842
# background	48271 ± 371	111060 ± 548
μ	5280.90 ± 0.02	5280.90 ± 0.02
σ_1	8.461 ± 0.059	8.427 ± 0.043
σ_2	17.099 ± 0.354	16.688 ± 0.238
fraction	0.757 ± 0.012	0.744 ± 0.009
slope	-0.0012 ± 0.0001	-0.0012 ± 0.0001

Table 3.18.: Parameters obtained from fits to the reconstructed B^+ mass resonance using $B^+ \to J/\psi\,K^+$ decays in 2011 and 2012 data.

(a) ANN1 response (no cut applied)

(b) ANN1 response (cut applied)

(c) IP significance (no cut applied)

(d) IP significance (cut applied)
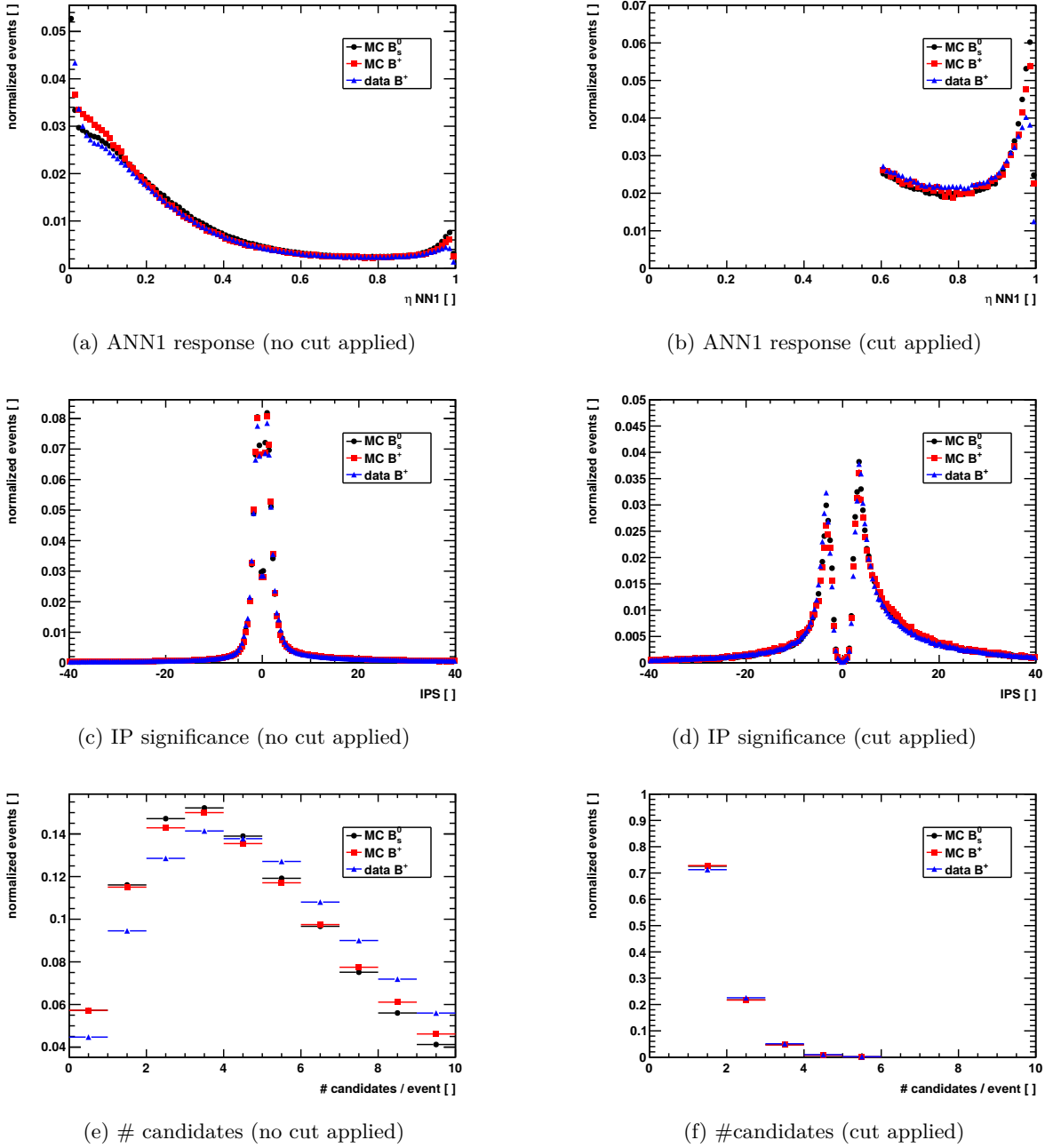
(e) # candidates (no cut applied)

(f) #candidates (cut applied)

Figure 3.22.: ANN1 applied to different samples: $B_s^0 \to J/\psi\,\phi$ decays (simulation) and $B^+ \to J/\psi\,K^+$ decays (simulation and data).
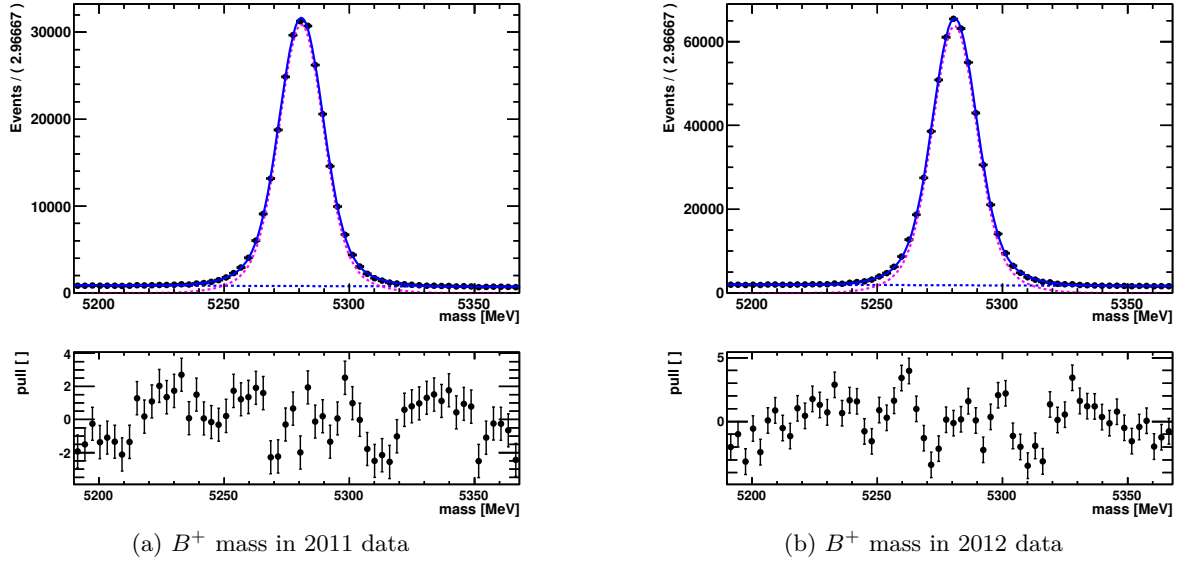
(a) B^+ mass in 2011 data

(b) B^+ mass in 2012 data

Figure 3.23.: Fits to the reconstructed B^+ mass resonance using $B^+ \to J/\psi K^+$ decays in 2011 (a) and 2012 (b) data.

Training

The training task of ANN2 is to distinguish between B and $\overline{B}$ mesons. Therefore, the response of ANN2 is identified with the mistag probability of the B event. A different approach is chosen compared to the cut-based OS kaon tagger. While the cut-based tagger uses tight requirements to select one tagging candidate in the event, the ANN-based tagger allows to select more tagging candidates which fulfill the requirements of ANN1. The advantage of this approach is, that ANN1 is not only used to filter tagging candidates, but also to assign a probability of being an OS kaon to each candidate. The probability is exploited in ANN2 to derive the tagging decision from one or multiple tagging candidates per B event. This inclusive approach allows to tag more B events. Thus, the tagging efficiency of the ANN-based tagger is higher compared to the cut-based tagger.

ANN2 uses up to three ANN1 high-ranked tagging candidates per B event. The candidates are ordered, the highest-ranked candidate comes first. The ANN1 response of each candidate is multiplied by its charge and used as input variable to ANN2. Additionally, a particle identification variable, the ANN-based PID [46], is used as input. The ANN-based PID variable is defined between 0 and 1. Values close to 1 are assigned to kaons to discriminate mainly pions. The ANN-based PID is also multiplied by the charge of the tagging candidate. In this way, additional information is provided to ANN2, which tagging candidate is most likely correlated to the opposite b hadron flavour. Figure 3.24 shows the variables, separately for B^+ events called signal and B^- events called background. In case, less than three high-ranked candidates per B event are present, the related variables are set to zero.

Besides tagging candidate variables, event variables are also used as inputs to ANN2. The signal B transverse momentum p_T, the number of primary vertices (PV) and the number of tagging
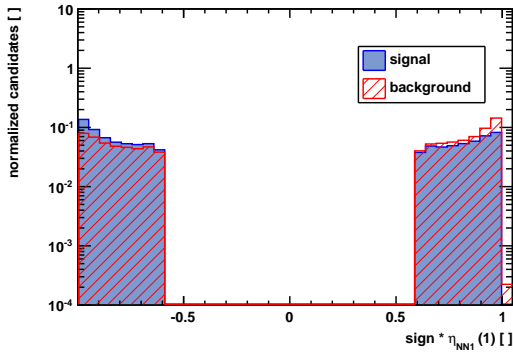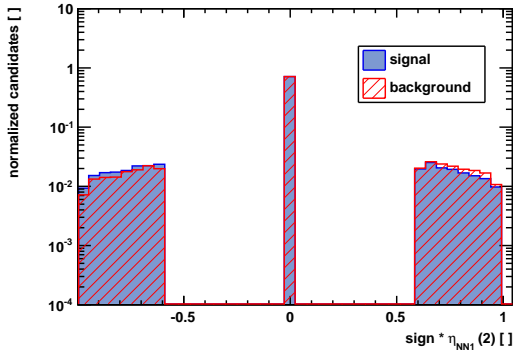
(a) Best candidate: charge $\times \, y_{ANN1}$
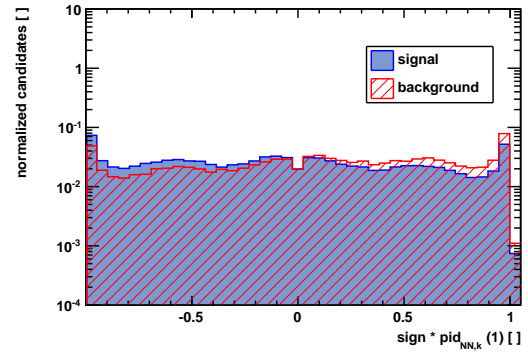
(b) Best candidate: charge $\times \, PID_{ANN,k}$

(c) 2nd best candidate: charge $\times \, y_{ANN1}$

(d) 2nd best candidate: charge $\times \, PID_{ANN,k}$

(e) 3rd best candidate: charge $\times \, y_{ANN1}$

(f) 3rd best candidate: charge $\times \, PID_{ANN,k}$

Figure 3.24.: ANN2 input variables separately drawn for signal (blue), which are arbitrarily chosen to be B^+ decays, and background (red), which are B^- decays. The ANN1 response (left) and the ANN-based PID variable (right) of the three high-ranked tagging candidates, best first, multiplied by the charge sign are taken.

candidates per event, counted after applying the cut on the ANN1 response. These quantities provide information to ANN2 of the general quality of the B event. If the event contains many primary interactions, it is unlikely to tag it correctly because the assignment of the PV to the decay might be wrong. Accordingly, if the signal B is low-energetic most likely the opposite b hadron is low-energetic, too. In this case the separation between the PV and the b hadron decay vertex is less significant. The number of tagging candidates per event provides information about the distinction of the tagging decision. The event variables are shown in figure 3.25. The same nomenclature is used: B^+ events are called signal, B^- events background.
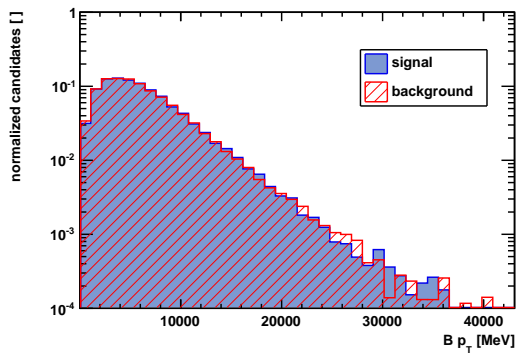
(a) signal B p_T
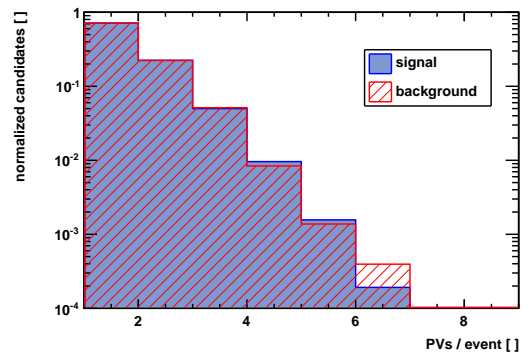
(b) # PVs / event
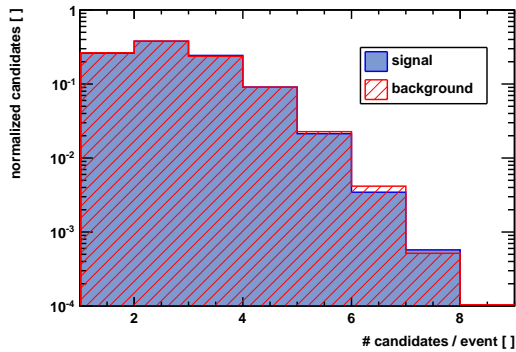
(c) # candidates / event

Figure 3.25.: ANN2 input variables separately drawn for signal (blue), which are arbitrarily chosen to be B^+ decays, and background (red), which are B^- decays: p_T of the signal B, # of PVs per event and # tagging candidates per event. The latter is evaluated after the cut on the ANN1 response is applied.

Before the training, the input variables are normalized and the linear decorrelation transformation is applied, analog to the treatment of input variables in ANN1. Figure 3.26 shows the tagging candidate related input variables after transformation. The separation power between the distributions is more visible. The event related input variables after transformation are basically unchanged.
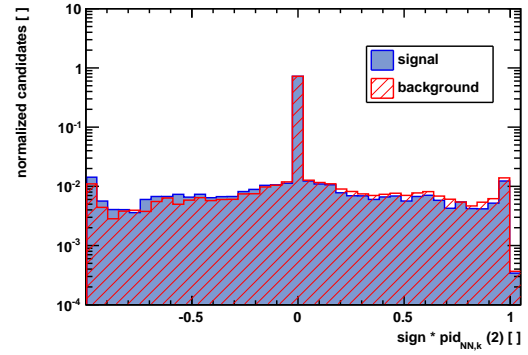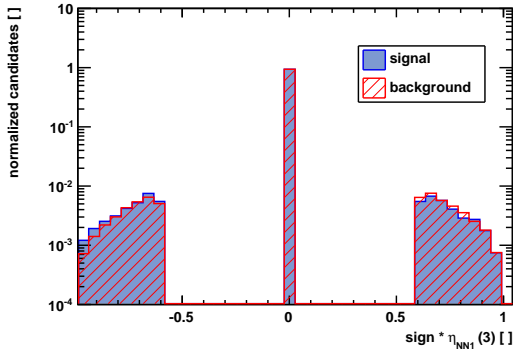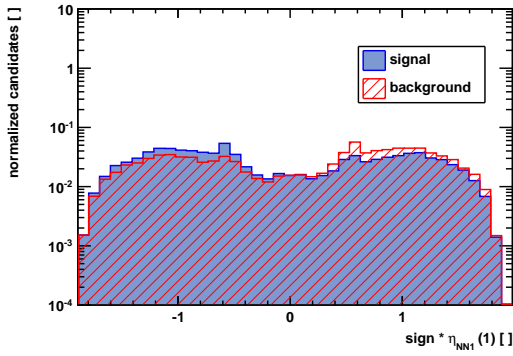
(a) Best candidate: charge $\times\, y_{ANN1}$
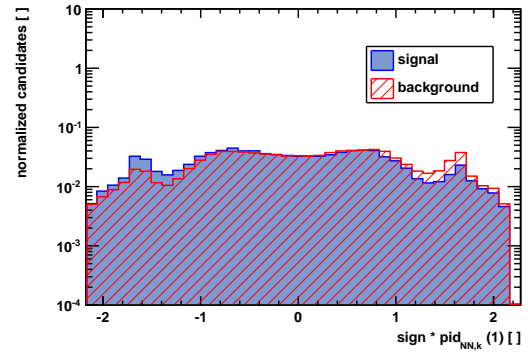
(b) Best candidate: charge $\times\, PID_{ANN,k}$
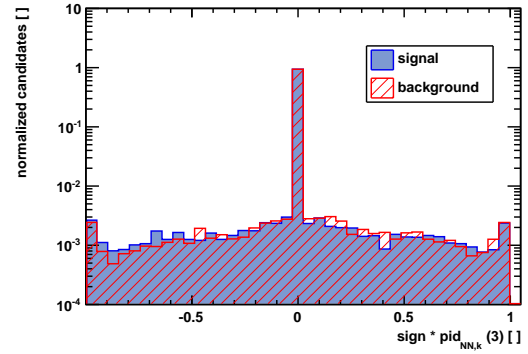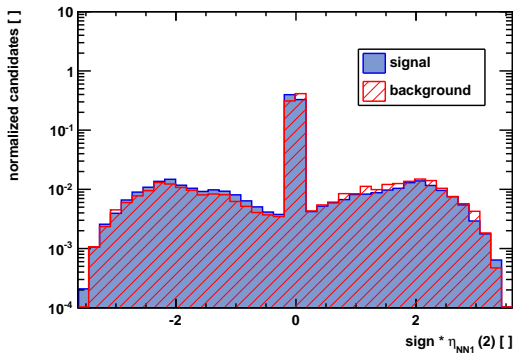

(c) 2nd best candidate: charge $\times\, y_{ANN1}$
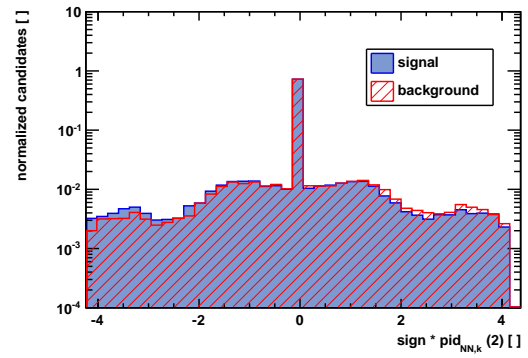
(d) 2nd best candidate: charge $\times\, PID_{ANN,k}$
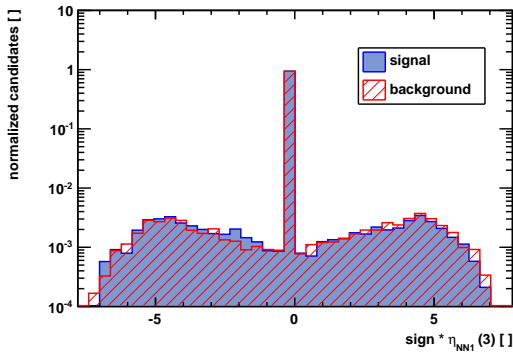
(e) 3rd best candidate: charge $\times\, y_{ANN1}$
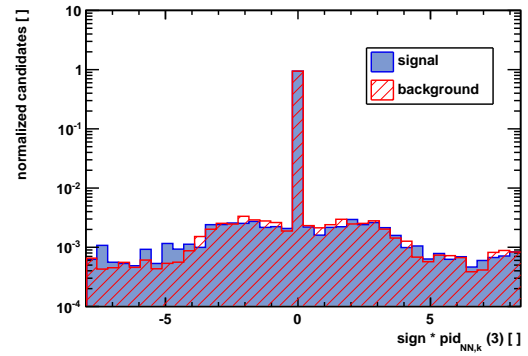
(f) 3rd best candidate: charge $\times\, PID_{ANN,k}$

Figure 3.26.: Linear decorrelation transformation applied to ANN2 input variables separately drawn for signal (blue) and background (red). The ANN1 response (left) and the ANN-based PID variable (right) of the three high-ranked tagging candidates, best first, multiplied by the charge sign are taken.

The separation power of the ANN2 input variables is ranked, see table 3.19. The same formula is used as in section 3.6.2. The ANN1 response of the highest-ranked tagging candidate provides the best separation power. It is directly followed by the ANN-based PID variable of that candidate. However, the signal B p_T variable provides more separation power than the ANN1 response of the 3rd best candidate.

Rank	Variable	Separation Power
1	best candidate charge $\times$ y_{ANN1}	2.270×10^{-2}
2	best candidate charge $\times$ $PID_{ANN,k}$	1.914×10^{-2}
3	2nd best candidate charge $\times$ $PID_{ANN,k}$	1.882×10^{-3}
4	2nd best candidate charge $\times$ y_{ANN1}	1.465×10^{-3}
5	3rd best candidate charge $\times$ $PID_{ANN,k}$	8.684×10^{-4}
6	$p_T(B)$	5.862×10^{-4}
7	3rd best candidate charge $\times$ y_{ANN1}	2.904×10^{-4}
8	#candidates	1.276×10^{-4}
9	#vertices	1.194×10^{-4}

Table 3.19.: Separation power of the input variables of ANN2.

Different combinations of input variables were used in the optimization process of ANN2. The overall trend is, that more information provided to the network leads to better classification. However, due to the tightened cut applied to the response of ANN1, to remove the fragmentation kaons of the B_s^0 production, the gain of information obtained by using more the than three high ranked candidates is vanished.

Prior to the training, the sample of $B^+ \rightarrow J/\psi K^+$ decays in 2012 data, that contain at least one tagging candidate per event, is split in two halves. The combinatorial background is removed using sWeights [79], obtained from the fits to the reconstructed B mass resonance. Thus, for training 55×10^3 B^+ and B^- sWeighted events are used, while the same number of events is used for testing the estimator of the ANN2.

The best network performance is achieved with the following configuration. A network with one hidden layer is created, where $n + 2$ neurons are present in the hidden layer. n is the number of input variables, nine in this case. Analog to ANN1, a Bayesian ANN, implemented in the TMVA framework is used. The neuron activation function is chosen to be a logistic sigmoid function, so that the network response is in the range between 0 and 1. The network topology of ANN2 is shown in figure 3.27.

The distribution of the ANN2 response is shown in figure 3.28. Ideally, the distributions would have a minimal overlap. However, this is not the case here. The signal distribution is relatively shifted to values closer to +1 and the background distribution to 0. There is a minimum at 0.5, where the network response is just a guess. This is a good sign, that the training task is well defined and reflected in the network response. The training convergence test is also shown in figure 3.28. The estimator improves over the number of used training cycles and becomes constant after 450 training cycles. However, the performance on the training sample is slightly better
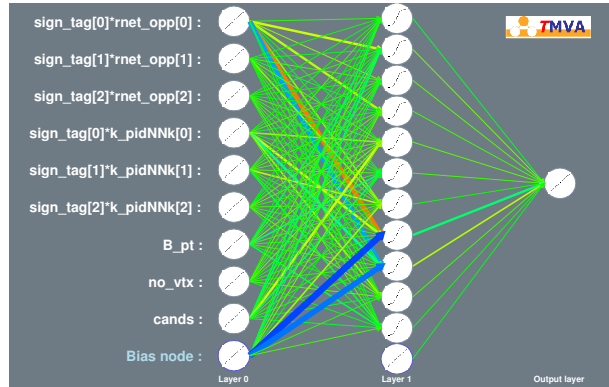
Figure 3.27.: Network Topology of ANN2.

compared to the one on the test sample. This shows, that the network becomes already sensitive to statistical fluctuation in the training data set. Further training cycles would probably lead to overtraining.
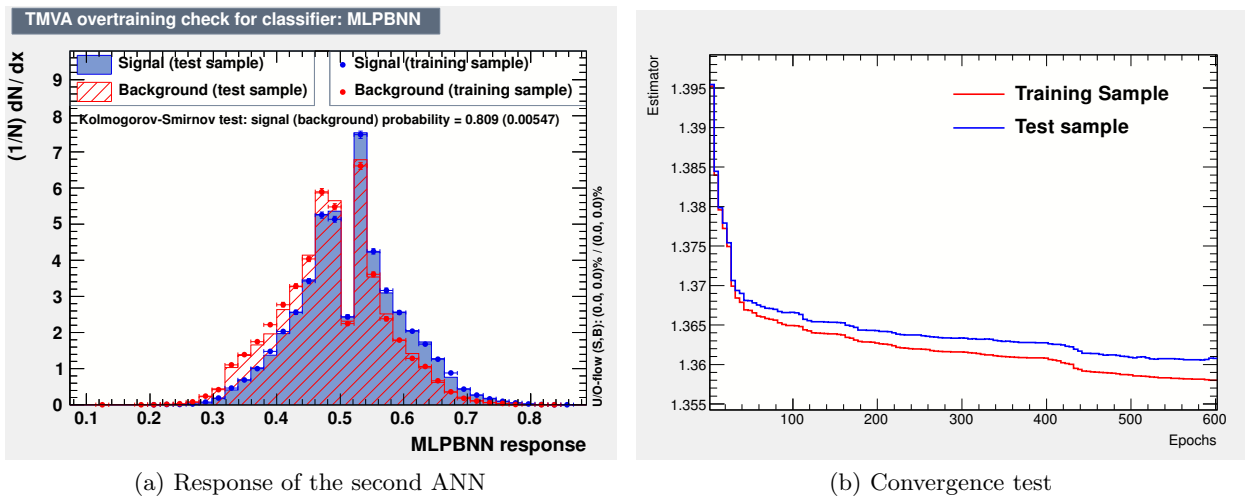
(a) Response of the second ANN

(b) Convergence test

Figure 3.28.: (a) Response of ANN2. (b) Convergence test of the training process of ANN2. The estimator becomes constant over a large number of epochs.

A ranking of the correlation of the ANN2 response to each input variable is given in table 3.20. The order is similar to the separation power of the input variables, shown in table 3.19. The best tagging candidate, according to the ANN1 response has the highest impact on the ANN2 response. Here, the correlations of the ANN-based PID variable and the ANN1 response to the ANN2 response are similar, while the former is slightly higher. The correlation between the third best tagging candidate and the ANN2 response is similar to the correlation of the signal B p_T with the ANN2 response. Hence, adding more tagging candidates per event does not improve the performance of the ANN2.

Rank	Variable	Correlation to y_{ANN2}
1	best candidate charge $\times$ $PID_{ANN,k}$	0.826
2	best candidate charge $\times$ y_{ANN1}	0.818
3	2nd best candidate charge $\times$ $PID_{ANN,k}$	0.232
4	2nd best candidate charge $\times$ y_{ANN1}	0.210
5	3rd best candidate charge $\times$ $PID_{ANN,k}$	0.058
6	3rd best candidate charge $\times$ y_{ANN1}	0.055
7	$p_T(B)$	0.052
8	#candidates	0.034
9	#vertices	0.015

Table 3.20.: Correlation of the ANN2 response to each input variable.

Analog to the optimization of ANN1, the separation power and separation significance were used in the iterative process to determine the optimal ANN2. Table 3.21 shows both quantities evaluated for networks, where the number of additional neurons in the hidden layer is varied. The configuration with 2 additional neurons was chosen for ANN2.

Additional Neurons	Separation Power	Separation Significance
2	0.026	0.227
3	0.025	0.224
4	0.025	0.224
5	0.024	0.219
6	0.025	0.222
7	0.025	0.227
8	0.024	0.220
9	0.021	0.206

Table 3.21.: Separation power and separation significance of networks, configured and trained as ANN2, but varying the number of additional neurons in the hidden layer.

This study shows, the task of finding OS kaons in $\sqrt{s} = 8\,\text{TeV}$ events at the LHC is not trivial. Nevertheless, the combination of two ANNs provides a powerful tool for the OS kaon tagger.

Derivation of the Tagging Decision

The tagging decision and the predicted mistag probability is derived from the response of ANN2, called nn_2 in the following. A B^+ event is tagged correctly, if nn_2 is greater than 0.5. While a B^- event is tagged correctly, if nn_2 is less than 0.5. Following the tagging convention, the tag decision is set to:

- +1, if $nn_2 > 0.5 \rightarrow$ a B^+ was produced,

- -1, if $nn_2 < 0.5 \rightarrow$ a B^- was produced.

The predicted mistag probability η is typically defined in the range between 0% (no mistag) and 50% (random tag). Therefore, nn_2 is folded from its initial range [0,1] to [0, 0.5] by:

- $\eta = 1 - nn_2$, if tag = +1 (the candidate is tagged as B^+),

- $\eta = nn_2$, if tag = -1 (the candidate is tagged as B^-).

Probabilistic Interpretation of the ANN Response

In order to translate the ANN2 response, nn_2, into a well defined mistag probability some additional considerations are necessary. By construction nn_2 gives the probability for a signal event (here a B meson) normalized to the number of signal plus background events, B and $\bar{B}$ events:

$$nn_2 = \frac{\#B(nn_2)}{\#B(nn_2) + \#\bar{B}(nn_2)} \tag{3.40}$$

Figure 3.29 illustrates this equality, determined on $B^+ \rightarrow J/\psi\,K^+$ data:
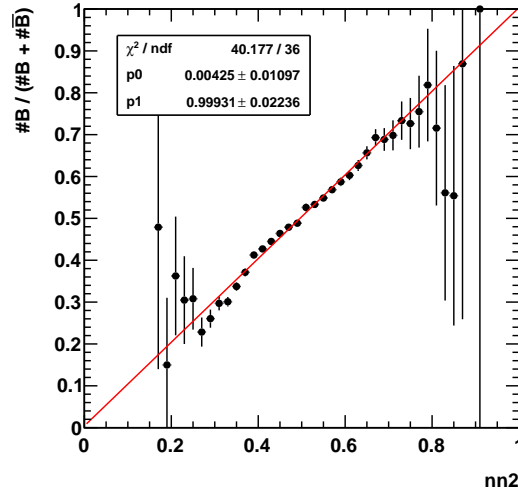
Figure 3.29.: Distribution of the ANN2 output nn_2 versus the normalized probability for a B meson: $\frac{\#B(nn_2)}{\#B(nn_2)+\#\bar{B}(nn_2)}$ in 2012 $B^+ \rightarrow J/\psi\,K^+$ data.

However, *CP* asymmetries in the opposite *b* hadron decay, detection asymmetries of positively and negatively charged kaons or possible statistical fluctuations in the training sample of ANN2 can cause differences in the ANN2 response for B or $\bar{B}$ mesons. Thus, the ANN2 response nn_2 is symmetrized

$$nn_2' = \frac{nn_2 + (1 - nn_{2,cc})}{2}, \tag{3.41}$$

where $nn_{2,cc}$ is the ANN2 response, evaluated after flipping the charge signs of the input variables. The linear dependence of $1 - nn_{2,cc}$ on nn_2 is shown in figure 3.30.
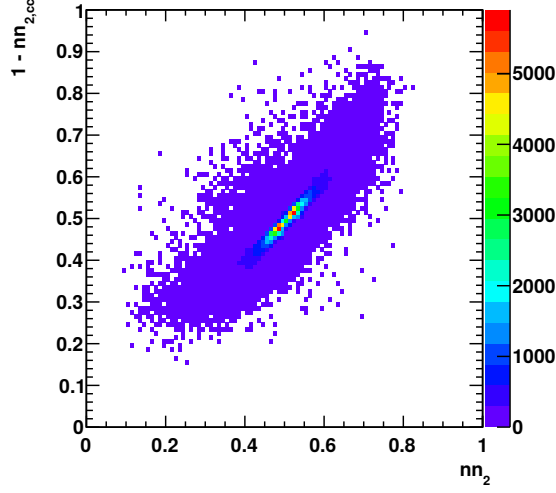
Figure 3.30.: Two dimensional distribution of the ANN2 response nn_2 versus $1 - nn_{2,cc}$, which is the ANN2 response simultaneously determined on the same events but with charge conjugated input variables.

This symmetrization is a prerequisite to interpret the ANN2 response as a well defined mistag probability η [90]. Consequently, the number of B mesons with the symmetrized ANN2 response nn_2', $N_B(nn_2')$, equals the number of $\overline{B}$ mesons with the network response $1 - nn_2'$, $N_{\overline{B}}(1 - nn_2')$. The derivation of the tagging decision is accordingly:

- tag = +1 (B), if $nn_2' > 0.5$

- tag = −1 ($\overline{B}$), if $nn_2' < 0.5$.

The predicted mistag probability is defined as

- $\eta = 1 - nn_2'$, if tag = +1 (the candidate is tagged as B),

- $\eta = nn_2'$, if tag = -1 (the candidate is tagged as $\overline{B}$).

This definition of η leads to a well defined predicted mistag probability, shown in the following. The measured mistag probability for a B meson with a predicted mistag probability η is

$$
\begin{aligned}
\omega_B(\eta) &= \frac{N_B(nn_2' = \eta)}{N_B(nn_2' = \eta) + N_B(nn_2' = 1 - \eta)} \\
&= \frac{\mathcal{N}(nn_2' = \eta)\eta}{\mathcal{N}(nn_2' = \eta)\eta + \mathcal{N}(nn_2' = 1 - \eta)(1 - \eta)} \\
&= \frac{\mathcal{N}(nn_2' = \eta)\eta}{\mathcal{N}(nn_2' = \eta)\eta + \mathcal{N}(nn_2' = \eta)(1 - \eta)} \\
&= \eta \,,
\end{aligned}
\tag{3.42}
$$

where $\mathcal{N}(nn'_2 = \eta) = N_B(nn'_2 = \eta) + N_{\bar{B}}(nn'_2 = \eta)$ and exploiting the symmetric definition of nn'_2 in equation 3.41. Accordingly, $\omega_{\bar{B}}(\eta) = \eta$ holds [90]. Hence, the predicted mistag probability is well defined by the symmetrized ANN2 response.

3.6.4. Calibration of the ANN-based OS Kaon Tagger

The predicted mistag probability, as defined in equation 3.41, is calibrated to the measured mistag probability, determined on $B^+ \to J/\psi K^+$ data. This is done for 2011 and 2012 data, separately, shown in figure 3.31. The linear calibration function

$$\omega_{\mathrm{measured}} = p_0 + p_1 \cdot (\eta_{\mathrm{predicted}} - \langle \eta \rangle) \tag{3.43}$$

is used, where $\langle \eta \rangle$ is the mean value of the $\eta_{\mathrm{predicted}}$ distribution. From the calibration fit the calibration parameters p_0 and p_1 are determined, given in table 3.22. Within the statistical uncertainties the calibration parameters agree between the two data sets.
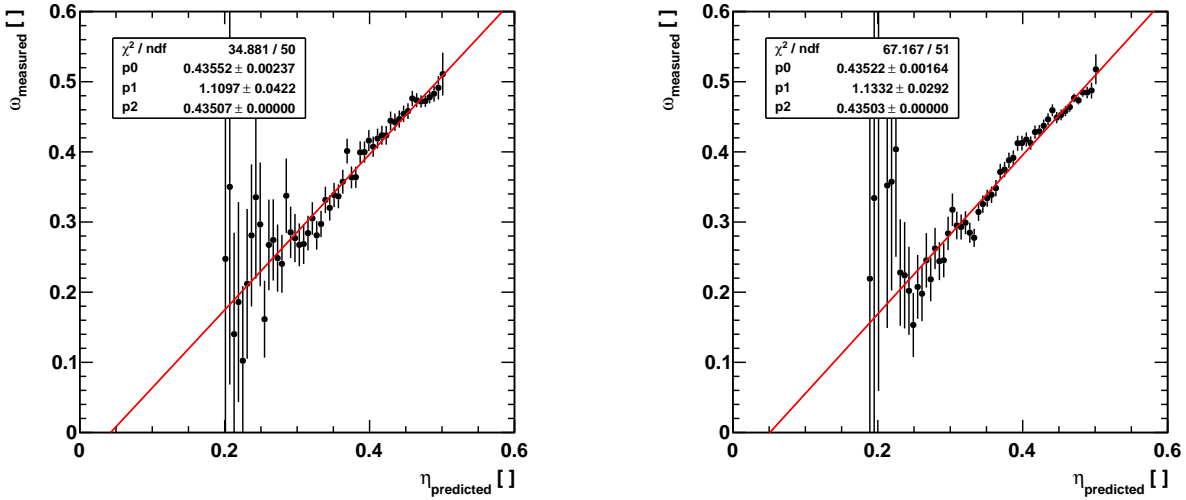
(a) Mistag probability calibration on 2011 data   (b) Mistag probability calibration on 2012 data

Figure 3.31.: Calibration of the predicted mistag probability $\eta_{\mathrm{predicted}}$ to the measured mistag probability $\omega_{\mathrm{measured}}(\eta_{\mathrm{predicted}})$ on 2011 data (a) and on 2012 data (b).

The obtained calibration is applied to the predicted mistag probability of the tagger. It moves the predicted mistag probabilities to slightly higher values, see figure 3.32.

The universality of the calibration is shown using different B meson decay channels in simulation. $B^+ \to J/\psi K^+$ decays are compared to $B^0 \to J/\psi K^{*0}$ and $B_s^0 \to J/\psi \phi$ decays. Figure 3.33 shows the fitted calibration distributions overlaid for each decay mode. The corresponding fit parameter values are given in table 3.23. There is no significant deviation between the parameters from the different calibration fits. Thus, the goal of the re-optimization of the ANN-based OS kaon tagger is achieved.

Parameter	Fit Value in 2011 Data	Fit Value in 2012 Data
p_0	0.436 ± 0.002	0.435 ± 0.002
p_1	1.110 ± 0.042	1.133 ± 0.029
$\langle \eta_{pred} \rangle = p_2$	0.435 ± 0.000	0.435 ± 0.000

Table 3.22.: Parameter obtained from calibration fits to the ANN-based OS kaon tagger measured mistag probability using 2011 and 2012 $B^+ \to J/\psi K^+$ data.
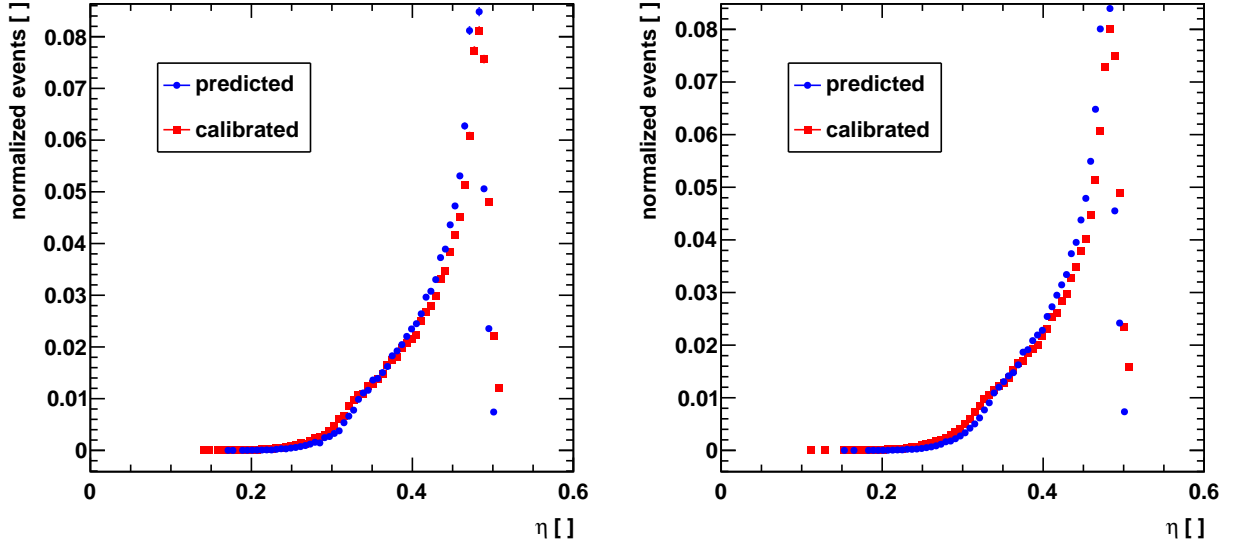
(a) Calibrated predicted mistag probability on 2011 data   (b) Calibrated predicted mistag probability on 2012 data

Figure 3.32.: Calibrated predicted mistag probability $\eta_{\text{predicted}}$ on 2011 (a) and on 2012 $B^+ \to J/\psi K^+$ data (b). The calibrated mistag distribution is slightly shifted to higher mistag probabilities with respect to the uncalibrated mistag distribution.

Parameter	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^{*0}$ MC	$B_s^0 \to J/\psi \phi$ MC
p_0	0.415 ± 0.003	0.411 ± 0.005	0.421 ± 0.002
p_1	1.264 ± 0.054	1.288 ± 0.074	1.312 ± 0.028
$\langle \eta_{pred} \rangle = p_2$	0.426 ± 0.000	0.422 ± 0.000	0.423 ± 0.000

Table 3.23.: Parameter obtained from calibration fits to the measured mistag probability of the ANN-based OS kaon tagger using $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$ and $B_s^0 \to J/\psi \phi$ decays in MC simulation. The calibration is portable among the different B decay channels.
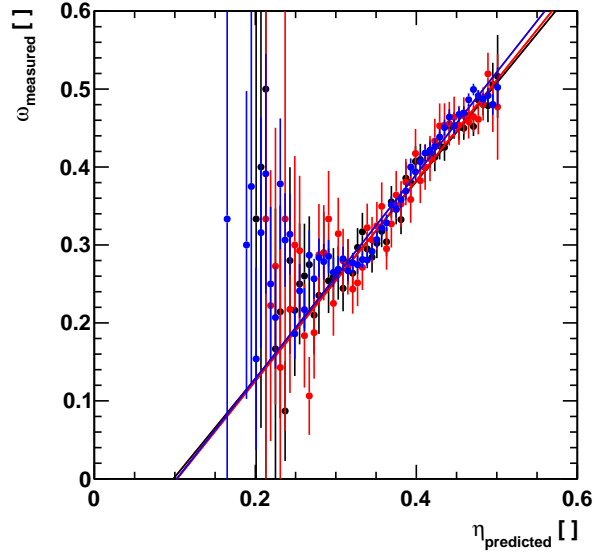
Figure 3.33.: Calibration of the ANN-based OS kaon tagger mistag probability applied to different MC channels: $B^+ \to J/\psi\,K^+$ decays in black, $B^0 \to J/\psi\,K^{*0}$ decays in red and $B_s^0 \to J/\psi\,\phi$ in blue. The calibration is portable among different B decay channels.

3.6.5. Performance of the ANN-based OS Kaon Tagger

The performance of the ANN-based OS kaon tagger is determined using $B^+ \to J/\psi\,K^+$ decays in 2011 and 2012 data. For comparison, the numbers before the re-optimization of the tagger to the B_s^0 production topology are also evaluated. Both, average and event-by-event mistag probabilities are computed, defined in equation 3.2 and equation 3.6, respectively. Table 3.24 shows all numbers.

The re-optimization of the tagger decreases the tagging power significantly. The tagging efficiency is 12% to 14% decreased. Comparing the performance on 2011 and 2012 data, tagging efficiency is slightly higher in 2012 data, where at larger center-of-mass energies events with higher track multiplicities are produced. The tagging power is compatible in both data taking years. This shows, that the ANN-based OS kaon tagger can treat higher track multiplicities. In the next section a comparison of the ANN-based OS kaon tagging performance with the performance of other OS tagging algorithms is given.

3.6.6. Comparison of Tagging Performances

This section shows the improvement in tagging power of the ANN-based OS kaon tagger compared to the cut-based OS kaon tagger. To quantify the effective improvement, combinations with all other OS taggers are evaluated, after separate calibrations of each combination. All numbers are evaluated for 2011 and 2012 data separately, using $B^+ \to J/\psi\,K^+$ data. The combination of taggers is done following the combination technique given in subsection 3.4.1.

data set	$\epsilon_{tag}(\%)$	$\omega(\%)$	$\epsilon_{eff}(\%)$
		average	
2011 B^+ optimized	57.97 ± 0.10	43.74 ± 0.13	0.91 ± 0.04
2011 B_s^0 optimized	44.33 ± 0.10	43.60 ± 0.15	0.73 ± 0.04
2012 B^+ optimized	59.21 ± 0.07	43.87 ± 0.09	0.89 ± 0.03
2012 B_s^0 optimized	45.11 ± 0.07	43.57 ± 0.10	0.75 ± 0.03
		event-by-event	
2011 B^+ optimized	57.97 ± 0.10	42.11 ± 0.11	1.44 ± 0.04
2011 B_s^0 optimized	44.33 ± 0.10	41.43 ± 0.11	1.30 ± 0.04
2012 B^+ optimized	59.21 ± 0.07	42.00 ± 0.08	1.52 ± 0.03
2012 B_s^0 optimized	45.11 ± 0.07	41.45 ± 0.11	1.32 ± 0.03

Table 3.24.: Tagging performance of the ANN-based OS kaon tagger using average mistag probabilities (upper part) and event-by-event mistag probabilities (lower part), determined on $B^+ \to J/\psi K^+$ 2011 and 2012 data. Both optimizations of the tagger are considered.

An overview of the tagging performances of the single cut-based OS taggers is given in table 3.25. The upper part of the table uses average mistag probabilities, the lower part event-by-event mistag probabilities.

Compared to the cut-based OS kaon tagger, which has a tagging power of 0.99 ± 0.02 using event-by-event mistag probabilities on 2012 data, the ANN-based OS kaon tagger achieves a relative improvement in tagging power of 54% using the B^+ optimization and 33% using the B_s^0 optimization. This reduces the statitical uncertainty on a mixing asymmetry measurement by 19% and 13%, respectively, without taking more data, see subsection 3.4.2. In the following, the transport of this improvement through the full OS tagger combination is evaluated.

Combination with the OS Vertex Charge Tagger

Overlap in tagged events is expected with the OS vertex charge tagger. Therefore, first a combination of the OS vertex tagger with the ANN-based or the cut-based OS kaon taggers is done. The combinations are calibrated, prior to the determination of tagging performance. Figure 3.34 shows the calibration fits and calibrated predicted mistag distributions for the ANN-based OS kaon and cut-based vertex charge tagger combination, separately for 2011 and 2012 $B^+ \to J/\psi K^+$ data. The calibrations are compatible among the data taking years. Figure 3.35 shows the same distributions for the combination of cut-based OS kaon and vertex charge tagger combination, separately for 2011 and 2012 $B^+ \to J/\psi K^+$ data.

The tagging performances for both combinations of taggers are given in table 3.26 using average mistag probabilities and event-by-event mistag probabilities. The ANN-based OS kaon tagger has a very steep distribution of predicted mistag values towards 50% mistag, see figure 3.32. This behaviour is transported through the combination with the OS vertex charge tagger, see figure 3.34.
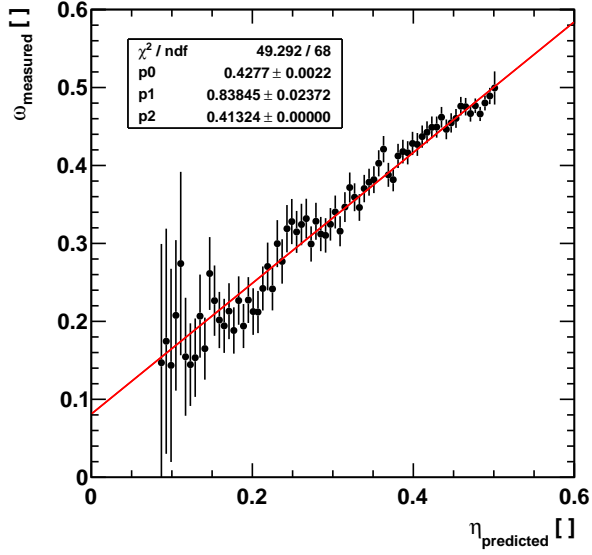
(a) 2011 data calibration
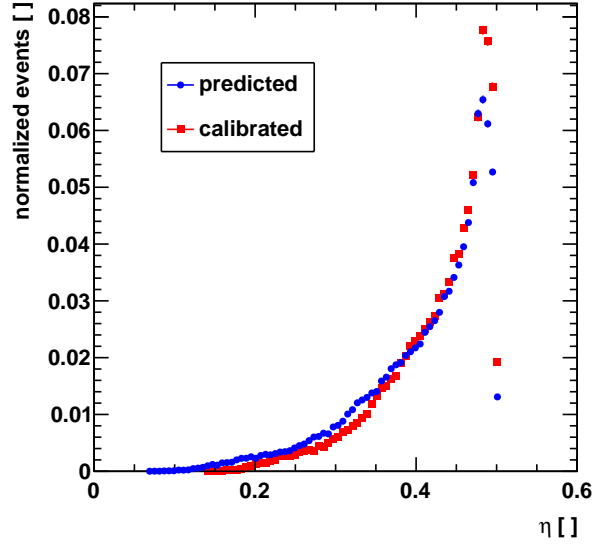
(b) 2011 data mistag distribution
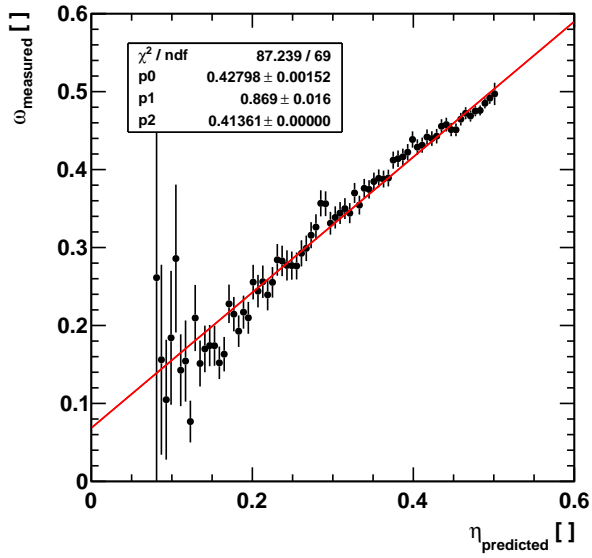
(c) 2012 data calibration
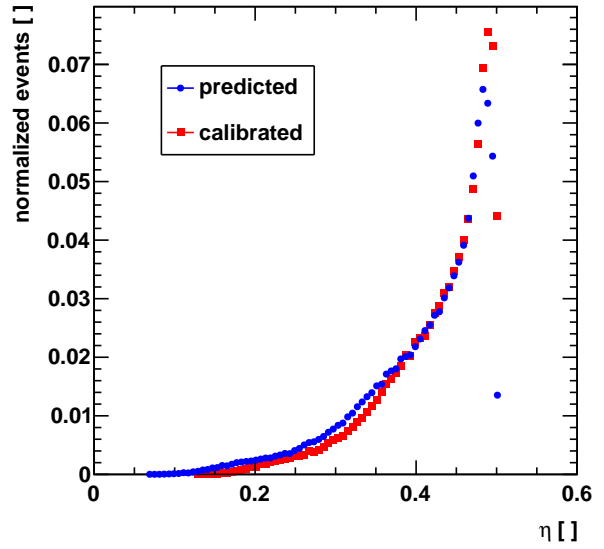
(d) 2012 data mistag distribution

Figure 3.34.: Calibration of the combination of the OS vertex tagger and the ANN OS kaon tagger on 2011 data (upper plots) and 2012 data (lower plots).
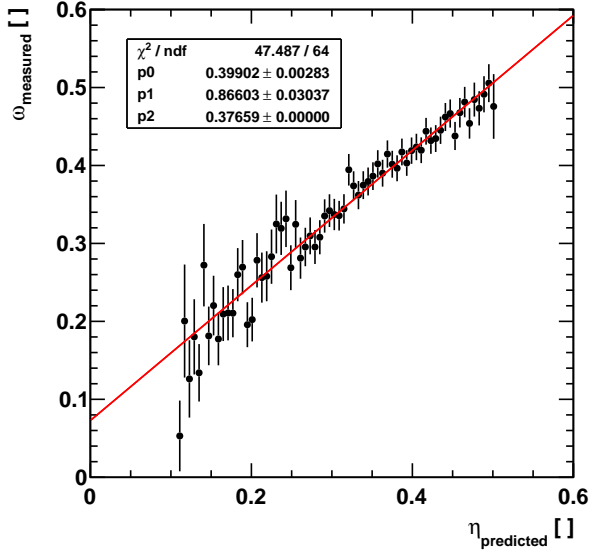
(a) 2011 data calibration
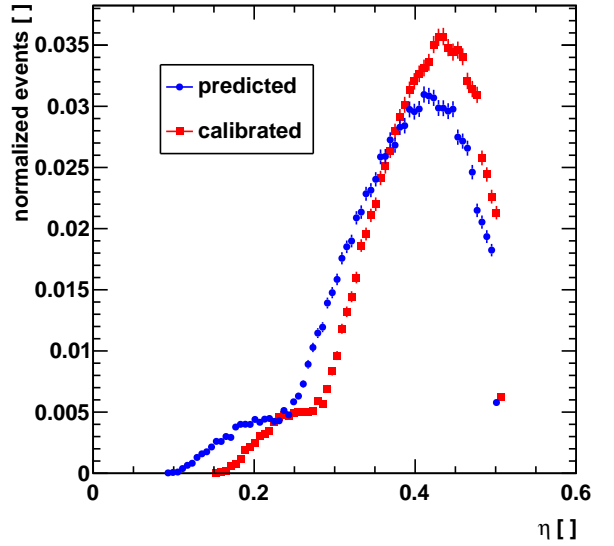
(b) 2011 data mistag distribution

(c) 2012 data calibration

(d) 2012 data mistag distribution

Figure 3.35.: Calibration of the combination of the OS vertex tagger and the OS kaon tagger on 2011 data (upper plots) and 2012 data (lower plots).

tagger	$\epsilon_{tag}(\%)$	$\omega(\%)$	$\epsilon_{eff}(\%)$
		average	
2011 vertex	17.27 ± 0.08	39.54 ± 0.24	0.76 ± 0.03
2011 kaon	16.75 ± 0.08	39.35 ± 0.25	0.76 ± 0.03
2011 muon	5.13 ± 0.05	30.37 ± 0.42	0.79 ± 0.03
2011 electron	1.56 ± 0.03	28.71 ± 0.74	0.28 ± 0.02
2012 vertex	18.08 ± 0.05	39.75 ± 0.16	0.76 ± 0.02
2012 kaon	16.98 ± 0.05	39.26 ± 0.16	0.78 ± 0.02
2012 muon	5.34 ± 0.03	29.95 ± 0.27	0.86 ± 0.02
2012 electron	1.56 ± 0.02	29.65 ± 0.51	0.26 ± 0.01
		event-by-event	
2011 vertex	17.27 ± 0.08	37.30 ± 0.24	1.11 ± 0.03
2011 kaon	16.75 ± 0.08	37.72 ± 0.25	1.01 ± 0.03
2011 muon	5.13 ± 0.05	27.51 ± 0.42	1.04 ± 0.03
2011 electron	1.56 ± 0.03	27.93 ± 0.74	0.30 ± 0.02
2012 vertex	18.08 ± 0.05	37.66 ± 0.16	1.10 ± 0.02
2012 kaon	16.98 ± 0.05	37.96 ± 0.16	0.99 ± 0.02
2012 muon	5.34 ± 0.03	27.65 ± 0.27	1.07 ± 0.02
2012 electron	1.56 ± 0.02	28.14 ± 0.51	0.30 ± 0.01

Table 3.25.: Preliminary tagging performance of the cut-based OS taggers using average mistag probabilities (upper part) and event-by-event mistag probabilities (lower part). The numbers are determined using $B^+ \to J/\psi K^+$ decays in 2011 and 2012 data.

More tagged events with predicted mistag probabilities close to 0.5 do not harm if event-by-event mistag probabilities are applied. However, they dilute the average mistag probability. Thus, in case of average mistag probabilities the cut-based combination shows a better tagging power, after the re-optimization to B_s^0 decays. This can be changed, by requiring an upper limit (below 0.5) on the calibrated predicted mistag probabilities of the ANN-based OS kaon tagger.

However, it reflects the realistic mistag prediction by the tagger, that in a large fraction of events only little information of the signal B flavour can be extracted. The tagging power from event-by-event mistag probabilities confirms this. It is highest, when all tagged events are considered but weighted according to their predicted mistag probability.

After combination, the event-by-event tagging power of the B^+ or B_s^0 optimized ANN-based OS kaon tagger is very similar. In total, an improvement of relative 16% and 14% in tagging power using event-by-event mistag probabilities is achieved, respectively.

tagger	$\epsilon_{tag}(\%)$	$\omega(\%)$	$\epsilon_{eff}(\%)$
		average	
2011 vertex + cut kaon	28.04 ± 0.09	40.00 ± 0.18	1.12 ± 0.04
2011 vertex + ANN kaon B^+	63.05 ± 0.10	43.14 ± 0.13	1.19 ± 0.05
2011 vertex + ANN kaon B_s^0	50.32 ± 0.10	42.83 ± 0.14	1.04 ± 0.04
2012 vertex + cut kaon	28.89 ± 0.06	40.09 ± 0.13	1.14 ± 0.03
2012 vertex + ANN kaon B^+	64.38 ± 0.07	43.22 ± 0.09	1.18 ± 0.03
2012 vertex + ANN kaon B_s^0	51.34 ± 0.07	42.84 ± 0.10	1.05 ± 0.03
		event-by-event	
2011 vertex + cut kaon	28.04 ± 0.09	37.78 ± 0.18	1.67 ± 0.05
2011 vertex + ANN kaon B^+	63.05 ± 0.10	41.16 ± 0.09	1.97 ± 0.04
2011 vertex + ANN kaon B_s^0	50.32 ± 0.10	40.31 ± 0.15	1.89 ± 0.06
2012 vertex + cut kaon	28.89 ± 0.06	37.72 ± 0.12	1.74 ± 0.04
2012 vertex + ANN kaon B^+	64.38 ± 0.07	41.15 ± 0.06	2.02 ± 0.03
2012 vertex + ANN kaon B_s^0	51.34 ± 0.07	40.19 ± 0.10	1.98 ± 0.04

Table 3.26.: Tagging performance of the combination of the vertex tagger with the cut-based or with the ANN-based OS kaon tagger using average mistag probabilities (upper part) and event-by-event mistag probabilities (lower part).

Combination with the Full OS Tagger

The full OS tagger combinations, where the OS vertex charge, OS muon and OS electron taggers are combined with the ANN-based or the cut-based OS kaon tagger, are shown in the following. The overlap of tagged events of the ANN-based OS kaon tagger and the OS lepton taggers is expected to be very small. Hence, the absolute improvement in tagging power after combination with the OS vertex charge tagger, $\sim 0.2\%$, is transported through the full OS combination. Figure 3.36 shows the calibration fits and calibrated predicted mistag distributions for the OS tagger combination including the ANN-based OS kaon tagger. Within the statistical uncertainties on the calibration fit parameters the calibrations are comparable between 2011 and 2012 $B^+ \to J/\psi K^+$ data. Figure 3.37 shows same distributions for the OS tagger combination including the cut-based OS kaon tagger.

After calibration the tagging power of each combination on each data set is evaluated using average and event-by-event mistag probabilities, given in table 3.27. The numbers are computed separately for the B^+ optimization of the ANN-based OS kaon tagger and for the B_s^0 optimization.

The tagging power using average mistag probability is lower for the combination with the ANN-based OS kaon tagger, while the tagging power using event-by-event mistag probabilities is higher. The absolute improvement in tagging power of 0.2% is unchanged in case of B_s^0 optimization of the ANN-based OS kaon tagger, while in case of the B^+ optimization, the absolute gain in tagging power is slightly higher. The relative improvement in tagging power is 8% and 10%, respectively.
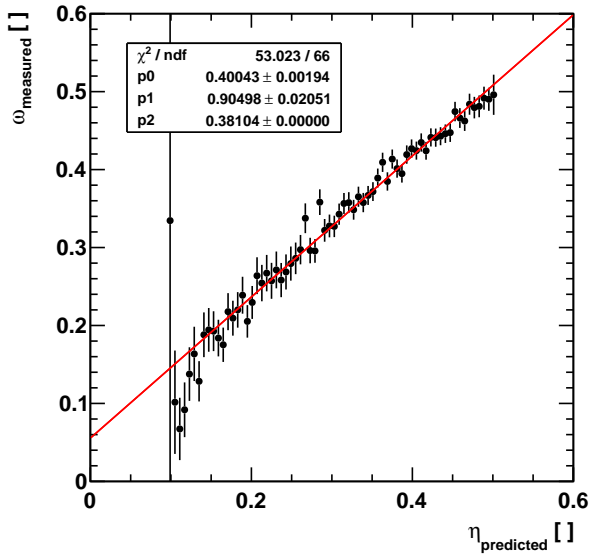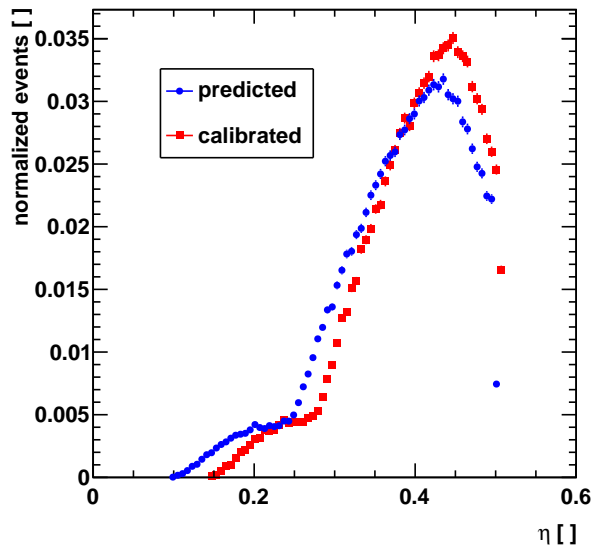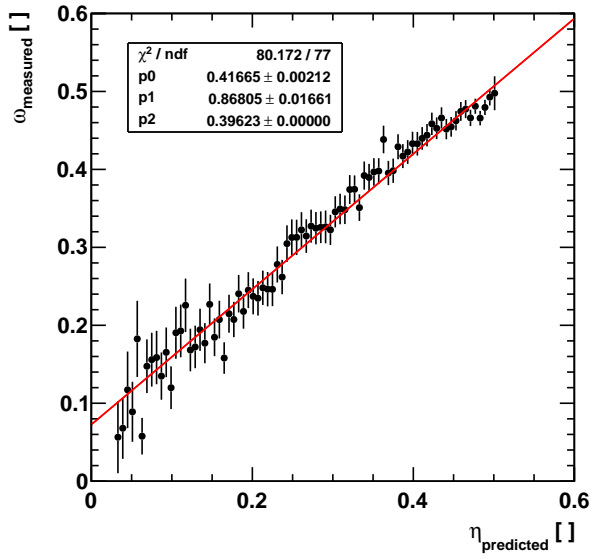
(a) 2011 data calibration
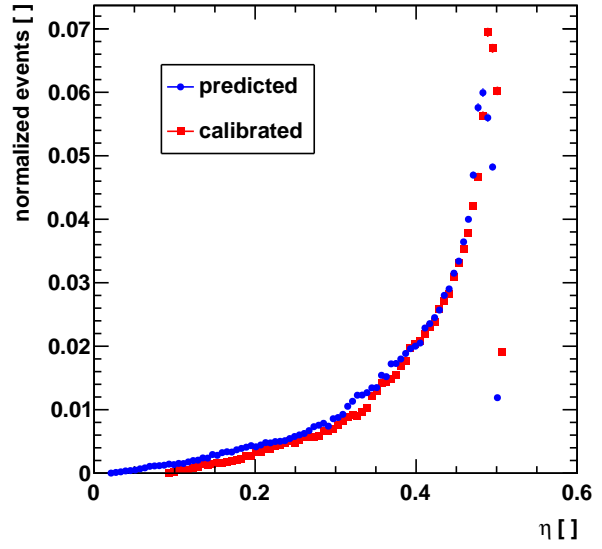
(b) 2011 data mistag distribution
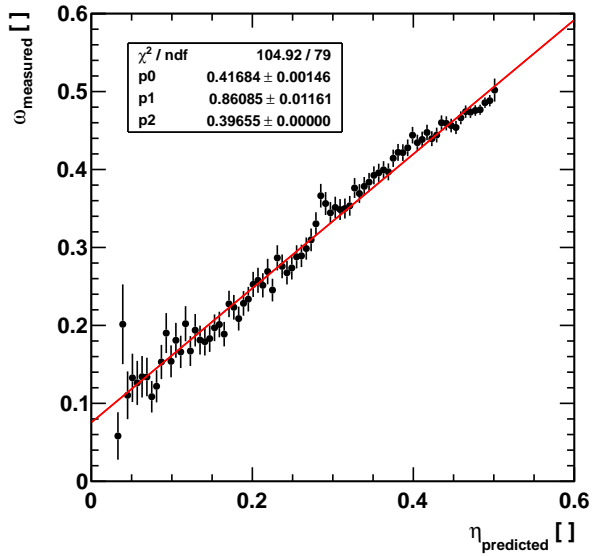
(c) 2012 data calibration
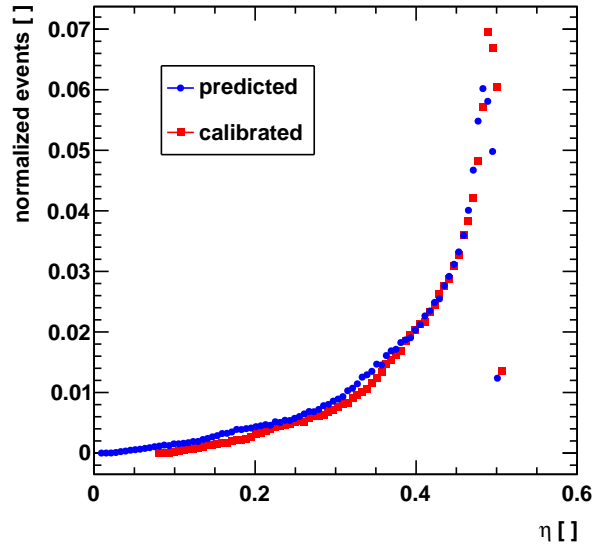
(d) 2012 data mistag distribution

Figure 3.36.: Calibration of the combination of the OS vertex charge, OS muon, OS electron and ANN-based OS kaon tagger on $B^+ \rightarrow J/\psi K^+$ 2011 data (upper plots) and on 2012 data (lower plots).
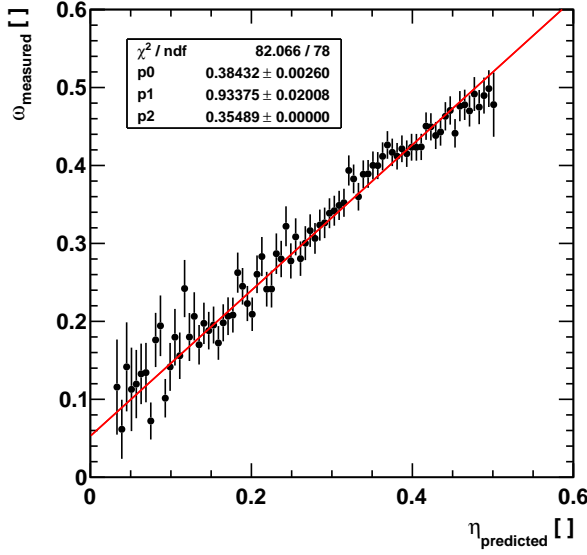
(a) 2011 data calibration
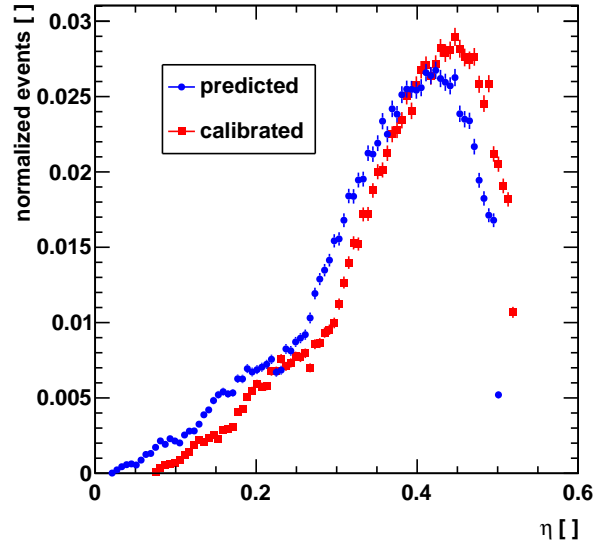
(b) 2011 data mistag distribution
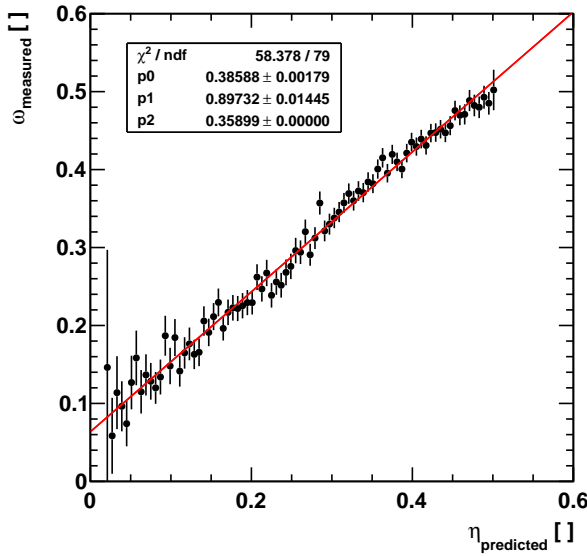
(c) 2012 data calibration
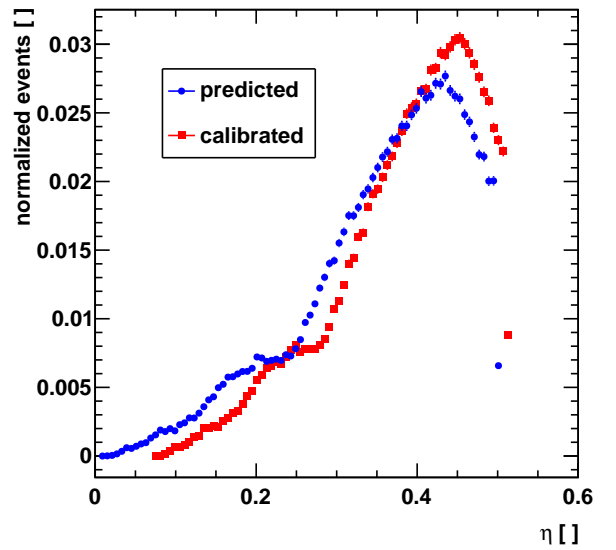
(d) 2012 data mistag distribution

Figure 3.37.: Calibration of the combination of the OS taggers on 2011 data (upper plots) and on 2012 data (lower plots).

tagger	$\epsilon_{tag}(\%)$	$\omega(\%)$	$\epsilon_{eff}(\%)$
		average	
2011 OS + cut kaon	31.75 ± 0.09	38.55 ± 0.17	1.67 ± 0.05
2011 OS + ANN kaon B^+	64.70 ± 0.10	42.25 ± 0.13	1.56 ± 0.05
2011 OS + ANN kaon B_s^0	52.63 ± 0.10	41.74 ± 0.14	1.44 ± 0.05
2012 OS + cut kaon	32.70 ± 0.07	38.64 ± 0.12	1.69 ± 0.03
2012 OS + ANN kaon B^+	66.02 ± 0.07	42.30 ± 0.08	1.56 ± 0.04
2012 OS + ANN kaon B_s^0	53.68 ± 0.07	41.73 ± 0.09	1.47 ± 0.03
		event-by-event	
2011 OS + cut kaon	31.75 ± 0.09	35.24 ± 0.16	2.77 ± 0.06
2011 OS + ANN kaon B^+	64.70 ± 0.10	39.34 ± 0.09	2.94 ± 0.05
2011 OS + ANN kaon B_s^0	52.63 ± 0.10	38.21 ± 0.14	2.93 ± 0.07
2012 OS + cut kaon	32.70 ± 0.07	35.56 ± 0.11	2.73 ± 0.04
2012 OS + ANN kaon B^+	66.02 ± 0.07	39.37 ± 0.06	2.99 ± 0.03
2012 OS + ANN kaon B_s^0	53.68 ± 0.07	38.29 ± 0.09	2.95 ± 0.05

Table 3.27.: Tagging performance of the combination of the OS vertex, electron and muon taggers and the cut-based OS kaon tagger compared to the combination of the OS taggers (vertex, electron and muon) with the ANN-based OS kaon tagger using average mistag probabilities (upper part) and event-by-event mistag probabilities (lower part).

**Tagger Correlations**

The before mentioned improvements in the combined OS tagging power due to the ANN-based OS kaon tagger are substantiated by lower correlations to the other OS taggers. This section evaluates two types of correlations between taggers, the correlations of mistag probabilities and of tagging efficiencies.

To compute the mistag probability correlations between tagger i and tagger j, the two dimensional distribution of the calibrated predicted mistag probabilities η_i and η_j is used. The mistag correlation factor $\rho(\eta_i, \eta_j)$ is defined as

$$\rho(\eta_i, \eta_j) = \frac{\mathcal{C}_{i,j}}{\sigma_{\eta_i}\sigma_{\eta_j}}, \tag{3.44}$$

where $\mathcal{C}_{i,j}$ is the covariance between the mistag distributions of tagger i and j and σ_{η_i} the width of the ith mistag probability distribution. The correlation values are shown in table 3.28 for the combination with the ANN-based OS kaon tagger and in table 3.29 for the combination with the cut-based OS kaon tagger. The mistag probability correlations between the ANN-based OS kaon tagger and the OS lepton taggers are a factor 2.5 to 4 lower compared to the ones with the cut-based OS kaon tagger. Between the OS vertex charge tagger and the ANN-based OS kaon tagger the correlation is 30% smaller compared to the one with the cut-based OS kaon tagger. Although the tagging efficiency of the ANN-based OS kaon tagger is a factor of 2.6 higher than

the one of the cut-based OS kaon tagger the mistag probability correlations are significantly lower. That emphasizes the different approach of using a two stepped ANN, which learns to classify events depending on the ANN driven combined probability of tracks to come from opposite b hadron decays instead of selecting one best tagging track per event.

	muon	electron	ANN-based kaon	vertex
muon	–	0.1351 / 0.1590	0.0485 / 0.0646	0.3406 / 0.3356
electron	–	–	0.0740 / 0.0785	0.3054 / 0.3169
ANN-based kaon	–	–	–	0.2198 / 0.2376
vertex	–	–	–	–

Table 3.28.: Mistag Probability correlations between two different OS taggers measured on 2011 / 2012 $B^+ \rightarrow J/\psi K^+$ data, respectively.

	muon	electron	cut-based kaon	vertex
muon	–	0.1351 / 0.1590	0.1729 / 0.1785	0.3406 / 0.3356
electron	–	–	0.1690 / 0.1921	0.3054 / 0.3169
cut-based kaon	–	–	–	0.3138 / 0.3390
vertex	–	–	–	–

Table 3.29.: Mistag Probability correlations between two different OS taggers measured on 2011 / 2012 $B^+ \rightarrow J/\psi K^+$ data, respectively.

The tagging efficiency correlation factor $\rho(\epsilon_i, \epsilon_j)$ between tagger i and tagger j is defined as

$$\rho(\epsilon_i, \epsilon_j) = \frac{\epsilon_{tag}^{i\&j}}{\epsilon_{tag}^i \epsilon_{tag}^j} - 1, \tag{3.45}$$

where $\epsilon_{tag}^{i\&j}$ is the tagging efficiency for a given sample of events, to be tagged by both taggers and $\epsilon_{tag}^i, \epsilon_{tag}^j$ are the tagging efficiencies of each tagger on this sample. If the overlap $\epsilon_{tag}^{i\&j}$ is zero, the correlation factor is -1. The correlation factor illustrates, whether the majority of events tagged by one tagger is contained in the set of events tagged by another tagger.

The tagging efficiency correlation factor values are given in table 3.30 for the combination with the ANN-based OS kaon tagger and in table 3.31 for the combination with the cut-based OS kaon tagger. The different working principle of the ANN-based OS kaon tagger is very prominent here. The tagging efficiency correlations between the OS lepton taggers and the OS vertex charge tagger are big, the events tagged by the muon or electron taggers are (almost) completely contained in the tagged events by the vertex charge tagger. The ANN-based OS kaon tagger and the OS vertex charge tagger share only 2/3 of the events tagged by the vertex charge tagger. In contrast, the cut-based OS kaon tagger tags more than 80% of the same events as the vertex charge tagger.

	muon	electron	ANN-based kaon	vertex
muon	–	-0.1520 / -0.1454	0.1525 / 0.1570	0.9867 / 0.9529
electron	–	–	0.3542 / 0.3151	1.4416 / 1.2579
ANN-based kaon	–	–	–	0.4743 / 0.4513
vertex	–	–	–	–

Table 3.30.: Tagging efficiency correlations between two different OS taggers measured on 2011 / 2012 $B^+ \to J/\psi K^+$ data, respectively.

	muon	electron	cut-based kaon	vertex
muon	–	-0.1520 / -0.1454	0.2428 / 0.2401	0.9867 / 0.9529
electron	–	–	0.4122 / 0.3401	1.4416 / 1.2579
cut-based kaon	–	–	–	1.0720 / 1.0122
vertex	–	–	–	–

Table 3.31.: Tagging efficiency correlations between two different OS taggers measured on 2011 / 2012 $B^+ \to J/\psi K^+$ data, respectively.

In summary, the OS tagger combination profits from the ANN-based OS kaon tagger. The mistag probabilities of the OS taggers with respect to the ANN-based OS kaon tagger are less correlated. However, the gain of 60% in tagging efficiency by the ANN-based OS kaon tagger is partially covered by the OS vertex charge tagger. The absolute tagging efficiency correlation factors between the OS vertex charge and ANN-based OS kaon tagger compared to the one with the cut-based OS kaon tagger are improved. While the relative improvement in these correlation factors is 20% smaller than the gain in tagging efficiency. This explains, why the absolute improvement of 0.3% in tagging power of the ANN-based OS kaon tagger compared to the cut-based OS kaon tagger is reduced by 1/3 to 0.2% absolute gain in tagging power after the combination with the vertex charge tagger. The additional combination with the OS lepton taggers does not lower the absolute gain in tagging power.

3.6.7. Systematic Effects on the Calibration

To perform precision *CP* violation measurements it is crucial that the tagger is properly calibrated for B and $\overline{B}$ decays separately. Charge dependent reconstruction asymmetries might result in a different tagging performance for B and $\overline{B}$ decays. Many dependencies are taken into account in the ANN input variables, however, in order to transport the tagger from one data set to another, one needs to test such dependencies explicitly. Thus, such effects will be investigated in this section. The tests can be classified into three different types.

First, the influence of possible charge and detector asymmetries on the mistag calibration are tested. Particles and anti-particles can interact differently with the detector, mainly due to two

reasons. The cross sections for material interaction are different for K^+ and K^- [72]. This effect on the calibration is tested by splitting the data sample into B^+ and B^- subsamples. On the other hand particles and anti-particles are bent into different halves of the detector by the magnetic field. Due to asymmetries in the detector material detection inefficiencies can arise. Here, the special LHCb feature, the reversal of the magnet polarity can be exploited. The effect on the calibration is tested by splitting the data sample into magnet polarity up and magnet polarity down subsamples.

Second, different event topologies can have an impact on the predicted mistag calibration and tagging performance. Heavier mesons require higher production energies. They are dominantly selected in events with more QCD background which means higher track multiplicities. The ANN-based OS kaon tagger was designed to cope with high track multiplicities, however, the influence of more tagging background tracks on the mistag probability can not be excluded. Also the fraction of events containing more PVs per event is higher in those events. The selection criteria for B_s^0 mesons typically also enhance the mean value of the selected B_s^0 p_T distribution compared to B^0 or B^+ selections. Hence, the effect on the mistag calibration is tested by splitting the sample into subsamples according to different track multiplicities, signal B p_T ranges and number of primary vertices per event.

Third, the running conditions may have changed over the full run period. The effect of different instantaneous luminosities may cause different predicted mistag probabilities. It is tested by splitting the sample into subsamples according to different run periods.

Charge and Detection Asymmetries

To test the influence of charge asymmetries on the calibration of the predicted mistag probability of the ANN-based OS kaon tagger, the $B^+ \to J/\psi K^+$ data samples are split according to the initial flavour of the B meson. 2011 and 2012 data is tested separately. The influence of possible detection asymmetries is tested by splitting the samples according to magnet polarities. Figure 3.38 shows the calibrations performed on each subsample. The fitted calibration parameters p_0 and p_1 are listed in table 3.32 for the B^+ B^- tests and in table 3.33 for the magnet polarity up and down tests. The subsamples are statistically independent. Hence, the difference between the subsamples in p_0 and p_1 is calculated divided by the combined uncertainty on the difference.

The mistag calibration p_0 offset parameter shows a significant deviation between the B^+ and B^- subsamples of 5 σ in 2011 data and 6.7σ in 2012 data. This difference has to be taken into account, when the ANN-based OS kaon tagger is applied to a *CP* asymmetry measurement.

For different magnet polarities no systematic effect on the mistag calibration is observed.

Different Event Topologies

The influence of different event topologies on the mistag calibration is tested by splitting the sample into subsamples according to different number of primary vertices per event, track multiplicities and signal B p_T ranges. To test the influence of the mean number of PVs per event the data set is split into three subsets: events that contain one reconstructed PV per event, events containing two PVs and events containing three or more PVs per event. The subsets according to different track multiplicities are chosen such that they contain roughly the same amount of events each. The first
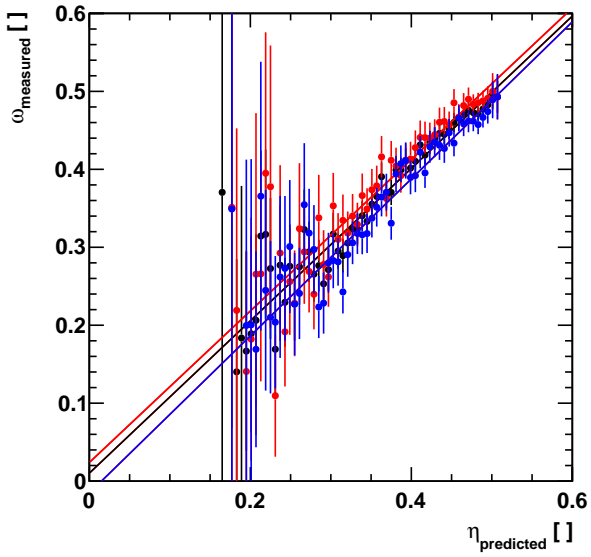
(a) initial flavour, 2011
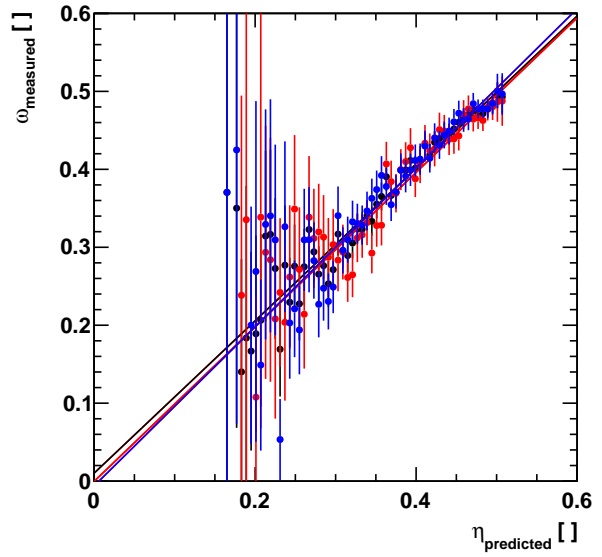
(b) magnet polarity, 2011
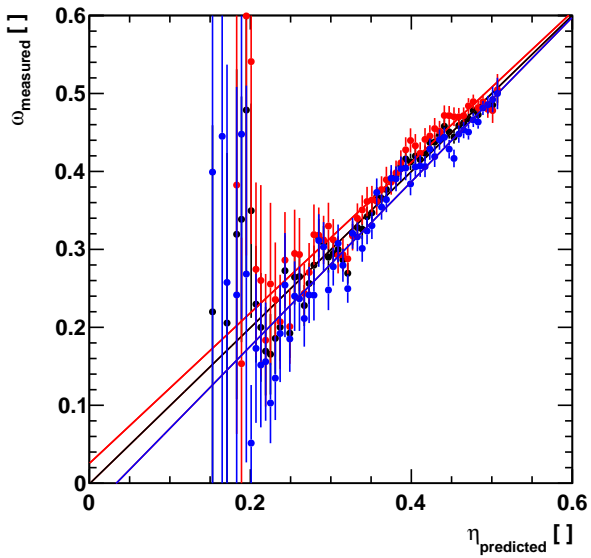
(c) initial flavour, 2012
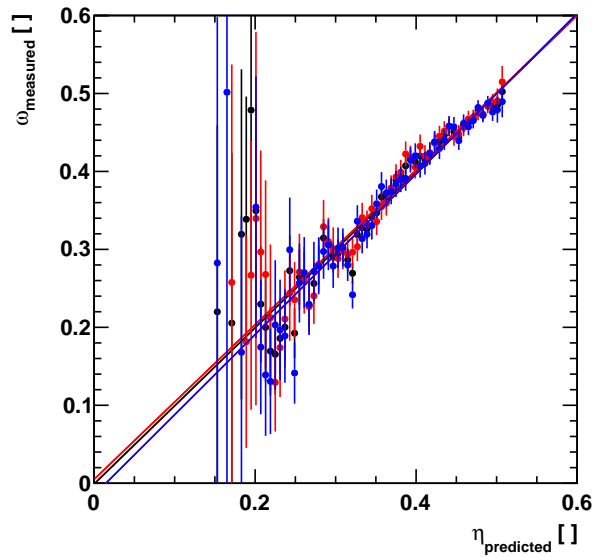
(d) magnet polarity, 2012

Figure 3.38.: Systematic effects on the mistag calibration (top 2011 data, bottom 2012 data) splitting the sample according to (a, c) the initial flavour of the signal B meson (red: B^+, blue: B^-), (b, d) the magnet polarity (red: up, blue: down). The calibration of the whole sample is shown in black.

sample	p_0	$\Delta p_0/\sigma(\Delta p_0)$	p_1	$\Delta p_1/\sigma(\Delta p_1)$
B^+	0.447 ± 0.003	–	0.972 ± 0.054	–
B^-	0.423 ± 0.003	–	1.007 ± 0.052	–
Δ cat. 1-2 ($\times 100$)	2.419 ± 0.475	5.1	3.545 ± 7.441	0.5
B^+	0.446 ± 0.002	–	0.965 ± 0.037	–
B^-	0.424 ± 0.002	–	1.055 ± 0.035	–
Δ cat. 1-2 ($\times 100$)	2.195 ± 0.327	6.7	8.972 ± 5.123	1.8

Table 3.32.: Systematic test on the mistag probability calibration for splitting the sample according to different initial flavours of the signal B meson using 2011 data (top) and 2012 data (bottom). The significance of the differences in p_0 and p_1 are calculated.

sample	p_0	$\Delta p_0/\sigma(\Delta p_0)$	p_1	$\Delta p_1/\sigma(\Delta p_1)$
magnet up	0.431 ± 0.004	–	0.991 ± 0.058	–
magnet down	0.438 ± 0.003	–	1.020 ± 0.048	–
Δ cat. 1-2 ($\times 100$)	0.679 ± 0.481	1.4	2.903 ± 7.509	0.4
magnet up	0.437 ± 0.002	–	0.992 ± 0.036	–
magnet down	0.433 ± 0.002	–	1.029 ± 0.037	–
Δ cat. 1-2 ($\times 100$)	0.427 ± 0.327	1.3	3.730 ± 5.132	0.7

Table 3.33.: Systematic test on the mistag probability calibration for splitting the sample according to different magnet polarities using 2011 data (top) and 2012 data (bottom). The significance of the differences in p_0 and p_1 are calculated.

of three categories contains events with track multiplicities smaller than 100, the second contains events with track multiplicities between 100 and 200, the third above 200 tracks per event. The calibration fits are shown in figure 3.39, separately for 2011 and 2012 data. The corresponding fit parameter values are given in table 3.34 for PVs per event splits and in table 3.35 for track multiplicity splits. In 2012 data a deviation of 3.9σ is observed in the mistag calibration offset parameter p_0, comparing events that contain 1 PV per event with events that contain 3 or more PVs per event. The different subsamples according to low and high track multiplicities yield in 2011 data a significant difference in the mistag calibration offset parameter p_0 of 5.5σ and in 2012 data of 9.2σ. However, this comparison of subsets probably overestimates the effect, because typically the data set where the mistag probability calibration is determined from is not that different in track multiplicities compared to the *CP* measurement data set. Though, the effect should be reevaluated for the given conditions when a systematic uncertainty is determined.

The data sets are split in three different sets according to the signal B p_T, containing roughly the same amount of events. In the first subsample the B p_T is smaller than $4\,\mathrm{GeV}$, in the second it is larger than $4\,\mathrm{GeV}$ but smaller than $8\,\mathrm{GeV}$ and in the third set it is larger than $8\,\mathrm{GeV}$. The
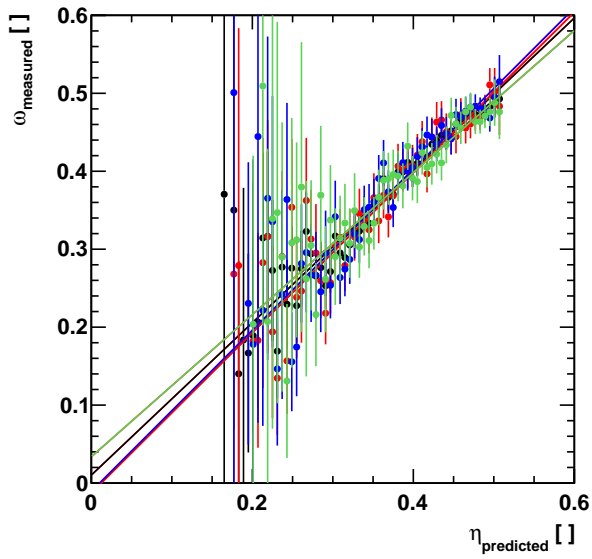
(a) number of vertices, 2011
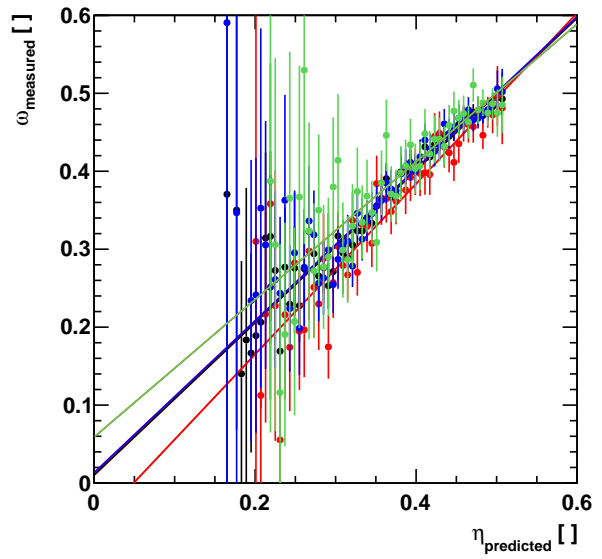
(b) track multiplicity, 2011
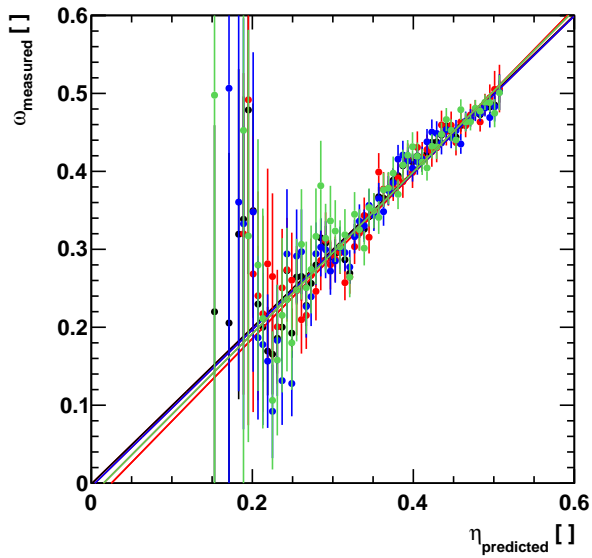
(c) number of vertices, 2012
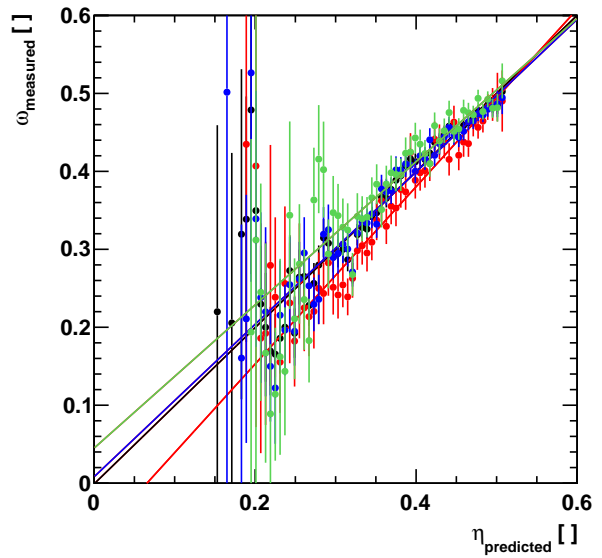
(d) track multiplicity, 2012

Figure 3.39.: Calibration systematics for (top 2011 data, bottom 2012 data) splitting the sample according to (a, c) the number of primary vertices in the event (red: #PVs = 1, blue: #PVs = 2, green: #PVs ≥ 3), (b, d) the track multiplicity of the event (red: #tracks < 100, blue: #tracks ∈ [100, 200], green: #tracks > 200). The calibration of the whole sample is shown in black.

sample	p_0	$\Delta p_0/\sigma(\Delta p_0)$	p_1	$\Delta p_1/\sigma(\Delta p_1)$
#vertices $= 1$	0.429 ± 0.004	–	1.028 ± 0.061	–
#vertices $= 2$	0.439 ± 0.004	–	1.033 ± 0.060	–
#vertices ≥ 3	0.437 ± 0.004	–	0.911 ± 0.076	–
Δ cat. 1-2 ($\times 100$)	1.023 ± 0.568	1.8	0.541 ± 8.591	0.1
Δ cat. 1-3 ($\times 100$)	0.820 ± 0.602	1.4	11.657 ± 9.756	1.2
Δ cat. 2-3 ($\times 100$)	0.203 ± 0.582	0.3	12.198 ± 9.671	1.3
#vertices $= 1$	0.427 ± 0.003	–	1.059 ± 0.047	–
#vertices $= 2$	0.432 ± 0.003	–	1.006 ± 0.041	–
#vertices ≥ 3	0.443 ± 0.003	–	1.038 ± 0.046	–
Δ cat. 1-2 ($\times 100$)	0.554 ± 0.413	1.3	5.296 ± 6.207	0.9
Δ cat. 1-3 ($\times 100$)	1.620 ± 0.419	3.9	2.156 ± 6.565	0.3
Δ cat. 2-3 ($\times 100$)	1.066 ± 0.382	2.8	3.141 ± 6.179	0.5

Table 3.34.: Systematic test on the mistag probability calibration for splitting the sample according to different number of primary vertices per event using 2011 data (top) and 2012 data (bottom). The significance of the differences in p_0 and p_1 are calculated.

sample	p_0	$\Delta p_0/\sigma(\Delta p_0)$	p_1	$\Delta p_1/\sigma(\Delta p_1)$
#tracks < 100	0.414 ± 0.005	–	1.094 ± 0.066	–
#tracks $\in [100, 200]$	0.437 ± 0.003	–	0.975 ± 0.054	–
#tracks > 200	0.451 ± 0.005	–	0.883 ± 0.086	–
Δ cat. 1-2 ($\times 100$)	2.363 ± 0.571	4.1	11.901 ± 8.506	1.4
Δ cat. 1-3 ($\times 100$)	3.696 ± 0.666	5.5	21.068 ± 10.810	1.9
Δ cat. 2-3 ($\times 100$)	1.332 ± 0.592	2.2	9.167 ± 10.121	0.9
#tracks < 100	0.409 ± 0.004	–	1.137 ± 0.053	–
#tracks $\in [100, 200]$	0.432 ± 0.002	–	0.978 ± 0.037	–
#tracks > 200	0.452 ± 0.003	–	0.919 ± 0.050	–
Δ cat. 1-2 ($\times 100$)	2.300 ± 0.436	5.3	15.963 ± 6.471	2.5
Δ cat. 1-3 ($\times 100$)	4.270 ± 0.463	9.2	21.864 ± 7.257	3.0
Δ cat. 2-3 ($\times 100$)	1.970 ± 0.373	5.3	5.901 ± 6.202	1.0

Table 3.35.: Systematic test on the mistag probability calibration for splitting the sample according to different different track multiplicities per event using 2011 data (top) and 2012 data (bottom). The significance of the differences in p_0 and p_1 are calculated.

mistag calibration fits are shown in figure 3.40. The corresponding fit parameter values are given in table 3.36. No significant deviation in p_0 or p_1 between the subsamples is observed.

(a) signal B p_T, 2011

(b) signal B p_T, 2012
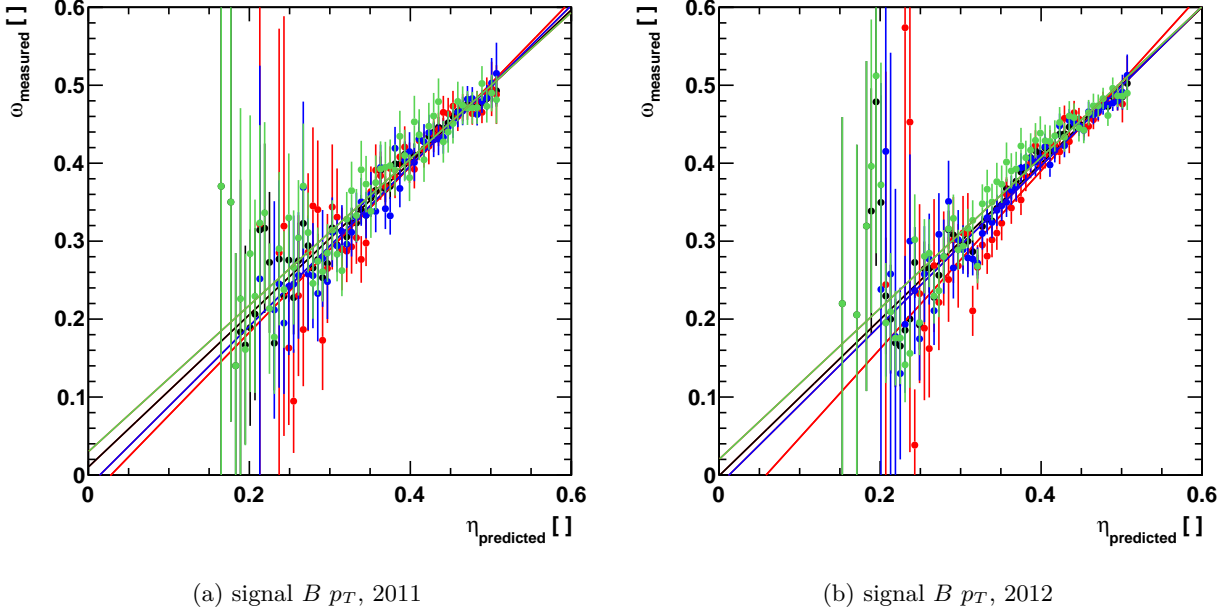
Figure 3.40.: Calibration systematics for ((a) 2011 data, (b) 2012 data) splitting the sample according to the p_T of the signal B meson (red: $p_T < 4\,\text{GeV}$, blue: $p_T \in [4, 8]\,\text{GeV}$, green: $p_T > 8\,\text{GeV}$). The calibration of the whole sample is shown in black.

Different Running Conditions

The influence of different running conditions on the predicted mistag probability calibration of the ANN-based OS kaon tagger is tested by splitting the sample in subsamples according to run periods. In LHCb every physics run is assigned an integer continuous run number. Hence, different run periods are encoded by this run number. 2011 data is split in three run periods containing roughly the same amount of events each: run numbers smaller than 9.5×10^4, run numbers between 9.5×10^4 and 1.01×10^5 and run numbers above 1.01×10^5. 2012 data is split in three run periods, accordingly: run numbers smaller than 1.19×10^5, run numbers between 1.19×10^5 and 1.29×10^5 and run numbers above 1.29×10^5. The calibration fits are shown in figure 3.41. The corresponding fit parameter values are given in table 3.37. No significant change in the fitted calibration parameters due to different run periods is found.

Summary

The influence of charge asymmetries on the predicted mistag probability calibration is significant. It is recommended to calibrate the ANN-based OS kaon tagger separately on B and $\overline{B}$ events

sample	p_0	$\Delta p_0/\sigma(\Delta p_0)$	p_1	$\Delta p_1/\sigma(\Delta p_1)$
$Bp_t < 4\,\text{GeV}$	0.440 ± 0.004	–	1.064 ± 0.070	–
$Bp_t \in [4,8]\,\text{GeV}$	0.434 ± 0.004	–	1.028 ± 0.062	–
$Bp_t > 8\,\text{GeV}$	0.429 ± 0.005	–	0.939 ± 0.063	–
Δ cat. 1-2 ($\times 100$)	0.609 ± 0.550	1.1	3.562 ± 9.411	0.4
Δ cat. 1-3 ($\times 100$)	1.030 ± 0.615	1.7	12.475 ± 9.422	1.3
Δ cat. 2-3 ($\times 100$)	0.421 ± 0.606	0.7	8.913 ± 8.844	1.0
$Bp_t < 4\,\text{GeV}$	0.438 ± 0.003	–	1.142 ± 0.051	–
$Bp_t \in [4,8]\,\text{GeV}$	0.433 ± 0.003	–	1.023 ± 0.043	–
$Bp_t > 8\,\text{GeV}$	0.433 ± 0.003	–	0.968 ± 0.042	–
Δ cat. 1-2 ($\times 100$)	0.495 ± 0.384	1.3	11.931 ± 6.639	1.8
Δ cat. 1-3 ($\times 100$)	0.443 ± 0.422	1.0	17.431 ± 6.584	2.6
Δ cat. 2-3 ($\times 100$)	0.052 ± 0.409	0.1	5.499 ± 5.987	0.9

Table 3.36.: Systematic test on the mistag probability calibration for splitting the sample according to different different signal B p_T ranges using 2011 data (top) and 2012 data (bottom). The significance of the differences in p_0 and p_1 are calculated.

(a) run number, 2011
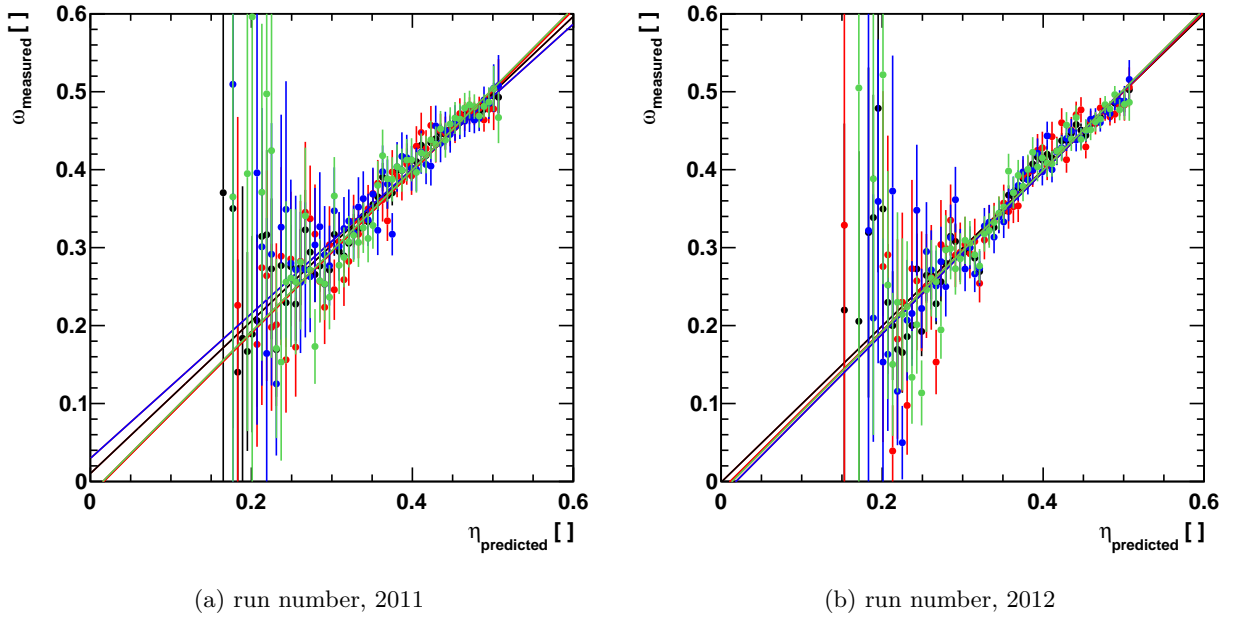
(b) run number, 2012

Figure 3.41.: Calibration systematics for (left 2011 data, right 2012 data) splitting the sample according to the run number, representing the time the data was taken (red: run# $< 9.5 \times 10^4$, blue: run# $\in [9.5 \times 10^4, 1.01 \times 10^5]$, green: run# $> 1.01 \times 10^5$), red: run# $< 1.19 \times 10^5$, blue: run# $\in [1.19 \times 10^5, 1.29 \times 10^5]$, green: run# $> 1.29 \times 10^5$, respectively for 2012 data. The calibration of the whole sample is shown in black.

sample	p_0	$\Delta p_0/\sigma(\Delta p_0)$	p_1	$\Delta p_1/\sigma(\Delta p_1)$
run # $< 9.5 \times 10^4$	0.433 ± 0.004	–	1.037 ± 0.060	–
run # $\in [9.5 \times 10^4, 1.01 \times 10^5]$	0.434 ± 0.004	–	0.929 ± 0.071	–
run # $> 1.01 \times 10^5$	0.437 ± 0.004	–	1.039 ± 0.064	–
Δ cat. 1-2 ($\times 100$)	0.071 ± 0.588	0.1	10.816 ± 9.268	1.2
Δ cat. 1-3 ($\times 100$)	0.362 ± 0.564	0.6	0.245 ± 8.737	0.0
Δ cat. 2-3 ($\times 100$)	0.291 ± 0.598	0.5	11.061 ± 9.521	1.2
run # $< 1.19 \times 10^5$	0.436 ± 0.003	–	1.023 ± 0.047	–
run # $\in [1.19 \times 10^5, 1.29 \times 10^5]$	0.433 ± 0.003	–	1.041 ± 0.042	–
run # $> 1.29 \times 10^5$	0.435 ± 0.003	–	1.033 ± 0.043	–
Δ cat. 1-2 ($\times 100$)	0.247 ± 0.407	0.6	1.745 ± 6.357	0.3
Δ cat. 1-3 ($\times 100$)	0.076 ± 0.408	0.2	0.978 ± 6.383	0.2
Δ cat. 2-3 ($\times 100$)	0.171 ± 0.390	0.4	0.767 ± 6.028	0.1

Table 3.37.: Systematic test on the mistag probability calibration for splitting the sample according to different different run numbers using 2011 data (top) and 2012 data (bottom). The significance of the differences in p_0 and p_1 are calculated.

before using it in *CP* asymmetry measurements. The influence of different track multiplicities and number of PVs per event on the calibration, shown in this study, is overestimated. It is recommended to study this effect in detail with the given realistic track multiplicity and PV multiplicity settings for a *CP* analysis. All other effects are found to be negligible.

3.7. Summary

A novel OS kaon tagging algorithm has been developed in this thesis. It is based on machine learning techniques. Two feed-forward artificial neural networks (ANN) have been chosen, where the first ANN serves as input for the second ANN. The first ANN is trained on MC simulation to identify kaons from opposite *b* hadron decays. The response of the first ANN, determined per tagging candidate, is used in two ways. A minimum cut is applied to suppress soft QCD background. Additionally, the ANN response is applied as weight to each tagging candidate. Up to three best-ranked tagging candidates, according to the response of the first ANN, are combined in the second ANN, which predicts the tagging decision accompanied by the mistag probability. It is trained on data.

The tagger was first optimized on $B^+ \to J/\psi K^+$ decays, only. However, fragmentation kaons present in the production of B_s^0 mesons, were selected by this optimization, once the tagger is applied to B_s^0 mesons. Hence, the first ANN was re-trained on a sample of $B_s^0 \to J/\psi \phi$ decays to remove this background.

Using ANNs, the tagging power of the OS kaon tagger improves by 33% using B_s^0 optimization and 50% using B^+ optimization. This reduces the statistical uncertainty of mixing asymmetry measurements by 13% and 19%, respectively.

After the re-optimization on the sample of $B_s^0 \to J/\psi\,\phi$ decays the portability to different B decay channels, $B^0 \to J/\psi\,K^{*0}$ or $B_s^0 \to J/\psi\,\phi$, is guaranteed. After combination with the OS vertex charge tagger the absolute improvement in tagging power of 0.3% is reduced to 0.2%. The tagging power of the full OS tagger combination, built of the ANN-based OS kaon tagger, the OS vertex charge tagger and the OS muon and electron taggers, is improved by absolute 0.2%, which corresponds to a relative improvement of 8% compared to the OS combination including the cut-based OS kaon tagger. A relative improvement of 8% in tagging power translates into a 4% improved statistical uncertainty on the mixing asymmetry.

The predicted mistag probability is calibrated against the measured mistag probability in $B^+ \to J/\psi\,K^+$ data. Systematic studies on the calibration showed, that the predicted mistag probability is affected by initial flavour asymmetries. It is recommended to calibrate the tagger separately for B and $\bar{B}$ decays before using it in *CP* asymmetry measurements.

4. The B^0–$\overline{B}^0$ Oscillation Frequency Δm_d

B^0–$\overline{B}^0$ mixing was discovered by the ARGUS collaboration in 1987 [23]. At the DORIS[1] II e^+e^- storage ring, $B^0\overline{B}^0$ pairs were produced from $\Upsilon(4S)$ decays in a coherent P-wave state. Hence, B^0B^0 or $\overline{B}^0\overline{B}^0$ events could only be observed, if one of the $B^0/\overline{B}^0$ mesons had undergone mixing. The mixing strength

$$r = \frac{\Gamma(B^0 \to \overline{B}^0 \to X')}{\Gamma(B^0 \to X)} = \frac{N(B^0B^0) + N(\overline{B}^0\overline{B}^0)}{N(B^0\overline{B}^0)} \tag{4.1}$$

was measured using three analysis methods. First, ARGUS searched for fully reconstructed $\Upsilon(4S) \to B^0B^0$ or $\overline{B}^0\overline{B}^0$ decays. Within 88×10^3 $\Upsilon(4S)$ decays, one $\Upsilon(4S) \to B^0B^0$ decay was fully reconstructed, see figure 4.1, compatible with the expected occurrence at a mixing strength of $r = 0.2$.
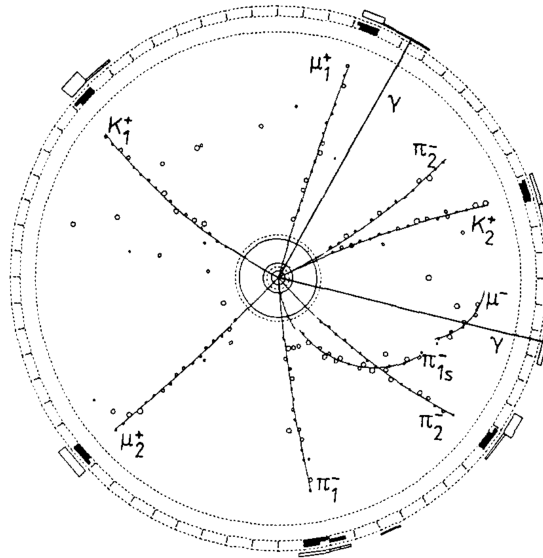
Figure 4.1.: Fully reconstructed $\Upsilon(4S) \to B^0B^0$ event in the ARGUS detector, part of the discovery of B^0–$\overline{B}^0$ mixing. One B^0 decays into $B_1^0 \to D_1^{*-}\mu_1^+\nu_1$, with $D_1^{*-} \to \pi_{1s}^-\overline{D}^0$ and $\overline{D}^0 \to K_1^+\pi_1^-$. The other B^0 decays into $B_2^0 \to D_2^{*-}\mu_2^+\nu_2$, with $D_2^{*-} \to \pi^0 D^-$, where $\pi^0 \to \gamma\gamma$ and $D^- \to K_2^+\pi_2^-\pi_2^-$. Taken from [23]

Second, lepton pairs originating from $\Upsilon(4S)$ decays were reconstructed. In this case, the charge of the lepton from the b decay identifies the decay of a B or $\overline{B}$ meson. Hence, like-sign lepton

---

[1]Doppel-Ring-Speicher, located at Deutsches Elektronen-Synchrotron (DESY), Hamburg.

pairs occur only in case of B^0–$\bar{B}^0$ mixing. Third, one B^0 from $\Upsilon(4S)$ was reconstructed, while the second B^0 was tagged with a high momentum lepton.

The B^0–$\bar{B}^0$ mixing strength was measured by combining the observation of one fully reconstructed $\Upsilon(4S) \to B^0 B^0$ event, 24.8 like-sign lepton pairs (at 4.0σ) and 4.1 tagged $B^0/\bar{B}^0$ events (at 3.0σ), resulting in

$$r = 0.21 \pm 0.08 \quad \text{or} \quad \chi = \frac{r}{1+r} = 0.17 \pm 0.05. \tag{4.2}$$

The B^0–$\bar{B}^0$ mixing discovery year 1987 was the breakthrough year of the B factory idea, i.e. $e^+ e^-$ annihilation is the cleanest way to study CP violation in B meson decays. The center-of-mass collision energy is set to the $\Upsilon(4S)$ mass resonance ($10.6\,\text{GeV}$) to produce large amounts of $B\bar{B}$ pairs[2]. In order to precisely determine the B decay time the center-of-mass frame of the $\Upsilon(4S)$ is Lorentz boosted. Therefore, the two beams are operated at asymmetric energies, one high-energy electron beam at $8 - 9\,\text{GeV}$ and one low-energy positron beam at $3.1 - 3.5\,\text{GeV}$. The most precise measurements of Δm_d today were performed at B factories by BaBar and Belle and at the LHC by LHCb. These three measurements are sketched in the following.

BaBar reconstructed 50×10^3 $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays[3] using a data sample of 88×10^6 $B\bar{B}$ events collected at the $\Upsilon(4S)$ resonance at the PEP-II asymmetric-energy $e^+ e^-$ storage ring at SLAC[4]. The flavour of the other B meson is determined from the charge of another high momentum lepton. A result of $\Delta m_d = (0.511 \pm 0.007\,(\text{stat})^{+0.007}_{-0.006}\,(\text{syst}))\,\text{ps}^{-1}$ is obtained [91].

Belle reconstructed 177×10^3 $B^0 \to D^{*-}\ell^+\nu_\ell$ and $B^0 \to D^-\pi^+$, $D^{*-}\pi^+$, $D^{*-}\rho^+$, $J/\psi K^{*0}$ decays using a data sample of 152×10^6 $B\bar{B}$ pairs collected at the $\Upsilon(4S)$ resonance at the asymmetric-energy $e^+ e^-$ KEKB collider at KEK[5]. To identify the flavour of the other B meson charged leptons, pions, kaons and Λ baryons are used. A result of $\Delta m_d = (0.511 \pm 0.005\,(\text{stat}) \pm 0.006\,(\text{syst}))\,\text{ps}^{-1}$ is obtained [92].

LHCb reconstructed 88×10^3 $B^0 \to D^-\pi^+$ decays and 39×10^3 $B^0 \to J/\psi K^{*0}$ decays using a data set, which corresponds to an integrated luminosity of $1\,\text{fb}^{-1}$ of proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$. With a production cross-section of $\sigma(pp \to B^0 + X) = 38.1 \pm 0.6\,(\text{stat}) \pm 3.7\,(\text{syst}) \pm 4.7(\text{norm.})\mu\text{b}$ [36] at $\sqrt{s} = 7\,\text{TeV}$ within the LHCb acceptance, this corresponds to 38×10^9 produced B^0 mesons. The production flavour of the B^0 is determined by the opposite side and the same side pion tagging algorithms. A result of $\Delta m_d = (0.5156 \pm 0.0051\,(\text{stat}) \pm 0.0033\,(\text{syst}))\,\text{ps}^{-1}$ is obtained [93]. Although reconstruction and tagging efficiencies of B mesons are significantly lower, the LHC serves as a competing "hadronic B factory" due to the high $b\bar{b}$ production rate.

From the first observation of B^0–$\bar{B}^0$ mixing by ARGUS until today, various experiments have measured the time-dependent oscillation frequency Δm_d, see figure 4.2. The Heavy Flavour Averaging Group (HFAG) combines all Δm_d measurements to obtain a world average value. The latest value, as of Fall 2014, is [22]

$$\Delta m_d = 0.510 \pm 0.003\,\text{ps}^{-1}. \tag{4.3}$$

---

[2]The fraction of $\Upsilon(4S)$ decaying into $B\bar{B}$ pairs is larger than 96% [22].

[3]Charge conjugation of particles is implied unless explicitly stated.

[4]Stanford Linear Accelerator Center, California, USA

[5]High Energy Accelerator Research Organization, Tsukuba, Ibaraki Prefecture, Japan

ALEPH (3 analyses) — $0.446 \pm 0.026 \pm 0.019$ ps^{-1}

DELPHI* (5 analyses) — $0.519 \pm 0.018 \pm 0.011$ ps^{-1}

L3 (3 analyses) — $0.444 \pm 0.028 \pm 0.028$ ps^{-1}

OPAL (5 analyses) — $0.479 \pm 0.018 \pm 0.015$ ps^{-1}

CDF1* (4 analyses) — $0.495 \pm 0.033 \pm 0.027$ ps^{-1}

D0 (1 analysis) — $0.506 \pm 0.020 \pm 0.016$ ps^{-1}

BABAR* (4 analyses) — $0.506 \pm 0.006 \pm 0.004$ ps^{-1}

BELLE* (3 analyses) — $0.509 \pm 0.004 \pm 0.005$ ps^{-1}

LHCb* (3 analyses) — $0.514 \pm 0.005 \pm 0.003$ ps^{-1}

Average of above after adjustments — 0.510 ± 0.003 ps^{-1}

CLEO+ARGUS (χ_d measurements) — 0.498 ± 0.032 ps^{-1}

World average Fall 2014 — 0.510 ± 0.003 ps^{-1}

Δm_d (ps^{-1})
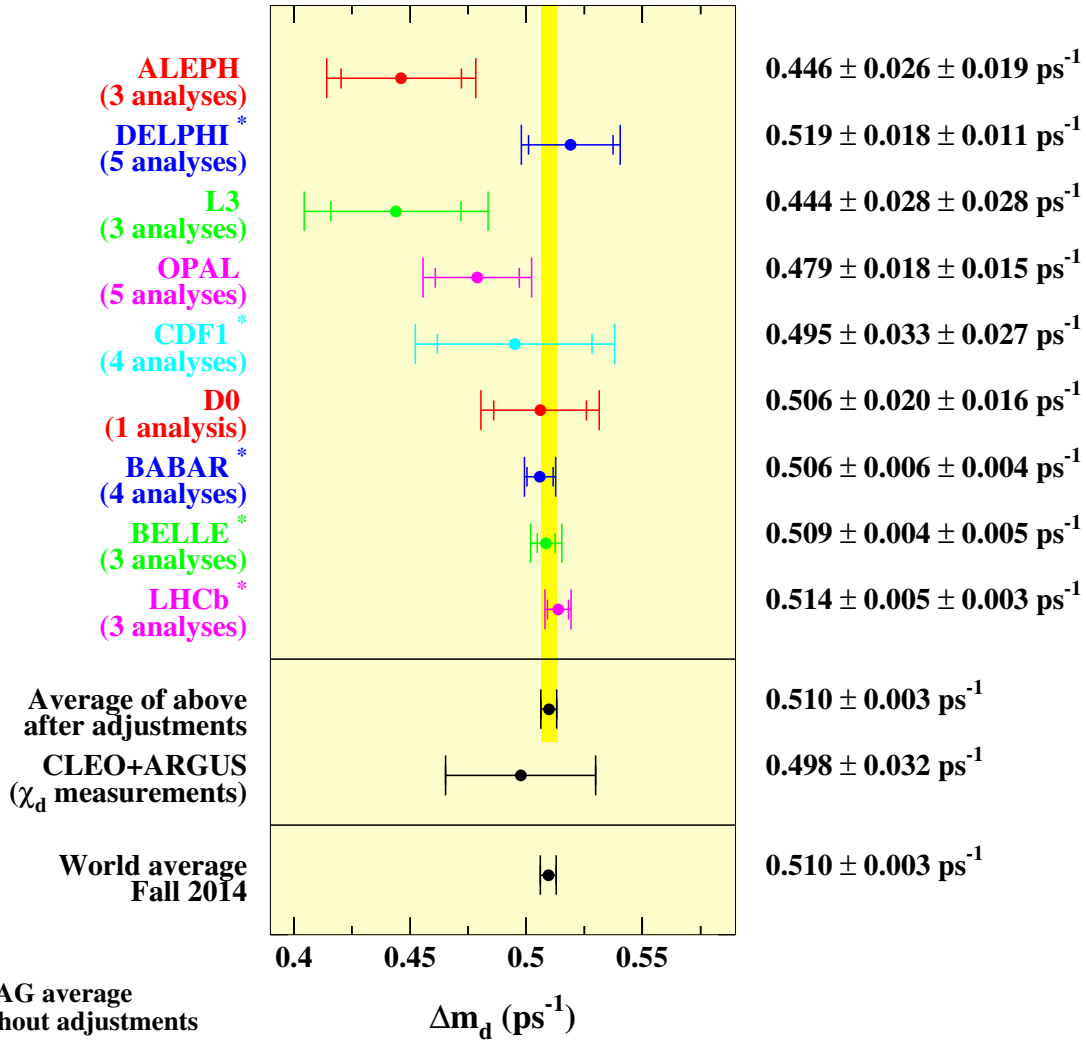
* HFAG average without adjustments

Figure 4.2.: HFAG Δm_d world average. The combination of measurements from almost 30 years of experimental determination results in a very precise Δm_d value, which has a relative uncertainty of 6‰. Taken from [22]

Hence, Δm_d is currently known with a relative experimental precision of 6‰. Now, what is the purpose of measuring Δm_d at LHCb more precisely? The experimental precision on *CP* asymmetries in semileptonic *B* decays offers room for New Physics contributions, see chapter 1. LHCb is on its way to measure these asymmetries precisely. The level of accuracy needed to understand these processes can be demonstrated by a very precise determination of well established processes like B^0–$\overline{B}^0$ mixing.

Four main ingredients are needed to measure Δm_d precisely: the B^0 production flavour, the B^0 decay flavour, the B^0 decay time, and a large number of reconstructed B^0 mesons. A large data set of semileptonic B^0 decays, $\mathcal{O}(10^6)$, is analysed, collected at $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ by the LHCb experiment. Semileptonic decays are flavour specific, thus, the decay flavour of the B^0 is given by the charge of the final state lepton. The production flavour is predicted by flavour tagging algorithms, see chapter 3.

The goal of the measurement is a precision comparable or even better than the world average. This can be achieved by exploiting the advantages of the LHC. Due to the large $b\bar{b}$ cross section, a high statistics data set of semileptonic B^0 decays is recorded, which is efficiently triggered by the final state lepton. An excellent B^0 decay time resolution is obtained, thanks to the large boost of the $b\bar{b}$ pair in proton-proton collisions and to the vertex detector of LHCb. A well performing particle identification is obtained, due to the ring imaging Cherenkov detectors of LHCb.

On the other hand, certain disadvantages have to be tackled at a hadron collider. Compared to e^+e^- annihilation at a *B* factory, proton-proton collisions at the LHC produce orders of magnitudes higher amounts of background, i.e. a large amount of low QCD background is produced in the primary interaction. In addition, multiple primary interactions per event are common. Cuts on the signal *B* decay length allow a clean signal reconstruction. However, flavour tagging algorithms are affected by large track multiplicities per event. The tagging power is drastically reduced, hence, optimized tagging algorithms and large statistics are needed to obtain the statistical power needed for a precision measurement of Δm_d.

To face this problem a novel tagging algorithm has been developed utilizing Artificial Neural Networks (ANN), that manages to deal with high track multiplicities. This ANN-based OS kaon tagger combined with other tagging algorithms is used in the precision measurement of Δm_d. The world average precision can only be outplayed, when every part of the analysis plays perfectly together.

The following sections document the approach and results in detail. In section 4.1 the analysis is introduced. The decay channels are shown, the experimental approach is explained and the different B^0–$\overline{B}^0$ oscillation damping factors are illustrated. Section 4.2 presents the analysed data set, the trigger and selection requirements. Section 4.3 is devoted to MC simulation. Simulation plays a major role in the measurement of Δm_d, because the B^0 is only partially reconstructed due to the missing neutrino. A simulation-driven method, the so called k factor method, first introduced by the CDF collaboration [94], is used to correct the B^0 momentum for the momentum carried away by the neutrino. Section 4.4 shows the choice of flavour tagging algorithms used to predict the production flavour of the B^0. Section 4.5 presents all relevant backgrounds that have to be considered for this analysis.

Section 4.6 introduces the Probability Density Function ($\mathcal{PDF}$) to fit Δm_d. Section 4.7 shows the decay time acceptance function and the measured decay time resolution, which is composed

of the flight distance resolution and the k factor resolution. The k factor method plays a key role in the analysis. The accuracy of the initial k factor correction is not sufficient for the aimed precision on Δm_d. Hence, an improved k factor method is developed. Section 4.8 demonstrates the validity of the fit procedure using MC simulation. Section 4.9 shows toy experiment studies to evaluate the systematic uncertainties. Finally, the result on Δm_d and the measured tagging performance are given in section 4.10.

The oscillation frequency is measured on four data sets: two semileptonic decay channels, each reconstructed in data sets of two data taking periods, 2011 at $\sqrt{s} = 7\,\text{TeV}$ and 2012 at $\sqrt{s} = 8\,\text{TeV}$. These four data sets are fitted separately, first. Second, the data sets are fitted simultaneously, floating the specific parameters per data set but fitting for one combined Δm_d value.

4.1. Measurement of Δm_d

The author's work presented here has also been made available LHCb internally [3] and is currently in the LHCb-wide review process for publication. The collaboration organizes analysis working groups, where people from different institutes and countries collaborate to set up a valuable, stable and robust physics analysis. The author has contributed to the semileptonic working group which has a subdivision devoted to the measurement of Δm_d. In this subgroup the main analysis strategy decisions were made, however the real analysis implementation was carried out by single persons separately. The choice was made to develop, maintain and use two different fitter frameworks, one in Heidelberg presented in this thesis and one in Milan, Italy. The Heidelberg fitter is based on the implementation of [95] used to measure the *CP* violating phase ϕ_s. It has been adapted since then for the measurement of the B_s^0–$\overline{B}_s^0$ oscillation frequency Δm_s [96] and the semileptonic *CP* asymmetry a_{sl}^d [1] within the Heidelberg LHCb group. This thesis shows a precision measurement of Δm_d carried out by the author, based on the Heidelberg fitter framework. When collaborative work inputs are referred to, they are explicitly marked.

4.1.1. Semileptonic Decay Channels

The knowledge of the decay flavour of the B meson is a crucial ingredient for the measurement of Δm_d. Here, two flavour specific semileptonic decay channels of B^0 mesons are chosen:

$$B^0 \to D^- \mu^+ \nu_\mu X \quad (\text{with} \quad D^- \to K^+ \pi^- \pi^-) \tag{4.4}$$

$$B^0 \to D^{*-} \mu^+ \nu_\mu X \quad (\text{with} \quad D^{*-} \to \overline{D}^0 \pi^- \quad \text{and} \quad \overline{D}^0 \to K^+ \pi^-) \tag{4.5}$$

X denotes an arbitrary number of additional non-reconstructed hadronic particles. The first decay channel, equation 4.4, is later on referred to as $B^0 \to D^- \mu^+ \nu_\mu X$ mode. The reconstructed decay topology is sketched in figure 4.3. The D^- decays into three hadronic final state particles. It is fully reconstructed, while the B^0 which decays into the D^- meson and a muon is partially reconstructed due to the non-reconstructible neutrino.

The second decay channel, equation 4.5, is later on referred to as $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode. The reconstructed decay topology is sketched in figure 4.4. Both D decays, the D^{*-} and the $\overline{D}^0$ decay, are fully reconstructed. The $\overline{D}^0$ decays into two hadronic final state particles. The D^{*-} decays
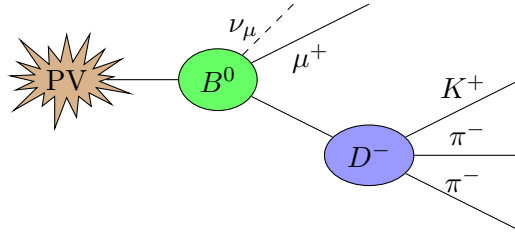
Figure 4.3.: Topology of the reconstructed $B^0 \to D^- \mu^+ \nu_\mu X$ (with $D^- \to K^+ \pi^- \pi^-$) decay. The ν_μ is not reconstructed. PV denotes the primary interaction vertex.

instantaneously into the $\bar{D}^0$ meson and a low-energetic pion. The B^0 decay into the D^{*-} and a muon is partially reconstructed.
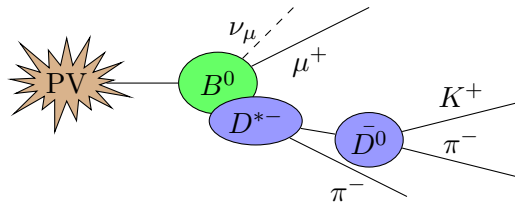

Figure 4.4.: Topology of the reconstructed $B^0 \to D^{*-} \mu^+ \nu_\mu X$ (with $D^{*-} \to \bar{D}^0 \pi^-$ and $\bar{D}^0 \to K^+ \pi^-$) decay. The ν_μ is not reconstructed. PV denotes the primary interaction vertex.

The charge of the muon tags the b quark content of the B^0 meson at decay. With the choice of using semileptonic decays both advantages and disadvantages come. The advantages are high branching ratios, $\mathcal{B}(B^0 \to D\ell^+\nu_\ell \text{ anything}) = (9.2 \pm 0.8)\%$ [18] and a clean trigger signature of the decay due to the muon, resulting in a high statistics data set. The hadronic background in the subsequent D decay can easily be removed by D mass fits. The major disadvantage of this choice is the non-reconstructible neutrino. Since the LHCb detector is a forward spectrometer, it is not possible to reconstruct missing energy carried away by the neutrino. The effect on the reconstructed B^0 decay time is corrected by the k factor method. However, the reconstructed B mass is also affected. Therefore, it is not possible to disentangle semileptonic feed down modes, which are higher D resonance decays, using a fit to the reconstructed B mass. Hence, also the sample composition needs to be understood, which relies mainly on simulation.

4.1.2. Analysis Overview

The oscillation frequency Δm_d is determined from a fit to the time-dependent mixing asymmetry distribution, as introduced in chapter 1:

$$\mathcal{A}(t) = \frac{N(t)_{\text{unmixed}} - N(t)_{\text{mixed}}}{N(t)_{\text{unmixed}} + N(t)_{\text{mixed}}} \simeq \cos(\Delta m_d t)$$

Several steps are needed to derive this distribution. An illustration of these steps is shown in figure 4.5 using a flow chart.
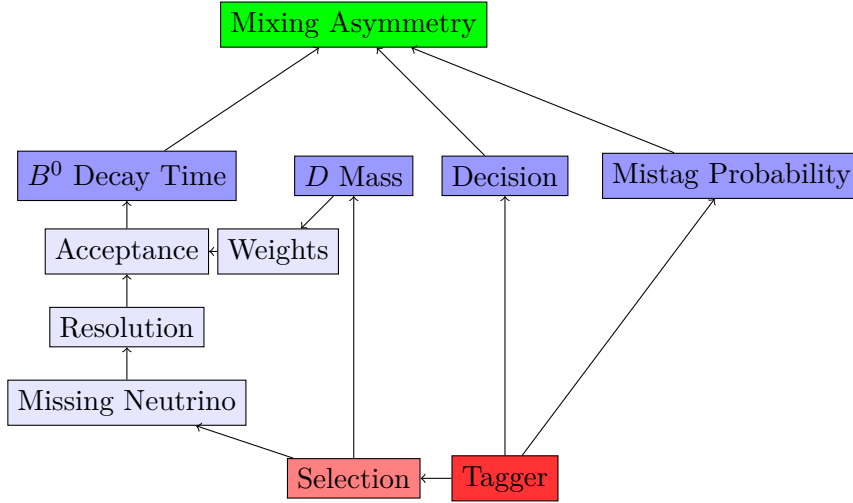
Figure 4.5.: Flow chart to illustrate the measurement of the oscillation frequency Δm_d extracted from the mixing asymmetry (from bottom to top). The tagging efficiency and selection requirements define the data set. The missing momentum of the neutrino and the decay time resolution and acceptance shape the B^0 decay time distribution. The D mass resonance is used to statistically remove hadronic background via weights. The tagging decision and mistag probability construct the mixing asymmetry out of the decay time distribution.

The mixing asymmetry depends on the B^0 decay time. Therefore, two B^0 decay time distributions are determined separately for two classes of reconstructed B^0 decays. One class contains B^0 decays, whose tagging decision equals the final state muon charge. These are unmixed decays, the production flavour equals the decay flavour, thus the B^0 meson did not oscillate. The other class contains B^0 decays, whose tagging decision is opposite to the final state muon charge. These are mixed decays, the production flavour is opposite to the decay flavour, thus the B^0 meson did oscillate. The affiliation into the two classes is distorted by the mistag probability, which damps the mixing asymmetry.

The tagging algorithm has a tagging efficiency, which is smaller than 100%. Hence, the data set used for the measurement is defined by the tagging efficiency and the additional selection requirements. The missing neutrino is the driving factor for the decay time determination. The B^0 decay time cannot be measured directly, because of its proportionality to the B^0 momentum, which is partially reconstructed here. The simulation-driven k factor method is used to correct the reconstructed B^0 decay time. Subsequently, the decay time resolution and acceptance functions are determined. The D mass resonances allow to separate between signal and mis-reconstructed background. They are fitted to statistically remove the background from the data set using sWeights [79, 78]. These sWeights are applied to the B^0 decay time distribution.

The measurement of Δm_d is developed using the blind analysis technique. A random offset and scale is applied to the Δm_d fit result, obtained by fitting data, until the analysis is finalized in

order not to bias the final result. The novel ANN-based OS kaon tagging algorithm is used in combination with other tagging algorithms to obtain the maximum tagging power.

4.1.3. Experimental approach

The decay time of the B^0 meson is reconstructed using the measured flight distance L and the momentum p of the B:

$$t = \frac{M(\text{PDG})L}{pc} \tag{4.6}$$

with $M(PDG)$ the mass of the B, taken from the PDG [18] and c the speed of light. However, the measured momentum of the B, which is nothing but the sum of the momenta of the visible daughter particles of the B, is underestimated due to the non-reconstructible neutrino. There are different experimental ways to cope with the missing momentum. Either, one may use the knowledge of the primary and secondary vertex positions to correct for the missing neutrino momentum up to a quadratic ambiguity [97]. On the other hand, one may follow the CDF approach and correct for the average momentum loss statistically based on simulation. Here, the so called k factor is defined as the ratio between the reconstructed and the true B momentum:

$$k = \frac{p_{\text{rec}}}{p_{\text{true}}}. \tag{4.7}$$

The k factor is folded into the description of the decay time distribution, described in section 4.7. A crucial point of this approach is to guarantee, that the simulation reflects properly the conditions present in the data set. Therefore, extensive studies on the sample composition, selection efficiencies and kinematics of simulation and data are performed. The uncertainties on these parts are treated as inputs for the evaluation of systematic uncertainties on Δm_d.

4.1.4. Damping factors

The measured mixing asymmetry is damped by several factors:

$$\mathcal{A}_{\text{meas}}(t) \simeq \mathcal{D}_k(t)\mathcal{D}_{\sigma_t}\mathcal{D}_{\text{tag}}\cos(\Delta m_d t). \tag{4.8}$$

First, the k factor, as explained above, enters directly the determination of the decay time. Its damping on the mixing asymmetry is time-dependent in contrast to the other damping factors. Second, the flight distance resolution enters the decay time distribution at small values close to zero. It is closely related to the decay time acceptance distribution, which carries a selection specific shape, and shows the inefficiency in the capability of reconstructing very small flight distances. Third and most important here, the tagging dilution damps the amplitude of the mixing asymmetry significantly.

To illustrate the different effects and to get an idea about the orders of magnitudes of the damping factors, a toy study is performed. A set of 1.6×10^7 toy events is generated and fitted under different measurement conditions. The generated data set is ten times larger than the largest real data set. Figure 4.6 compares two cases: realistic conditions and idealistic conditions for a measurement of Δm_d. The difference in the amplitude of the mixing asymmetry is remarkable.
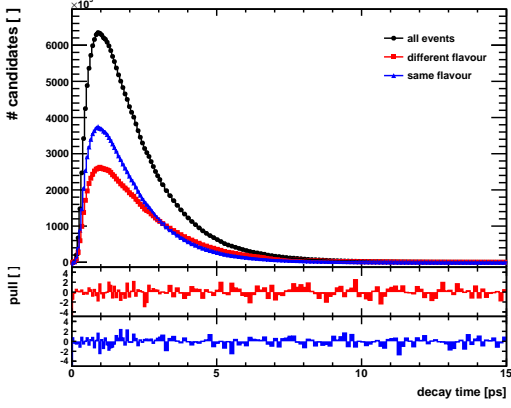
(a) Decay time distribution under realistic conditions
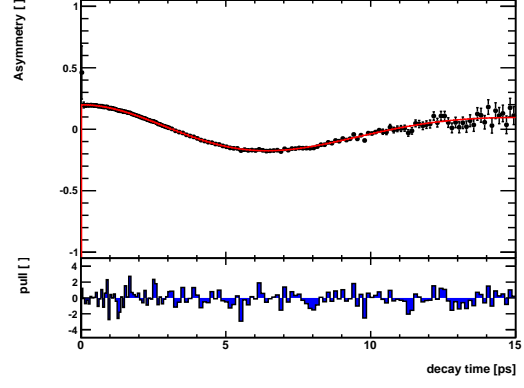
(b) Corresponding mixing asymmetry distribution under realistic conditions
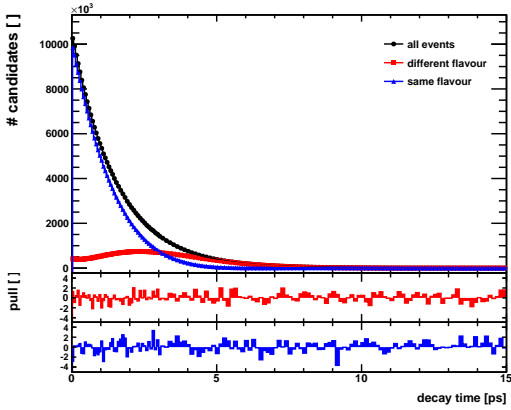
(c) Decay time distribution under idealistic conditions
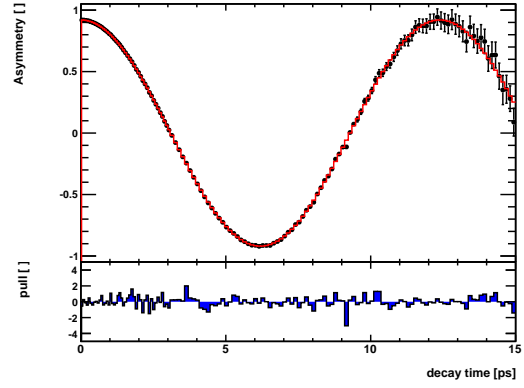
(d) Corresponding mixing asymmetry distribution under idealistic conditions

Figure 4.6.: Illustration of the damping factors on the mixing asymmetry using 1.6×10^7 simulated B^0 decays, generated under realistic conditions (upper plots) and under idealistic conditions (lower plots). Left: the B^0 decay time distributions for all events (black) and split up according to mixed (red, different production and decay flavour) and unmixed (blue, same production and decay flavour) events. The corresponding mixing asymmetry distributions are shown (right). Idealistic conditions contain 4% mistag probability, therefore, $\mathcal{A}|_{t=0} = 0.92$.

Under realistic conditions all damping factors are of the order of the later data fit conditions. In detail, this means the k factor distribution has a mean value of 1 and a width of 0.14, the flight distance resolution distribution has Gaussian shape with a mean of 0 fs and width of 100 fs and the mistag probability ω is of order 40%. Going from realistic to idealistic conditions, without changing the structure of the fit model is done by exchanging the k factor distribution by a δ function, which peaks at 1, reducing the flight distance distribution width by a factor 100, from 100 fs to 1 fs, and reducing ω from 40% to 4%. Additionally, the influence of the decay time acceptance distribution is tested by exchanging the turn-on distribution by a constant distribution. A truly ideal mixing asymmetry distribution would start at 1. The remaining dilution comes from the mistag probability of 4%. Figure 4.7 shows the four different scenarios, starting from the idealistic case and turning on one effect at a time, only.

Quantitatively, the influence of the different damping factors on the mixing asymmetry is reflected in the fitted statistical uncertainty on Δm_d, given in table 4.1.

Fit Model	$\sigma(\Delta m_d)\ \left[\mathrm{ns}^{-1}\ \right]$
idealistic	0.14
with decay time acceptance	0.12
with flight distance resolution	0.14
with k factor resolution	0.16
with mistag probability ω	0.76

Table 4.1.: The influence of different damping factors on the precision of the oscillation frequency Δm_d using 1.6×10^7 simulated B^0 decays.

By far the largest impact on the precision of Δm_d has the tagging dilution. It damps the mixing asymmetry by a factor of $(1 - 2\omega)$, see section 3.4.2. The realistic decay time acceptance shape improves the uncertainty, because the total number of generated events is kept constant, while the realistic decay time acceptance increases the relative amount of valuable long-lived to short-lived B^0 mesons for the measurement. The realistic flight distance resolution has no significant impact on the statistical uncertainty of Δm_d. The k factor resolution increases the statistical uncertainty on Δm_d slightly. However, it is decay time-dependent. It damps the mixing asymmetry of long-lived B^0 mesons more than the asymmetry of short-lived B^0 mesons. Thus, the k factor correction has to be determined accurately in order not to bias the oscillation frequency.

4.2. Data set and Selection

The measurement of the semileptonic *CP* asymmetry in B^0– $\overline{B}^0$ mixing at LHCb, a_{sl}^d [1], has been carried out using the same semileptonic decay channels and similar k factor and fitter frameworks. In the $B^0 \to D^{*-}\mu^+\nu_\mu X$ a fit to the mixing asymmetry has been used since a couple of years by the flavour tagging group to calibrate the OS taggers on B^0 data [76]. However, in the mistag probability calibration fits, the mixing frequency Δm_d was fixed to the PDG value. In a combined effort of the author and colleagues from Milan and Ferrara to measure Δm_d in
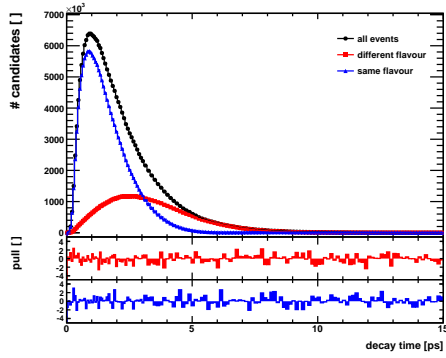
(a) Effect of the decay time acceptance
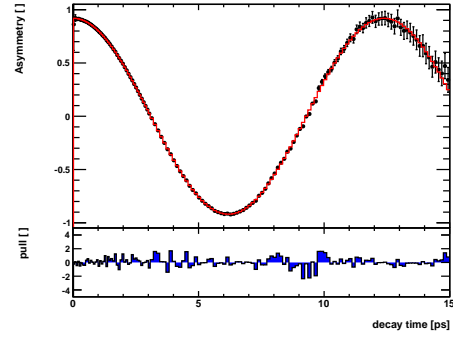
(b) Effect of the decay time acceptance
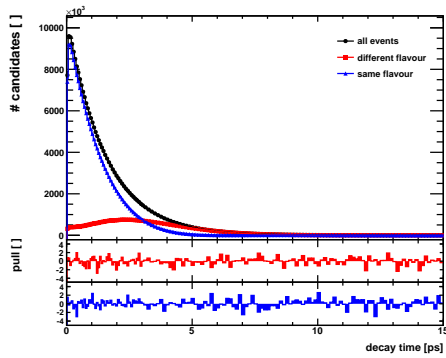
(c) Effect of the flight distance resolution
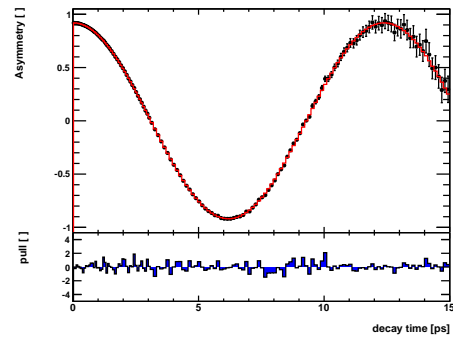
(d) Effect of the flight distance resolution
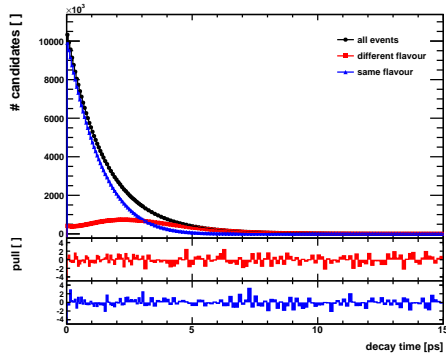
(e) Effect of the k factor correction
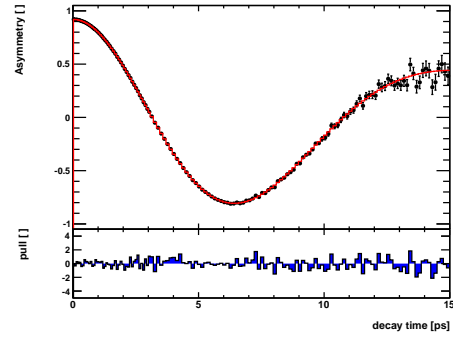
(f) Effect of the k factor correction
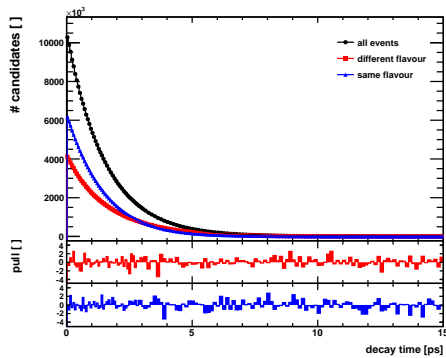
(g) Effect of the mistag probability ω
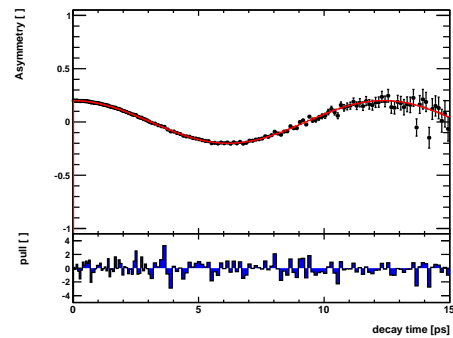
(h) Effect of the mistag probability ω

Figure 4.7.: One-by-one illustration of the different damping factors on the mixing asymmetry, as in figure 4.6 explained. Left plots: effect on the decay time distribution; right plots: effect on the mixing asymmetry distribution.

semileptonic B^0 decays at LHCb, the $B^0 \to D^- \mu^+ \nu_\mu X$ selection was adopted by the author from the a_{sl}^d measurement, while the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ selection was set by the Italian colleagues.

The full LHC Run I data set is used, collected by LHCb in the years 2011 and 2012. It corresponds to approximately $1\,\text{fb}^{-1} + 2\,\text{fb}^{-1}$ of data. The 2011 data was taken at a center-of-mass energy of $7\,\text{TeV}$, the 2012 data at $8\,\text{TeV}$.

Although semileptonic B^0 decays are only partially reconstructed, the decay topology has characteristic properties, that allow an efficient selection. The most useful features are depicted in figure 4.8. There are two displaced vertices of good quality, the B^0 decay vertex and the D decay vertex. There is one track that is associated to a muon candidate, detected in the muon stations of LHCb. The trajectories of the final state particles have large impact parameters with respect to the Primary Vertex (PV) and a harder transverse momentum spectrum than low QCD background from the PV.
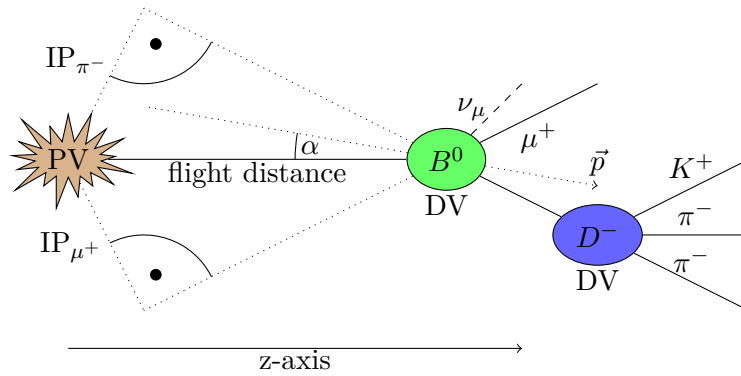
Figure 4.8.: Visualization of selection variables using the $B^0 \to D^- \mu^+ \nu_\mu X$ decay. The B^0 decay vertex (DV) is drawn in green, the D^- DV is drawn in blue. The distance between the primary vertex (PV) and the B^0 DV is measured as flight distance (fD) of the B^0. The impact parameter (IP) is measured as the perpendicular connection of a track with respect to the PV. The B^0 momentum ($\vec{p}$) does not point back to the PV due to the missing momentum of the neutrino. The angle between the B^0 $\vec{p}$ and the connection of the PV and the B^0 DV ($\alpha = $ DIRA) is used as pointing constraint. The distance of closest approach (DOCA) between two tracks is measured as the perpendicular connection between those tracks.

4.2.1. Trigger Selection

The final state muon triggers the hardware L0 trigger. This high momentum muon is also used to trigger the HLT1[6]. HLT2 selects topologically 2, 3 or 4 body B decays[7], where one of the decay daughters is a muon. Tight track quality cuts and multivariate selection criteria are applied [98, 99]. The advantage of such an inclusive trigger is, that not all B daughters have to be reconstructed to trigger an event. Hence, no requirements on the invariant B or D masses are

---

[6]The trigger line HLT1TrackMuon is used.

[7]The trigger lines HLT2TopoMu2,3,4BodyBBDT are used.

applied, but on the displacement of the decay of the long-lived particles, e.g. impact parameter. Boosted decision trees are used to perform this task efficiently [100].

4.2.2. Stripping and Off-line Selection

Events that pass the trigger requirements are fully reconstructed and written to storage. Particle candidates are formed out of tracks and particle identification information. Those candidates are combined to reconstruct the semileptonic B^0 decay. As a next step a pre-selection, called stripping selection, is applied LHCb-centrally to the full data set. Its purpose is to reduce the data set using selection requirements, defined by the analysis groups. Hence, the data analyst accesses a fully reconstructed, stripped data set[8]. Additional off-line selection requirements are used to suppress further background.

The vertex fit to form a B^0 candidate has to be of good quality. To ensure, that the B^0 comes from the assigned primary vertex (PV), the angle between the direction of the B^0 momentum and its flight direction (DIRA), given by the distance between the PV and the decay vertex (DV) of the B^0, is used. The B^0 DV has to be subsequently to the PV and the D DV subsequently to the B^0 DV, which is ensured by a minimum requirement on the difference of the z-positions of the vertices, where the z-axis is defined from the interaction point into the LHCb spectrometer. Minimum momentum requirements or tight cuts on the reconstructed mass are not applied to the B^0 candidate due to the missing momentum of the non-reconstructed neutrino.

Muons from semileptonic B decays are high energetic. To select those candidates, minimum transverse momentum (p_T) and momentum (p) requirements are used. The muon track fit and the track extrapolation through the LHCb spectrometer have to be of good quality. The latter is assured by a cut on the so-called ghost probability. A minimum requirement on particle identification (DLL) ensures to select real muons. The impact parameter (IP) of the muon track with respect to the PV is used to suppress muons from the primary interaction. Both, the IP and the uncertainty on the IP are used.

The D decay is fully reconstructed. The D^- and D^0 mesons decay into hadronic final states of kaons and pions. As many hadronic particles are produced in the primary proton-proton interaction, random combinations of particles to form the D candidate have to be suppressed. Therefore, the D mass resonance is required to be close to the PDG value. Additionally, the sum of the transverse momenta of the daughter particles has to fulfill a minimum requirement. The distance of closest approach (DOCA) divided by its uncertainty σ_{DOCA} of the daughter tracks has to be below a maximum requirement and the fit quality of the D DV has to be good. A minimum requirement on the D decay time ensures to select the relatively long-lived D meson candidate. The D DV has to be well separated from the PV, ensured by a large IP, while the correct PV assignment is ensured by a DIRA requirement.

In $B^0 \to D^{*-}\mu^+\nu_\mu X$ decays, the intermediate D^* resonance is very short-lived. It decays immediately into a D^0 and a low-energetic pion. The reconstructed mass of the D^* and the difference of the reconstructed D^* and D^0 masses are used to reduce combinatorial background. To

---

[8]For the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode the stripping line *StrippingBd2DstarMuNuTight* is used, for the $B^0 \to D^-\mu^+\nu_\mu X$ mode *Strippingb2DpMuXB2DMuNuXLine*.

ensure that the slow pion comes from the D^* decay, not from the PV, a minimum IP requirement is used.

The kaon and pion candidates from the D^- or D^0 decays are required to have a minimum momentum and transverse momentum. The track fits have to be of good quality and the ghost probability has has to be low. DLL requirements enhance the separation of kaon and pion candidates. A minimum requirement on the IP significance ensures that the particles do not originate from the PV, but from the displaced D vertex.

The stripping and off-line selection requirements used to select $B^0 \to D^- \mu^+ \nu_\mu X$ decays are summarized in table 4.2, the ones used to select $B^0 \to D^{*-} \mu^+ \nu_\mu X$ decays are summarized in table 4.3. Additional veto cuts are applied to suppress specific background decays, see next subsections.

candidate	variable	requirement
B^0	mass	$\in [3, 5.2]\,\mathrm{GeV}$
	vertex χ^2/ndf	< 6
	cos(DIRA)	> 0.999
	$z(DV_D) - z(DV_B)$	$> 0\,\mathrm{mm}$
μ	p_{T}	$> 800\,\mathrm{MeV}$
	p	$> 3.0\,\mathrm{GeV}$
	ghost probability	< 0.5
	track χ^2/ndf	< 4
	$\mathrm{DLL}_{\mu\pi}$	> 0
	$\mathrm{IP}/\sigma_{\mathrm{IP}}$	> 3
D^-	$|M(D) - M_{PDG}(D)|$	$< 80\,\mathrm{MeV}$
	$\sum p_{\mathrm{T}}$ D daughters	$> 1.8\,\mathrm{GeV}$
	$\mathrm{DOCA}/\sigma_{\mathrm{DOCA}}$ D daughters	< 4.5
	vertex χ^2/ndf	< 6
	vertex separation significance	> 10
	cos(DIRA)	> 0.99
	τ	$> 0.1\,\mathrm{ps}$
	log(IP)	$> -3\log(\,\mathrm{mm})$
K, π	p_{T}	$> 300\,\mathrm{MeV}$
	p	$> 2.0\,\mathrm{GeV}$
	ghost probability	< 0.5
	track χ^2/ndf	< 4
	$\mathrm{DLL}_{K\pi}$	$> 4(K)\ \ < 10(\pi)$
	$\mathrm{IP}/\sigma_{\mathrm{IP}}$	> 3

Table 4.2.: Stripping and off-line requirements to select $B^0 \to D^- \mu^+ \nu_\mu X$ (with $D^- \to K^+ \pi^- \pi^-$) decays.

candidate	variable	requirement
B^0	mass	$\in [3, 5.2]$ GeV
	vertex $\chi^2/$ndf	< 6.6
	$\cos(\text{DIRA})$	> 0.999
	$z(DV_D) - z(DV_B)$	> -2.5 mm
	$z(DV_B) - z(PV)$	> 0.5 mm
μ	p_T	> 0.8 GeV
	p	> 2.0 GeV
	ghost probability	< 0.5
	$\text{DLL}_{\mu\pi}$	> -4
	$\text{IP}/\sigma_\text{IP}$	> 2
	IP	> 0.05 mm
D^{*-}	p_T	> 1.25 GeV
	vertex $\chi^2/$ndf	< 10
	$|M(D^{*-}) - M_{PDG}(D^{*-})|$	< 100 MeV
	$|M(D^{*-}) - M(D^0)|$	< 165 MeV
π from D^{*-}	p_T	> 110 MeV
	IP	> 0.04 mm
D^0	p_T	> 1.8 GeV
	$|M(D^0) - M_{PDG}(D^0)|$	< 60 MeV
	vertex $\chi^2/$ndf	< 10
	vertex separation significance	> 7.1
	$\log(\text{IP})$	$> -3 \log(\text{mm})$
K, π from D^0	p_T	> 400 MeV
	p	> 2.0 GeV
	ghost probability	< 0.5
	IP	> 0.04 mm
	$\text{DLL}_{K\pi}$	$> 0 \ (-)$
	$\text{IP}/\sigma_\text{IP}$	> 2

Table 4.3.: Stripping and off-line requirements to select $B^0 \to D^{*-}\mu^+\nu_\mu X$ (with $D^{*-} \to \overline{D}^0\pi^-$ and $\overline{D}^0 \to K^+\pi^-$) decays.

4.2.3. Λ_b^0 Veto

About 10% of the produced b quarks hadronize to b-baryons, mainly Λ_b^0 baryons. About 10% of Λ_b^0 decay semileptonically [18] into $\Lambda_c^+ \ell^- \bar{\nu}_\ell$, from which 5% of the Λ_c^+ decay hadronically into $pK^-\pi^+$ [18]. If the final state proton is misidentified as a pion the Λ_b^0 decay looks very similar to the $B^0 \to D^- \mu^+ \nu_\mu X$ decay mode. Figure 4.9 shows the $K\pi\pi$ invariant mass distribution in $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data, computed under the proton mass hypothesis for the lowest momentum pion. There is a prominent peak at the mass of the Λ_c^+ baryon ($m_{\Lambda_c^+} = 2286.46 \pm 0.14\,\text{MeV}$ [18]) in the $B^0 \to D^- \mu^+ \nu_\mu X$ mode. Hence, a veto cut at these decays is applied to the $B^0 \to D^- \mu^+ \nu_\mu X$ mode, rejecting events with a proton to pion particle identification likelihood larger than 10 and an invariant mass, recalculated after exchanging the proton mass hypothesis, within 2260 MeV and 2310 MeV.
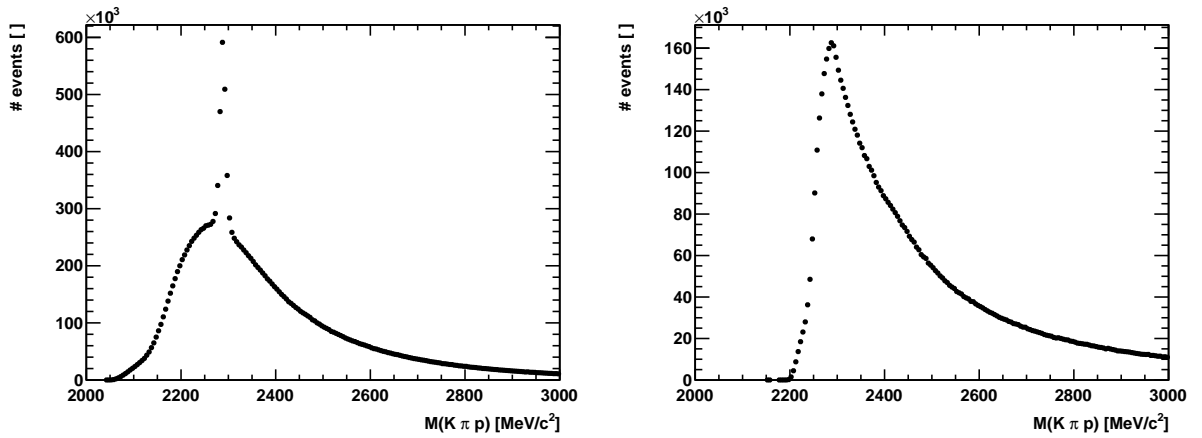

Figure 4.9.: $K\pi\pi$ invariant mass distributions in the $B^0 \to D^- \mu^+ \nu_\mu X$ mode (left) and the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode (right) on 2012 data, computed under the proton mass hypothesis for the lowest momentum pion. In the $B^0 \to D^- \mu^+ \nu_\mu X$ mode a clear mass peak at the Λ_c^+ mass, $m_{\Lambda_c^+} = 2286.46 \pm 0.14\,\text{MeV}$ [18], is visible.

4.2.4. J/ψ Veto

$B \to J/\psi X$ decays, where the J/ψ decays into two muons, is a source of background due to misidentification of one of the muons as a high momentum pion. Figure 4.10 shows the $\mu\pi$ invariant mass distribution in $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data, where the mass hypothesis of the highest momentum pion is exchanged by a muon mass hypothesis. In both modes there is a clear mass peak at the J/ψ mass visible. To remove this background a veto cut is applied in both modes, rejecting the event if the pion carries the *IsMuon* flag and if the invariant mass, recalculated after exchanging the muon mass hypothesis, is between 3070 MeV and 3150 MeV.
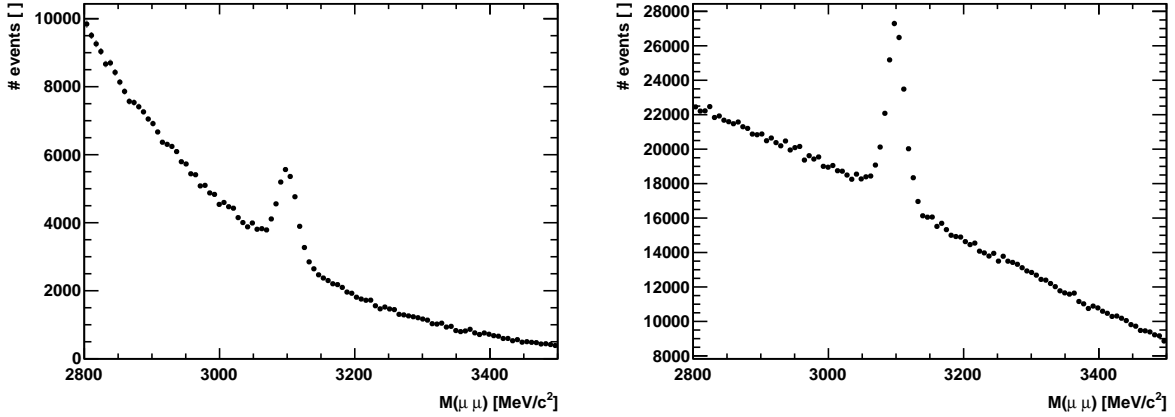
Figure 4.10.: $\mu\pi$ invariant mass distributions in the $B^0 \to D^-\mu^+\nu_\mu X$ (left) and the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode (right) on 2012 data, computed under the muon mass hypothesis for the highest momentum pion. In both distributions a clear peak at the J/ψ mass, $m_{J/\psi} = 3096.916 \pm 0.011\,\mathrm{MeV}$ [18], is visible.

4.2.5. B^+ Veto

Of the order of 10% of the reconstructed decays which pass the selection are semileptonic B^+ decays. Semileptonic B^+ decays have a similar decay topology as semileptonic B^0 signal decays. However, an additional charged pion is not reconstructed in case of the B^+ decay: $B^+ \to D^-\mu^+\nu_\mu X$ and $B^+ \to D^{*-}\mu^+\nu_\mu X$, here contained in the X. Due to the missing momentum, carried away by the neutrino, the B^0 mass resonance cannot be not reconstructed. Instead, a rather broad B mass shape is present. Hence, B^0 and B^+ candidates cannot be separated by their invariant mass.

Therefore, a different approach is followed, initially developed by rare decays searches, like $\tau \to \mu\mu\mu$ [101] and rare decay measurements like $B_s^0 \to \mu\mu$ [102]. A multivariate classifier is trained on track isolation.

Track isolation is computed by combining every track from the event with one of the signal decay daughter tracks. If those pairs of tracks form a good vertex, the signal decay daughter track is badly isolated and the decay is most likely background. The multivariate classifier, here a boosted decision tree (BDT), is trained on this classification task and assigns a probability to each event as a measure for this track isolation. The latest optimisation study for this classifier was performed in [103] based on the work, documented in [104]. In this optimised status, the isolation BDT was successfully used in the $\tau \to \mu\mu\mu$ analysis [105].

For the Δm_d analysis, a second BDT was trained following closely the approach of the isolation BDT mentioned above, but including additional semileptonic decay variables: the reconstructed broad B mass, the pointing angle, which is the angle between the B momentum and the vector connecting the primary and the B decay vertex, the corrected mass, which is defined as $m_{corr} = \sqrt{m_B^2 + p_{T\nu}^2} + p_{T\nu}$. The transverse momentum of the neutrino $p_{T\nu}$ is not reconstructed. Similar to [106] it is defined as the sine of the pointing angle times the absolute value of the B momentum.

A 'best' pion is defined as the one having the smallest impact parameter with respect to the D^0/ D^+ decay vertex. The invariant mass of the D^*/ D^+ and the 'best' pion gives a handle on D^{**} resonances, that are more contained in B^+ decays than in B^0 decays. The cone isolation is defined as the fraction of the B p_T divided by the B p_T plus the sum of the p_T of all tracks, that fulfill $\sqrt{\delta\eta^2 + \delta\phi^2} < 1$, where $\delta\eta, \delta\phi$ are the difference in pseudo-rapidity and in ϕ coordinate between the track and the B candidate. Additionally, the Δm_d isolation BDT uses the output of the isolation BDT per decay daughter track as used in the $\tau \to \mu\mu\mu$ analysis [105] and the isolation track counter from [104] as inputs. The Δm_d isolation BDT is trained on $B^0 \to D^{*-}\mu^+\nu_\mu X$ simulation, using the $B^0 \to D^{*-}\mu^+\nu_\mu X$ sample as signal class and the $B^+ \to D^{*-}\mu^+\nu_\mu X$ simulation as background class. Before training, the track multiplicities in simulation are re-weighted to match the track multiplicities in data, which are on average higher.

The development of the Δm_d isolation BDT was done within the Δm_d semileptonic subgroup, documented in [3]. The author has applied the Δm_d isolation BDT to all data sets used for the analysis in data and simulation. Figure 4.11 shows the Δm_d isolation BDT response in simulation for both modes. The B^0 signal response peaks at $+1$, while the B^+ background response peaks at -1.
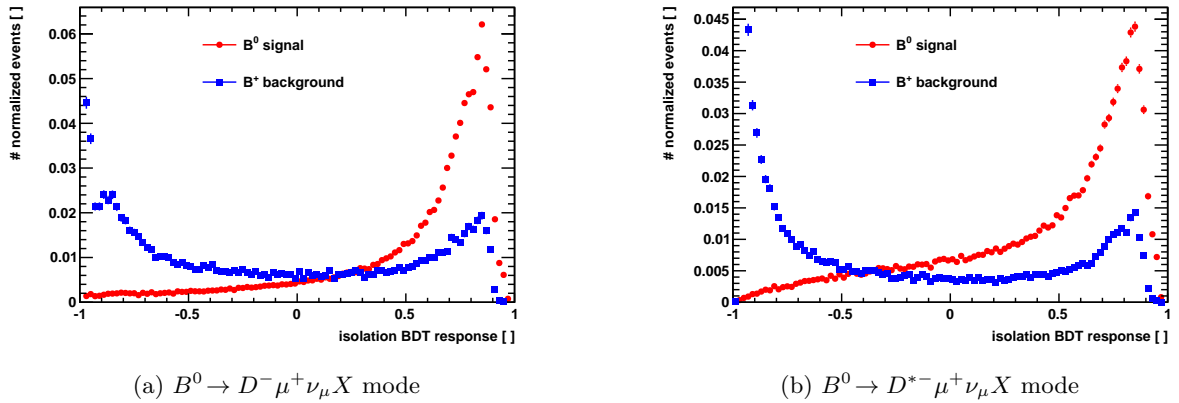

(a) $B^0 \to D^-\mu^+\nu_\mu X$ mode　　　　　　(b) $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode

Figure 4.11.: Isolation BDT response in B^0 signal MC and B^+ background MC in the $B^0 \to D^-\mu^+\nu_\mu X$ mode (left) and $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode (right).

To reduce the amount of B^+ background, a cut is applied to the Δm_d isolation BDT output. The cut value is chosen to loose 10% of the signal B^0 MC events. A scan through tighter cut values showed, that the ratio of B^0 to B^+ is not improving. The sample specific cut values are shown in table 4.4. The fraction of B^+ background is determined data driven by fitting the isolation BDT distribution in data. The isolation BDT distribution histograms obtained from B^0 signal and B^+ background simulation are used hereby as templates. The only free parameter of the fit is the B^+ fraction, shown in section 4.4. For comparison the B^+ fraction is also computed on simulation, only, taking into account the selection efficiencies normalised to the branching ratios of the decays, shown in section 4.3.

MC sample	isolation BDT
2011 $B^0 \rightarrow D^- \mu^+ \nu_\mu X$	> -0.12
2012 $B^0 \rightarrow D^- \mu^+ \nu_\mu X$	> -0.16
2011 $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$	> -0.36
2012 $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$	> -0.38

Table 4.4.: Cut values on the isolation BDT response to veto B^+ decays at 10% B^0 signal loss.

The measurement of Δm_d requires flavour tagging, which posses an algorithm specific tagging efficiency. Hence, the selection efficiencies are given in section 4.4, together with the tagging efficiencies.

4.3. Monte Carlo Simulation

MC simulation is used to understand the composition of semileptonic decays in the data set. The reconstruction of semileptonic decays is inclusive, i.e. the decay can contain more particles, that are not reconstructed. Due to the missing momentum carried away by the neutrino, the decay chain is kinematically underconstrained. Hence, additional non-reconstructed tracks from the B^0 decay are possible, that originate from higher D resonances. Those events are called feed down modes, which are included in the MC "cocktails".

The k factor, used to correct the B^0 momentum for the missing neutrino momentum, is derived from MC. The impact of branching ratio uncertainties, that limit the knowledge of the sample composition, is treated as systematic uncertainty on Δm_d, tested in section 4.9.

For each decay mode a signal B^0 decay sample and a background B^+ decay sample is generated for 2011 and 2012 conditions, see table 4.5. In the simulated signal samples the B^0 is forced to decay semileptonically either into a D^- or a D^{*-} meson, including possible higher D resonances. Subsequently, the D^- meson is forced to decay into the $K^+\pi^-\pi^-$ final state and the D^{*-} meson is forced to decay into the $\overline{D}^0(\to K^+\pi^-)\pi^-$ final state. Analogously, in the B^+ background samples, the B^+ is forced to decay either into a D^- μ^+ ν_μ π^+, or a D^* μ^+ ν_μ π^+, with subsequent D decays as in the signal B^0 samples. In both cases, the pion is not reconstructed.

Type	Decay	Generated events [$\times 10^6$]	
		2011	2012
Signal	$B^0 \to D^- \mu^+ \nu_\mu X$	10.0	20.2
Signal	$B^0 \to D^{*-} \mu^+ \nu_\mu X$	20.2	40.1
Background	$B^+ \to D^- \mu^+ \nu_\mu X$	2.6	7.6
Background	$B^+ \to D^{*-} \mu^+ \nu_\mu X$	2.0	4.1
Background	inclusive $D^+ \to K^-\pi^+\pi^+$	5.0	10.1
Background	inclusive $D^{*-} \to \overline{D}^0\pi^-$ with $\overline{D}^0 \to K^+\pi^-$	6.0	12.1
Background	$B_s^0 \to D_{s1}^- \mu^+ \nu_\mu X,\ D_{s1} \to D^- K_S^0$	-	2.6
Background	$B_s^0 \to D_{s1}^- \mu^+ \nu_\mu X,\ D_{s1} \to D^{*-} K_S^0$	-	2.6
Background	$\Lambda_b^0 \to \mathrm{n}^0 D^- \mu^+ \nu_\mu X$	-	2.6
Background	$\Lambda_b^0 \to \mathrm{n}^0 D^{*-} \mu^+ \nu_\mu X$	-	2.6

Table 4.5.: Generated samples. The number of generated events is given. It includes both magnet polarities and Pythia versions (6 and 8). The X in each decay mode indicates that a cocktail of decay modes has to be considered.

After the Monte Carlo production was finished a bug was discovered in the cocktail decay files. Wrong form factors were contained in the decay files for D^* decays. Therefore, the Δm_d semileptonic subgroup developed event by event kinematic correction weights, applied to all D^* decays in every simulation sample.

For both modes and years inclusive D background samples are generated to study the inclusive background composition. This is necessary, because B_s^0 and Λ_b^0 mesons can also decay semileptonically into charmed mesons. The possible fractions of these decays are estimated after applying the full off-line selection and isolation BDT requirements. B_s^0 decays occur with a fraction of $1.5 \pm 0.4\%$ in the $B^0 \to D^- \mu^+ \nu_\mu X$ mode and $1.6 \pm 0.3\%$ in the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode. Due to the limited statistics of the MC samples, these predictions are not very precise. Introducing a B_s^0 component into the Δm_d fit is not practicable since the Δm_s oscillation is too fast to be resolved in the semileptonic decay. Hence, the B_s^0 component is treated as a systematic. The Λ_b^0 component is found to be negligible in both modes having applied all selection requirements. Additionally, dedicated B_s^0 and Λ_b^0 decay background samples are generated, to study the mixing asymmetry distributions of these backgrounds.

The specifications of the MC generation process are defined by the Δm_d semileptonic subgroup using the relative branching fractions given in appendix B.1. They are based on the work documented in [107]. The current settings are summarized in [108]. To the simulation samples, the same stripping, trigger and offline selection requirements are applied as on real data.

4.4. Flavour Tagging

The production flavour of B^0 mesons can be determined by opposite side taggers and by the same side pion tagger, which exploits a charged pion track from the signal b hadronization. Using all available tagging algorithms and combining their responses results in the highest possible tagging power, thus, the best sensitivity on Δm_d. However, the actual choice of tagging algorithms must be carefully made due to backgrounds.

A non negligible amount of B^+ background, $\mathcal{O}(6\%)$, is present in the data set. B^+ mesons are charged, hence, they do not oscillate. Particularly at low decay times, the mixing asymmetry distribution of B^+ events tagged by a certain tagger has to be constant. It can have a non-zero offset, due to a non-zero mistag probability, but the tagging algorithm should not introduce a time dependence into the mixing asymmetry.

Figure 4.12 shows the mixing asymmetry distribution of B^0 signal simulation events overlaid by B^+ background events. It is shown in the $B^0 \to D^- \mu^+ \nu_\mu X$ mode, where the higher fraction of B^+ background is expected. The B^+ mixing asymmetry should be constant, but it is clearly not. The reason is, that the same side pion tagger selects the additional charged pion track from semileptonic B^+ decays, which is not reconstructed in the signal decay. However, the charged pion track from the signal decay is of opposite charge than the fragmentation pion. This reverts the tagging decision of the same side pion tagger, thus, non-oscillating B^+ decays appear to have undergone mixing. In order to avoid the introduction of an artificial background oscillation the same side pion tagger cannot be used in the Δm_d analysis.

Figure 4.13 shows the mixing asymmetry distributions of B^0 and B^+ simulated events tagged by opposite side taggers: the OS muon, electron, vertex charge and ANN-based kaon tagger. Unfortunately, the B^+ simulation statistics are low. The B^+ mixing distributions tagged by lepton taggers, that have low tagging efficiencies, are dominated by statistical variations. However, lepton taggers select leptonic tagging tracks, that leave clean signatures in the detector. It is very unlikely, that they pick up the additional pion from the semileptonic B^+ decay.
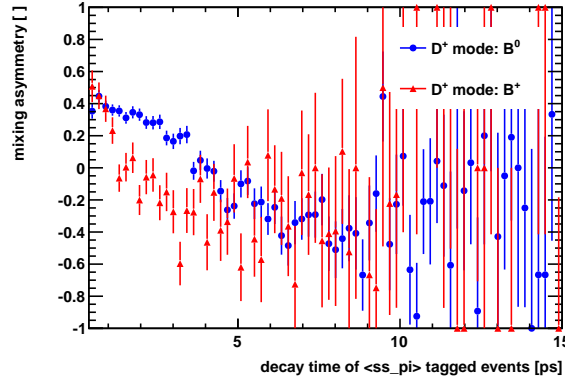
Figure 4.12.: Mixing asymmetry distribution of B^0 signal and B^+ background simulated events tagged by the same side pion tagger. The mixing asymmetry of B^+ events should be constant, especially at low decay times, but it shows oscillating behaviour here. The SS pion tagger is not used.

The mixing asymmetry distribution of B^+ decays tagged by the vertex charge tagger shows a non-constant behaviour at low decay times. It is possible, that the additional pion track from semileptonic B^+ decays is contained in the inclusive secondary vertex reconstruction of the vertex charge tagger at low B decay times. Thus, the OS vertex charge tagger is not used in the Δm_d analysis.

The mixing asymmetry distribution for B^+ decays tagged by the ANN-based kaon tagger is constant. Hence, a combination of the OS muon, electron and ANN-based kaon taggers is chosen for the Δm_d analysis.

4.4.1. Definition of Tagging Categories

A maximum tagging power is obtained when the events are weighted with the predicted event-by-event mistag probability. This can be done in an unbinned fit. However, several million semileptonic B^0 decays are selected and tagged in data. Hence, the fit to the mixing asymmetry must be performed as a binned fit, because an unbinned fit would demand too many computing resources.

To profit from the predicted mistag probability the events are categorized in tagging categories of increasing predicted mistag probability. The average measured mistag probability is fitted per category. Indeed a similar tagging power can be obtained in a binned fit using four tagging categories compared to an unbinned fit.

A study, to find the optimal set of tagging categories, was determined in the Δm_d semileptonic subgroup, see [3]. The goal of the optimization is to find the minimum number of categories, that yield a tagging power close to the event-by-event tagging power. A minimum number of categories is desirable, because each category introduces an additional dimension in the Δm_d fit. The boundaries of the categories are chosen such, that each category is properly populated.
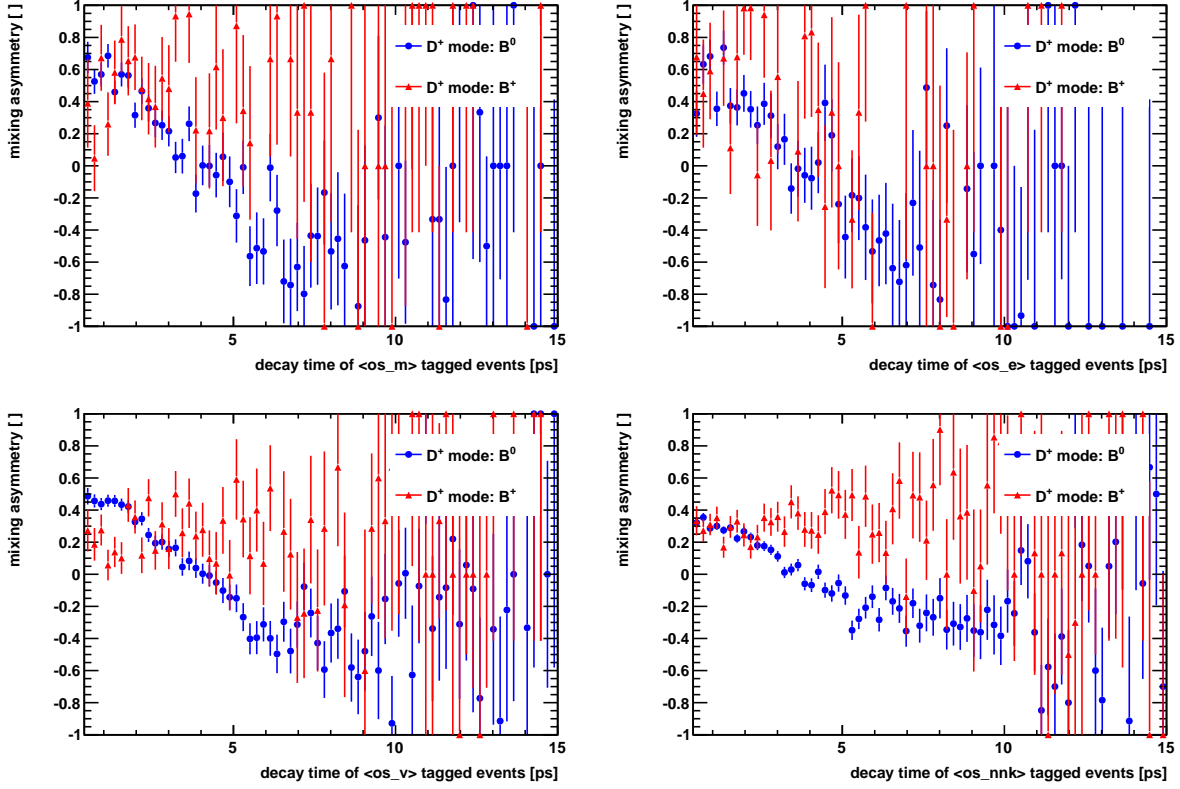
Figure 4.13.: Mixing asymmetry distributions of B^0 signal and B^+ background simulated events tagged by single opposite side taggers. Top left: OS muon tagger, right: OS electron tagger. Bottom left: vertex charge tagger, right: ANN-based kaon tagger. At low decay times the mixing asymmetry of B^+ events tagged by the OS vertex charge tagger is not constant. Hence, a combination of the OS muon, electron and ANN-based kaon taggers is used.

A single category provides a tagging power of 82% of the event-by-event tagging power. 98% of the event-by-event tagging power is reached with four categories. Introducing more tagging categories does not significantly improve the reached tagging power. The tagging category boundaries are shown in table 4.6.

The range of the predicted mistag probability is 0% to 50%. Here, the upper boundary of the highest mistag category is chosen at 47.5% to exclude badly tagged events. This increases the average measured mistag probability in that category and the combined mistag probability of all categories. The cost in tagging efficiency does not undercut the win in mistag probability, because $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1 - 2\omega)^2$.

category	lower boundary	upper boundary
0	0.00	0.25
1	0.25	0.33
2	0.33	0.41
3	0.41	0.475

Table 4.6.: Tagging category boundary values. The events are categorized with respect to the event-by-event predicted mistag probability. Using four tagging categories yields 98% of the maximum tagging power, which is the event-by-event tagging power.

4.4.2. Selection Efficiencies

The final selection efficiencies are shown in table 4.7 for 2011 and 2012, $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ tagged data. The events are tagged by the combination of OS muon, electron and ANN-based kaon taggers. In total 5.8×10^6 of semileptonic B^0 data events are selected and tagged. These events include combinatoric background of the subsequent D decays. This background will be removed by fits to the D mass resonances.

Selection	$B^0 \to D^- \mu^+ \nu_\mu X$		$B^0 \to D^{*-} \mu^+ \nu_\mu X$	
	2011	2012	2011	2012
# Post-stripping events [10^6]	7.35	20.74	5.41	15.16
# Selected events [10^6]	2.17	5.82	1.21	2.80
Selection efficiency	29.5%	28.1%	22.3%	18.5%
# Tagged events [10^6]	1.02	2.84	0.58	1.38
Tagging efficiency	47.2%	48.7%	47.8%	49.4%

Table 4.7.: Tagged events in 2011 and 2012, $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode data samples. A combination of OS muon, electron and ANN-based kaon taggers is used. Efficiencies are calculated with respect to the events that pass the previous selection step.

4.5. Backgrounds

This section discusses all relevant backgrounds for the measurement of Δm_d. There are two types of background. One type is related to the reconstruction of the D decay. The D decays hadronically into $K\pi\pi$ in case of the D^- and $K\pi$ in case of the D^0. The D^0 comes from a $D^* \to D^0 \pi$ decay. Possible mis-reconstruction occurs, if hadronic tracks are combined randomly. This combinatoric background is determined by a mass fit to the D mass resonance. In case of the $B^0 \to D^- \mu^+ \nu_\mu X$ mode, a one-dimensional fit to the D^- mass resonance is done, while in case

of the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode a two-dimensional fit to the D^0 mass resonance and to the mass difference of $m(D^*) - m(D^0) = \delta m$ is done.

The second type of background are mis-reconstructed B decays. As already discussed, semileptonic B^+ decays are not distinguishable from semileptonic B^0 decays due to the broad mass distribution of the reconstructed semileptonic B decays. The neutrino carries away momentum, hence the momentum of the reconstructed B^0 is underestimated. On the other side, momentum of possible additional non-reconstructed tracks, like the charged pion in case of the semileptonic B^+ decay, cannot be separated from the missing momentum of the neutrino. In order to determine the fraction of B^+ decays on data, a different quantity than the B^0 mass is exploited, the isolation BDT distribution. Using the shapes of B^0 and B^+ isolation BDT distributions from MC in a fit to the isolation BDT distribution in data, the B^+ fraction is extracted. As a cross check, the B^+ fraction is also calculated using simulation, only.

Besides semileptonic B^+ decays, also semileptonic B_s^0 and Λ_b^0 decays are possible. The fractions of B_s^0 and Λ_b^0 decays, estimated from MC, are 1.5% and 0.1%, respectively. They are treated as systematic uncertainty on Δm_d. To evaluate this systematic, the dedicated B_s^0 and Λ_b^0 MC samples are used to study the behaviour of the mixing asymmetry distribution.

4.5.1. Hadronic Combinatoric Background in the $B^0 \to D^-\mu^+\nu_\mu X$ Mode

In the $B^0 \to D^-\mu^+\nu_\mu X$ mode the reconstructed D^- mass distribution is fitted by a binned extended maximum likelihood[9] fit to extract the signal and background yields. Here, signal includes all types of semileptonic B decays. Two components are considered: D^- signal and D^- combinatorial background. The D^- mass distribution is binned in 80 uniform bins between 1790 MeV and 1950 MeV. The signal mass model $\mathcal{S}$ is composed of two Gaussian distributions and a Crystal Ball distribution:

$$
\begin{aligned}
\mathcal{S}(m) = {} & G(m; \mu_m, \sigma_{m,G1}) + f_{m,G} \cdot G(m; \mu_m, \sigma_{m,G2}) \\
& + (1 - f_{m,G}) \cdot f_{m,CB} \cdot CB(m; \mu_m, \sigma_{m,CB}, n_{m,CB}, \alpha_{m,CB}) \,.
\end{aligned}
\tag{4.9}
$$

The Gaussian distributions, $G(m; \mu_m, \sigma_{m,G1})$ and $G(m; \mu_m, \sigma_{m,G2})$, account for mass resolution effects. One mass mean value μ_m, but different mass widths, $\sigma_{m,G1}, \sigma_{m,G2}$ are used. The fraction $f_{m,G}$ connects the two distributions. The asymmetric Crystal Ball distribution $CB(m; \mu_m, \sigma_{m,CB}, n_{m,CB}, \alpha_{m,CB})$ accounts for radiative tails towards lower reconstructed D^- masses. It shares the same mass mean value as the Gaussian distributions, but has a different width, $\sigma_{m,CB}$, and parameters $n_{m,CB}, \alpha_{m,CB}$. The fraction $f_{m,CB}$ connects the Crystal Ball distribution with the double Gaussian distributions. The background mass model is a single exponential distribution $\mathcal{C}$:

$$
\mathcal{C}(m) = e^{m_0 \cdot m},
\tag{4.10}
$$

with the parameter m_0. The total fit to the D^- mass distribution is an extended likelihood fit, i.e. the number of signal ($nSig$) and background ($nBkg$) events are fitted:

$$
\mathcal{P}(m) = nSig \cdot \mathcal{S}(m) + nBkg \cdot \mathcal{C}(m).
\tag{4.11}
$$

---

[9]The definition of a maximum likelihood fit is given in section 4.6.

Parameter	2011 data	2012 data
$\mu_m\,[\,\mathrm{MeV}]$	1871.04 ± 0.01	1871.103 ± 0.007
$\sigma_{m,G,1}\,[\,\mathrm{MeV}\,]$	7.1 ± 0.1	7.10 ± 0.07
$\sigma_{m,G,2}\,[\,\mathrm{MeV}\,]$	12.1 ± 0.3	11.9 ± 0.2
$\sigma_{m,CB}\,[\,\mathrm{MeV}\,]$	4.70 ± 0.07	4.72 ± 0.04
$\alpha_{m,CB}$	2.37 ± 0.03	2.37 ± 0.02
$n_{m,CB}$	0.00001	0.00001
$f_{m,G}$	0.65 ± 0.02	0.63 ± 0.01
$f_{m,CB}$	0.21 ± 0.02	0.22 ± 0.01
$m_0\,[\,\mathrm{MeV}^{-1}]$	-0.00227 ± 0.0001	-0.002340 ± 0.00009
$\#nSig$	$572\,579 \pm 1820$	$1\,597\,639 \pm 3\,045$
$\#nBkg$	$184\,041 \pm 1709$	$513\,034 \pm 2\,862$

Table 4.8.: $B^0 \to D^- \mu^+ \nu_\mu X$ mode: 2011 and 2012 data fitted parameters determined in fits to the reconstructed D^- mass distribution. For the fit parameter definitions see text.

All parameters are floated except the Crystal Ball parameter $n_{m,CB}$. Figure 4.14 shows the fits to reconstructed D^- mass distributions in 2011 and 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ data. Tagged events are considered, only. The corresponding fit parameters are listed in table 4.8. In 2011 data 5.7×10^5
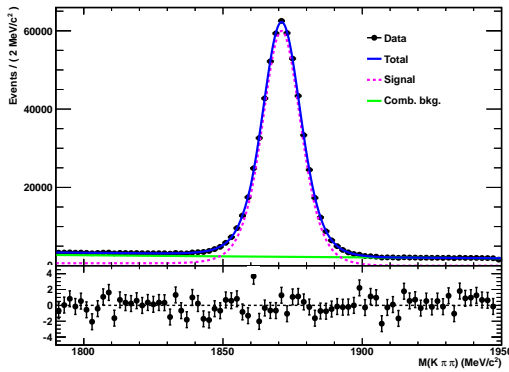
(a) D^- 2011 data
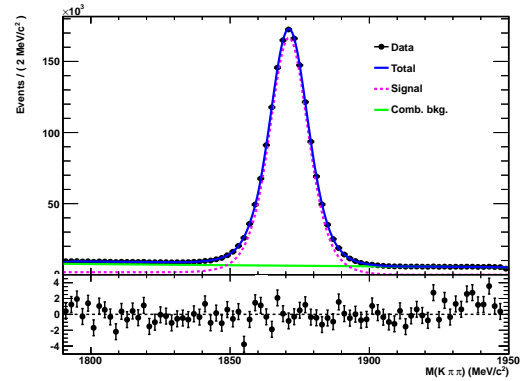
(b) D^- 2012 data

Figure 4.14.: Fit to the D^- mass distributions in 2011 data (left) and 2012 data (right), tagged events, only. The signal component (magenta dashed line) and the hadronic combinatoric background component (green line) are separately shown. The full mass fit model is shown as blue line.

signal events are extracted, in 2012 data 1.6×10^6, where signal includes all semileptonic B decays.

4.5.2. Hadronic Combinatoric Background in the $B^0 \to D^{*-}\mu^+\nu_\mu X$ Mode

In the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode the reconstructed D^0 mass and the mass difference $\delta m = m(D^*) - m(D^0)$ are fitted by a two dimensional extended maximum likelihood fit. Three components are considered: $D^* \to D^0(K\pi)\pi$ signal, D^0 directly from B decays and combinatoric background. The signal components describes decays, where the full reconstruction was successful: the $D^* \to D^0\pi$ decay and the D^0 to $K\pi$ decay. The D^0 from B component describes real D^0 to $K\pi$ decays, but without an intermediate D^* resonance, instead the D^0 mesons come directly from B decays. The last component describes hadronic combinatoric background, where non of the two D decays is reconstructed correctly. Also here, "signal" includes all types of semileptonic B decays. The reconstructed D^0 mass and δm distributions are binned in 80 equally sized bins, between 1820 MeV and 1912 MeV the former and between 139. MeV and 165.5 MeV the latter.

The signal component peaks in the D^0 mass and in the δm distribution. Its D^0 mass model is identical to the D^- mass model $\mathcal{S}(m)$, separating signal $D^0 \to K\pi$ decays from combinatoric background. Its δm model is composed of two Gaussian distributions and a Johnson distribution:

$$\mathcal{S}(\delta m) = G(\delta m; \mu_{\delta m}, \sigma_{\delta m, G1}) + f_{\delta m, G} \cdot G(\delta m; \mu_{\delta m}, \sigma_{\delta m, G2})$$
$$+ (1 - f_{\delta m, G}) \cdot f_{\delta m, RJ} \cdot \mathcal{RJ}(\delta m; \mu_{\delta m}, \sigma_{\delta m, RJ}, \gamma_{\delta m}, \delta_{\delta m}). \tag{4.12}$$

The two Gaussian distributions $G(\delta m; \mu_{\delta m}, \sigma_{\delta m, G1}), G(\delta m; \mu_{\delta m}, \sigma_{\delta m, G2})$, share the same mean δm parameter $\mu_{\delta m}$, but different widths $\sigma_{\delta m, G1}, \sigma_{\delta m, G2}$ to account for mass resolution effects. They are connected by the fraction $f_{\delta m, G}$. The Johnson distribution $\mathcal{RJ}(\delta m; \mu_{\delta m}, \sigma_{\delta m, RJ}, \gamma_{\delta m}, \delta_{\delta m})$ also shares the same mean parameter $\mu_{\delta m}$, but has a different width parameter $\sigma_{\delta m, RJ}$ and additional $\gamma_{\delta m}, \delta_{\delta m}$ parameters. It is defined using the *RooJohnson* parametrization[10] from the RooFit package and the substitution $x = (\delta m - \mu_{\delta m})/\sigma_{\delta m}$:

$$\mathcal{RJ}(x; \sigma, \gamma, \delta) = \delta/\sqrt{2\pi\sigma^2 \cdot (1 + x^2)} \cdot \exp\left(-0.5 \cdot (\gamma + \delta \cdot \log(x + \sqrt{1 + x^2}))^2\right) \tag{4.13}$$

The parameter δ is set to 1. The two-dimensional signal model is then:

$$\mathcal{S}(m, \delta m) = \mathcal{S}(m) \cdot \mathcal{S}(\delta m). \tag{4.14}$$

The D^0 from B component peaks in the D^0 mass distribution modeled by the signal mass model $\mathcal{S}(m)$, but does not peak in the δm distribution. It's shape is modeled by the *RooDstD0BG* distribution:

$$\mathcal{D}^0(\delta m; m_{\delta m}^{th}, a_{\delta m}, b_{\delta m}, c_{\delta m}) = \mathcal{R}ooDstD0BG(\delta m; m_{\delta m}^{th}, a_{\delta m}, b_{\delta m}, c_{\delta m}). \tag{4.15}$$

The threshold parameter $m_{\delta m}^{th}$ determines the threshold of the δm distribution start, it is set constant. The parameters $a_{\delta m}, b_{\delta m}, c_{\delta m}$ are defined as follows:

$$\mathcal{R}ooDstD0BG(\delta m; m_{\delta m}^{th}, a_{\delta m}, b_{\delta m}, c_{\delta m}) =$$
$$(1 - e^{-(\delta m - m_{\delta m}^{th})/c_{\delta m}}) \cdot (\delta m/m_{\delta m}^{th})^{a_{\delta m}} + b_{\delta m}(\delta m/m_{\delta m}^{th} - 1), \tag{4.16}$$

---

[10] For better readability, the subscripts δm are not written.

for $\delta m - m_{\delta m}^{th} > 0$. The two-dimensional D^0 from B model is then:

$$\mathcal{D}^0 fB(m, \delta m) = \mathcal{S}(m) \cdot \mathcal{D}^0(\delta m). \tag{4.17}$$

The D^0 combinatoric background does not peak in the D^0 mass nor the δm distribution. The former part is modeled by the D mass background exponential model $\mathcal{C}(m)$, the latter is described analogue to the D^0 from B component by the $\mathcal{R}ooDstD0BG(\delta m)$ distribution. Thus, the two-dimensional combinatoric model is

$$\mathcal{C}(m, \delta m) = \mathcal{C}(m) \cdot \mathcal{R}ooDstD0BG(\delta m; m_{\delta m}^{th}, a_{\delta m}, b_{\delta m}, c_{\delta m}). \tag{4.18}$$

The same parameters $m_{\delta m}^{th}, a_{\delta m}, b_{\delta m}, c_{\delta m}$ are used for the fit to both backgrounds.

The total two-dimensional $m_{D^0}, \delta m$ extended fit model is then:

$$\mathcal{P}(m, \delta m) = nSig \cdot \mathcal{S}(m, \delta m) + nBkg_{D^0 fB} \cdot \mathcal{D}^0 fB(m, \delta m) + nBkg \cdot \mathcal{C}(m, \delta m). \tag{4.19}$$

The number of signal ($nSig$), D^0 from B ($nBkg_{D^0 fB}$) and combinatoric background ($nBkg$) events are extracted from the fit. All shape parameters are floated, except the Johnson parameter $\delta_{\delta m}$, the $\mathcal{R}ooDstD0BG$ threshold parameter $m_{\delta m}^{th}$ and the Crystal Ball parameter $n_{m,CB}$. The fits to the m_{D^0} and Δm distributions in the $B^0 \to D^{*-}\mu^+\nu_\mu X$ with 2011 and 2012 data are shown in figure 4.15. Also here tagged events only are considered.

The corresponding fit parameters are listed in table 4.9. In 2011 data 3.1×10^5 signal events are extracted, in 2012 data 7.4×10^5, where signal includes all semileptonic B decays.

4.5.3. sWeights

From the above shown mass fits sWeights [78, 79] can be extracted. These event-by-event weights are used to statistically remove the background components from the data which is fitted to extract Δm_d. This approach is followed to stabilize the Δm_d fit. It is advantageous in three ways. First, the fits to disentangle the D background components are done prior to the Δm_d fit. This speeds up the Δm_d fit significantly. Second, the D background components do not have to be modeled for the decay time and mixing asymmetry distributions. Indeed, the D background components show oscillating and non-oscillating behaviour, because they do not consist of pure mis-reconstructed combinatoric background, but also of mis-reconstructed D coming from B decays. A second oscillation frequency would have to be introduced in this case. Hence, using sWeights to statistically remove the D background components reduces the Δm_d fit dimensions by a factor of two. Third, the sWeights are used to determine fit inputs, like B^+ fractions directly from sWeighted data, as explained in the next section.

The sWeights are extracted from fits to the reconstructed $m(D^-)$ and $m(D^0), \delta m$ distributions per tagging category. The single fits are shown in the appendix B.2 for the $B^0 \to D^-\mu^+\nu_\mu X$ and $B^0 \to D^{*-}\mu^+\nu_\mu X$ modes in 2011 and in 2012 data. As a cross check, the Δm_d fit is also performed on an sWeighted data sample, where the sWeights are determined without splitting the sample in tagging categories. The deviation between the nominal fit and this cross check fit is taken as systematic uncertainty on Δm_d.
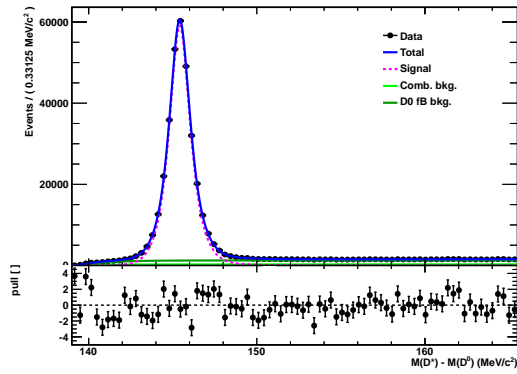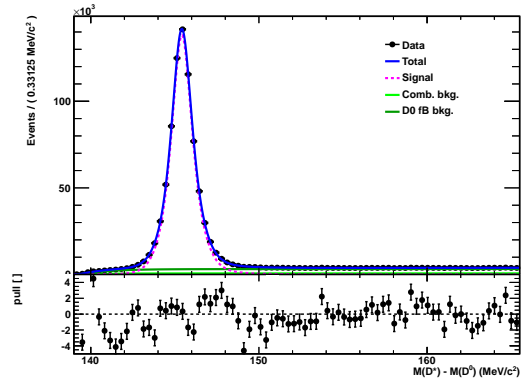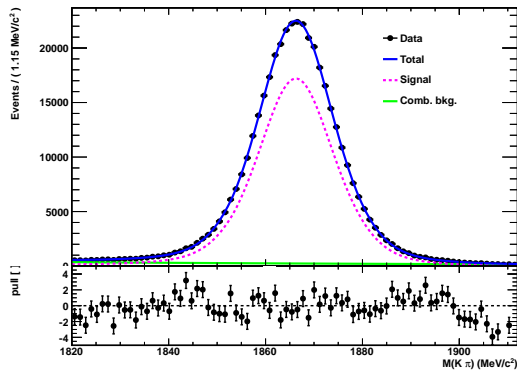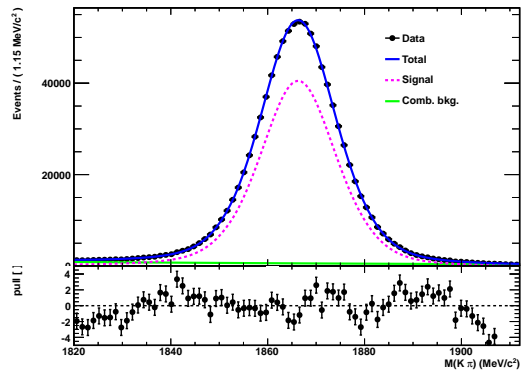
(a) D^* 2011 data

(b) D^* 2012 data

(c) D^* 2011 data

(d) D^* 2012 data

Figure 4.15.: Two-dimensional fit to the reconstructed $\delta m = m(D^*) - m(D^0)$ distributions (upper plots) and the reconstructed $m(D^0)$ distributions (lower plots) for 2011 data (left) and 2012 data (right), tagged events, only. The signal component (magenta dashed line), the hadronic combinatoric background component (light green line) and the D^0 from B component (dark green line) are shown separately. The latter is separable in the δm distribution, only. The full mass fit model is shown as blue line.

4.5.4. B^+ background

The dominant B^+ background is controlled by the isolation BDT response, introduced in section 4.2. The fraction of B^+ events in data is determined data-driven.

The full isolation BDT response distribution is fitted on sWeighted data. Note, no cut is applied to the isolation BDT response here. Two components are modeled: B^0 signal decays and B^+ background decays. The shapes of the two components are taken from MC as histograms. The

Parameter	2011	2012
$\mu_m[\text{MeV}]$	1866.26 ± 0.01	1866.288 ± 0.010
$\sigma_{m,G1}[\text{MeV}]$	8.27 ± 0.07	8.32 ± 0.04
$\sigma_{m,G2}[\text{MeV}]$	16.2 ± 0.3	16.6 ± 0.2
$f_{m,G}$	0.717 ± 0.007	0.726 ± 0.004
$\sigma_{m,CB}[\text{MeV}]$	5.14 ± 0.05	5.07 ± 0.03
$\alpha_{m,CB}$	2.52 ± 0.02	2.56 ± 0.01
$n_{m,CB}$	0.00001	0.00001
$f_{m,CB}$	0.152 ± 0.009	0.150 ± 0.004
$m_0[\text{MeV}^{-1}]$	-0.00875 ± 0.0004	-0.00895 ± 0.0002
$\mu_{\delta m}[\text{MeV}]$	145.368 ± 0.006	145.375 ± 0.006
$\sigma_{\delta m,G1}[\text{MeV}]$	0.54 ± 0.02	0.52 ± 0.03
$\sigma_{\delta m,G2}[\text{MeV}]$	0.97 ± 0.02	0.94 ± 0.01
$f_{\delta m,G}$	0.23 ± 0.02	0.21 ± 0.01
$\sigma_{\delta m,RJ}[\text{MeV}]$	0.64 ± 0.02	0.67 ± 0.02
$\delta_{\delta m}$	1	1
$\gamma_{\delta m}$	-0.248 ± 0.02	-0.2197 ± 0.010
$f_{\delta m,RJ}$	0.36 ± 0.02	0.37 ± 0.02
$m_{\delta m}^{th}[\text{MeV}]$	139	139
$a_{\delta m}$	3.70 ± 0.08	2.78 ± 0.09
$b_{\delta m}$	-5.000 ± 0.07	-3.23 ± 0.2
$c_{\delta m}$	1.81 ± 0.05	1.66 ± 0.03
$\#nSig$	314093 ± 686	743612 ± 1049
$\#nBkg$	19643 ± 280	49261 ± 440
$\#nBkg_{D^0fB}$	92463 ± 518	232737 ± 803

Table 4.9.: $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode: 2011 and 2012 data fitted parameters determined in two-dimensional fits to the reconstructed δm and $m(D^0)$ mass distributions. For the fit parameter definitions see text. Parameters of the fits to the D^0 mass distribution are listed in the upper part of the table, parameters from δm fits are listed in the middle part. The extracted number of signal, combinatoric background and D^0 from B decays background events are listed in the lower part.

only free fit parameter is the fraction of B^+ events with respect to all events:

$$f_{B^+} = \frac{N_{B^+}}{N_{B^0} + N_{B^+}} \ , \tag{4.20}$$

where N_{B^0} is the number of B^0 events and N_{B^+} is the number of B^+ events.

The isolation BDT response distribution in sWeighted data is fitted by a maximum likelihood fit. The data is binned in 60 uniformly sized bins between -1 and +1. The signal and background shape histograms are prior to the fit reweighted to match the track multiplicity distribution in

data. Kinematic weights, as introduced in section 4.2, to correct for the wrong form factors in the D^* simulation are also applied.

Due to the limited statistics in simulation, a second reweighting is applied to the histograms in simulation after performing the fit to the data distribution. With this first estimate for f_{B^+} the sum of the B^0 signal and B^+ background histograms is reweighted to match the sWeighted data histogram. Then, the fit is performed again to extract the final f_{B^+} value. The second reweighting step changes the fit result on the 0.1% level, which indicates the robustness of the method against statistics induced differences in the distribution shapes.

Fits to the isolation BDT distributions in 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data are shown in figure 4.16. Tagged events only are considered.

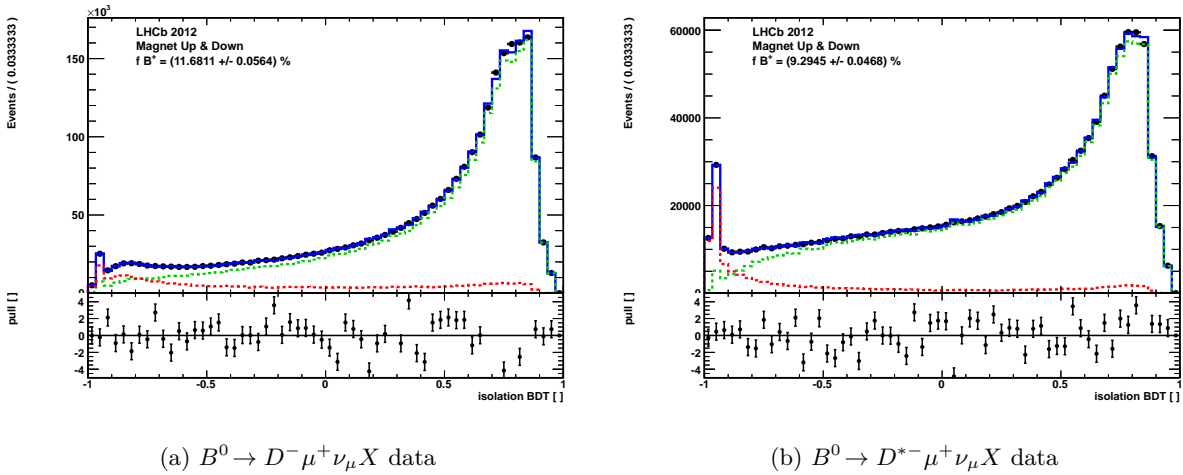(a) $B^0 \to D^- \mu^+ \nu_\mu X$ data

(b) $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data

Figure 4.16.: Isolation BDT response distributions fitted on 2012 sWeighted $B^0 \to D^- \mu^+ \nu_\mu X$ data (left) and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data (right).

B^+ fractions are determined per tagging category per data set. The results of all fits are given in table 4.10 for the $B^0 \to D^- \mu^+ \nu_\mu X$ mode and in table 4.11 for the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode.

The fitted isolation BDT distribution shape is integrated, to determine the B^+ fraction within the range of the cut applied to the isolation BDT distribution, as shown in table 4.4. The results are also shown in table 4.10 ($B^0 \to D^- \mu^+ \nu_\mu X$ mode) and in table 4.11 ($B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode). These B^+ fractions per tagging category within the cut range, are taken as fit inputs for the Δm_d fit. Possible deviations induced by the uncertainties on the branching ratios which enter the simulation cocktail composition are taken as systematic uncertainty on Δm_d, shown in section 4.9.

In $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data compared to $B^0 \to D^- \mu^+ \nu_\mu X$ data, the B^+ fraction is lower. Also the amount of combinatoric D background is lower in $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data. The reason is mostly given by the additional $\delta m = m(D^*) - m(D^0)$ constraint in this mode. The cut applied to the isolation BDT distribution reduces the B^+ fraction by a factor of ~ 3, while in $B^0 \to D^- \mu^+ \nu_\mu X$ data, this is only a factor ~ 2. In the end, the B^+ fraction is about 5 to 6% in $B^0 \to D^- \mu^+ \nu_\mu X$ data and about 3% in $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data.

Events	f_{B^+} [%] without cut	f_{B^+} [%] with cut
Cat. 0	12.59 ± 0.36	5.65 ± 0.16
Cat. 1	13.52 ± 0.29	5.96 ± 0.13
Cat. 2	13.05 ± 0.20	5.75 ± 0.09
Cat. 3	11.31 ± 0.11	4.82 ± 0.05
All	12.11 ± 0.09	5.25 ± 0.04
Cat. 0	13.48 ± 0.24	6.32 ± 0.11
Cat. 1	13.97 ± 0.19	6.47 ± 0.09
Cat. 2	12.67 ± 0.12	5.93 ± 0.06
Cat. 3	10.78 ± 0.07	4.89 ± 0.03
All	11.68 ± 0.06	5.37 ± 0.03

Table 4.10.: $B^0 \to D^- \mu^+ \nu_\mu X$ mode: B^+ fractions determined in 2011 data (top) and 2012 data (bottom) in four tagging categories, Cat. 0 - 3, and on the whole sample, All. Left: no cut is applied to the isolation BDT response, right: the cut is applied.

4.5.5. Evaluation of the B^+ Fraction in Simulation

As a cross check the B^+ fraction f_{B^+} is also calculated on simulation. Large uncertainties on semileptonic branching fractions, as shown in appendix B.1, propagate directly into the determination of f_{B^+}. Therefore, the method to measure f_{B^+} directly from data was developed.

Under the assumption of isospin symmetry, B^+ and B^0 have the same production cross section. Using the number of generated events from table 4.5 the total selection efficiencies for B^0 and B^+

Events	f_{B^+} [%] without cut	f_{B^+} [%] with cut
Cat. 0	9.08 ± 0.28	2.97 ± 0.09
Cat. 1	9.96 ± 0.23	3.25 ± 0.07
Cat. 2	9.30 ± 0.15	2.87 ± 0.05
Cat. 3	9.28 ± 0.09	3.00 ± 0.03
All	9.28 ± 0.07	2.97 ± 0.02
Cat. 0	9.23 ± 0.18	2.95 ± 0.06
Cat. 1	10.16 ± 0.15	3.20 ± 0.05
Cat. 2	9.54 ± 0.10	3.05 ± 0.03
Cat. 3	8.94 ± 0.06	2.89 ± 0.02
All	9.29 ± 0.05	2.99 ± 0.02

Table 4.11.: $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode: B^+ fractions determined in 2011 data (top) and 2012 data (bottom) in four tagging categories, Cat. 0 - 3, and on the whole sample, All. Left: no cut is applied to the isolation BDT response, right: the cut is applied.

events can be calculated, see table 4.12. The full selection is applied, including the cut on the isolation BDT response. Tagged events are considered, only.

Decay	2011	2012
$B^0 \to D^- \mu^+ \nu_\mu X$	0.557 ± 0.002 %	0.550 ± 0.002 %
$B^0 \to D^{*-} \mu^+ \nu_\mu X$	0.762 ± 0.002 %	0.756 ± 0.001 %
$B^+ \to D^- \mu^+ \nu_\mu X$	0.196 ± 0.003 %	0.205 ± 0.002 %
$B^+ \to D^{*-} \mu^+ \nu_\mu X$	0.186 ± 0.003 %	0.206 ± 0.002 %

Table 4.12.: Total efficiencies of each MC sample for selected, tagged events. Both magnet polarities and Pythia versions are added. The full off-line selection is applied, including the cut on the isolation BDT response.

In 2011 data the estimated B^+ fractions are:

$$f_{B^+}(B^0 \to D^- \mu^+ \nu_\mu X) = (6.44 \pm 1.15(1.15(\mathcal{B}), 0.09\,(\text{stat})))\% \tag{4.21}$$

$$f_{B^+}(B^0 \to D^{*-} \mu^+ \nu_\mu X) = (3.38 \pm 0.78(0.77(\mathcal{B}), 0.05\,(\text{stat})))\% \tag{4.22}$$

and in 2012 data:

$$f_{B^+}(B^0 \to D^- \mu^+ \nu_\mu X) = (6.81 \pm 1.21(1.21(\mathcal{B}), 0.05\,(\text{stat})))\% \tag{4.23}$$

$$f_{B^+}(B^0 \to D^{*-} \mu^+ \nu_\mu X) = (3.74 \pm 0.85(0.85(\mathcal{B}), 0.04\,(\text{stat})))\%. \tag{4.24}$$

The total uncertainty is split up into a systematic uncertainty induced by the large uncertainties on semileptonic branching ratios ($\mathcal{B}$) and into a statistical component (stat). The branching ratio uncertainty dominates the total uncertainty on f_{B^+}. If f_{B^+} is determined on samples, generated with different Pythia versions, the variation is small, less than 15% of the uncertainty on f_{B^+}. Given the large uncertainties estimated here, the B^+ fractions extracted from simulation and the ones from data agree perfectly.

4.5.6. B_s^0 and Λ_b^0 background

B_s^0 background occurs on a level of $(1.5 \pm 0.4)\%$. While, the amount of Λ_b^0 background is negligible $((0.1 \pm 0.1)\%)$, mainly due to the Λ_b^0 veto cut. Figure 4.17 shows a comparison of mixing asymmetry distributions, comparing B^0 signal, B^+, B_s^0 and Λ_b^0 background decays.

The mixing frequency in the B_s^0–$\bar{B}_s^0$ system, Δm_s, is a factor 34 faster than the B^0 mixing frequency Δm_d. Given the k factor damping, Δm_s is not resolvable in the semileptonic Δm_d fit. A constant mixing asymmetry at zero is present, see figure 4.17. The Λ_b^0 mixing asymmetry shows the same offset as the B^+ mixing asymmetry distribution. In the Δm_d fit the Λ_b^0 and B_s^0 components are absorbed in the non-oscillating B^+ component. A systematic uncertainty on Δm_d related to the limited knowledge of the B_s^0 and Λ_b^0 fraction in data is assigned, see section 4.9.
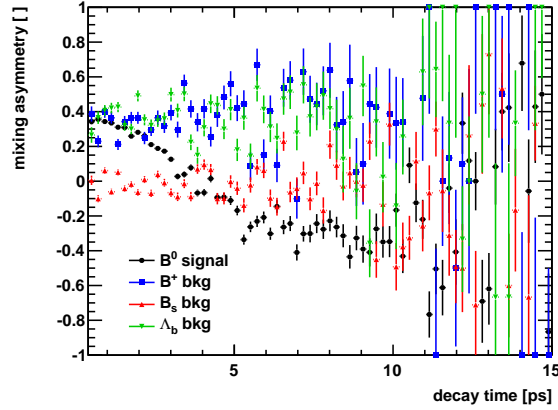
Figure 4.17.: Comparison of different b hadron samples: B^0 signal versus B^+, B_s^0 and Λ_b^0 background semileptonic decays. The mixing asymmetry is constant for all backgrounds. A mistag offset is present in the B^+ and Λ_b^0 background decays, while the B_s^0 sample does not show a mistag offset. It averages out due to the fast oscillation frequency Δm_s. To illustrate the difference between the distributions within the limited MC statistics, the bin errors are scaled by a factor of 0.5.

4.6. Probability Density Function of Δm_d

This section introduces the Probability Density Function ($\mathcal{PDF}$) used to determine Δm_d. A $\mathcal{PDF}$ is a prediction for the distribution of the measured quantities $\vec{x}$ depending on the parameters $\vec{\lambda}$, $\mathcal{PDF}(\vec{x}; \vec{\lambda})$. It is normalized to one, $\int \mathcal{PDF}(\vec{x}; \vec{\lambda}) d\vec{x} = 1$, to conserve the total probability. Experimentally, N measurements of $\vec{x}$ are obtained, $\vec{x}_e$ with $e = 1, ..., N$. The aim of the fit procedure is to find the parameters $\vec{\lambda}$, that maximize the probability of $\vec{x}_e$ to occur, given by the likelihood $\mathcal{L} = \prod_{e=1}^{N} \mathcal{PDF}(\vec{x}_e; \vec{\lambda})$. Numerically, it is more stable to minimize the negative logarithm of the likelihood

$$-\ln \mathcal{L} = -\sum_{e=1}^{N} \mathcal{PDF}(\vec{x}_e; \vec{\lambda}) \tag{4.25}$$

than to maximize it. The fit procedure is implemented in the ROOT MINUIT minimization package [109]. It is controlled by a customized interface, written in C++. The minimization procedure can be either run optimized on speed or on precision. Here, the choice has been set to precision[11]. After minimization the errors of the fit parameters are estimated using Gaussian approximation, implemented in the HESSE algorithm of the MINUIT package. Due to large statistics the data is binned prior to the fitting procedure. Hence, a binned maximum likelihood fit is performed to determine the oscillation frequency Δm_d.

---

[11] The MIGRAD minimization technique is run with strategy 2. This allows MINUIT to use more function calls, which improves the reliability of the minimization.

Two physical components are modeled by the $\mathcal{PDF}$: the B^0 signal component, which performs oscillations with the frequency Δm_d and the B^+ background component, which does not oscillate. The hadronic combinatorial D mass component is removed on a statistical basis following the sFit technique [78]. Prior to the fit procedure sWeights are determined in separate fits to the D mass, following the sPlot technique [79]. The sWeights are used to unfold the hadronic combinatorial D mass background component in the Δm_d fit, without modeling the decay time and mixing asymmetry distributions of this background. After the statistical removal of D backgrounds, the two decay channels are described by the same $\mathcal{PDF}$. All background components have been evaluated, shown in section 4.5.

The decay time distribution is binned in 140 bins of increasing width between 0 ps and 15 ps. The increasing bin width is chosen to describe properly the steep turn-on of the decay time distribution due to selection inefficiencies a small decay times. The events are categorized in four tagging categories of increasing mistag probability, as introduced in section 4.4. Two additional dimensions are defined, given by the product of the tagging decision and the muon charge: $q = +1$, where the b flavour has changed between production and decay, also called mixed decay, and $q = -1$, where the flavour has not changed between production and decay, also called unmixed decay.

There are nine parameters floated in the Δm_d fit:

- Δm_d

- four mistag probabilities $\omega_{B^0, i}$, with $i = 0, ..3$ for category i

- four mistag probabilities $\omega_{B^+, i}$, with $i = 0, ..3$ for category i.

One average measured mistag probability is fitted per predicted mistag probability category for both B^0 and B^+ components, separately. All other parameters, the B^0 and the B^+ lifetimes, the decay time resolution parameters and the decay time acceptance shapes, are fixed.

The $\mathcal{PDF}$ of B^0 decays depends on the state q and the decay time t. It is proportional to the exponential decay rate, determined by the B^0 lifetime $\tau_{B^0} = 1/\Gamma_d$, and to the mixing term, determined by the decay width difference $\Delta \Gamma_d$ and the mixing frequency Δm_d, see chapter 1:

$$\mathcal{P}_{B^0}(q, t) \propto e^{-\Gamma_d t}(\cosh(\frac{\Delta \Gamma_d}{2}t) - q\cos(\Delta m_d t)). \tag{4.26}$$

The state $q = +1$ indicates mixed decays and $q = -1$ unmixed decays. Since $\Delta \Gamma_d \sim 0$, equation 4.26 simplifies to:

$$\mathcal{P}_{B^0}(q, t) \propto e^{-\Gamma_d t}(1 - q\cos(\Delta m_d t)). \tag{4.27}$$

The tagging decision carries a certain mistag probability ω, which damps the mixing asymmetry by the tagging dilution factor $\mathcal{D}_{tag} = (1 - 2\omega)$, see section 3.4.2:

$$\mathcal{P}_{B^0}(q, t) \propto e^{-\Gamma_d t}(1 - q\mathcal{D}_{tag}\cos(\Delta m_d t)). \tag{4.28}$$

In case of the non-oscillating B^+ background, the mixing term vanishes and the $\mathcal{PDF}$ of B^+ decays simplifies to

$$\mathcal{P}_{B^+}(q, t) \propto \begin{cases} e^{-\Gamma_u t}(1 - \omega_{B^+}) & q = 1 \\ e^{-\Gamma_u t}(\omega_{B^+}) & q = -1 \end{cases} \tag{4.29}$$

where the mistag probability also appears, because the tagging algorithm does not know about the signal B meson type.

In real world experiments one has to include additional resolution effects into the $\mathcal{PDF}$. The effects, which are specific for semileptonic decays, are introduced in the following.

4.6.1. Collecting all resolution effects

As already introduced in section 4.1 the decay time of the B^0 candidate is not directly reconstructible due to the missing neutrino momentum. This enters over the reconstructed B^0 momentum as an uncertainty into the reconstructed decay time, given by

$$t = \frac{M(PDG)L}{pc},$$
(4.30)

where $M(PDG)$ is the B mass taken from the PDG [110]. The reconstructed flight distance L is determined by the distance of the primary and the B^0 decay vertices. The B momentum p, which is the sum of the D and μ four-momenta is incomplete. Both, the uncertainty on the flight distance and the uncertainty on the momentum have to be taken into account. Hence, the uncertainty on the decay time is computed by error propagation applied to equation 4.30:

$$\sigma_t = \sqrt{\left(\frac{M}{pc}\sigma_L\right)^2 + \left(t\frac{\sigma_p}{p}\right)^2}$$
(4.31)

The first term reflects the uncertainty on the reconstructed flight distance, while the second term reflects the uncertainty on the reconstructed momentum, which is fully determined by the k factor resolution in semileptonic decays. The k factor is a measure of the momentum fraction carried away by the neutrino: $k = \frac{p_{rec}}{p_{true}}$ [94].

These effects are folded into the $\mathcal{PDF}$ with a double convolution. First, the $\mathcal{PDF}$ is convolved with a triple Gaussian resolution function with a mean width of about $100\,\text{fs}$ to account for the reconstructed flight distance resolution. This convolution is done analytically:

$$\mathcal{P}_B(q,t) \xrightarrow{\text{decay time resolution}} \mathcal{P}_B(q,t-t') \otimes_t R(t').$$
(4.32)

The flight distance resolution $R(t')$ is determined from simulation, separately for each decay mode. Second, the $\mathcal{PDF}$ is multiplicatively convoluted with the k factor resolution $F(k)$, which is obtained from simulation as a binned histogram. The determination of both resolutions is shown in section 4.7.

$$\mathcal{P}_B(q,t-t') \otimes_t R(t') \xrightarrow{\text{k factor resolution}} [\mathcal{P}_B(q,t-t') \otimes_t R(t')] \otimes_k F(k)$$
(4.33)

Practically, this is just a sum over all bins in the k factor histogram:

$$[\mathcal{P}_B(q,t-t') \otimes_t R(t')] \otimes_k F(k) = \sum_i [\mathcal{P}_B(q,k_it-t') \otimes_t R(t')]k_iF_i \ .$$
(4.34)

The decay time t is multiplied by the k factor in each bin k_i and the total $\mathcal{PDF}$ is multiplied with the normalized bin content F_i. Another multiplication with $k_i = \mathrm{d}t_{\text{true}}/\mathrm{d}t_{\text{rec}}$ is necessary to

normalize the k factor convolution correctly. A detailed derivation of equation 4.34 is given in [111] calculating the convolution integrals.

To account for selection inefficiencies, that influence the shape of the decay time distribution at small decay times, e.g. IP cuts, and reconstruction effects, that cause inefficiencies at large decay times, a decay time acceptance distribution $a(t)$ is multiplied to the $\mathcal{PDF}$ from Eq. 4.34:

$$\mathcal{P}_B(q,t) = \mathcal{N} \times ([\mathcal{P}_B(q,t-t') \otimes_t R(t')] \otimes_k F(k)) \times a(t). \tag{4.35}$$

The term $a(t)$ is a normalized decay time acceptance histogram, which is obtained directly from sWeighted data. $\mathcal{N}$ accounts for the normalization of the $\mathcal{PDF}$. It is calculated by integrating over the decay time and summing over the k factors k_i:

$$\mathcal{N}(t_{min}, t_{max}) = \int_{t_{min}}^{t_{max}} dt \sum_{k_i} a(t)[\mathcal{P}_B(q,t-t') \otimes_t R(t')]F_i k_i \Delta k_i. \tag{4.36}$$

Finally, the full $\mathcal{PDF}$ is made of two parts, one for the oscillating B^0 signal component and one for the non-oscillating B^+ background component. The two parts are connected by the B^+ fraction f_{B^+}:

$$\mathcal{PDF}(q,t) = \mathcal{P}_{B^0}(q,t) + f_{B^+}\mathcal{P}_{B^+}(q,t). \tag{4.37}$$

The flight distance resolution parameters and the k factor histograms are determined separately for the B^0 and B^+ components, see next section.

4.7. Decay Time Description

The reconstruction of the decay time in semileptonic decays, without the momentum of the neutrino, is a challenging task. A statistical correction method, the k factor method, is used to obtain the decay time from the underestimated B momentum, as introduced in section 4.1.3. Consequently, the k factor resolution impacts the decay time resolution. The k factor method has been improved to account for dependencies on the true B decay time, that appear in the present precision of the Δm_d measurement. Both methods are explained here, first the standard k factor method, as introduced by CDF [94], then the improved method. The reconstructed B^0 flight distance also impacts the decay time resolution, explained below.

The decay time distribution is also affected by selection efficiency effects. The LHCb trigger uses mainly cuts on the particles' impact parameter to select B decays with the topology of displaced decay vertices. This leads to a less efficient reconstruction of B decays with small decay times. Additionally, due to the imperfect implementation of the track reconstruction in the LHCb vertex detector, a loss in reconstruction efficiency for high decay times is also present [77]. A decay time acceptance distribution accounts for both effects, given below.

The k factor and flight distance resolutions are determined separately for B^0 and B^+ decays using MC simulation. The decay time acceptance is obtained directly from sWeighted data. These three effects are modeled in the fit to extract Δm_d.

4.7.1. k Factor Method

The k factor, defined as quotient of the reconstructed B momentum divided by the true B momentum:

$$k = \frac{p_{rec}}{p_{true}}, \tag{4.38}$$

is a measure of the momentum fraction carried away by the neutrino, shown in figure 4.18. In B^+ background decays at least one additional track, a charged pion, is not reconstructed. This affects the k factor distribution. Hence, the k factor is determined separately for B^0 and B^+ decays, as shown in figure 4.18.
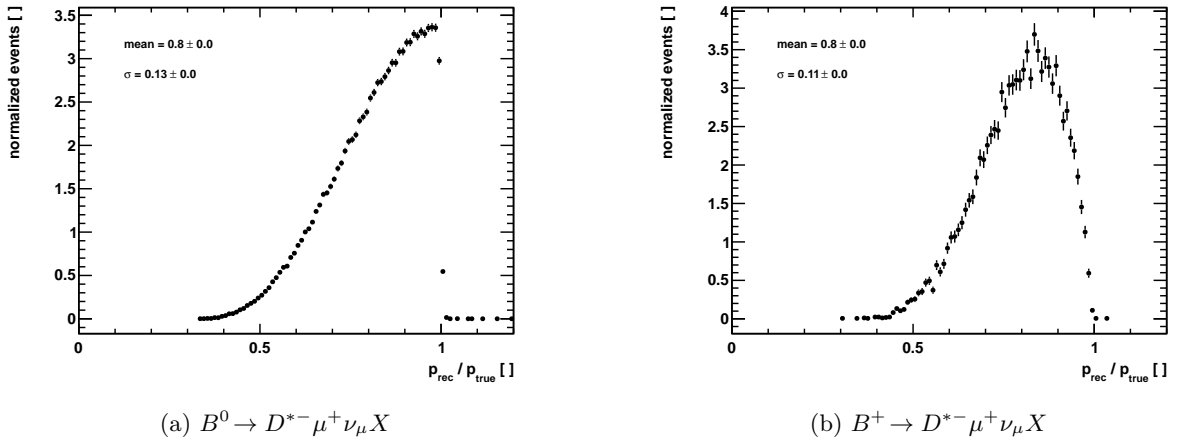

(a) $B^0 \to D^{*-}\mu^+\nu_\mu X$

(b) $B^+ \to D^{*-}\mu^+\nu_\mu X$

Figure 4.18.: K factor distributions determined in the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode 2012 MC (left) and the $B^+ \to D^{*-}\mu^+\nu_\mu X$ 2012 MC (right). These distributions are averaged over all reconstructed B masses M_{rec}.

The method was first introduced by the CDF collaboration [94], to evaluate a correction for the partially reconstructed semileptonic B decay time. Besides the reconstructed decay time, the missing neutrino momentum also affects the reconstructed B mass. Therefore, the k factor is determined in bins of the reconstructed B mass, shown in figure 4.19. The mean of each k factor histogram per B mass bin is plotted. The uncertainty per bin in y direction is the RMS value of each k factor histogram, not the uncertainty of the mean value, to illustrate the widths of the k factor distributions in different bins of the reconstructed B mass.

The k factor dependence on the reconstructed B mass M_{rec} is described by an empirical second order polynomial:

$$k_{\text{fit}}(M_{\text{rec}}) = p_0 + p_1\left(M_{\text{rec}}\right) + p_2\left(M_{\text{rec}}\right)^2 , \tag{4.39}$$

where $p_{0,1,2}$ are the fit parameters and k_{fit} is the expected k factor based on the visible B mass. The fit parameters are given in table 4.13 for the B^0 samples and in table 4.14 for the B^+ samples.

The k factor dependence on the reconstructed B mass, given in equation 4.39, is directly multiplied to the reconstructed B decay time. The result is illustrated in figure 4.20, showing the corrected B decay time and the corrected k factor histogram. The reconstructed B momentum

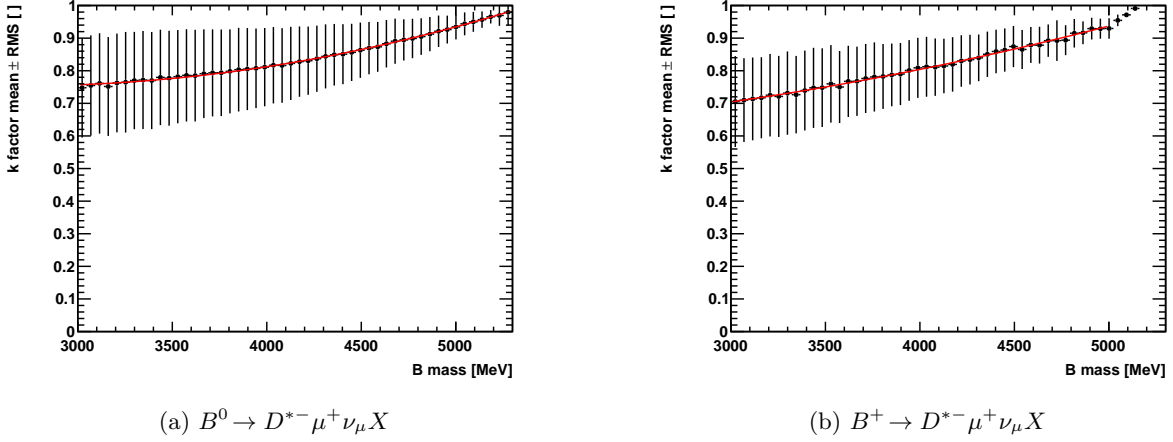(a) $B^0 \to D^{*-}\mu^+\nu_\mu X$ (b) $B^+ \to D^{*-}\mu^+\nu_\mu X$

Figure 4.19.: Mean k factor in bins of the visible B mass for tagged events in $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode 2012 MC. Left: signal component, right: $B^+ \to D^{*-}\mu^+\nu_\mu X$ background component. The plotted k factor mean error is given as RMS of each k factor histogram per bin.

is too small, hence the reconstructed B decay time too large. By rescaling the decay time with the k factor depending on the reconstructed B mass, the corrected reconstructed decay time distribution matches the true decay time distribution. However, not only the mean value but also the width of the k factor distribution needs to be taken into account. The width after correction for $k_{fit}(M_{\rm rec})$ on event-by-event basis is shown in figure 4.20. By doing so, the mean of the scaled k factor distribution is now at one, the width is slightly increased, from 0.13 to 0.15. In this

$B^0 \to D^-\mu^+\nu_\mu X$	2011	2012
$p_0\ [10^{-1}]$	9.7 ± 5.0	9.6 ± 5.2
$p_1\ [10^{-4}]$	-1.8 ± 2.2	-1.7 ± 2.4
$p_2\ [10^{-8}]$	3.3 ± 2.5	3.3 ± 2.6

$B^0 \to D^{*-}\mu^+\nu_\mu X$	2011	2012
$p_0\ [10^{-1}]$	10.3 ± 5.2	9.9 ± 5.8
$p_1\ [10^{-4}]$	-2.0 ± 2.4	-1.8 ± 2.7
$p_2\ [10^{-8}]$	3.5 ± 2.6	3.3 ± 3.0

Table 4.13.: Parameters of the fit to the mean k factor in bins of the visible B mass for tagged events in $B^0 \to D^-\mu^+\nu_\mu X$ mode (upper table) and $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode (lower table), 2011 MC (left) and 2012 MC (right).

$B^+ \to D^- \mu^+ \nu_\mu X$	2011	2012
p_0 $[10^{-1}]$	6.8 ± 6.5	7.8 ± 6.7
p_1 $[10^{-4}]$	-0.5 ± 3.1	-1.1 ± 3.2
p_2 $[10^{-8}]$	2.1 ± 3.8	2.7 ± 3.9

$B^+ \to D^{*-} \mu^+ \nu_\mu X$	2011	2012
p_0 $[10^{-1}]$	6.8 ± 7.0	6.0 ± 6.9
p_1 $[10^{-4}]$	-0.6 ± 3.4	-0.2 ± 3.4
p_2 $[10^{-8}]$	2.2 ± 4.1	1.7 ± 4.0

Table 4.14.: Parameters of the fit to the mean k factor in bins of the visible B mass for all events in $B^+ \to D^- \mu^+ \nu_\mu X$ mode (upper table) and $B^+ \to D^{*-} \mu^+ \nu_\mu X$ mode (lower table), 2011 B^+ MC (left) and 2012 B^+ MC (right).

way, the dependence of the k factor on the B mass can be transported into the Δm_d fit, without adding an extra B mass dimension.

(a) corrected decay time

(b) corrected k factor
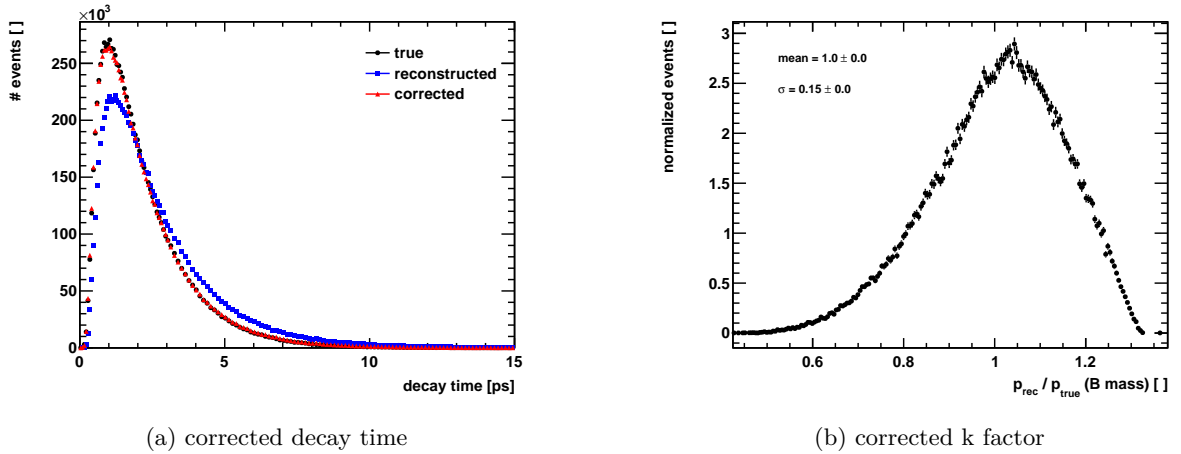
Figure 4.20.: Left: the reconstructed B decay time (blue squares) is corrected by the mean k factor as a function of the B mass, $k_{fit}(M_{\rm rec})$, (red triangles) to match the true decay time distribution (black circles). Right: spread of the k factor distribution after event-by-event correction to the mean value $k_{fit}(M_{\rm rec})$. Tagged events in $B^0 \to D^{*-} \mu^+ \nu_\mu X$ 2012 MC are shown.

The above presented k factor method is analogue to the original method introduced by the CDF experiment. CDF has measured Δm_d with a precision between 0.05 and $0.1\,\mathrm{ps}^{-1}$. However, this analysis aims for a precision of the order of $0.003\,\mathrm{ps}^{-1}$. It turned out, that the k factor method introduces a bias on Δm_d of the order of $0.004\,\mathrm{ps}^{-1}$. The reason for this bias is, that the
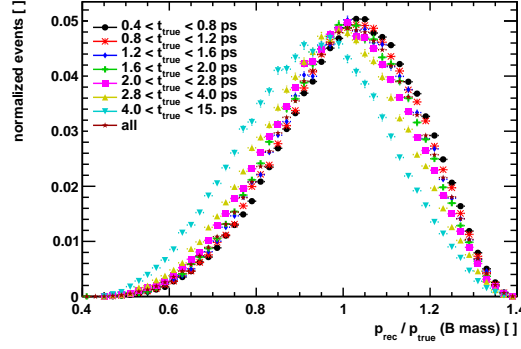
Figure 4.21.: Corrected k factor distributions in different true B decay time (t_{true}) bins.

k factor correction depends also on the true B decay time, see figure 4.21. The corrected k factor distributions are plotted for different true B decay time bins. Introducing this dependence in a toy MC and ignoring the dependence in the fit reproduces the observed bias seen in fits to large statistics MC samples. As a result of this study, the k factor method is improved to avoid the bias on Δm_d, shown in the next section.

4.7.2. Improved k Factor Method

The main idea of the advancement to the k factor method is to take the dependence of the k factor on the true decay time into account. Similar to the standard k factor method, where the k factor was determined in bins of the reconstructed B mass, the k factor is now determined in bins of the true decay time. The dependence on the reconstructed B mass is dropped. A schematic shown in figure 4.22 gives an overview of the basic concept of the improved k factor method.

One k factor histogram is created per true decay time bin. The binning is chosen such that it matches the binning of the Δm_d fit, 140 bins with increasing widths. Note, that the k factor histograms are filled per true decay time bin. Figure 4.23 shows 2×10^4 factor histograms filled in the first 20 true decay time bins.

Due to the exponential shape of the decay time distribution, the statistics fade away in the high decay time bins. It is not possible to get sufficient information on the k factor in these upper bins. Therefore, neighboring bins are added, depending on the decay time bin. Technically, the following scheme is used: t_{true} bin $\pm n$ neighbor bins, where $n = 1, 2, 3, 6, 12, 18, 24$ for $t_{true} > 1.38, 2.1, 2.85, 3.95, 5.21, 6.46, 7.72$ ps. This results in overlapping k factor histograms, but the relative information on the position in the true decay time distribution is kept. Analogue, to the standard k factor method, the mean values of the single k factor histograms are plotted versus the true decay time, see figure 4.24.

The distribution is fitted by an empirical function, consisting of 4 exponential distributions:

$$k_{\text{fit}}(t_{\text{true}}) = p_0 + p_1 e^{p_2 t_{\text{true}}} + p_3 e^{p_4 t_{\text{true}}} + p_5 e^{p_6 t_{\text{true}}} + p_7 e^{p_8 t_{\text{true}}}. \tag{4.40}$$

The $k_{\text{fit}}(t_{\text{true}})$ value is used to correct the reconstructed decay time and to shift the k factor histograms to be centered around 1. To keep the normalization of the Δm_d $\mathcal{PDF}$ it is also folded
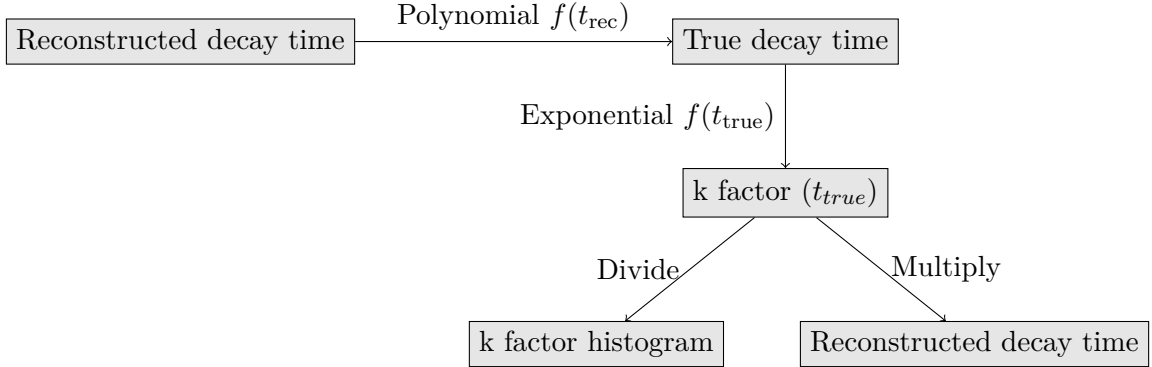
Figure 4.22.: Schematic to illustrate the working principle of the improved k factor method. The basic concept of the standard k factor method is kept. The mean k factor is determined in bins of the true B decay time, before it was the reconstructed B mass. This dependence is multiplied as a correction to the reconstructed decay time. The k factor histogram is corrected by the average $k_{fit}(t_{true})$. The novel point is the translation between the true decay time and the reconstructed decay time via a polynomial function $f(t_{rec})$.

into the 1.40×10^5 factor histograms per true decay time interval. The first 20 of these corrected k factor histograms are shown in figure 4.25.

However, fitting Δm_d on real data requires a matching between the true decay time bin and the reconstructed decay time, which is the only quantity available in data. Only after this matching it is possible to apply $k_{\text{fit}}(t_{\text{true}})$ to the reconstructed decay time. Therefore, a translation function between the reconstructed and the true decay time is created. It is determined from the 2 dimensional histogram, shown in figure 4.26, which is projected onto the y axis to obtain the true decay time as a function of the reconstructed decay time.

It is fitted by a second order polynomial:

$$t_{\text{true}}(t_{\text{rec}}) = p_0 + p_1 \left(t_{\text{rec}}\right) + p_2 \left(t_{\text{rec}}\right)^2 \ . \tag{4.41}$$

The robustness of this reconstructed to true decay time translation is shown in section 4.8.

Another novelty compared to the standard k factor method is, that one $k_{\text{fit}}(t_{\text{true}})$ scaled k factor histogram is used per corresponding time bin by the Δm_d fit $\mathcal{PDF}$. Figure 4.27 highlights the difference between the standard k factor method and the improved k factor method at this point. With this two stepped k factor lookup correction method and usage of 140 properly scaled k factor histograms, Δm_d is fitted without bias on signal B^0 simulation, shown in section 4.8.
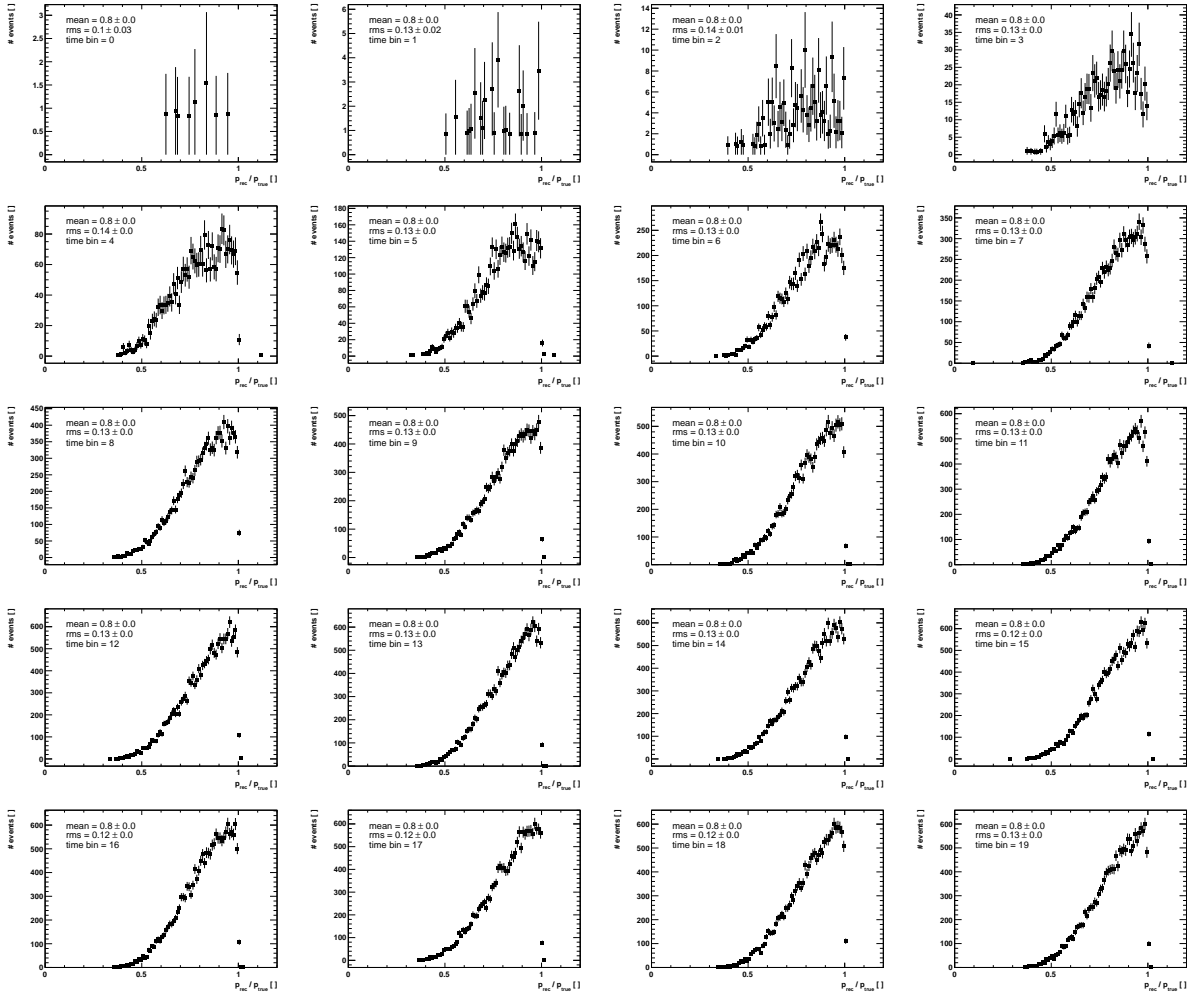
Figure 4.23.: 2×10^4 factor histograms filled in the first 20 bins out of 140 bins of the true decay time for 2012 $B^0 \to D^{*-}\mu^+\nu_\mu X$ MC.
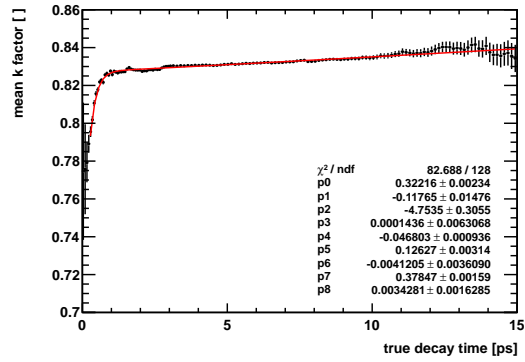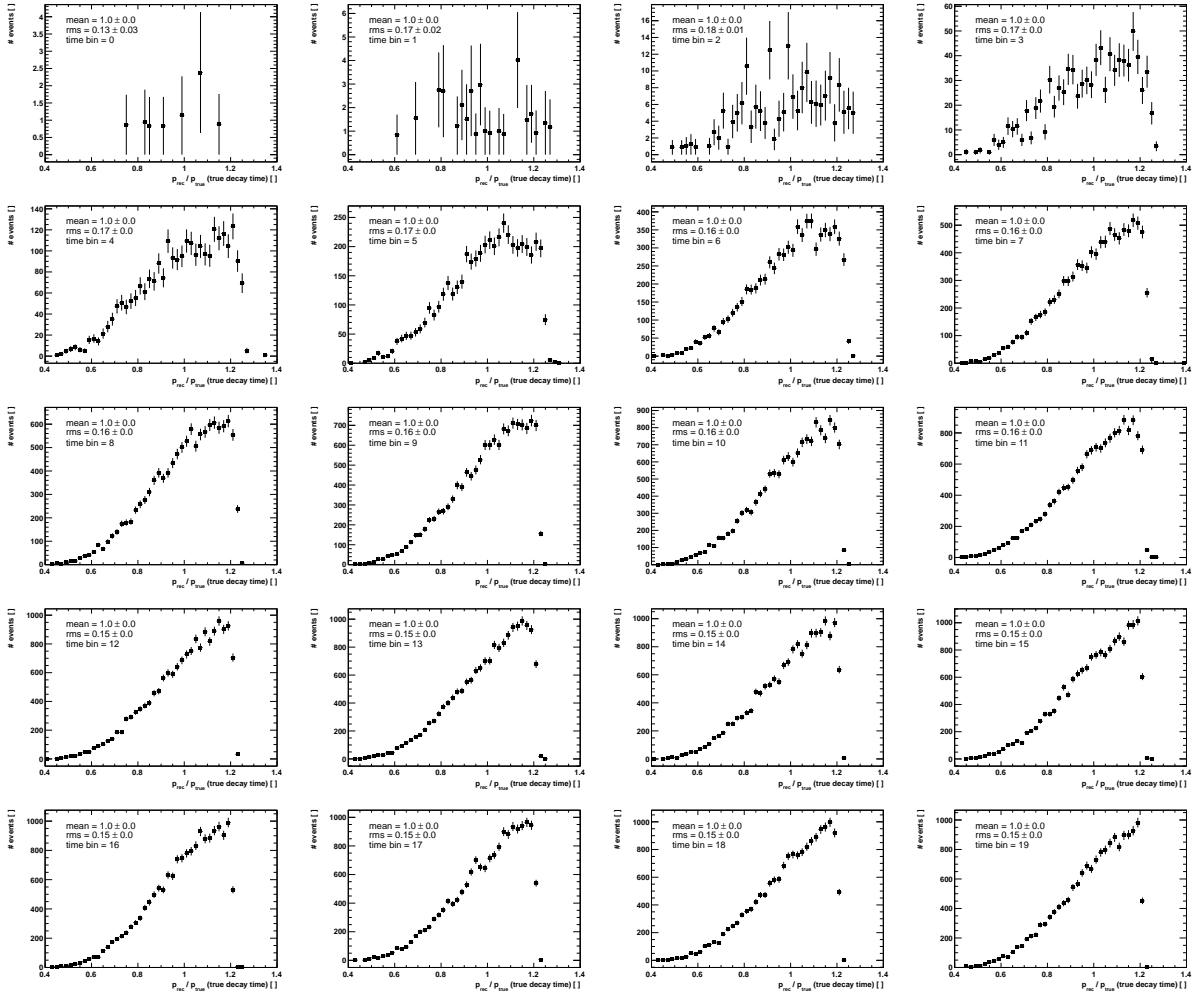
Figure 4.24.: Fitted mean k factor distribution in MC true decay time bins of tagged events in $B^0 \to D^{*-}\mu^+\nu_\mu X$ 2012 MC. The distribution is fitted by an empirical function given in equation 4.40. The uncertainty per bin in y direction is given by the uncertainty on the mean value of the k factor histogram.

Figure 4.25.: 20 corrected k factor histograms in the first 20 bins out of 140 bins of the true decay time for 2012 $B^0 \to D^{*-}\mu^+\nu_\mu X$ MC.

(a) 2 dimensional t_{true} vs t_{rec} distribution

(b) projection on the y-axis
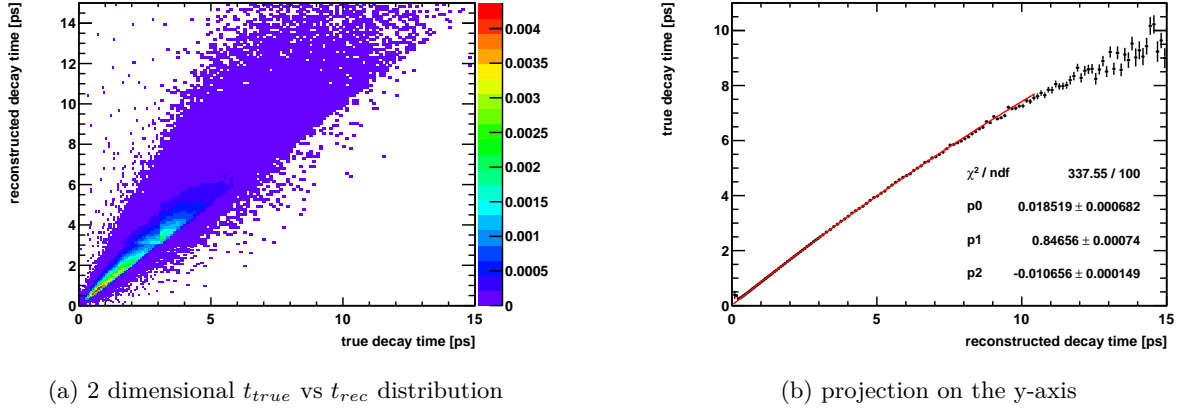
Figure 4.26.: MC true versus reconstructed decay time of tagged events in $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode 2012 MC. Left: the 2 dimensional distribution, right: projected on the y-axis. The dependence of the reconstructed decay time on the true decay time is fitted by a second order polynomial.

Standard method   Improved method
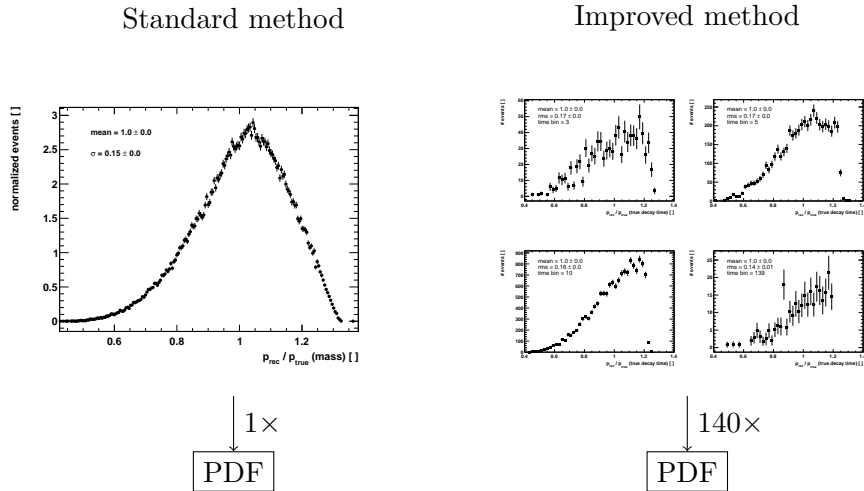
Figure 4.27.: Schematic to highlight the difference between the standard k factor method and the improved k factor method. Formerly, one average k factor histogram was put into the Δm_d $\mathcal{PDF}$, now, 1.40×10^5 factor histograms, filled per true decay time bin, are put into the Δm_d $\mathcal{PDF}$. The k factor correction is determined in bins of the true B decay time in the new method, while it was done as a function of the reconstructed B mass before.

4.7.3. Flight Distance Resolution

The flight distance resolution is determined from simulation. It is calculated as the difference of the reconstructed flight distance entering the decay time minus the true flight distance entering the decay time determination:

$$\sigma_L = \frac{L_{rec} \cdot M(PDG)}{p_{MC} \cdot c} - \frac{L_{MC} \cdot M(PDG)}{p_{MC} \cdot c}. \tag{4.42}$$

L_{rec} is the reconstructed, while L_{MC} is the MC true flight distance of the B^0. $M(PDG)$ is the PDG value of the B^0 mass. c is the speed of light and p_{MC} is the MC true B^0 momentum. The effect of the reconstructed momentum is taken out here to disentangle the different resolution effects.

The flight distance resolution is fitted by a triple Gaussian distribution:

$$\begin{aligned} \mathcal{S}(t) = & f_1(1 - f_2) \cdot G(t; \mu, \sigma_1) \\ & + (1 - f_1)(1 - f_2) \cdot G(t; \mu, \sigma_2) \\ & + (1 - f_1)f_2 \cdot G(t; \mu, \sigma_3). \end{aligned} \tag{4.43}$$

All three Gaussian distributions share the same mean value μ, but have different width values $\sigma_{1,2,3}$. They are connected by the fractions f_1 and f_2. The flight distance resolution is determined separately for B^0 signal decays and B^+ background decays, see figure 4.28 showing the $B^0 \rightarrow D^{*-}\mu^+\nu_\mu X$ mode in 2012 MC. For B^0 decays, tagged events, only, are considered, however, due to lack of statistics in B^+ decays, all events, tagged and untagged, are considered.

(a) $B^0 \rightarrow D^{*-}\mu^+\nu_\mu X$ (b) $B^+ \rightarrow D^{*-}\mu^+\nu_\mu X$
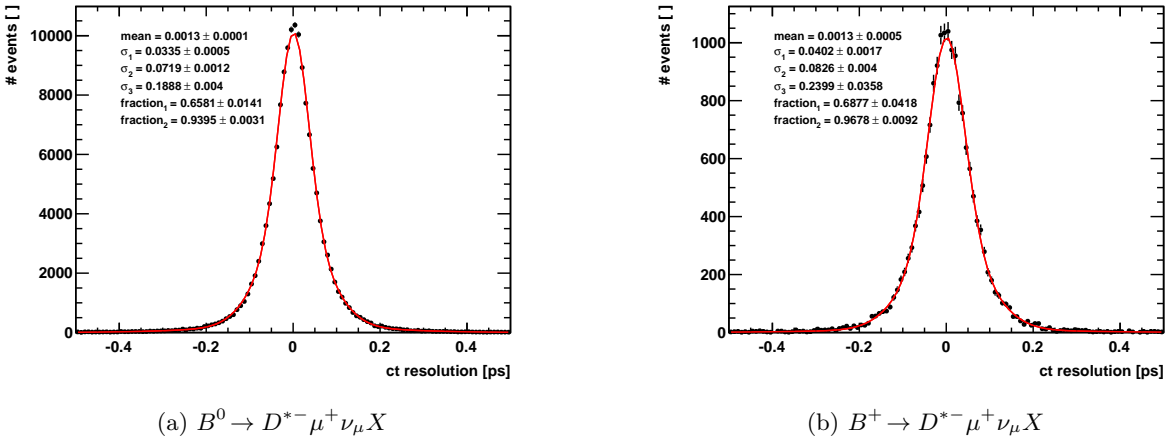
Figure 4.28.: Decay time resolution for tagged events in $B^0 \rightarrow D^{*-}\mu^+\nu_\mu X$ mode 2012 MC (left) and for $B^+ \rightarrow D^{*-}\mu^+\nu_\mu X$ 2012 MC (right). The effect of the reconstructed flight distance on the decay time resolution is evaluated. It is fitted by a triple Gaussian.

The flight distance resolution is determined on every sample, separately: $B^0 \rightarrow D^-\mu^+\nu_\mu X$ 2011 and 2012, $B^0 \rightarrow D^{*-}\mu^+\nu_\mu X$ 2011 and 2012, as well as for the corresponding B^+ background

samples. The obtained fit parameters are given in appendix B.3. An average flight distance resolution per sample is determined using

$$< \sigma >= \sqrt{f_1(1 - f_2) \cdot \sigma_1^2 + (1 - f_1)(1 - f_2) \ \cdot \sigma_2^2 + (1 - f_1)f_2 \cdot \sigma_3^2}. \tag{4.44}$$

The uncertainties of the single fit parameters are propagated using Gaussian error propagation. Table 4.15 shows the average width values for the different samples.

mode	year	$< \sigma >$ [ps]
$B^0 \to D^- \mu^+ \nu_\mu X$	2011	0.102 ± 0.006
$B^0 \to D^- \mu^+ \nu_\mu X$	2012	0.112 ± 0.005
$B^0 \to D^{*-} \mu^+ \nu_\mu X$	2011	0.119 ± 0.003
$B^0 \to D^{*-} \mu^+ \nu_\mu X$	2012	0.108 ± 0.003
$B^+ \to D^- \mu^+ \nu_\mu X$	2011	0.146 ± 0.035
$B^+ \to D^- \mu^+ \nu_\mu X$	2012	0.103 ± 0.011
$B^+ \to D^{*-} \mu^+ \nu_\mu X$	2011	0.085 ± 0.002
$B^+ \to D^{*-} \mu^+ \nu_\mu X$	2012	0.132 ± 0.022

Table 4.15.: Mean width of the flight distance resolution for tagged events in $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ signal (upper part) and $B^+ \to D^- \mu^+ \nu_\mu X$ and $B^+ \to D^{*-} \mu^+ \nu_\mu X$ background (lower part) for 2011 and 2012 MC, separately.

As can be seen in figure 4.7, the flight distance resolution has practically no impact on the determination of Δm_d. However, it is properly modeled to obtain a fit projection on the decay time distribution. Therefore, the obtained flight distance resolution fit parameters are directly used as inputs for the Δm_d fit. A systematic uncertainty on Δm_d caused by an underestimated flight distance resolution from MC is evaluated in section 4.9. An increase of the resolution width by 50% has only a small impact on Δm_d.

4.7.4. Decay Time Acceptance

The decay time acceptance is implemented in the Δm_d fit by a normalized histogram obtained from sWeighted data. It is determined as the quotient of the decay time distribution of events that pass the full selection, here sWeighted data, divided by the theory decay time distribution, according to a k factor convoluted true decay rate:

$$a(t) = \frac{T_{\text{reconstructed}}(t)}{T_{\text{theoretical}}(t)}. \tag{4.45}$$

Figure 4.29 shows the decay time acceptance distribution for $B^0 \to D^- \mu^+ \nu_\mu X$ 2012 sWeighted data.

Decay time acceptance distributions are obtained separately for 2011 and 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode sWeighted data. The same decay time acceptance is used for the
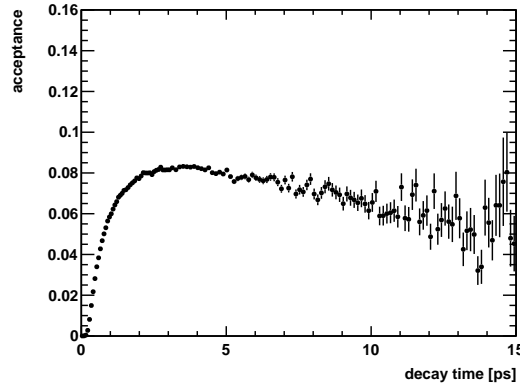
Figure 4.29.: Decay time acceptance histogram determined on tagged events in $B^0 \to D^- \mu^+ \nu_\mu X$ mode 2012 sWeighted data.

B^0 and the B^+ components. Possible differences due to the slightly different lifetimes of the B^0 and the B^+ meson and different k factor resolutions are expected to be negligible An according systematic uncertainty on Δm_d is evaluated.

The measurement of Δm_d is not sensitive to the exact fit of the decay time distribution in first order, because the mixing asymmetry is calculated by the quotient of decay time distributions, where the nominator is the difference of decay time distributions of mixed and unmixed events, while the denominator is the sum of both. As long as decay time acceptance effects impact both decay time distributions in the same way, all mis-modeling effects of the decay time distribution cancel. However, the decay time acceptance model is mainly implemented to obtain nice fit projections on the decay time distribution.

4.8. Validation of the Δm_d Fit Procedure using MC Simulation

To validate the fitting procedure B^0 signal MC is fitted. The statistical precision on the fitted Δm_d value is enhanced by cheating the flavour tagging. The MC true production flavour is used as "tag". In 4% of the events the "tag" is reversed to introduce a mistag probability of 4%. The statistical precision is further increased by fitting all events, not tagged events only. The k factor correction, as well as the flight distance resolution and decay time acceptance, are determined for each sample separately, depending on the requirement to fit all events or tagged events only. The Δm_d fit projections are shown in figure 4.30 for the $B^0 \to D^- \mu^+ \nu_\mu X$ and in figure 4.31 for the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode.

The corresponding fitted Δm_d values in both decay modes, both tagging requirement settings and both years are given in table 4.16. The standard k factor method gives a significant bias on Δm_d to higher frequency values. On $B^0 \to D^{*-} \mu^+ \nu_\mu X$ 2012 MC the highest statistics are available. With a statistical precision of $0.6 \, \text{ns}^{-1}$ Δm_d is fitted about 3 to $4 \, \text{ns}^{-1}$ too high, compared to $\Delta m_d = 0.507 \, \text{ps}^{-1}$ used as generation value in the MC simulation. The improved k factor method
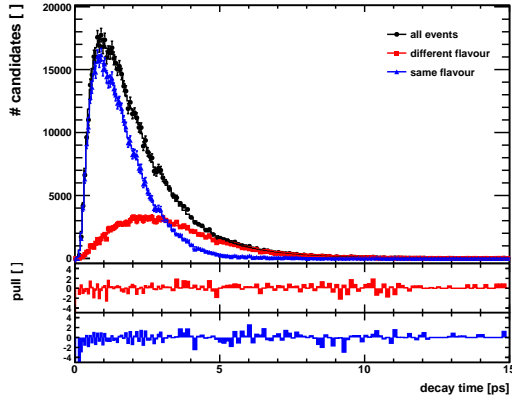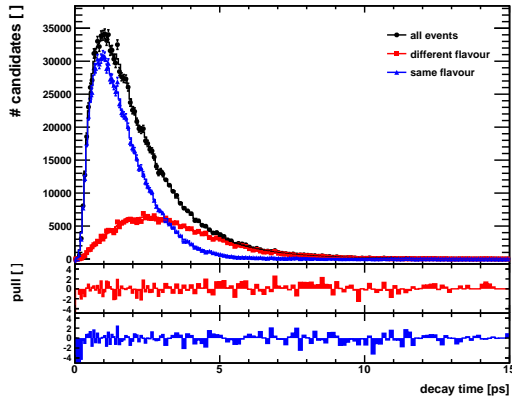
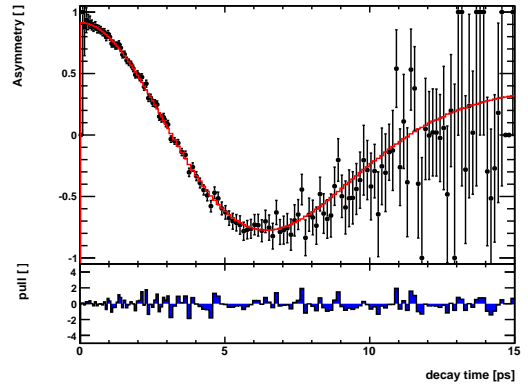(a) 2011 $B^0 \to D^- \mu^+ \nu_\mu X$ decay time
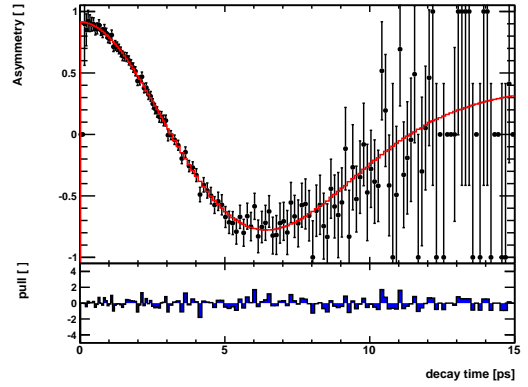
(b) 2011 $B^0 \to D^- \mu^+ \nu_\mu X$ mixing asymmetry

(c) 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ decay time

(d) 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ mixing asymmetry

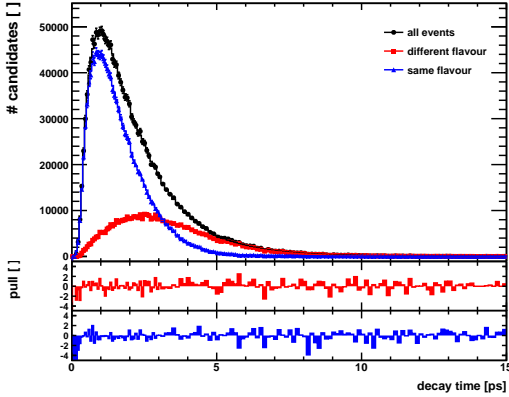Figure 4.30.: Fit projections of the Δm_d fit to 2011 (upper row) and 2012 (lower row) $B^0 \to D^- \mu^+ \nu_\mu X$ signal MC with cheated tagging.

removes this bias. However, the picture is not fully consistent. On 2011 MC, there are quite large variations.

In order to be able to judge, how significant the actual variations are and if they are actually related to the k factor treatment at all, Δm_d is also determined from fits to the MC true decay time. The k factor correction is not applied in this case. Both tests are performed on each sample: fitting all events of the sample and fitting the subset of tagged events, only, while cheating the mistag probability in both cases. Figure 4.32 shows the mixing asymmetry projections of all events in all four samples.

The mixing asymmetry distributions are not significantly damped due to the absence of the k factor correction. However, the mixing frequency is obtained from a fit to one oscillation period, only. The relatively slow oscillation in the B^0–$\overline{B}^0$ system compared to the B^0 lifetime does not allow to average over multiple periods to obtain a precisely determined frequency. Additionally,
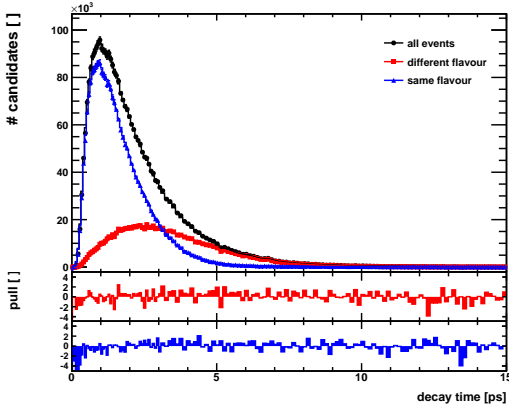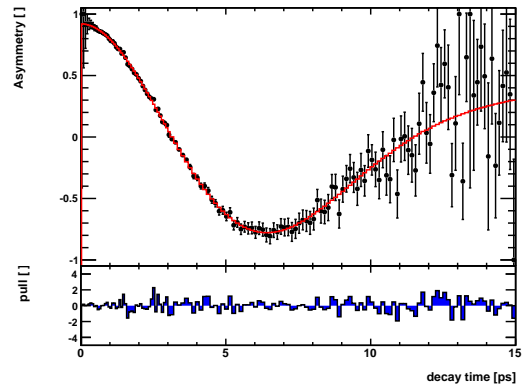
(a) 2011 $B^0 \to D^{*-}\mu^+\nu_\mu X$ decay time
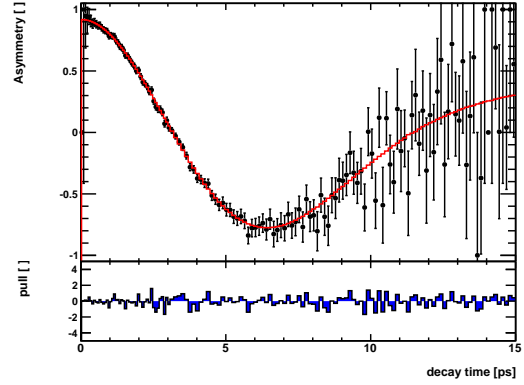
(b) 2011 $B^0 \to D^{*-}\mu^+\nu_\mu X$ mixing asymmetry

(c) 2012 $B^0 \to D^{*-}\mu^+\nu_\mu X$ decay time

(d) 2012 $B^0 \to D^{*-}\mu^+\nu_\mu X$ mixing asymmetry

Figure 4.31.: Fit projections of the Δm_d fit to 2011 (upper row) and 2012 (lower row) $B^0 \to D^{*-}\mu^+\nu_\mu X$ signal MC with cheated tagging.

the second maximum of the oscillation is less constrained than the first maximum due to the exponential decay. Hence, the fit becomes sensitive to the statistics of the sample. The obtained fit values are also given in table 4.16.

For comparison, the MC samples are additionally fitted under data fit conditions using real tagging. Each sample is fitted for one average mistag probability and in four tagging categories, fitting for each category one average mistag probability. The results obtained using the standard k factor method are compared to the ones using the improved k factor method, see table 4.17. The statistical precision on Δm_d obtained from these fits is not sensitive to the observed bias.

This gives valuable additional information of two kinds. First, the statistical sensitivity to fit Δm_d from the given MC sample is obtained. Comparing the statistical uncertainties on Δm_d between the fits using a k factor correction and the fits to t_{true}, that do not require any k factor correction, shows, that they are slightly larger in the first case. Hence, the k factor resolution

(a) 2011 $B^0 \to D^- \mu^+ \nu_\mu X$ all

(b) 2011 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ all

(c) 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ all
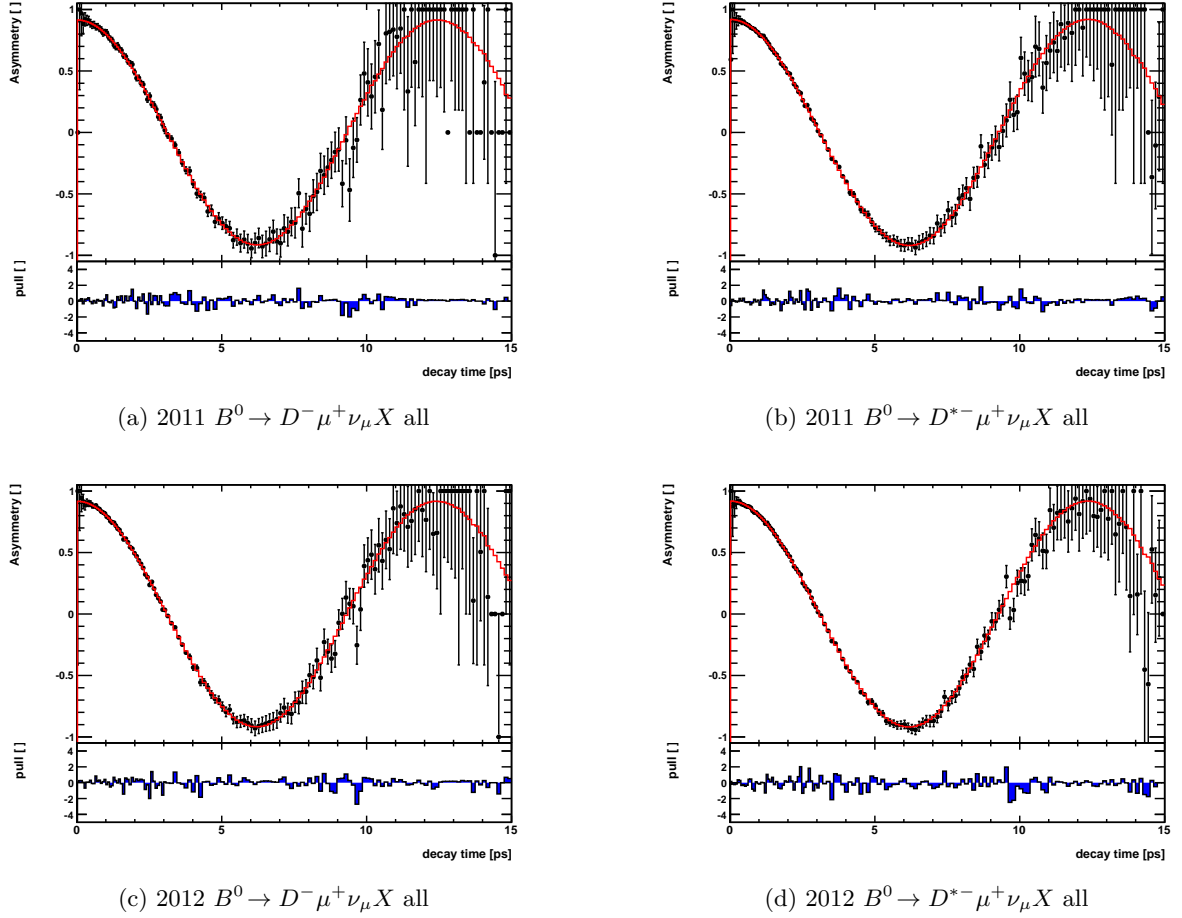
(d) 2012 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ all

Figure 4.32.: Mixing asymmetry fit projections of the Δm_d fit to the true decay time of all events in the sample using cheated tagging (without any k factor correction). $B^0 \to D^- \mu^+ \nu_\mu X$ (left) and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ (right) in 2011 signal MC (upper row) and 2012 signal MC (lower row).

does not have a big impact on the statistical uncertainty of Δm_d. Second, the MC samples show some statistical fluctuations in the determination of Δm_d. Therefore, the different bias values for 2011 and 2012 MC are within this statistically allowed range.

The k factor correction is determined from these statistically limited MC samples. Hence, a test is performed to determine the effect of the statistical limitation of the k factor correction. In this test, the bin contents of the k factor histograms are varied according to Gaussian fluctuations. To each k factor histogram bin content a Gaussian random number is added, which is drawn with a mean value of zero and a width value of the statistical uncertainty of the bin content. This procedure is repeated a hundred times, each time fitting the same sample of events. The results are shown in table 4.18, evaluated for each decay mode and year. Each fit result, obtained

mode	year	events	standard k factor $[\,\mathrm{ps}^{-1}\,]$	improved k factor $[\,\mathrm{ps}^{-1}\,]$	fit to t_{true} $[\,\mathrm{ps}^{-1}\,]$
D^-	2011	all	0.50619 ± 0.00143	0.50212 ± 0.00145	0.50509 ± 0.00127
D^-	2011	tagged	0.51190 ± 0.00248	0.50753 ± 0.00254	0.50961 ± 0.00220
D^-	2012	all	0.51011 ± 0.00103	0.50684 ± 0.00104	0.50638 ± 0.00091
D^-	2012	tagged	0.50835 ± 0.00173	0.50493 ± 0.00176	0.50473 ± 0.00151
D^{*-}	2011	all	0.51116 ± 0.00087	0.50587 ± 0.00088	0.50828 ± 0.00077
D^{*-}	2011	tagged	0.50987 ± 0.00149	0.50423 ± 0.00150	0.50709 ± 0.00130
D^{*-}	2012	all	0.51211 ± 0.00063	0.50841 ± 0.00064	0.50853 ± 0.00055
D^{*-}	2012	tagged	0.51131 ± 0.00104	0.50792 ± 0.00106	0.50781 ± 0.00091

Table 4.16.: Fitted values of Δm_d obtained from fits to all and to tagged events, only, using cheated tagging, in both decay modes and both data taking periods 2011 MC and 2012 MC. The $\Delta m_d = 0.507\,\mathrm{ps}^{-1}$ value was used during MC generation. For comparison Δm_d is also obtained from fits to the true decay time.

mode	year	mistag probability	standard k factor $[\,\mathrm{ps}^{-1}\,]$	improved k factor $[\,\mathrm{ps}^{-1}\,]$
D^-	2011	average	0.50028 ± 0.00920	0.49736 ± 0.00945
D^-	2011	per category	0.49998 ± 0.00816	0.49537 ± 0.00822
D^-	2012	average	0.50787 ± 0.00710	0.50596 ± 0.00721
D^-	2012	per category	0.51279 ± 0.00621	0.50896 ± 0.00627
D^{*-}	2011	average	0.50616 ± 0.00556	0.49799 ± 0.00563
D^{*-}	2011	per category	0.50893 ± 0.00491	0.50362 ± 0.00496
D^{*-}	2012	average	0.51457 ± 0.00421	0.51143 ± 0.00427
D^{*-}	2012	per category	0.51316 ± 0.00368	0.50889 ± 0.00372

Table 4.17.: Fitted values of Δm_d obtained from fits to tagged events, only, using real tagging, in both decay modes and both data taking periods 2011 MC and 2012 MC. The statistical precision on Δm_d is not sensitive to the bias on Δm_d.

with a statistically variated k factor histogram, is subtracted from the nominal fit result on that sample. The width of this distribution indicates the uncertainty on Δm_d related to the statistical limitation of the k factor correction.

This test reassures, that the statistical fluctuations on Δm_d are not primarily caused by the k factor correction. The variations on Δm_d are of the order $2 - 4\,\mathrm{ns}^{-1}$, while the statistical uncertainty on the k factor correction causes variations on Δm_d below $0.5\,\mathrm{ns}^{-1}$. Table 4.19 gives an overview of the variations on Δm_d obtained by the different methods.

The Δm_d results from the fits to t_{true} deviate from the generation value of $\Delta m_d = 0.507\,\mathrm{ps}^{-1}$ due to the before mentioned statistical reasons. Therefore, the differences between the Δm_d fits to the reconstructed decay time, using either the standard k factor method or the improved method,

| mode | year | standard k factor | | improved k factor | |
		mean $[\,\mathrm{ns}^{-1}\,]$	width $[\,\mathrm{ns}^{-1}\,]$	mean $[\,\mathrm{ns}^{-1}\,]$	width $[\,\mathrm{ns}^{-1}\,]$
D^-	2011	-0.06	0.37	$+0.06$	0.45
D^-	2012	-0.06	0.25	-0.00	0.25
D^{*-}	2011	$+0.02$	0.22	-0.34	0.21
D^{*-}	2012	-0.01	0.17	$+0.01$	0.14

Table 4.18.: Determination of the uncertainty on Δm_d induced by the limited statistics of the sample, where the k factor correction is determined from. The sample is restricted to tagged events, only, because this reflects the final Δm_d fit conditions on data. 100 fits are performed per sample, each fit result is subtracted from the nominal fit result. The mean and width values of the resulting distributions are given.

mode	year	events	fit t_{true} - 0.507 $[\,\mathrm{ns}^{-1}\,]$	improved k - t_{true} $[\,\mathrm{ns}^{-1}\,]$	standard k - t_{true} $[\,\mathrm{ns}^{-1}\,]$	(stand. - impr.)k $[\,\mathrm{ns}^{-1}\,]$
D^-	2011	all	-1.91	-2.97	1.10	4.07
D^-	2011	tagged	$+2.61$	-2.08	2.29	4.37
D^-	2012	all	-0.62	$+0.46$	3.73	3.27
D^-	2012	tagged	-2.27	$+0.20$	3.62	3.42
D^{*-}	2011	all	1.28	-2.41	2.88	5.29
D^{*-}	2011	tagged	0.09	-2.86	2.78	5.64
D^{*-}	2012	all	1.53	-0.12	3.58	3.70
D^{*-}	2012	tagged	0.81	$+0.11$	3.50	3.39

Table 4.19.: Comparison of the Δm_d fit values obtained using different methods. The first column shows the statistical goodness of the MC sample by computing the difference between Δm_d obtained by fitting the true decay time minus the generation value $\Delta m_d = 0.507\,\mathrm{ps}^{-1}$. The improved k factor method gives very close results for Δm_d on the 2012 MC samples. However, in 2011 MC the results are limited by statistical fluctuations. A comparison of the standard k factor method with the fit to t_{true} shows, that this method gives a biased result for Δm_d of $\sim 3.5\,\mathrm{ns}^{-1}$ for 2012 MC. The deviations obtained from 2011 MC are dominated by statistical fluctuations. The direct comparison of the Δm_d results obtained by the standard and the improved k factor methods gives the real size of the bias on Δm_d independent of the statistical fluctuations in the 2011 MC samples.

are computed with respect to the fits to the true decay time. On 2012 MC, these values agree very well, at the $0.2\,\mathrm{ns}^{-1}$ level for $B^0 \to D^- \mu^+ \nu_\mu X$ and at the $0.1\,\mathrm{ns}^{-1}$ level for the $B^0 \to D^{*-} \mu^+ \nu_\mu X$. However, the 2011 MC samples, show a deviation of $2.1\,\mathrm{ns}^{-1}$ for $B^0 \to D^- \mu^+ \nu_\mu X$ here and $2.9\,\mathrm{ns}^{-1}$ for $B^0 \to D^{*-} \mu^+ \nu_\mu X$. These deviations are taken as uncertainty on the knowledge, that the bias on Δm_d is removed by the improved k factor method. It is determined on statistically independent samples. Hence, it will be treated accordingly in the estimate of the uncertainties of the weighted average of the combined Δm_d measurement on data.

4.8.1. Demonstration of the Robustness of the t_{rec} to t_{true} Translation

A crucial point of the improved k factor method is the translation between the reconstructed and the true decay time by the second order polynomial. During MC generation the B^0 decay width $\Gamma_d = 0.658\,\mathrm{ps}^{-1}$, which is the inverse of the B^0 lifetime, was used as input. What happens to the determination of Δm_d, if this value is different in real data? Does the improved k factor method still work? To test this, Γ_d is changed within the current PDG uncertainty, $\tau_{B^0} = (1.519 \pm 0.005)\,\mathrm{ps}$ [72], by changing the fraction of short- to long-lived decays within a MC sample. Then, this variated sample is fitted with the default improved k factor correction. The difference between the Δm_d fit result from this test and the default fit result gives the systematic uncertainty on Δm_d.

The largest MC sample ($B^0 \to D^{*-} \mu^+ \nu_\mu X$, 2012) is used for this test. The B^0 decay width is fitted with $\Gamma_d = 0.658 \pm 0.002\,\mathrm{ps}^{-1}$, shown in figure 4.33. A very simple exponential decay model is taken here, choosing the minimum and maximum fit range borders such, that no resolution or reconstruction inefficiencies affect the decay width fit result. Of course, real lifetime measurements are a lot more difficult. True short-lived B mesons can be reconstructed as long-lived B mesons by mismatches in the reconstruction, such as wrong decay vertex assignments or randomly combined tracks. The probability to reconstruct true long-lived B mesons as short-lived B mesons instead is smaller, due to the better reconstructible signature in the detector of significantly displaced vertices.
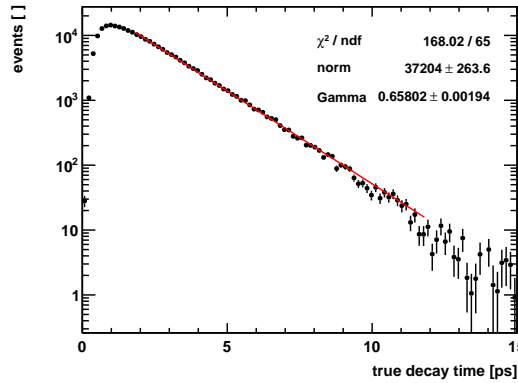

Figure 4.33.: True decay time distribution of 2012 MC events in the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode. A single exponential fit gives the B^0 decay width $\Gamma_d = (0.658 \pm 0.002)\,\mathrm{ps}^{-1}$.

First, the sample is enriched with long-lived events by randomly rejecting short-lived events until the mean B^0 decay width is smaller by 1σ of the PDG uncertainty. Second, the sample is enriched with short-lived events by randomly rejecting long-lived events until the mean B^0 decay width is larger by 1σ of the PDG uncertainty. Both scenarios are shown in figure 4.34, the corresponding B^0 decay widths fits are overlaid.
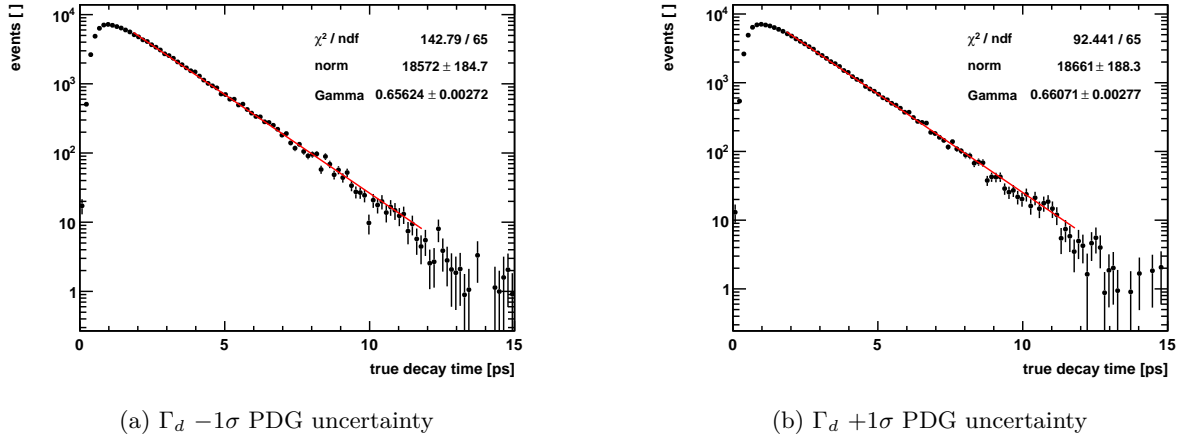

(a) Γ_d -1σ PDG uncertainty

(b) Γ_d $+1\sigma$ PDG uncertainty

Figure 4.34.: True decay time distribution of 2012 MC events in the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode. Left: the sample is enriched with long-lived events by randomly rejecting short-lived events, the mean B^0 decay width is 1σ of the PDG uncertainty smaller, right: the other way around.

It is necessary to reject half of the events, in order to increase or decrease the B^0 decay width by 1σ. Δm_d is fitted on these samples. The fit projections are shown in figure 4.35. The corresponding fit values are given in table 4.20.

$B^0 \to D^{*-}\mu^+\nu_\mu X$ 2012 MC sample	Γ_d [ps^{-1}]	Δm_d [ps^{-1}]
original	0.658 ± 0.002	0.50841 ± 0.00064
$\Gamma_d - 1\sigma$ short-lived	0.656 ± 0.003	0.50899 ± 0.00090
$\Gamma_d + 1\sigma$ long-lived	0.661 ± 0.003	0.50762 ± 0.00091

Table 4.20.: Parameters of the Δm_d fit applied to 2012 MC events in the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode, using cheated tagging. The sample is artificially enriched with long-lived or short-lived true B decay times. To increase or reduce the B^0 decay width by 1σ of the PDG uncertainty, the sample size is reduced by a factor of 2. The translation between the reconstructed and the true decay time is not sensitive to the exact knowledge of Γ_d.

The deviation in Δm_d is $-0.79\,\mathrm{ns}^{-1}$ for the long-lived enhanced sample and $+0.58\,\mathrm{ns}^{-1}$ for the short-lived enhanced sample. To judge, whether this is a significant deviation the statistical
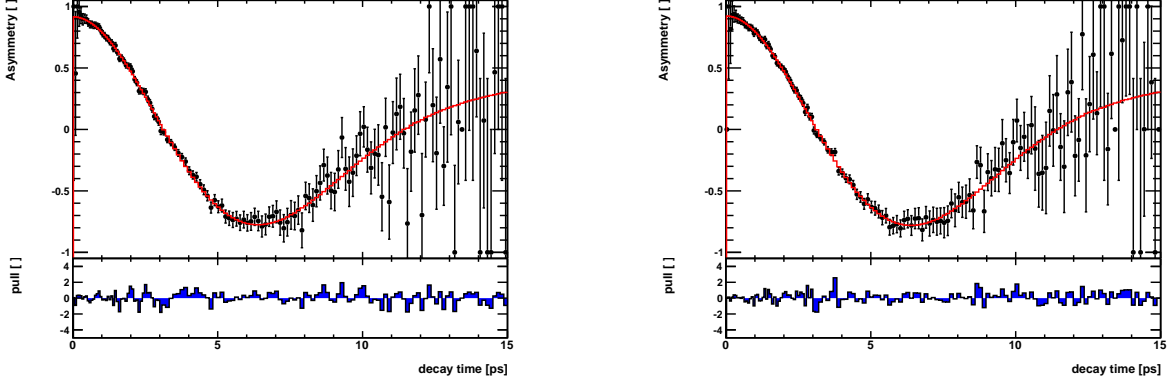
(a) Oscillation projection on events with $\Gamma_d - 1\sigma$

(b) Oscillation projection on events with $\Gamma_d + 1\sigma$

Figure 4.35.: Mixing asymmetry fit projections of the Δm_d fit to 2012 MC events in the $B^0 \rightarrow D^{*-}\mu^+\nu_\mu X$ mode using cheated tagging. Left: Γ_d is lowered by 1σ, right: Γ_d is increased by 1σ by changing the sample composition.

component of the deviation is evaluated. Given A with uncertainty σ_A is the measurement on the whole sample. B with uncertainty σ_B is the measurement that gives the larger deviation. Then C with uncertainty σ_C is a hypothetical measurement on the subsample of sample A, that is not contained in sample B. B and C are statistically independent measurements, so that

$$\frac{A}{\sigma_A^2} = \frac{B}{\sigma_B^2} + \frac{C}{\sigma_C^2}, \tag{4.46}$$

or

$$\frac{1}{\sigma_A^2} = \frac{1}{\sigma_B^2} + \frac{1}{\sigma_C^2}. \tag{4.47}$$

Then

$$\sigma_A^2 = \frac{\sigma_B^2 \sigma_C^2}{\sigma_B^2 + \sigma_C^2}. \tag{4.48}$$

The statistical uncertainty $\sigma(A - B)$ is obtained by

$$
\begin{aligned}
A - B &= \left(\frac{B}{\sigma_B^2} + \frac{C}{\sigma_C^2} \right) \sigma_A^2 - B \\
&= \frac{B\sigma_C^2 + C\sigma_B^2}{\sigma_B^2 + \sigma_C^2} - \frac{B(\sigma_B^2 + \sigma_C^2)}{\sigma_B^2 + \sigma_C^2} \\
&= \frac{(C - B)\sigma_B^2}{\sigma_B^2 + \sigma_C^2}
\end{aligned}
\tag{4.49}
$$

From line 1 to 2, equation 4.48 has been used. Given, that B and C are statistically independent measurements, the statistical uncertainty $\sigma(A - B)$ is

$$\sigma(A - B) = \sqrt{\sigma_B^2 + \sigma_C^2} \frac{\sigma_B^2}{\sigma_B^2 + \sigma_C^2} = \frac{\sigma_B^2}{\sqrt{\sigma_B^2 + \sigma_C^2}} \tag{4.50}$$

The statistically independent uncertainty σ_C is determined using equation 4.47:

$$\sigma_C = \frac{\sigma_A \sigma_B}{\sqrt{\sigma_B^2 - \sigma_A^2}} = 0.00090\,\mathrm{ps}^{-1}. \tag{4.51}$$

Inserting $\sigma_B = 0.91\,\mathrm{ns}^{-1}$ and $\sigma_C = 0.90\,\mathrm{ns}^{-1}$ yields $\sigma(A - B) = 0.65\,\mathrm{ns}^{-1}$. Hence, the upper variation on Δm_d is compatible with a statistical variation of 1.2σ and the lower variation on Δm_d is 0.9σ. This is not significant. Thus, the improved k factor method is not limited by the knowledge of the B^0 lifetime.

4.9. Systematic Uncertainty on Δm_d

This section shows the determination of the systematic uncertainty on the measurement of Δm_d. All contributing effects are summarized in table 4.23, given in the end of this section. Systematic uncertainties can bias the result of Δm_d on data. In contrast to the statistical uncertainty, given by the statistical power of the data set, they are evaluated on pseudo experiment data. These so called toy experiments are generated according to the Δm_d fit $\mathcal{PDF}$ using a random number generator [12]. Then, the toy data sets are fitted. By generating and fitting $\mathcal{O}(1000)$ of toy data sets the statistical component of the fit uncertainty vanishes but any systematic effect remains.

To evaluate the systematic uncertainty on Δm_d due to a given uncertainty, either the generation $\mathcal{PDF}$ is modified to account for this uncertainty or the fit $\mathcal{PDF}$ is modified. Both approaches are equivalent. Modifying the fit $\mathcal{PDF}$ has the advantage, that the generated set of toy samples can be used in multiple systematic tests. That saves significantly CPU hours and carbon dioxide. This approach is mainly used here.

A visualization method for systematic effects is to plot three types of histograms. First, the fitted Δm_d values are subtracted from the generated Δm_d values. This difference is filled in a histogram. If a systematic effect is present, a shift will show up. Second, the statistical uncertainties of each toy sample fit are filled in another histogram. The mean value of this histogram corresponds to the statistical precision of the Δm_d fit on data. By construction, the generation statistics, generated events per toy data set, are chosen such, that this condition is fulfilled. Third, the pull quantity is calculated, which is the shift value per toy data set divided by its statistical uncertainty. The pull distribution will take Gaussian shape, when the number of toy data sets is large enough. It has a mean value of 0 and a width of 1. When a significant systematic effect occurs, it will show up as shift of $1, 2, ..., n\sigma$ in the pull distribution by moving the pull mean value to $\pm 1, 2, ..., n$.

The Δm_d fit on data requires for every decay mode and data taking year different input parameters, such as flight distance resolution, decay time acceptance, k factor correction and B^+ fractions per tagging category. The measured mistag probabilities per category per signal or background component are determined individually as well. Therefore, the toy data sets are generated with the fit conditions of one of the 4 data sets, the one giving the best statistical precision on Δm_d, which is the $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ 2012 data set. The $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ mode is also chosen, because it carries larger uncertainties of the sample composition in terms of semileptonic feed down decay modes. Analogue to the data fit, 9 parameters are floated: Δm_d and four

---

[12]The ROOT TRandom3 class is used.

parameters for the mistag ω in B^0 and in B^+ decays, respectively. All other parameters are fixed. This reflects the setup of an sWeighted fit on data, with a statistical precision of $2.8\,\mathrm{ns}^{-1}$, where 1.6×10^6 events are generated.

After a sanity test, generating and fitting events with the same $\mathcal{PDF}$, four classes of systematic tests are performed. First, the general $\mathcal{PDF}$ settings are validated, such as the binning, the D mass sWeights and B^0 flight distance resolution and decay time acceptance. The uncertainties on the z-scale and momentum scale are also considered. Second, the B^+ background component is tested. A crucial input to the fit is the B^+ fraction per category, $f_{B^+,i}$, determined externally by a fit to the isolation BDT distribution on sWeighted data. The uncertainty on $f_{B^+,i}$ is determined by varying the branching ratios of the decay modes entering the MC cocktail and refitting the isolation BDT distribution on sWeighted data with the BR variated MC templates. One σ of this uncertainty is propagated on Δm_d using toy experiments. Additionally, the B^+ flight distance resolution and decay time acceptance are varied. In the default fit, the decay time acceptance is assumed to be the same for B^0 and B^+ decays. The impact of this assumption is tested here. Third, other types of background are introduced such as B_s^0 and $B^0 \to DD$ decays. A Λ_b^0 component is found to be negligible. The fourth class of tests is about the k factor distribution. The impact on Δm_d by possible deviations in the sample composition is tested. The fully reconstructed mode $B^+ \to J/\psi K^+$ is used to compare B momentum differences between data and MC. Differences are propagated on the B^0 momentum within the k factor distribution.

4.9.1. Validation of the Δm_d Fit Procedure using Toy Experiments

The fit procedure is validated by generating toy experiments according to the Probability Density Function $\mathcal{PDF}$, used for the fit. As introduced above, shift, statistical uncertainty and pull distributions for all floating parameters are obtained. Figure 4.36 shows the corresponding distributions for Δm_d. The shift and pull distributions are fitted by Gaussian distributions to extract the mean and width values. The pull distributions of the other eight floated fit parameters are shown in figure B.1 (the four B^0 ω parameters) and in figure B.2 (the four B^+ ω parameters), given in appendix B.4.

A small systematic shift on Δm_d of $(-0.5 \pm 0.1)\,\mathrm{ns}^{-1}$ is observed. The mean of the pull distribution is shifted by -0.24 ± 0.03, while the pull width is compatible with one, 0.98 ± 0.03. Thus, the statistical uncertainties of the single Δm_d fits are evaluated correctly, however the fitted values of Δm_d are slightly shifted. This effect is significant. Several explanations are possible.

First, the limited statistics of the generated samples are tested as source of this systematic shift. However, a statistical source is excluded by generating another set of toy samples with 10 times more statistics per sample: 1.6×10^6 events $\to 1.6 \times 10^7$ events. The shift stays identical in size.

Second, the effect of the binning choice is considered as reason for this systematic intrinsic shift. The toy data is generated unbinned by drawing random numbers according to the shape of the fit $\mathcal{PDF}$. However, the fit is performed on binned data. Going from continuous to binned data, the initial dimensionality is reduced, information is lost. Additionally, bin borders can have an impact on the frequency measurement via the zero crossings of the oscillation. The effect of the choice of binning is evaluated below. It partially explains the shift on Δm_d.

Thus, another possibility is, that there is a difference in the technical implementation of the generation model and the fit model. Therefore, the absolute size of the observed shift is added

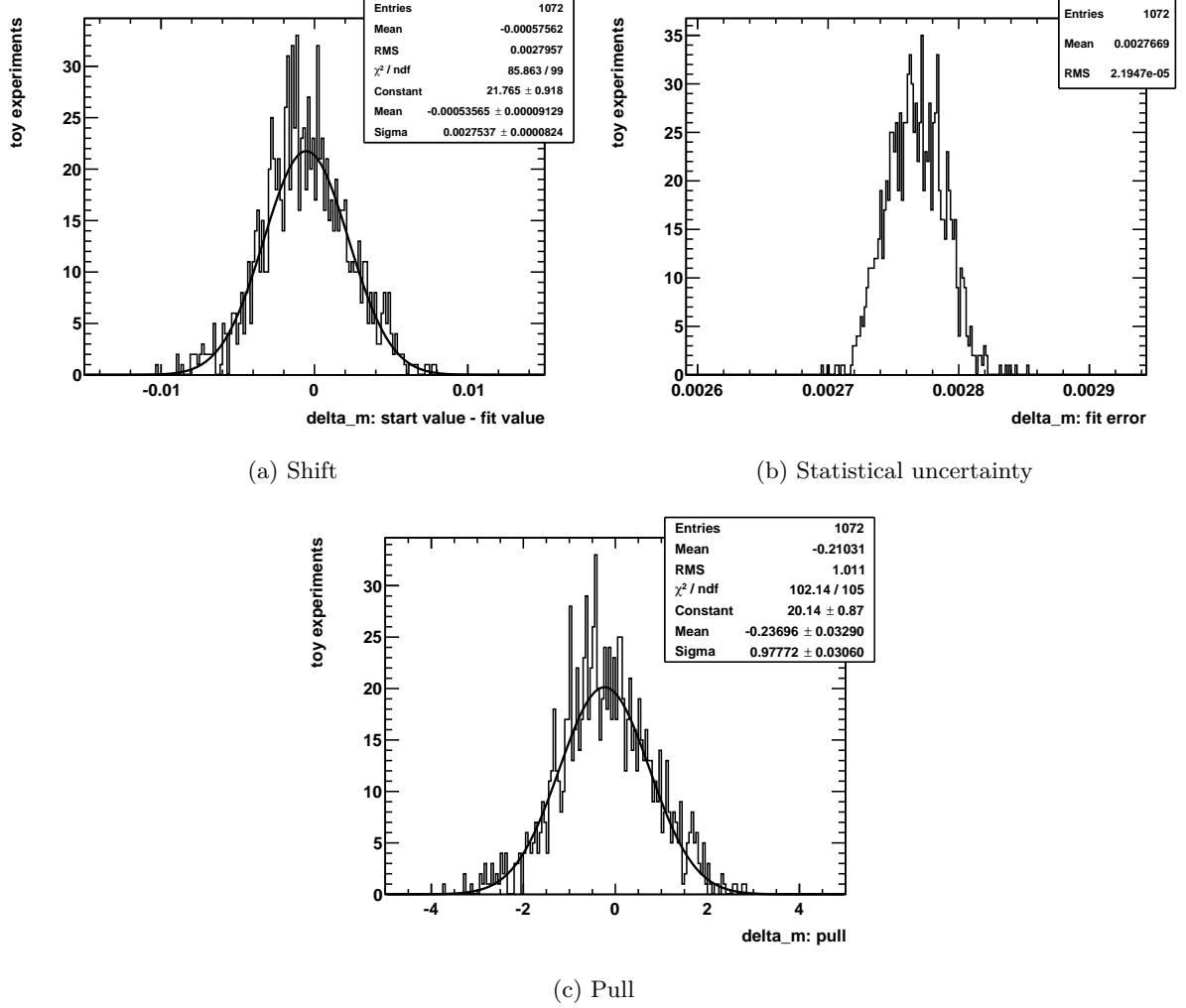(a) Shift

(b) Statistical uncertainty

(c) Pull

Figure 4.36.: Shift (a), statistical uncertainty (b) and pull (c) distributions on Δm_d of a set of toys samples, generated and fitted with 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ data set conditions. On Average, the statistical uncertainty of a single toy fit is $2.8\,\text{ns}^{-1}$. The shift and pull distributions are fitted by Gaussian distributions. A systematic shift on Δm_d of $-0.5 \pm 0.1\,\text{ns}^{-1}$ is observed. The reason for this shift is evaluated in the following. It is treated as a systematic. All following systematic variations are evaluated relative to this shift.

to the list of systematic uncertainties on Δm_d. All following systematic variations are evaluated relative to this intrinsic shift.

Choice of Binning

The default binning scheme uses 140 bins in the decay time with increasing widths to display the fast turn on, induced by cuts on the flight distance of the B^0, of the decay time distribution properly. Compared to the B^0 lifetime the B^0–$\bar{B}^0$ oscillation frequency is relatively slow. Thus, the mixing asymmetry is reconstructed for one oscillation period, only. The frequency measurement, given by the two zero crossings of one period, can be sensitive to the choice of bin borders close to these points. Naturally, the bin widths are chosen to be smaller in the high statistics beginning of the decay time distribution and wider, when statistics fade away, given by the exponential decay law.

Initially, the analysis was performed on 70 bins in the decay time. However, it soon turned out to be a non-optimal choice. The widths of the higher decay time bins were too wide to display the needed sensitivity to the second zero crossing of the oscillation. Hence, the number of bins was increased. The optimal sensitivity is reached with 140 bins.

The systematic uncertainty on Δm_d due to the choice of binning in the decay time is evaluated by changing the bin borders. The 140 bins of increasing bin widths are exchanged by 140 bins of uniform bin widths. The fit to the validation toy samples is repeated using these bins of uniform widths. The intrinsic shift on Δm_d is slightly increased to $(-0.6 \pm 0.1)\,\mathrm{ns}^{-1}$. The absolute difference between both intrinsic shifts of $0.1\,\mathrm{ns}^{-1}$ is taken as systematic uncertainty on Δm_d.

4.9.2. Effects of the B^0 Signal Modeling

The description of the B^0 signal component is tested in the following. Possible systematic uncertainties on Δm_d are evaluated.

B^0 Decay Time Acceptance

The default fit is done using a decay time acceptance histogram obtained from sWeighted data. As a systematic check, a set of toys is generated with the default acceptance histogram from data, but fitted with an acceptance histogram obtained from signal B^0 MC. No significant effect on Δm_d is observed. Hence, the mixing asymmetry is not sensitive to the modeling of the B^0 decay time distribution. Possible modeling differences are common for the nominator and denominator of the mixing asymmetry distribution, hence, they cancel.

B^0 Flight Distance Resolution

The effect on Δm_d caused by a worse flight distance resolution in data compared to MC simulation is tested. Therefore, a set of toy samples is generated with the nominal flight distance resolution, but fitted with an alternative resolution, which is a factor of 1.5 wider than the nominal one. A shift on Δm_d of $0.3\,\mathrm{ns}^{-1}$ is observed. This is taken as a systematic uncertainty on Δm_d.

D Mass sWeights

Δm_d is determined on sWeighted data in order to suppress the combinatoric background from mis-reconstructed D decays. The sWeights are determined from D mass resonance fits on data, see section 4.5. To account for possible differences in the amount of combinatorial background in different tagging categories, the data sample is split into tagging categories before sWeights are extracted. As a systematic test, the 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ data sample is fitted with sWeights, that were determined on the whole sample, without splitting in tagging categories. An absolute difference in Δm_d of $(0.1 \pm 3.0)\,\mathrm{ns}^{-1}$ is observed. Since the fits are performed on the same data sample, both measurements are 100% correlated. Hence, the small shift is significant and taken as a systematic uncertainty on Δm_d.

z Scale Uncertainty

The decay time is proportional to the flight distance. The uncertainty on the flight distance measurement is given by the VELO module position precision along the z-axis. For the measurement Δm_s at LHCb [96], two main contributions to the z scale uncertainty are evaluated: the overall length of the VELO (0.01%) and possible relative shifts of the individual modules within the VELO (0.02%). These effects are also considered here. The effects are added in quadrature to yield $\sigma_{z-scale} = 0.022\%$. This relative uncertainty translates directly into an absolute uncertainty on Δm_d of $0.1\,\mathrm{ns}^{-1}$, which is taken as a systematic uncertainty on Δm_d.

Momentum Scale Uncertainty

The decay time is proportional to the quotient of reconstructed B mass and momentum. In fully reconstructed decays most of the uncertainties on the momentum scale cancel, because the momentum measurement also enters the mass measurement. However, in semileptonic decays, which are partially reconstructed decays, the B mass is fixed to the PDG value. Thus, effects on the momentum scale do not cancel here. In [96] the uncertainty on the momentum scale is determined to be 0.15%. This relative uncertainty translates directly into an absolute uncertainty on Δm_d of $0.8\,\mathrm{ns}^{-1}$, which is taken as a systematic uncertainty on Δm_d.

4.9.3. Effects of the B^+ Background Modeling

B^+ background is non-oscillating. A possible mis-modeling of this background can change the mixing frequency of B^0 signal events. The mistag probability per category of B^+ background is floated in the Δm_d fit, however, the amount of B^+ background per category is fixed to values obtained from fits to the isolation BDT distribution in sWeighted data. The shape of the B^0 signal and B^+ background isolation BDT distributions are taken from MC simulation. Possible mis-modeling can enter here. The B^+ decay time acceptance is modeled in the Δm_d fit by the B^0 decay time acceptance determined on sWeighted data. Possible differences due to the different B^0 and B^+ lifetimes and reconstruction inefficiencies, are taken into account with the help of MC simulation. The following subsections show the effects in detail.

Mode	fraction [%]	weights (-1σ)	weights $(+1\sigma)$
$D^+\mu$	54.80 ± 3.03	0.94	1.06
$D^*\mu$	37.76 ± 0.84	0.98	1.02
$D_0^*\mu$	2.14 ± 0.79	0.63	1.37
$D_1'\mu$	0.71 ± 0.25	0.65	1.35
$D_1\mu$	1.84 ± 0.28	0.85	1.15
$D_2^*\mu$	1.98 ± 0.82	0.59	1.41
non-resonant	0.78 ± 0.41	0.47	1.53

Table 4.21.: MC sample composition in $B^0 \to D^-\mu^+\nu_\mu X$ mode after full selection. The $\mp 1\sigma$ columns show values for weights to decrease or enhance the given mode by $\mp 1\sigma$ of the fraction uncertainty, which is mainly determined by the uncertainty of the branching fraction.

B^+ Fractions

The externally fitted B^+ fractions per category are varied here. The sample composition, in terms of feed down modes, of the MC cocktail determines the shape of the B^0 signal and B^+ background isolation BDT distributions. Relatively large uncertainties on the semileptonic branching ratios dominate the uncertainty on the fraction of each component in the sample. The statistical component of the uncertainty on the B^+ fraction is negligible, as found out in the determination of the B^+ fraction in simulation, see section 4.5. Thus, the effect of the branching ratio uncertainties on the determination of the B^+ fractions from data is tested here.

Therefore, relative sample composition weights are calculated to enhance or decrease the given feed down mode by 1σ of the branching ratio uncertainty, shown in table 4.21. Table 4.22 shows the branching ratio uncertainties that enter the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode MC simulation cocktail. This sample consists of 95% D^* μ decays, in contrast to the $B^0 \to D^-\mu^+\nu_\mu X$ mode, where D^+ μ decays make only 55% of the full sample and D^* μ decays 38%.

The weights are applied one after another to the isolation BDT shapes in B^0 signal MC, see figure 4.37 for the $B^0 \to D^-\mu^+\nu_\mu X$ mode. By eye, there is no significant difference in the weighted isolation BDT distributions visible. For each weighted MC distribution the fit to sWeighted data is repeated extracting a B^+ fraction.

Figure 4.38 shows the results of the weighted isolation BDT distribution fits. In the $B^0 \to D^-\mu^+\nu_\mu X$ mode, where the sample consists of only 55% signal decays, an almost linear dependence of the fitted B^+ fractions on the relative branching ratio uncertainty weighting is present. Whereas in the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode, where the sample consists of 95% signal decays, the fitted B^+ fractions are rather constant except for the variation of the non resonant branching ratio uncertainties. The maximum difference in the fitted B^+ fractions of absolute 0.5% is larger in the $B^0 \to D^-\mu^+\nu_\mu X$, than in the $B^0 \to D^{*-}\mu^+\nu_\mu X$, where it is 0.3%.

The effect on Δm_d of this variation is tested in toys by generating toy events with the nominal B^+ fractions, but fitting with 0.5% up and downwards varied B^+ fractions. The ω parameters of the B^0 and B^+ components are left floating. For the up-type variation of the B^+ fractions,

Mode	fraction [%]	weights (-1σ)	weights $(+1\sigma)$
$D^*\mu$	94.99 ± 2.07	0.98	1.02
$D_0^*\mu$	0.35 ± 0.12	0.65	1.35
$D_1'\mu$	0.89 ± 0.34	0.63	1.37
$D_1\mu$	2.30 ± 0.34	0.85	1.15
$D_2^*\mu$	0.68 ± 0.31	0.55	1.45
non-resonant	0.79 ± 0.25	0.69	1.31

Table 4.22.: MC sample composition in $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode after full selection. The $\mp 1\sigma$ columns show values for weights to decrease or enhance the given mode by $\mp 1\sigma$ of the fraction uncertainty, which is mainly determined by the uncertainty of the branching fraction.

(a) Branching ratio uncertainties -1σ (b) Branching ratio uncertainties $+1\sigma$
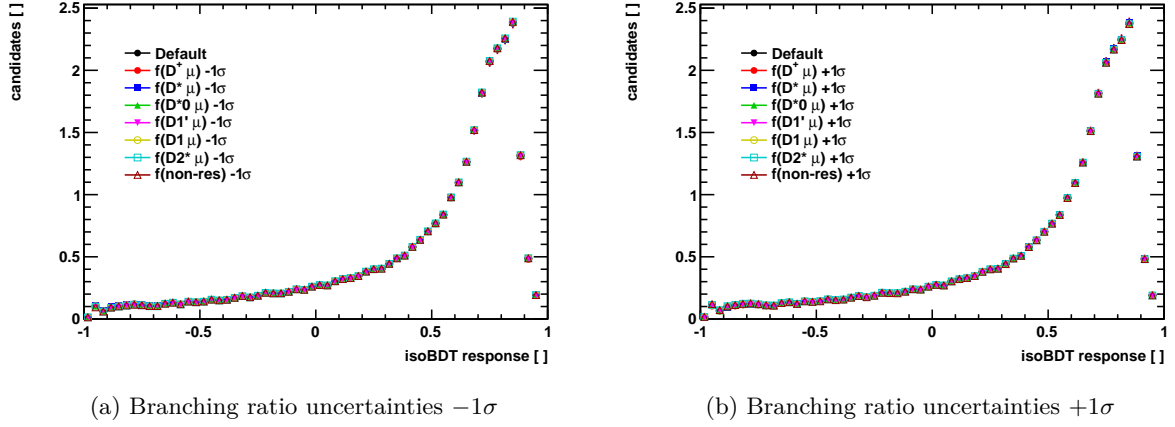
Figure 4.37.: Weighted B^0 isolation BDT response distribution in $B^0 \to D^-\mu^+\nu_\mu X$ mode to account for semileptonic branching ratio uncertainties varying (a) -1σ and (b) $+1\sigma$.

a negligible shift on Δm_d of $0.04\,\text{ns}^{-1}$ is found. For the down-type variation, a shift on Δm_d of $0.14\,\text{ns}^{-1}$ is found. The down-type variation effect on Δm_d of $0.1\,\text{ns}^{-1}$ is taken as systematic uncertainty on Δm_d.

B^+ Decay Time Acceptance

In the nominal Δm_d fit the same decay time acceptance distribution is used for the B^0 and the B^+ component, which is obtained from sWeighted data. Due to the different lifetimes of B^0 and B^+ mesons and possibly different reconstruction efficiencies, the effect of this assumption on Δm_d is tested.

Figure 4.38.: Fitted B^+ fractions (f_{B^+}) in $B^0 \to D^- \mu^+ \nu_\mu X$ mode (a) and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode (b), where the signal B^0 MC isolation BDT distribution template is weighted according to the semileptonic branching ratio uncertainties.

Toy samples are generated with the default fit setup but fitted with a different decay time acceptance distribution for the B^+ background component. Due to known data MC simulation differences in the decay time acceptance distribution, the B^+ decay time acceptance obtained from MC simulation is not directly taken here. However, the acceptance histogram obtained from sWeighted data is reweighted to the relative difference of the B^0 and B^+ decay time acceptance, both obtained from simulation, shown in figure 4.39. A shift on Δm_d of $0.4\,\text{ns}^{-1}$ is observed. This is taken as systematic uncertainty on Δm_d.

B^+ Flight Distance Resolution

The nominal Δm_d fit uses different flight distance resolutions for the B^0 and the B^+ component, both obtained from MC simulation. However, the statistics of the B^+ MC sample are limited, which causes large uncertainties on the fitted B^+ flight distance resolution parameters. Hence, the effect on Δm_d is determined by using different flight distance resolution parameters for the B^+ component. A set of toy samples is generated under nominal conditions but fitted using the B^0 flight distance resolution for both components. A shift on Δm_d of $0.1\,\text{ns}^{-1}$ is observed. This is taken as a systematic uncertainty on Δm_d.

4.9.4. Effects of the Modeling of other Backgrounds

This section evaluates other backgrounds, besides B^+ background, affecting the systematic uncertainty on Δm_d. Three types of additional backgrounds are identified: B_s^0, Λ_b^0 and $B \to D^{(*)-} D_s^+$ decays. They are investigated in the following.

Figure 4.39.: Nominal decay time acceptance distribution determined on sWeighted data (red) used in the Δm_d fit for both components, B^0 signal and B^+ background. This acceptance distribution is reweighted to the relative B^+/B^0 acceptance distribution differences obtained from MC simulation (blue).

B_s^0 and Λ_b^0 Background Component

B_s^0 mesons oscillate a factor 34 faster than B^0 mesons. Given the decay time resolution of semileptonic decays the fast B_s^0 oscillation is not resolvable, see figure 4.40. The oscillation averages out and appears to have a constant zero mixing asymmetry. Λ_b^0 decays do not oscillate. Therefore, they can be treated as the non-oscillating B^+ background. The Λ_b^0 mixing asymmetry is non-zero but constant over decay time.

The fraction of Λ_b^0 events in the sample is $0.1 \pm 0.1\%$ in the $B^0 \to D^- \mu^+ \nu_\mu X$ mode and 0.0% in the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode. A variation of that order of magnitude is covered by the variation of the B^+ fractions of absolute $\pm 0.5\%$ up and down. The fraction of B_s^0 decays is estimated to be $1.5 \pm 0.4\%$ in the $B^0 \to D^- \mu^+ \nu_\mu X$ mode and $1.6 \pm 0.3\%$ in the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode. A contribution of B_s^0 decays of that size is generated in toy MC samples, but the fit is performed without this extra component. A shift on Δm_d of $0.2 \, \text{ns}^{-1}$ is observed. This is taken as a systematic uncertainty on Δm_d.

$B \to D^{(*)-} D_s^+ X$ or $B \to D^{(*)-} \tau X$ Background Decays

A small contribution from $B \to D^{(*)-} D_s^+$ decays, where the D_s decays semileptonically, is also tested. $B \to D^{(*)-} \tau X$ decays, where $\tau \to \mu \nu_\mu \nu_\tau$, are kinematically similar, thus, have a similar k factor distribution. Therefore, the effect of this contribution is estimated by enlarging the contribution of the τ decays from 0.1% to 1.7% in the k factor distribution. Toy samples are generated without this effect, but fitted with the τ enlarged k factor distribution. A shift on Δm_d of $0.2 \, \text{ns}^{-1}$ is observed. This is taken as a systematic uncertainty on Δm_d.
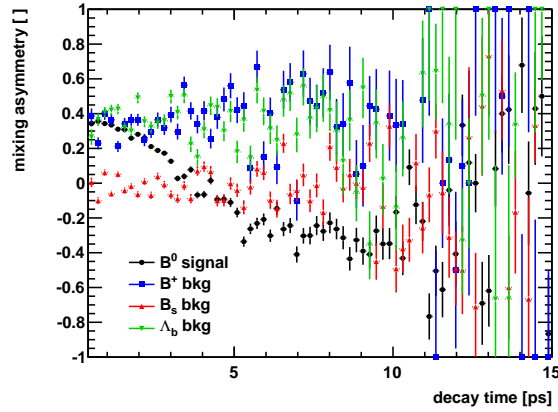
Figure 4.40.: Mixing asymmetry for different MC samples: B^0 signal (black), B^+ background (red), B_s^0 background (blue) and Λ_b^0 background (green).

4.9.5. Effects of the k Factor Correction

The k factor correction plays a central role in the analysis. Since it is determined using MC simulation, possible MC data differences have to be accounted for. The effect on Δm_d is determined, caused by differences in the B^0 momentum distribution between data and MC. Also the semileptonic branching ratio uncertainties are propagated into the k factor distributions to determine a systematic effect on Δm_d. The presented studies are done using the standard k factor method. However, the effects on Δm_d are identical, when using the improved k factor method.

Data MC Differences in the B Momentum Distribution

Possible data MC differences in the B momentum distribution are checked with the help of fully reconstructed $B^+ \to J/\psi K^+$ decays. The momentum distributions of B^+ data and B^+ MC are divided to obtain momentum weights. Those weights are applied to the k factor distribution, shown in figure 4.41. Both MC production Pythia versions (6 and 8) are tested. The agreement between the data and MC B^+ momentum distributions is slightly different for different Pythia versions.

To evaluate the systematic effect on Δm_d a set of toy samples is generated according to the nominal fit conditions, but fitted with the weighted k factor distribution. The k factor distribution, weighted by Pythia 8 momentum weights, yields the larger effect on Δm_d of $0.3 \, \text{ns}^{-1}$. The Pythia 6 weighted k factor distribution yields an almost negligible shift on Δm_d of $0.05 \, \text{ns}^{-1}$. The larger shift value is taken as systematic uncertainty on Δm_d.

Semileptonic Branching Ratio Uncertainties

Figure 4.42 shows the k factor histogram as a stacked histogram of all different feed down modes for both channels. The different sample compositions in $B^0 \to D^- \mu^+ \nu_\mu X$ compared to

(a) B momentum weights

(b) weighted k factor distributions
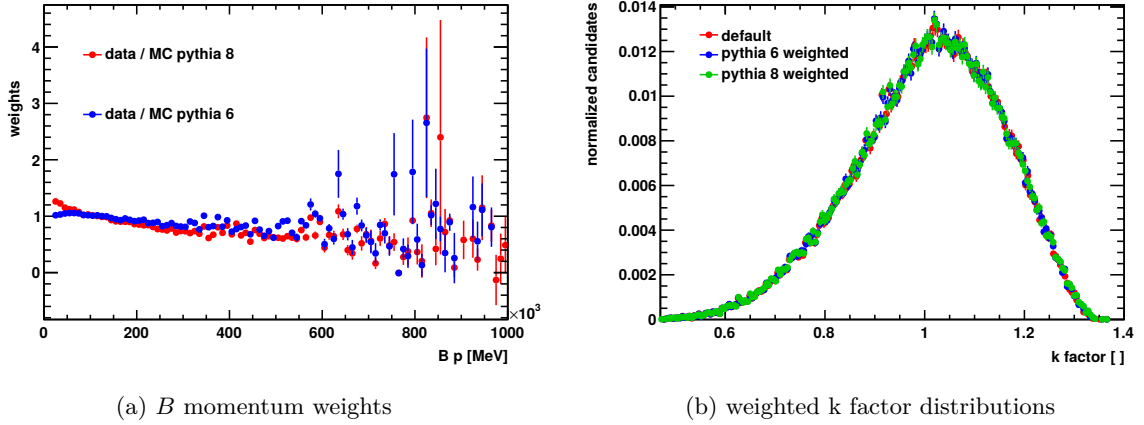
Figure 4.41.: (a) B momentum weights obtained by dividing the B^+ momentum distribution of data by the one obtained from MC using the fully reconstructed $B^+ \to J/\psi\,K^+$ decay mode. Pythia 6 and 8 versions are considered separately. (b) Default k factor distribution compared to the weighted k factor distributions using the weights from (a).

$B^0 \to D^{*-}\mu^+\nu_\mu X$ mode are well visible. Hence, semileptonic branching ratio uncertainties affect the $B^0 \to D^-\mu^+\nu_\mu X$ more.

As the semileptonic branching ratio uncertainties affect the determination of the B^+ fractions, they also affect the k factor distribution. Hence, the same $\pm 1\sigma$ weights are used here, see table 4.21 for the $B^0 \to D^-\mu^+\nu_\mu X$ mode and table 4.22 for the $B^0 \to D^{*-}\mu^+\nu_\mu X$ mode. They are applied one after another to the k factor distribution. A comparison of the default and weighted k factor distributions is shown in figure 4.43 for the $B^0 \to D^-\mu^+\nu_\mu X$ mode.

A single high statistics toy sample is generated with the default k factor histogram and fitted with each of the weighted k factor histograms. The fit results are shown in figure 4.44. The number of generated toy events is increased by a factor of 10 from 1.6×10^6 to 1.6×10^7 events. That decreases the statistical uncertainty on Δm_d from $2.8\,\mathrm{ns}^{-1}$ to $1.7\,\mathrm{ns}^{-1}$. No large shift on Δm_d for any of the weighted k factor histograms is observed. The quadratic sum of the differences between the fitted value and the default value of Δm_d is calculated. It yields a mean deviation $0.4\ \mathrm{ns}^{-1}$. This is taken as systematic uncertainty on Δm_d.

4.9.6. Combination of the Systematic Uncertainty on Δm_d

Table 4.23 shows all systematic uncertainties considered for the measurement of Δm_d. The different uncertainties are grouped according to five types of tests. First, the fit procedure is validated using toy experiments. Second, the B^0 signal component modeling is tested. Third, all uncertainties related to the B^+ background modeling are evaluated. Fourth, uncertainties related to possible contributions of other backgrounds, B_s^0, Λ_b^0 or double D or τ decays, are given. Fifth, uncertainties induced by the k factor correction method are evaluated.
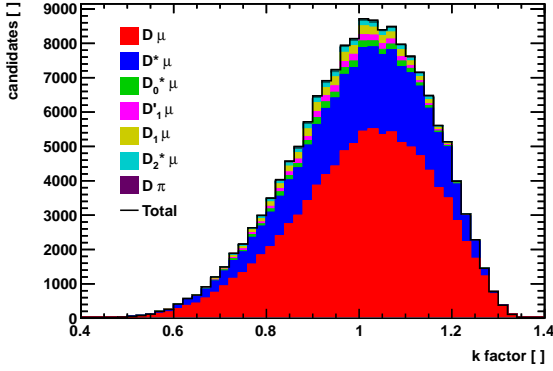
(a) $B^0 \to D^- \mu^+ \nu_\mu X$ mode
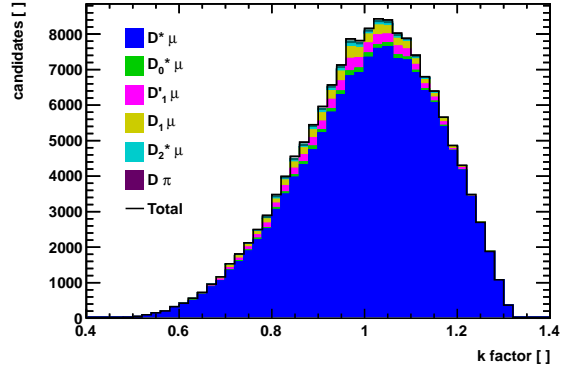
(b) $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode

Figure 4.42.: Sample composition of $B^0 \to D^- \mu^+ \nu_\mu X$ (a) and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ (b) displayed in stacked k factor histograms. The corresponding fraction numbers are given in table 4.21 and table 4.22.
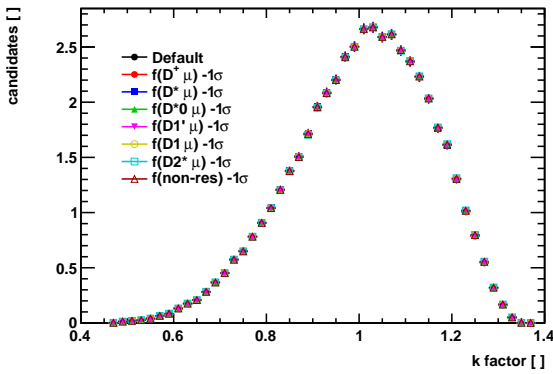
(a) Branching ratio uncertainty -1σ
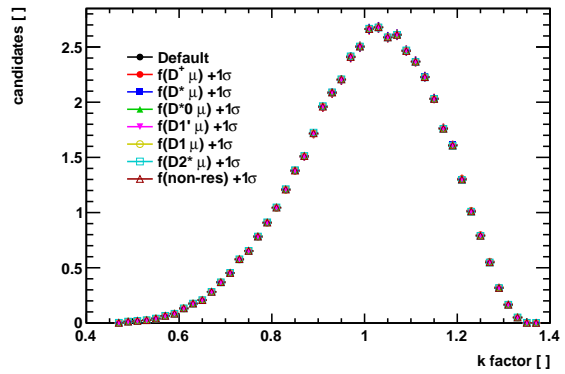
(b) Branching ratio uncertainty $+1\sigma$

Figure 4.43.: Comparison of k factor distributions in the $B^0 \to D^- \mu^+ \nu_\mu X$ mode after applying weights corresponding to a variation of (a) -1σ and (b) $+1\sigma$ of the semileptonic branching ratio uncertainties.
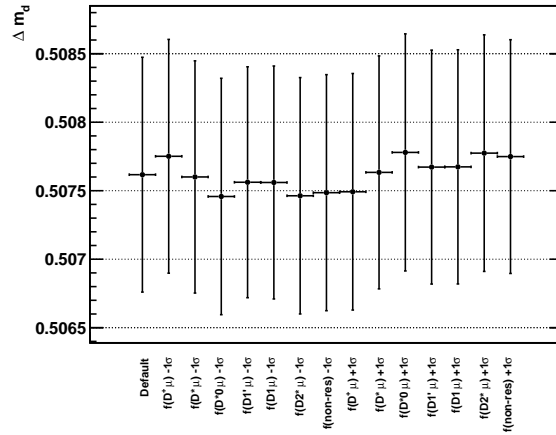
Figure 4.44.: A high statistics toy sample (1.6×10^7 events) is generated with nominal fit conditions. The sample is fitted with weighted k factor histograms corresponding to $\pm 1\sigma$ variations in the semileptonic branching ratio uncertainties.

The largest systematic uncertainty on Δm_d comes from the knowledge of the momentum scale in LHCb. In all fully reconstructed decay modes, this uncertainty cancels at first order. However, in partially reconstructed semileptonic B decays, the determination of the B^0 decay time is fully affected by this uncertainty due to the absence of the reconstructed B^0 mass resonance. All other systematic uncertainties are of similar size. The total systematic uncertainty on Δm_d is obtained by the square root of the quadratic sum of the single effects. An overall systematic uncertainty on Δm_d of $1\,\mathrm{ns}^{-1}$ is evaluated. For comparison, the combined statistical precision, fitting all four data sets, is $2\,\mathrm{ns}^{-1}$.

Source of uncertainty	Uncertainty $[\,\mathrm{ns}^{-1}\,]$
Validation:	
bias of nominal fit	0.5
binning	0.1
B^0 **signal:**	
B^0 acceptance	neg.
B^0 flight distance resolution	0.3
D mass sWeights	0.1
z-scale uncertainty	0.1
momentum scale uncertainty	0.8
B^+ **background:**	
B^+ fraction	0.1
B^+ acceptance	0.4
B^+ flight distance resolution	0.1
Other backgrounds:	
B_s^0 component	0.2
Λ_b^0 component	-
$B^0 \to D^{(*)-}D_s^+X$ or $B^0 \to D^{(*)-}\tau X$ decays	0.2
k factor distribution:	
Momentum distributions	0.3
Branching ratio from $D^{*(*)}$	0.4
Quadratic sum	1.2

Table 4.23.: Overview of all contributions to the systematic uncertainty on Δm_d. The variations are determined on sets of toy samples ($\mathcal{O}(1000)$), with $B^0 \to D^- \mu^+ \nu_\mu X$ 2012 data conditions (1.6×10^6 signal events). The statistical precision on Δm_d of a single toy sample fit is $2.8\,\mathrm{ns}^{-1}$. Entries marked with - are found to be non present. Entries marked with 'neg.' are found to be negligible.

4.10. Results

The method to measure Δm_d has been developed using the blind analysis technique. The blinding offset is removed from the fitted value of Δm_d, after all systematic checks and data crosschecks are finished. The systematic studies were presented in the previous section. This section shows the crosschecks evaluated on data.

Four non-overlapping data sets are used to measure Δm_d: 2011 and 2012 data in the $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-}\mu^+ \nu_\mu X$ modes. These data sets are fitted separately, first. Δm_d is calculated as the weighted average from the four results. Second, one combined fit is performed on all four data sets simultaneously. The simultaneous fit picks up all input parameters data set specifically, as well as the mistag parameters per category are fitted data set specific. The Δm_d

subsample name	events requirement
All	—
mag Up	magnet polarity up
mag Dn	magnet polarity down
μ^+	positively charged final state muon
μ^-	negatively charged final state muon
#PV $= 1$	number of primary vertices per event $= 1$
#PV $= 2$	number of primary vertices per event $= 2$
#PV > 2	number of primary vertices per event at least 3
Mult 1	track multiplicity per event <150
Mult 2	track multiplicity per event >150 and <220
Mult 3	track multiplicity per event >220
Run 1	run number $<9.5 \times 10^4$ (1.19×10^5)
Run 2	run number $>9.5 \times 10^4$ (1.19×10^5) and $<1.01 \times 10^5$ (1.29×10^5)
Run 3	run number $>1.01 \times 10^5$ (1.29×10^5)

Table 4.24.: Definitions of data set split categories. The run number is a continuous number assigned to every data taking run, hence different values are given for 2011 and 2012 data, respectively.

parameter, however, is shared. A comparison of Δm_d obtained from the two fitting methods is given.

4.10.1. Crosscheck using Data Set Splits

The data sets are split into different subsamples, divided by magnet polarities, charge of the final state μ, number of primary vertices, track multiplicities and data taking periods. The definitions of the subsample splits are given in table 4.24. The splitting categories ensure that the final result is independent of systematic effects in data. The splits should give, besides statistical fluctuations, the same (blinded) fit value for Δm_d. A variation of up to two σ of the statistical uncertainty on Δm_d is tolerated.

Figure 4.45 shows the data set splits on all four data samples. A blinded value for Δm_d is given for every subsample fit. Within 2σ all subsamples yield a compatible blinded Δm_d fit value compared to the fit to the full sample.

4.10.2. Fit Projections

This section shows the Δm_d fit projections on the decay time distribution. The data is sWeighted to remove the combinatoric background from mis-reconstructed D decays. For every data set three types of projections are plotted: the total fit to the decay time distribution with a separation into the B^0 and B^+ fit components, the total fit to the decay time distribution with a separation into the decay times distributions of mixed and unmixed decays and the mixing asymmetry as
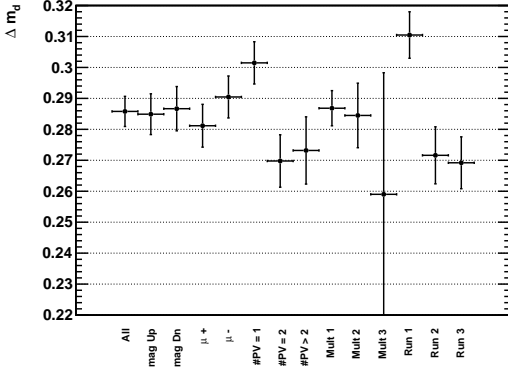
(a) 2011 $B^0 \to D^- \mu^+ \nu_\mu X$ data
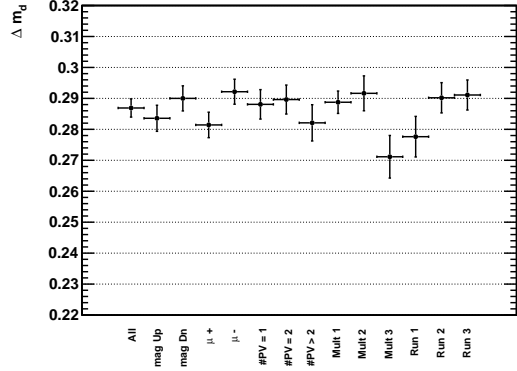
(b) 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ data
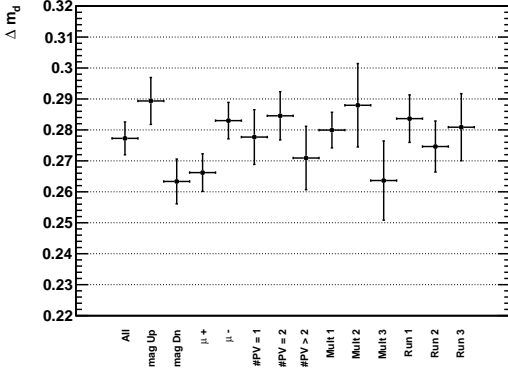
(c) 2011 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data
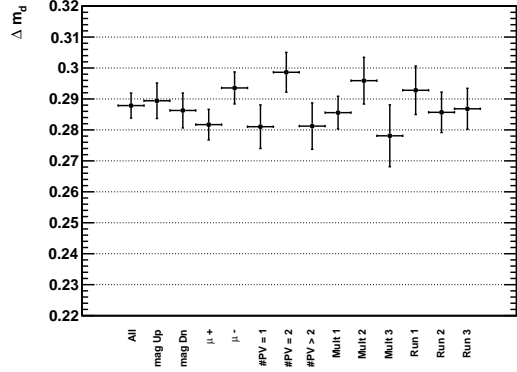
(d) 2012 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data

Figure 4.45.: Data set splits of blinded Δm_d fits for 2011 and 2012 data in $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ modes. The split definitions are given in table 4.24. All subsample fit results are compatible with the result obtained on the full sample within 2σ of the statistical uncertainty on Δm_d.

a function of the decay time. Every type is additionally plotted in the four tagging categories. Figure 4.46 shows the fit projection plots in 2011 $B^0 \to D^- \mu^+ \nu_\mu X$ mode data and figure 4.47 for 2012 data. Figure 4.48 shows the fit projection plots in 2011 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode data and figure 4.49 for 2012 data.

Pull distributions given below the uncategorized projections indicate the good description of the data distribution by the fit $\mathcal{PDF}$. The categorized mixing asymmetries show the different damping effects of different mistag probabilities.
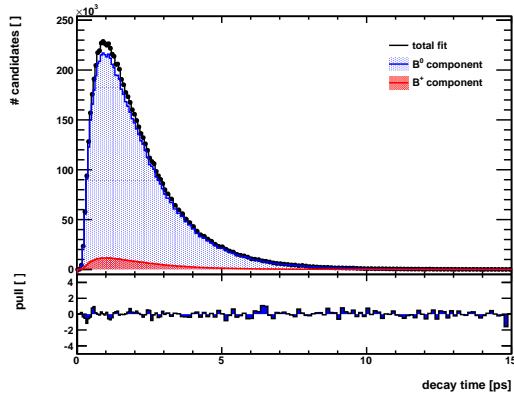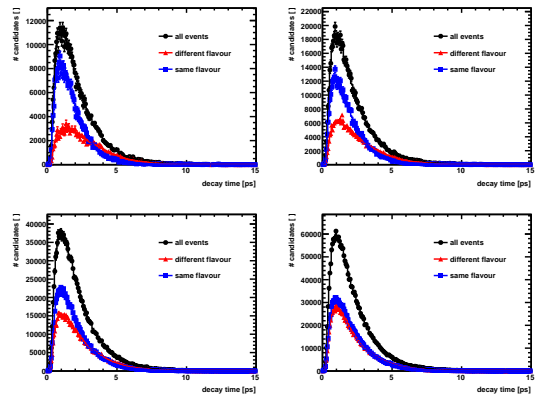
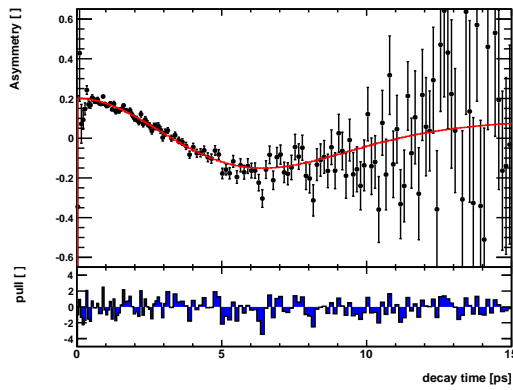(a) Decay time of B^0 and B^+ components
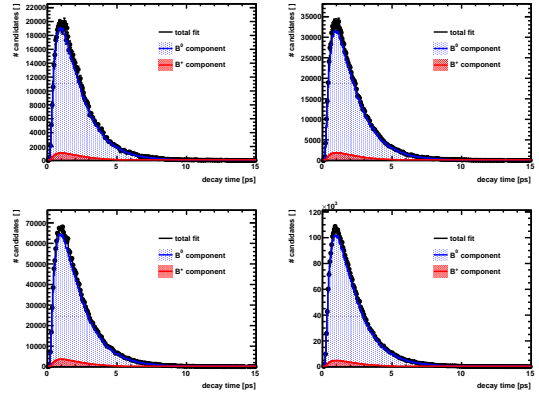
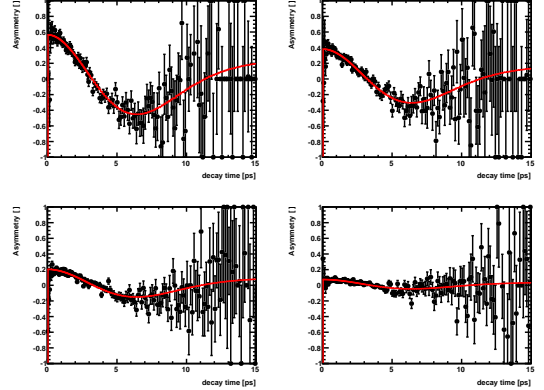(b) Decay time distributions in categories
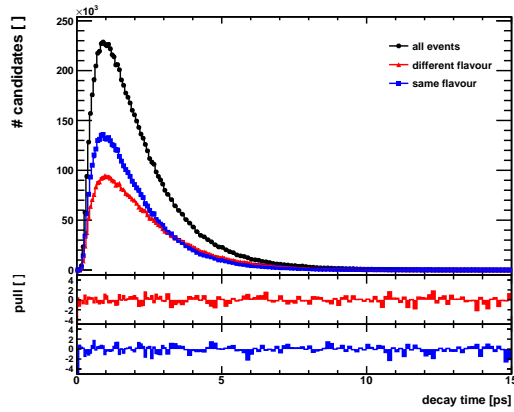
(c) Decay time of mixed and unmixed components
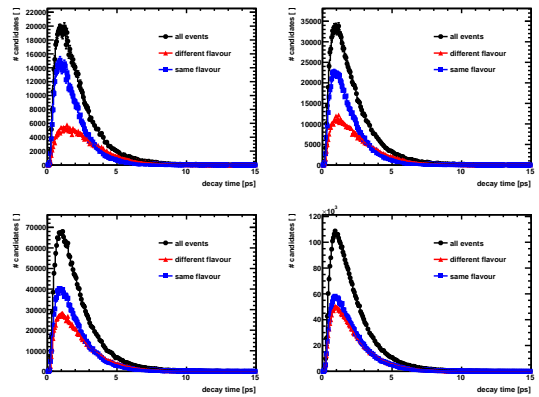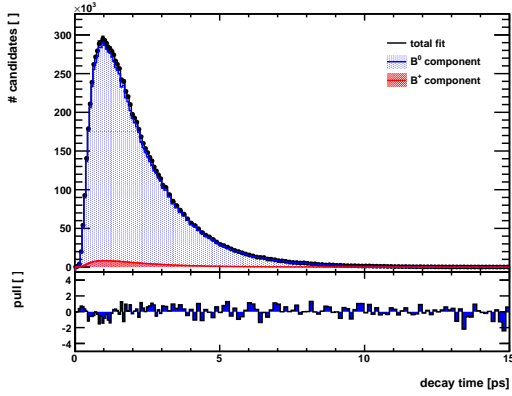
(d) Decay time distributions in categories
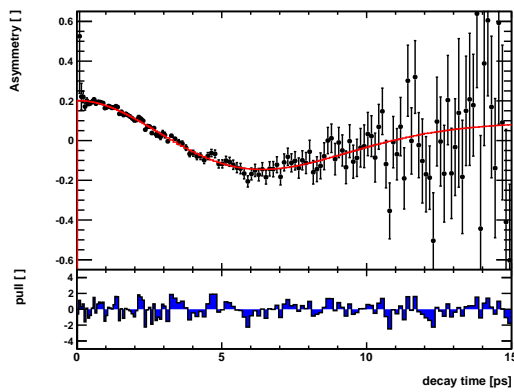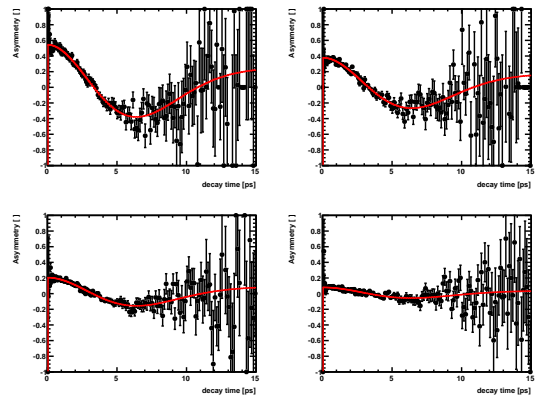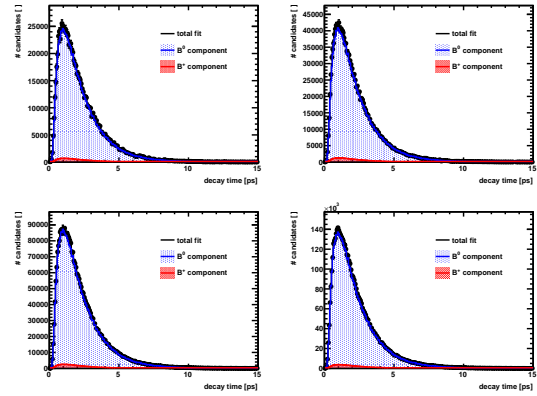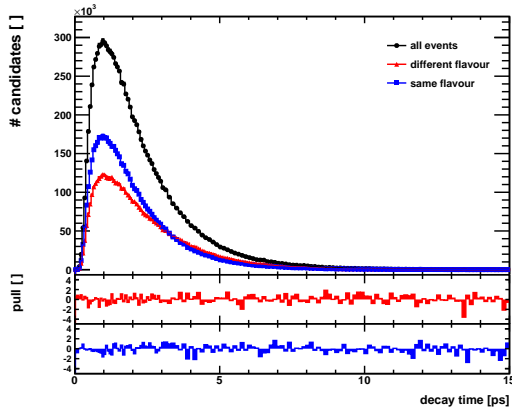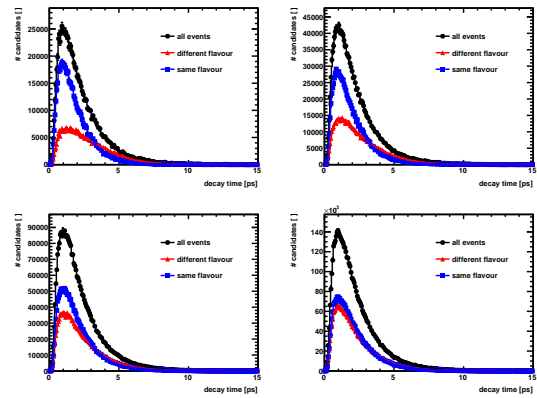
(e) Mixing asymmetry

(f) Mixing asymmetry in categories

Figure 4.46.: Fit projections of the Δm_d fit to 2011 $B^0 \to D^- \mu^+ \nu_\mu X$ data. Upper row: fitted decay time distributions separated into B^0 signal and B^+ background components. Middle row: fitted decay time distributions separated into decays with different production and decay flavour and decays with same production and decay flavour. Lower row: fitted mixing asymmetry distributions. Left column: all events, right column: events split in mistag categories of increasing mistag.
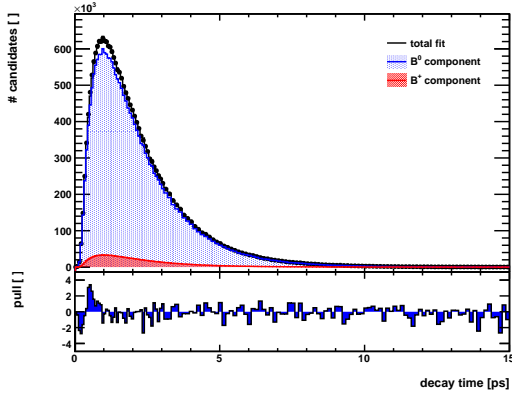
(a) Decay time of B^0 and B^+ components
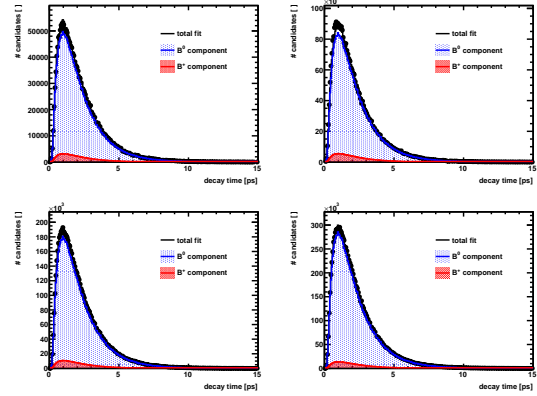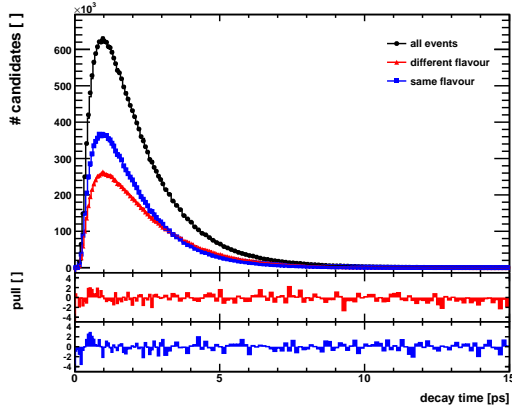
(b) Decay time distributions in categories

(c) Decay time of mixed and unmixed components
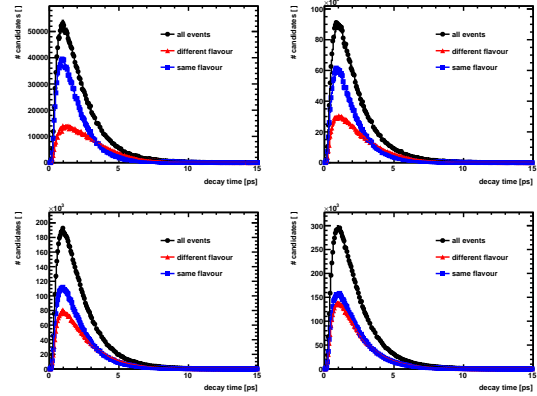
(d) Decay time distributions in categories
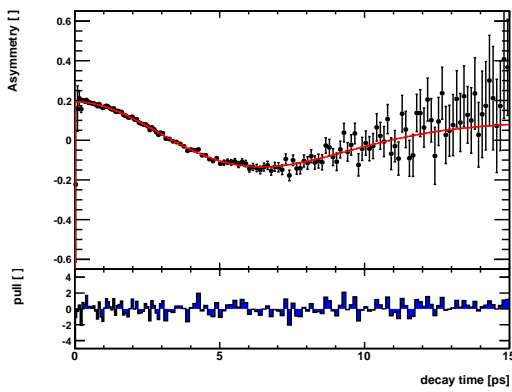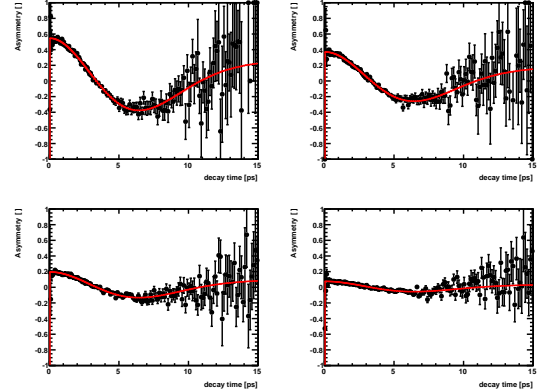
(e) Mixing asymmetry

(f) Mixing asymmetry in categories

Figure 4.47.: Fit projections of the Δm_d fit to 2012 $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ data, as in figure 4.46 explained.

179
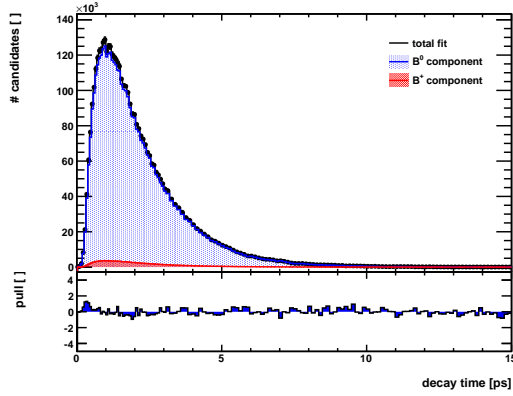
(a) Decay time of B^0 and B^+ components
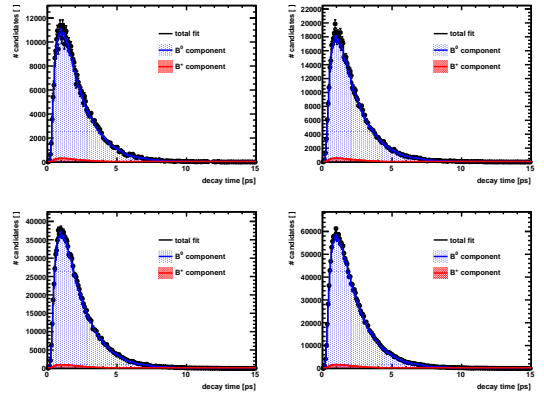
(b) Decay time distributions in categories
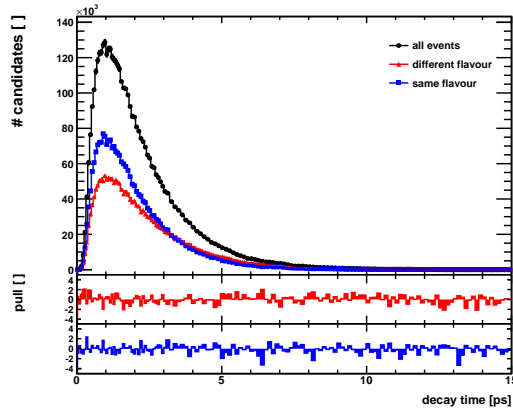
(c) Decay time of mixed and unmixed components

(d) Decay time distributions in categories

(e) Mixing asymmetry

(f) Mixing asymmetry in categories

Figure 4.48.: Fit projections of the Δm_d fit to 2011 $B^0 \rightarrow D^{*-}\mu^+\nu_\mu X$ data, as in figure 4.46 explained.

(a) Decay time of B^0 and B^+ components

(b) Decay time distributions in categories

(c) Decay time of mixed and unmixed components

(d) Decay time distributions in categories
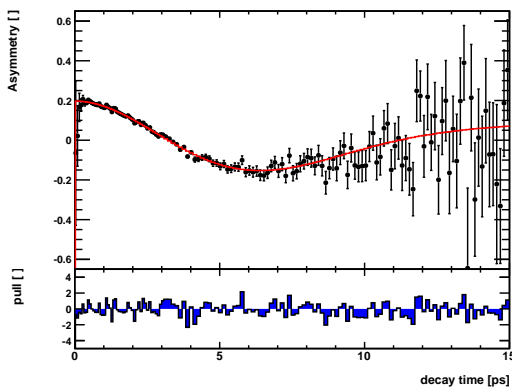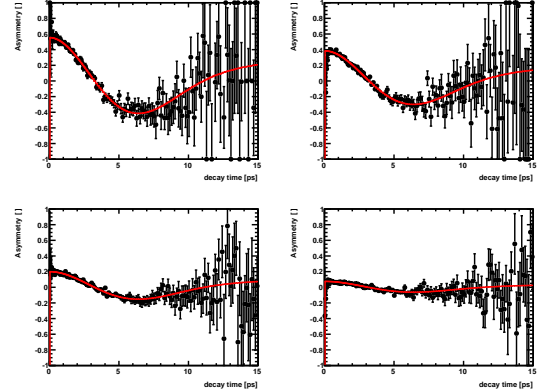
(e) Mixing asymmetry

(f) Mixing asymmetry in categories

Figure 4.49.: Fit projections of the Δm_d fit to 2012 $B^0 \to D^{*-}\mu^+\nu_\mu X$ data, as in figure 4.46 explained.

4.10.3. Tagging Performance

The tagging performance is evaluated for each data set. For $B^0 \to D^- \mu^+ \nu_\mu X$ mode table 4.25 shows the performance measured in 2012 data and table 4.26 shows the performance measured in 2011 data. Accordingly, for $B^0 \to D^{*-} \mu^+ \nu_\mu X$ mode the tagging performance for 2012 data is given in table 4.27 and for 2011 data in table 4.28. The tagging performance is determined separately for the B^0 signal and the B^+ background components.

quantity	cat. 0 [%]	cat. 1 [%]	cat. 2 [%]	cat. 3 [%]	Σ [%]
$B^0 \varepsilon_{\text{tag}}$	3.65 ± 0.08	6.36 ± 0.11	12.24 ± 0.17	15.11 ± 0.16	37.36 ± 0.27
$B^0\ \omega$	23.06 ± 0.36	31.62 ± 0.30	40.46 ± 0.12	46.22 ± 0.16	35.34 ± 0.51
$B^0\ \varepsilon_{\text{eff}}$	1.06 ± 0.05	0.86 ± 0.05	0.45 ± 0.03	0.09 ± 0.01	2.46 ± 0.08
$B^+ \varepsilon_{\text{tag}}$	3.53 ± 0.07	6.13 ± 0.09	11.99 ± 0.14	15.45 ± 0.13	37.10 ± 0.22
$B^+\ \omega$	17.45 ± 2.87	29.17 ± 2.11	35.96 ± 1.55	46.13 ± 1.45	32.18 ± 4.15
$B^+\ \varepsilon_{\text{eff}}$	1.50 ± 0.27	1.06 ± 0.22	0.95 ± 0.21	0.09 ± 0.07	3.60 ± 0.41

Table 4.25.: Tagging performance in 2012 $B^0 \to D^- \mu^+ \nu_\mu X$ data, separately for the B^0 signal and the B^+ background component. From the tagging efficiency (ε_{tag}) and the measured mistag probability (ω) the tagging power ($\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1 - 2\omega)^2$) is calculated per category. The last column gives the overall tagging performance, summed over the four categories.

quantity	cat. 0 [%]	cat. 1 [%]	cat. 2 [%]	cat. 3 [%]	Σ [%]
$B^0 \varepsilon_{\text{tag}}$	3.34 ± 0.12	5.90 ± 0.17	11.52 ± 0.25	14.51 ± 0.24	35.27 ± 0.40
$B^0\ \omega$	23.17 ± 0.34	31.25 ± 0.27	39.57 ± 0.20	46.06 ± 0.27	35.01 ± 0.55
$B^0\ \varepsilon_{\text{eff}}$	0.96 ± 0.07	0.83 ± 0.07	0.50 ± 0.06	0.09 ± 0.02	2.38 ± 0.12
$B^+ \varepsilon_{\text{tag}}$	3.29 ± 0.10	5.74 ± 0.13	11.29 ± 0.20	14.83 ± 0.19	35.15 ± 0.32
$B^+\ \omega$	19.55 ± 4.87	31.04 ± 3.76	44.79 ± 2.64	45.36 ± 2.50	35.19 ± 7.15
$B^+\ \varepsilon_{\text{eff}}$	1.22 ± 0.39	0.83 ± 0.33	0.12 ± 0.13	0.13 ± 0.14	2.30 ± 0.55

Table 4.26.: Fitted tagging parameters in 2011 $B^0 \to D^- \mu^+ \nu_\mu X$ data, following the same structure as explained in table 4.25.

The tagging efficiency (ε_{tag}) per category is determined by the number of sWeighted data events in that category, divided by the number of all sWeighted data events. Separate tagging efficiencies are evaluated for the B^0 and the B^+ component due to different B^+ fractions per category, see table 4.10 and table 4.11 in section 4.5. The uncertainties on the B^+ fractions are propagated into ε_{tag} as well as the statistical uncertainty on the number of events.

The measured mistag probability (ω) per category is determined by the Δm_d fit. Additionally, the tagging power ($\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1 - 2\omega)^2$) is calculated per category. The last column of each table gives the overall tagging performance of the sample by summing over all tagging categories.

quantity	cat. 0 [%]	cat. 1 [%]	cat. 2 [%]	cat. 3 [%]	Σ [%]
$B^0 \varepsilon_{\text{tag}}$	3.00 ± 0.10	5.64 ± 0.14	11.20 ± 0.20	16.21 ± 0.24	36.05 ± 0.36
$B^0 \ \omega$	22.66 ± 0.30	30.64 ± 0.42	40.06 ± 0.17	46.16 ± 0.23	34.88 ± 0.59
$B^0 \ \varepsilon_{\text{eff}}$	0.90 ± 0.06	0.85 ± 0.07	0.44 ± 0.04	0.10 ± 0.02	2.29 ± 0.10
$B^+ \varepsilon_{\text{tag}}$	3.02 ± 0.07	5.46 ± 0.09	11.09 ± 0.13	16.5 ± 0.16	36.07 ± 0.24
$B^+ \ \omega$	16.90 ± 8.10	39.79 ± 6.08	41.69 ± 4.30	50.9 ± 3.52	37.32 ± 11.55
$B^+ \ \varepsilon_{\text{eff}}$	1.32 ± 0.65	0.23 ± 0.27	0.31 ± 0.32	0.01 ± 0.04	1.87 ± 0.77

Table 4.27.: Fitted tagging parameters in 2012 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data, following the same structure as explained in table 4.25.

quantity	cat. 0 [%]	cat. 1 [%]	cat. 2 [%]	cat. 3 [%]	Σ [%]
$B^0 \varepsilon_{\text{tag}}$	3.09 ± 0.15	5.77 ± 0.19	9.83 ± 0.29	16.53 ± 0.31	35.22 ± 0.49
$B^0 \ \omega$	21.24 ± 0.56	30.63 ± 0.64	40.05 ± 0.46	46.57 ± 0.35	34.62 ± 1.03
$B^0 \ \varepsilon_{\text{eff}}$	1.02 ± 0.09	0.87 ± 0.09	0.39 ± 0.07	0.08 ± 0.03	2.36 ± 0.15
$B^+ \varepsilon_{\text{tag}}$	3.09 ± 0.10	5.53 ± 0.13	10.01 ± 0.19	16.45 ± 0.20	35.08 ± 0.32
$B^+ \ \omega$	44.62 ± 9.79	41.2 ± 12.8	31.68 ± 6.61	38.09 ± 7.58	38.90 ± 19.00
$B^+ \ \varepsilon_{\text{eff}}$	0.04 ± 0.13	0.17 ± 0.50	1.34 ± 0.97	0.93 ± 1.19	2.48 ± 1.62

Table 4.28.: Fitted tagging parameters in 2011 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data, following the same structure as explained in table 4.25.

The tagging power of B^0 signal is compatible between decay modes and years. Combining all four measurements, the tagging power measured in semileptonic B^0 decays is $2.39 \pm 0.05\%$, achieved by the combination of the ANN-based OS kaon tagger and the OS muon and electron taggers.

However, the tagging power of B^+ background differs significantly from the performance of B^0 signal and within different years and modes. The reason is, that the assignment to categories, which depends on the predicted mistag probability, is not properly defined for B^+ background. The momentum measurement of semileptonic B^+ decays, which are reconstructed as B^0 decays, is underestimated due to the additional non-reconstructed pion track. This shifts the predicted mistag probability to higher values, because the tagging algorithms predict the lowest mistag probabilities for high B momentum events. Thus, events that should be categorized in the best mistag category are propagated into the second best mistag category and so on. That decreases the measured mistag probability of B^+ background in the best mistag category significantly, while the effect is less pronounced in the subsequent categories. Hence, the average tagging power of the B^+ component is overestimated.

4.10.4. Results

First, the four data sets are fitted separately. All fits are converged. The extracted mistag probability parameters of the B^0 signal and B^+ background components are in a physical range between 0% and 50%. They increase for increasing tagging categories, according to predicted mistag probabilities. In the first category, the lowest mistag probability is measured, in the last category, the highest mistag probability is measured, see table 4.25, table 4.26, table 4.27 and table 4.28 for 2012 and 2011 $B^0 \to D^- \mu^+ \nu_\mu X$ data and 2012 and 2011 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data, respectively. However, the measured mistag probabilities of the B^+ component in 2011 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data in the best two categories do not follow this behavior. While the measured mistag probabilities of the B^0 signal component do not show this peculiarity here.

The B^+ fractions of this data set are very small in the best two categories, see table 4.11. Effectively, 840 B^+ events are present in the best category and 1480 B^+ events in the second best category. Due to low statistics the fit is not able to determine the correct B^+ mistag probability here. The fit converges, but tends to overestimate the statistical precision on Δm_d. Relative to the statistical precision on Δm_d given by the fit to 2012 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data, the uncertainty on Δm_d obtained from 2011 data should scale with the square root of the quotient of the number of sWeighted events in 2012 data divided by the number of sWeighted events in 2011 data:

$$\sigma(\Delta m_d(2011)) \propto \sigma(\Delta m_d(2012)) \cdot \sqrt{\frac{\#\text{events}(2012)}{\#\text{events}(2011)}} \qquad (4.52)$$

The fitted Δm_d value on 2011 $B^0 \to D^{*-} \mu^+ \nu_\mu X$ data is trusted, but the fit error estimate is calculated according to equation 4.52. This increases the error estimate effectively by 20% and reduces the impact on the combined Δm_d value of this measurement.

The combined Δm_d value is determined as weighted average of the four independent Δm_d measurements $i = 1, 2, 3, 4$ ($\Delta m_{d,i}$) and the corresponding uncertainty (σ_i), which is the combination of the statistical and the k factor bias uncertainties:

$$\Delta m_d = \left(\sum_{i=1}^{4} \frac{\Delta m_{d,i}}{\sigma_i^2} \right) \cdot \sigma_{comb}, \qquad (4.53)$$

where the combined uncertainty on all four measurements is calculated as quadratic sum of the single uncertainties:

$$\frac{1}{\sigma_{comb}} = \sqrt{\sum_{i=1}^{4} \frac{1}{\sigma_i^2}}. \qquad (4.54)$$

The fit results are given in table 4.29.

The k factor method was improved to remove the bias on Δm_d. However, due to limited statistics in MC simulation, the absolute knowledge on the bias removal is limited. Thus, an additional uncertainty on Δm_d is assumed, given in the third column of table 4.29. In order to avoid over-counting of statistical fluctuations, the difference of two methods to extract Δm_d from MC simulation is considered here. As reference the fit to the true decay time, without k factor correction, is taken. The result of the improved k factor correction applied to the reconstructed decay time in MC is tested. To increase the statistical sensitivity, the tagging performance is

data sample	Δm_d [ps^{-1}]	σ (stat)	σ(k factor)	σ(combined)
2012 $B^0 \rightarrow D^- \mu^+ \nu_\mu X$	0.5047	±0.0029	+0.0002	± 0.0029
2011 $B^0 \rightarrow D^- \mu^+ \nu_\mu X$	0.5012	±0.0049	-0.0021	± 0.0053
2012 $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$	0.5029	±0.0036	+0.0001	± 0.0036
2011 $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$	0.4981	±0.0054	-0.0029	± 0.0062
combined	0.5030	±0.0019	∓0.0006	±0.0020
simultaneous fit	0.5025	±0.0020	∓0.0006	±0.0020

Table 4.29.: Results of the Δm_d fit. The results of the single fits performed separately on each data set are given. The combination of the four separate fits ("combined") is calculated following equation 4.53. The result of one simultaneous fit to all four data sets is also given ("simultaneous fit"). The uncertainty on the 2011 $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ data result has been scaled according to equation 4.52.

cheated with 4% mistag probability, as explained in section 4.8. The difference of the latter minus the reference value is assumed to be the absolute knowledge of the bias removal on Δm_d. Hence, this value is taken as additional k factor uncertainty.

The statistical precision on Δm_d and the uncertainty of the k factor method are uncorrelated. They are added in quadrature per data set, given in the fourth column of table 4.29. This combined uncertainty per data set is used to derive the combined result of Δm_d.

Finally, one fit is performed to all four data sets simultaneously. The mistag probability parameters are fitted per data set, but Δm_d is a shared fit parameter, minimized on all data sets simultaneously. The fit input parameters, i.e. flight distance resolution, decay time acceptance and k factor correction, are set for each data set separately. The result of this simultaneous fit is also given in table 4.29. It is in agreement with the combined fit result, in terms of central value and of statistical precision. This reassures that the above explained treatment of the underestimated statistical uncertainty on the 2011 $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ data sample is correct.

The final result on Δm_d measured on 3.2×10^6 tagged semileptonic $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ decays, extracted from the full Run I LHCb data of $3\,\text{fb}^{-1}$, is

$$\Delta m_d = 0.503 \pm 0.002\,(\text{stat}) \pm 0.001\,(\text{syst})\,\text{ps}^{-1}. \tag{4.55}$$

The measurement agrees within 2σ with the current world average from Fall 2014 [22] of $\Delta m_d = 0.510 \pm 0.003\,\text{ps}^{-1}$. This is the world's best measurement of this quantity.

4.11. Summary

The B^0–$\bar{B}^0$ oscillation frequency Δm_d has been precisely measured using semileptonic B^0 decays. Two decay channels were reconstructed, $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$. The full LHCb Run I data, corresponding to $3\,\text{fb}^{-1}$ or 26×10^{10} $b\bar{b}$ pairs, has been exploited. A precision of $2\,\text{ns}^{-1}$

due to statistical uncertainties and of $1\,\mathrm{ns}^{-1}$ due to systematic uncertainties has been achieved. This results in a measurement of

$$\Delta m_d = 0.503 \pm 0.002\,(\text{stat}) \pm 0.001\,(\text{syst})\,\mathrm{ps}^{-1},$$

which is the world's best single measurement to date and supersedes the current world average from Fall 2014 [22] of $\Delta m_d = 0.510 \pm 0.003\,\mathrm{ps}^{-1}$. The two values agree within 2σ.

This precision can only be obtained, when every part of the analysis performs at its best. The crucial ingredients were a maximum statistical power of the tagged data sample, provided by the ANN-based OS kaon tagger combined with the OS muon and electron taggers, and an excellent reconstruction of the decay.

The k factor method was used to correct the measured B^0 decay time on a statistical basis for the missing momentum of the non-reconstructed neutrino. One of the two main challenges of the analysis was the understanding of the k factor dependence on the true B^0 decay time, which caused a $4\,\mathrm{ns}^{-1}$ shift on Δm_d. The k factor method was improved to account for this dependence and to remove the shift on Δm_d.

The other main challenge of the analysis was the treatment of the non-oscillating B^+ background. Due to the missing momentum carried away by the neutrino, the B^0 mass resonance cannot be reconstructed. Hence, a separation of B^0 and B^+ decays using a fit to the B^0 mass resonance is not possible. Therefore a multivariate analysis technique was used, exploiting the decay topology of the semileptonic B^+ decay. The absolute amount of B^+ background was reduced and the remaining fraction of B^+ decays in data was measured.

Conclusion

A novel flavour tagging algorithm and a precision measurement of the oscillation frequency in the B^0–$\overline{B}^0$ system Δm_d have been presented. The full LHC Run I data set was analysed corresponding to $1\,\mathrm{fb}^{-1}$ of data taken in 2011 by the LHCb experiment and $2\,\mathrm{fb}^{-1}$ taken in 2012. With $26 \cdot 10^{10}$ recorded $b\bar{b}$ pairs LHCb is an ideal place to perform precision measurements in the B-system.

Flavour tagging algorithms exploit the $b\bar{b}$ pair production and the signal b-hadron hadronization process to determine the production flavour of neutral B mesons. This is challenging at the LHC, because inelastic proton-proton interactions are accompanied by a large amount of soft QCD background. Tagging algorithms have to distinguish between this background and tagging particles, which are correlated to the b production flavour. In the standard LHCb approach tagging particles are selected using a cut-based selection optimized on Monte Carlo (MC) simulation.

In case of the opposite side kaon tagging algorithm a better approach can be chosen, because the tagging performance is significantly worse in data compared to MC. Two main reasons for that have been determined. The amount of soft QCD background is not properly modelled in MC. Additionally, the impact parameter resolution in MC is too good compared to data.

Thus, a novel opposite side kaon tagging algorithm has been developed. The cut-based selection of tagging particles was replaced by a multi-objective optimizing algorithm based on machine learning. Two artificial neural networks (ANN) have been trained. The first ANN classifies tagging particles to distinguish them from soft QCD background efficiently. It assigns higher weights to those particles, which are likely to be correlated to the b flavour, and lower weights to those, which are less likely to be correlated to the flavour of the b hadron. The particles with the highest weights according to the response of the first ANN are combined in a second ANN to derive the tagging decision on the b production flavour. The first ANN is trained on MC simulation, while the second ANN is trained on data. An increase of the performance of the opposite side kaon tagging algorithm of 50% and 30% is achieved on $B^+ \to J/\psi\, K^+$ data. The second number corresponds to a readjustment of the algorithm to the B_s^0 production topology. For the first time, a tagging algorithm with sufficient tagging performance has been developed, to demonstrate, that the optimal working point of opposite side tagging algorithms does depend on the signal B type.

The mixing frequency Δm_d has been measured precisely using semileptonic B^0 decays and the new ANN-based opposite side kaon tagging algorithm combined with other tagging algorithms. The B^0 decays into a charmed meson, a muon and a muon neutrino. The advantage of semileptonic decays is high statistics, a consequence of large branching fractions of the decay modes and a clean trigger signature provided by the muon. However, the muon neutrino is not reconstructible. Therefore, the B^0 decay time is statistically corrected for the missing momentum carried away by the neutrino. This correction is a crucial ingredient for the time-dependent measurement of Δm_d. It introduces a time-dependent damping on the mixing asymmetry. In this thesis an improved correction method has been developed to avoid a bias on Δm_d, which was beforehand dominating the precision of the analysis. Another challenge of the measurement is non-oscillating

B^+ background. The missing neutrino does not allow to reconstruct a B^0 mass resonance. Thus, B^+ background cannot be separated from B^0 signal using a fit to the reconstructed B^0 mass distribution. A multivariate classifier has been used instead, which exploits track isolation. The absolute amount of B^+ background has been reduced and the fraction of remaining background has been measured on data. A result of

$$\Delta m_d = 0.503 \pm 0.002 \,(\text{stat}) \pm 0.001 \,(\text{syst}) \,\text{ps}^{-1} \tag{4.56}$$

has been obtained. The precision of this measurement supersedes the HFAG world average from Fall 2014 [22] of $\Delta m_d = 0.510 \pm 0.003 \,\text{ps}^{-1}$. The mean values agree within $2\,\sigma$.

The experimental precision on Δm_d is a factor of 30 smaller than the uncertainty of the SM prediction. Thus, New physics (NP) that would show up as deviations from SM predictions are not detectable in the measurement of Δm_d. However, the closely related semileptonic *CP* asymmetry in the B^0–$\overline{B}^0$ system, a_{sl}^d, is suitable to detect NP. Here, the current experimental precision, measured by the LHCb experiment [1], is a factor 32 larger than the SM prediction uncertainty. The measurement of a_{sl}^d is time-dependent, thus, the measurement of a_{sl}^d also depends on the B^0–$\overline{B}^0$ oscillation frequency Δm_d. With the precision measurement of Δm_d LHCb has shown to control the decay time reconstruction of semileptonic B^0 decays at a very high level. Hence, LHCb will be sensitive to NP in the B^0–$\overline{B}^0$ system once more LHC collision data is taken.

The improved flavour tagging algorithm will be applied to future *CP* asymmetry measurements, that rely on flavour tagging, e.g. the *CP* violating phase ϕ_s in $B_s^0 \to J/\psi\,\phi$ decays. It will enhance the sensitivity of the measurements on top of the reduced statistical uncertainty due to larger data sets collected in Run II of the LHC. In this way, LHCb will be capable to rule out, whether there are NP contributions in the B^0–$\overline{B}^0$ and B_s^0–$\overline{B}_s^0$ systems or not.

Bibliography

[1] R. Aaij et al. Measurement of the Semileptonic CP Asymmetry in B^0– $\bar{B}^0$ Mixing. *PRL*, 114:041601–9, 2015, `arXiv:1409.8586 [hep-ex]`.

[2] K. Kreplin, G. Krocker, and S. Hansmann-Menzemer. *The Opposite-side kaon tagger: Data MC performance comparison and optimization using neural networks.* LHCb-INT-2013-014, 2014.

[3] C. Bozzi, M. Calvi, M. Fiore, M. Frosini, L. Grillo, B. Khanji, K. Kreplin, A. Lupato, S. Stahl, J. van Tilburg, S. Vecchi, M. Vesterinen, J. de Vries, and S. Wandernoth. Measurement of the mixing frequency Δm_d using semileptonic B^0 decays. Internal note LHCb-ANA-2013-079, 2015.

[4] I. Brock and T. Schörner-Sadenius. *Physics at the Terascale.* Wiley-VCH, Weinheim, Germany, 2001.

[5] G. Altarelli. The Standard Model of Particle Physics. 2005, `arXiv:0510281v1 [hep-ph]`.

[6] U. Nierste. Three Lectures on Meson Mixing and CKM phenomenology. 2009, `arXiv:0904.1869 [hep-ph]`.

[7] D. Boutigny et al. The BABAR Physics Book: Physics at an Asymmetric B Factory. page 1056, 1998.

[8] K. Anikeev et al. B Physics at the Tevatron: Run II and Beyond. page 583, 2001, `arXiv:0201071v2 [hep-ph]`.

[9] S. Glashow. Partial-Symmetries of Weak Interactions. *Nucl.Phys.*, 22:579, 1961.

[10] S. Weinberg. A Model of Leptons. *Phys.Rev.Lett.*, 19:1264, 1967.

[11] A. Salam and J. Ward. Electromagnetic and weak interactions. *Phys.Lett.*, 13:168, 1964.

[12] P. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys.Rev.Lett.*, 13(16): 508–509, 1964.

[13] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys.Rev.Lett.*, 13(9):321–323, 1964.

[14] G. Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716(716):1–29, 2012, `arXiv:1207.7214 [hep-ex]`.

Bibliography

[15] S. Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716(716):30–61, 2012, `arXiv:1207.7235 [hep-ex]`.

[16] M. Kobayashi and T. Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Prog.Theor.Phys.*, 49:652–657, 1973.

[17] L. Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Phys.Rev.Lett.*, 51: 1945–1947, 1983.

[18] K. A. Olive et al. Review of Particle Physics. *Chin. Phys. C*, 38(090001), 2014.

[19] La thuile conference, 2015. 1-7 March La Thuile, Aosta Valley, Italy.

[20] CKMfitter group, J. Charles, et al. Updated results and plots available at: `http://ckmfitter.in2p3.fr`, 2015.

[21] U. Nierste. Flavour physics at LHCb beyond Run I: a theorist's view on challenges and opportunities, 2015. Presentation at the 75th LHCb Week on 27 Feb 2015.

[22] Y. Amhis et al. Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2014. 2014, `arXiv:1412.7515 [hep-ex]`. Updated results and plots available at: `http://www.slac.stanford.edu/xorg/hfag/`.

[23] ARGUS collaboration, H. Albrecht *et al.*. *Observation of B^0 - $\overline{B}^0$ mixing*. Phys.Lett **B192** 245, 1987.

[24] A. Lenz, U. Nierste, et al. Anatomy of new physics in B–$\overline{B}$ mixing. *Phys.Rev.D.*, 83:036004, 2011, `arXiv:1008.1593 [hep-ph]`.

[25] A. Lenz. B-mixing in and beyond the Standard model. 2014, `arXiv:1409.6963 [hep-ph]`.

[26] A. Lenz and U. Nierste. Numerical Updates of Lifetimes and Mixing Parameters of B Mesons. 2011, `arXiv:1102.4274 [hep-ph]`.

[27] V. M. Abazov et al. Measurement of the anomalous like-sign dimuon charge asymmetry with 9 fb^{-}1 of $p\bar{p}$ collisions. *Phys.Rev.*, D84:052007, 2011, `arXiv:1106.6308 [hep-ex]`.

[28] G. Borissov and B. Hoeneisen. Understanding the like-sign dimuon charge asymmetry in p $\overline{p}$ collisions. 2013, `arXiv:1303.0175 [hep-ph]`.

[29] A. Lenz, U. Nierste, et al. Constraints on new physics in B–$\overline{B}$ mixing in the light of recent LHCb data. *Phys.Rev.D.*, 86:033008, 2012, `arXiv:1203.0238 [hep-ph]`.

[30] E. L. and B. P. LHC Machine. *JINST*, 3:S08001, 2008.

[31] LHC machine outreach. `http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/lhc_in_pictures.htm`, 2015.

[32] L. M. Status of the LHC. *J. Phys. Conf. Ser.*, 455:012001, 2013.

[33] ATLAS Collaboration. *ATLAS Experiment – Public Results.* https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults, 2015.

[34] LHCb Collaboration. Material for presentations, LHCb webpage. `http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material_for_Presentations.html`, 2015.

[35] E. Norrbin and T. Sjöstrand. Production and Hadronization of Heavy Quarks. *Eur.Phys.J.*, C17:137–161, 2000, `arXiv:0005110v1 [hep-ph]`.

[36] R. Aaij et al. Measurement of the B meson production cross-sections in proton proton collisions at $\sqrt{s}$ =7 TeV. *JHEP*, 08:17, 2013, `arXiv:1306.3663 [hep-ex]`.

[37] A. A. Alves Jr. et al. The LHCb detector at the LHC. *JINST*, 3:S08005, 2008.

[38] R. Aaij et al. LHCb Detector Performance. *IJMPA*, 30(07):1530022, 2015, `arXiv:1412.6352 [hep-ex]`.

[39] J. Brehmer, J. Albrecht, and P. Seyfert. Ghost probability: an efficient tool to remove background tracks. LHCb-INT-2012-025, September 2012.

[40] The LHCb Collaboration. *The LHCb Detector at the LHC.* 2008 JINST 3 S08005, 2008.

[41] R. Aaij et al. Performance of the LHCb Vertex Locator. *JINST*, 9:P09007, 2014.

[42] The LHCb Collaboration. LHCb Silicon Tracker - Material for Publications. `http://lhcb.physik.uzh.ch/ST/public/material/index.php`, 2015.

[43] The LHCb Collaboration. *LHCb Inner Tracker design report.* CERN-LHCC/2002-029, 2003.

[44] R. Arink et al. Performance of the LHCb Outer Tracker. *JINST*, 9:P01002, 2014.

[45] The LHCb Collaboration. *LHCb Outer Tracker Technical Design Report.* CERN/LHCC-2001-024, 2001.

[46] C. Jones. ANN PID Retuning for Reco 14 Data, 2013. Presentation in the Physics Performance, Trigger & Stripping Meeting on 6 May.

[47] The LHCb Collaboration. *LHCb RICH technical design report.* CERN-LHCC/2000-037, 2000.

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

[49] The LHCb Collaboration. *LHCb Calorimeters Technical Design Report.* CERN/LHCC-2000-036, 2000.

[50] F. Archilli et al. Performance of the Muon Identification at LHCb. *J. Instrum.*, 8:P10020, 2013.

[51] The LHCb Collaboration. *LHCb Muon Technical Design Report.* CERN/LHCC-2000-037, 2000.

[52] A. A. Alves Jr. et al. Performance of the LHCb muon system. *JINST.*, 8:P02022, 2013.

[53] R. Aaij et al. The LHCb trigger and its performance in 2011. *JINST.*, 8:P04022, 2013.

[54] J. Albrecht et al. Performance of the LHCb High Level Trigger in 2012. *J. Phys. Conf. Ser.*, 513:012001, 2013.

[55] M. Frank et al. Deferred High Level Trigger in LHCb: A Boost to CPU Resource Utilization. *J. Phys. Conf. Ser.*, 513:012006, 2013.

[56] G. Corti et al. Software for the LHCb Experiment. *IEEE Transactions on Nuclear Science*, 53(3), 2006.

[57] G. Barrand et al. GAUDI - A software architecture and framework for building HEP data processing applications. *Comput. Phys. Commun.*, 140(45), 2001.

[58] M. Clemencic et al. Recent Developments in the LHCb Software Framework Gaudi. *J. Phys.: Conf. Ser.*, 219:042006, 2010.

[59] LHCb Collaboration. *The Gauss Project.* http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/, 2015.

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

[61] T. Sjöstrand, S. Mrenna, and P. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006, `arXiv:0603175 [hep-ph]`.

[62] T. Sjöstrand, S. Mrenna, and P. Skands. A brief introduction to pythia 8.1. *Comput. Phys.Commun.*, 178:852–867, 2008, `arXiv:0710.3820v1 [hep-ph]`.

[63] I. Belyaev et al. Handling of the generation of primary events in Gauss, the LHCb simulation framework. *Nuclear Science Symposium Conference Record (NSS/MIC)*, IEEE:1155, 2010.

[64] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[65] P. Golonka and Z. Was. PHOTOS Monte Carlo: a precision tool for QED corrections in *Z* and *W* decays. *Eur.Phys.J.*, C45:97–107, 2006, `arXiv:0506026 [hep-ph]`.

[66] S. Agostinelli et al. Geant4: a simulation toolkit. *Nucl. Instrum. Meth.*, A506:250, 2003.

[67] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Dubois, et al. Geant4 developments and applications. *IEEE Trans.Nucl.Sci.*, 53:270, 2006.

[68] LHCb Collaboration. *The Boole Project.* http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/, 2015.

[69] LHCb Collaboration. *The Brunel Project.* http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/, 2015.

[70] LHCb Collaboration. *The DaVinci Project.* http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/, 2015.

[71] LHCb Collaboration. *The Moore Project.* http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/, 2015.

[72] K. A. Olive et al. Review of particle physics. *Chin.Phys.*, C38:090001, 2014.

[73] G. Krocker. *Development and Calibration of a Same Side Kaon Tagging Algorithm and Measurement of the B_s^0– $\bar{B}_s^0$ Oscillation Frequency Δm_s at the LHCb Experiment.* PhD thesis, Heidelberg U., Nov 2013. Presented 20 Nov 2013.

[74] J. Weatherall. *CP* Violation in the B^0 meson system with BaBar. page 9, 2001.

[75] M. Grabalosa. *Flavour Tagging developments within the LHCb experiment.* PhD thesis, Barcelona U., Mar 2012. Presented 15 May 2012.

[76] R. Aaij et al. Opposite-side flavour tagging of b mesons at the LHCb experiment. oai:cds.cern.ch:1426509. *Eur. Phys. J. C*, 72:2022, February 2012, `arXiv:1202.4979`. CERN-PH-EP-2012-039. LHCb-PAPER-2011-027.

[77] R. Aaij et al. Measurement of the B^+, B^0, B_s^0 meson and Λ_b^0 baryon lifetimes. *JHEP*, 04: 10.1007–25, 2014, `arXiv:1402.2554 [hep-ex]`.

[78] Y. Xie. sFit: a method for background subtraction in maximum likelihood fit. 2009.

[79] M. Pivk and F. R. Le Diberder. sPlot: A statistical tool to unfold data distributions. *Nucl. Instrum. Meth.*, A555(1-2):356–369, 2005.

[80] A. Powell et al. PIDCalib: packages for extracting pid performance results from both collision data and mc. `https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage`.

[81] C. M. Bishop. *Pattern Recognition and Machine Learning.* Springer, New York, NY 10013, USA, 9th edition, 2006.

[82] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 2:303–314, 1989.

[83] C. G. Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA J Appl Math*, 6(1):76–90, 1970.

[84] R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3): 317–322, 1970.

[85] D. Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–26, January 1970.

[86] D. F. Shanno. Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656, July 1970.

[87] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. TMVA 4, toolkit for multivariate data analysis with ROOT, users guide. 040, November 2009, `arXiv:physics/0703039`.

[88] R. Aaij et al. Measurement of the *CP*-Violating Phase ϕ_s in the decay $B_s^0 \to J/\psi \phi$. *PRL*, 108:101803–8, 2012, `arXiv:1304.2600 [hep-ex]`.

[89] J. Zhong, R.-S. Huang, and S.-C. Lee. A program for the bayesian neural network in the ROOT framework. *Computer Physics Communications*, 182(12):2655–2660, December 2011, `arXiv:1103.2854`.

[90] K. Kreplin, G. Krocker, G. Tellarini, S. Hansmann-Menzemer, S. Vecchi, and M. Dorigo. *Neural-network-based kaon tagging: development of new same side and opposite side tagging algorithms*. LHCb-ANA-2014-003, 2015.

[91] B. Aubert et al. Measurement of the $\overline{B}^0$ lifetime and the $B^0 \overline{B}^0$ oscillation frequency using partially reconstructed $\overline{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays. *Phys.Rev.*, D73:012004, 2006, `arXiv:0507054 [hep-ex]`.

[92] K. Abe et al. Improved Measurement of *CP*-Violation Parameters $\sin 2\phi_1$ and $|\lambda|$, *B* Meson Lifetimes, and B^0– $\overline{B}^0$ Mixing Parameter Δm_d. *Phys.Rev.*, D71:072003, 2005, `arXiv:0408111 [hep-ex]`.

[93] R. Aaij et al. Measurement of the B^0– $\overline{B}^0$ oscillation frequency Δm_d with the decays $B^0 \to D^- \pi^+$ and $B^0 \to J/\psi K^{*0}$. *PLB*, 719:318–325, 2013, `arXiv:1210.6750 [hep-ex]`.

[94] F. Abe et al. Measurement of the $B^0 \overline{B}^0$ oscillation frequency using dimuon data in $p \bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev.*, D 60:051101–6, 1999.

[95] C. Langenbruch. *Measurement of the B_s^0 mixing phase in the decay $B_s^0 \to J/\psi \phi$ with the LHCb Experiment*. PhD thesis, Heidelberg U., Nov 2011. Presented 9 Nov 2011.

[96] S. Wandernoth. *Measurement of the B_s^0– $\overline{B}_s^0$ Oscillation Frequency at LHCb using $1\,fb^{-1}$ of data taken in 2011*. PhD thesis, Heidelberg U., Feb 2014. Presented 04 Feb 2014.

[97] T. Bird, S. Easo, U. Kerzel, R. W. Lambert, and K. Vervink. Time dependent semileptonic studies: correcting for missing momentum and the effect of multiple interactions. Technical Report LHCb-INT-2011-004. CERN-LHCb-INT-2011-004, CERN, Geneva, Feb 2011.

[98] M. Williams, V. Gligorov, C. Thomas, H. Dijkstra, J. Nardulli and P. Spradlin. *The HLT2 Topological Lines*. CERN/LHCb-PUB-2011-002, 2011.

[99] V. Gligorov, C. Thomas, and M. Williams. The HLT inclusive *B* triggers. *CERN-LHCb-PUB-2011-016*, 2011.

[100] V. V. Gligorov and M. Williams. Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree. *JINST*, 8:P02013, 2013, `arXiv:1210.6861 [physics.ins-det]`.

[101] R. Aaij et al. Search for the lepton flavour violating decay $\tau^- \to \mu m \, \mu^+ \, \mu m$. *JHEP*, 2015, `arXiv:1409.8548 [hep-ex]`.

[102] R. Aaij et al. Measurement of the $B_s^0 \to \mu^+ \, \mu m$ branching fraction and search for $B^0 \to \mu^+ \, \mu m$ decays at the LHCb experiment. *PRL*, 111:101805, 2013, `arXiv:1307.5024 [hep-ex]`.

[103] Gavardi, Laura. *Search for lepton flavour violation in τ decays at the LHCb experiment.* CERN-THESIS-2013-259, 2013.

[104] Giampiero Mancinelli and Justine Serrano. *Study of Muon Isoltaion in the $B_s^0 \to \mu^+ \, \mu m$ Channel.* LHCb-ANA-2010-011, 2010.

[105] The LHCb Collaboration. *Search for the lepton flavour violating decay $\tau^- \to \mu^+ \, \mu m \, \mu m$.* LHCb-ANA-2014-005, 2014.

[106] S. Dambach, U. Langenegger, and A. Starodumov. Neutrino reconstruction with topological information. *Nuclear Instruments and Methods in Physics Research*, A 569, 2006, `arXiv:0607294 [hep-ex]`.

[107] P. Urquijo. Semileptonic B decay branching fractions. Unpublished internal note.

[108] C. Bozzi. LHCb MC cocktails. Twiki page.

[109] F. James. MINUIT Function Minimization and Error Analysis. *CERN Program Library Long Writeup*, D506(94.1), 1994.

[110] J. Beringer et al. Review of particle physics. *Phys. Rev.*, D86:010001, 2012.

[111] C. Bozzi, M. Calvi, M. Fiore, M. Frosini, L. Grillo, B. Khanji, M. Kolpin, K. Kreplin, A. Lupato, S. Stahl, J. van Tilburg, M. Vesterinen, J. de Vries, and S. Wandernoth. *Measurement of the semileptonic CP asymmetry a_{sl}^d.* LHCb-ANA-2013-050, 2014.

A. Appendix to Flavour Tagging

A.1. Pid Calibration Reweighting
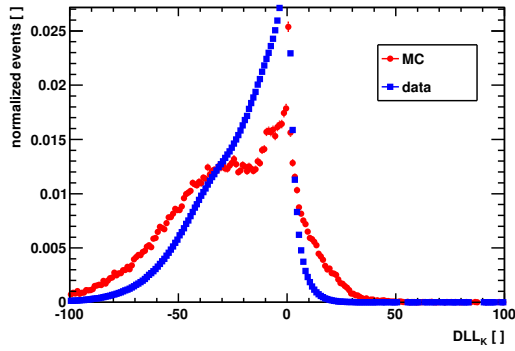
(a) before reweighting
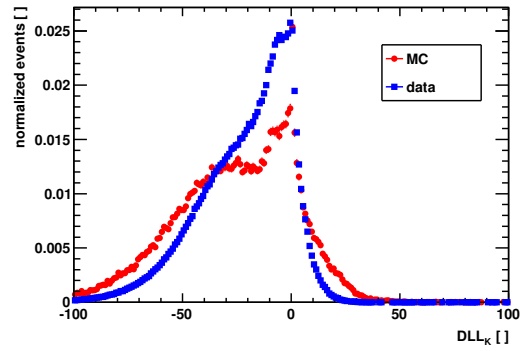
(b) after reweighting
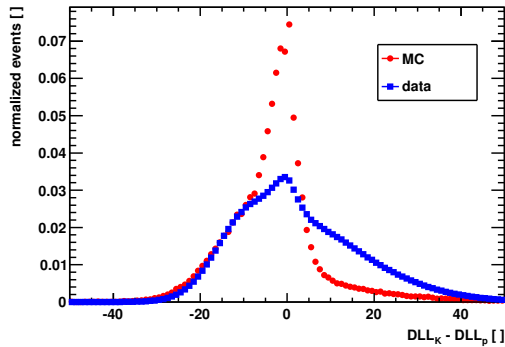
(c) before reweighting
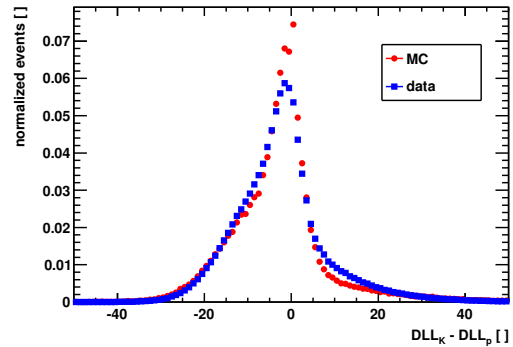
(d) after reweighting

Figure A.1.: $DLL_{K-\pi}$ distributions of true pions in MC and in calibration data (a) before reweighting, (b) after reweighting. DLL_{K-p} distributions of true pions in MC and in calibration data (c) before reweighting, (d) after reweighting.
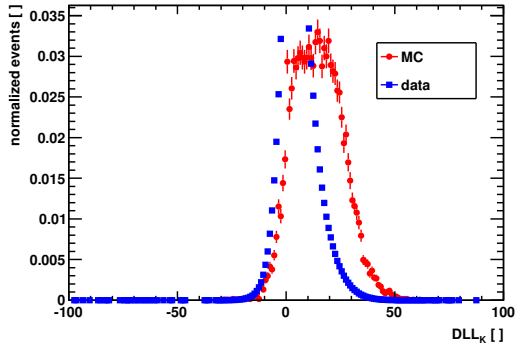
(a) before reweighting
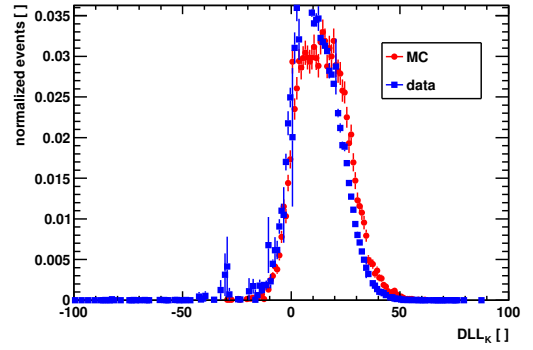
(b) after reweighting
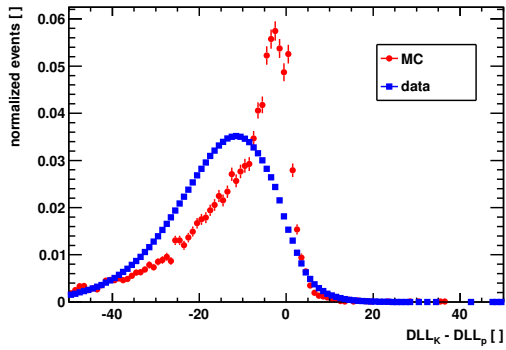
(c) before reweighting
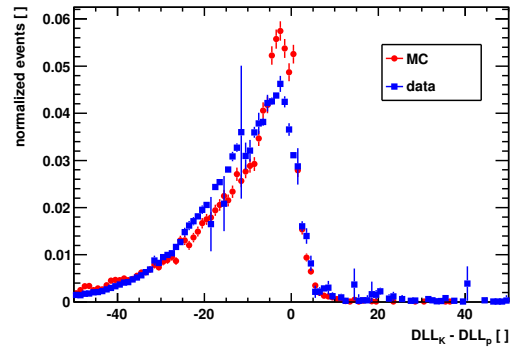
(d) after reweighting

Figure A.2.: $DLL_{K-\pi}$ distributions of true protons in MC and in calibration data (a) before reweighting, (b) after reweighting. DLL_{K-p} distributions of true protons in MC and in calibration data (c) before reweighting, (d) after reweighting.

A.2. Track Multiplicity Correction
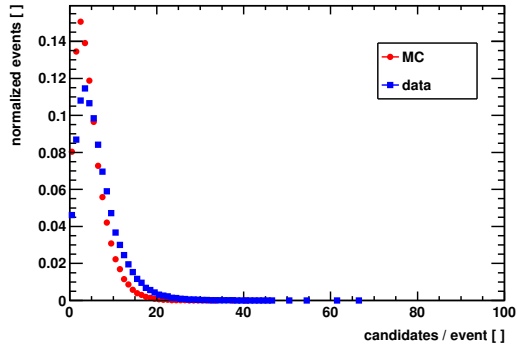
(a) before correction
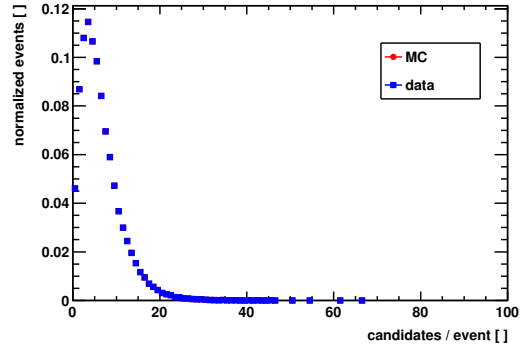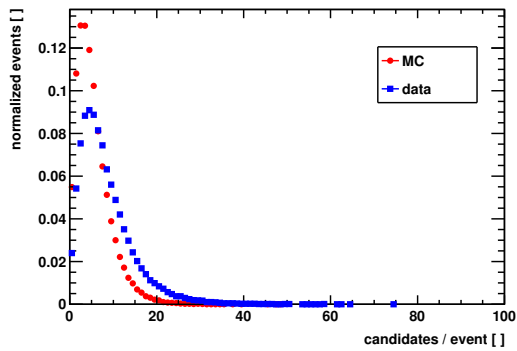
(b) after correction

(c) before correction
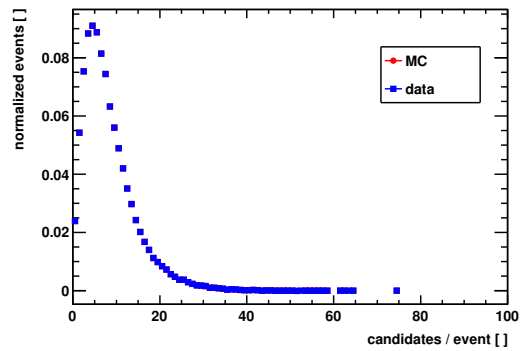
(d) after correction

Figure A.3.: Track multiplicity distribution for events with 1 PV per event (a) before and (b) after correction and with 2 PVs per event (c) before and (d) after correction.
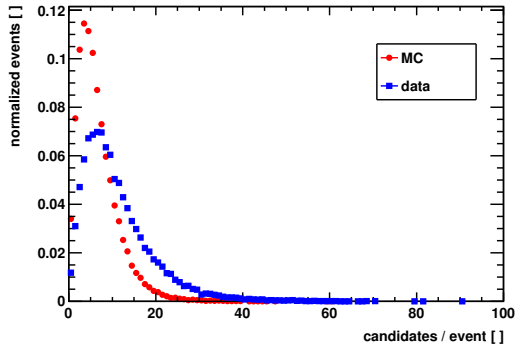
(a) before correction
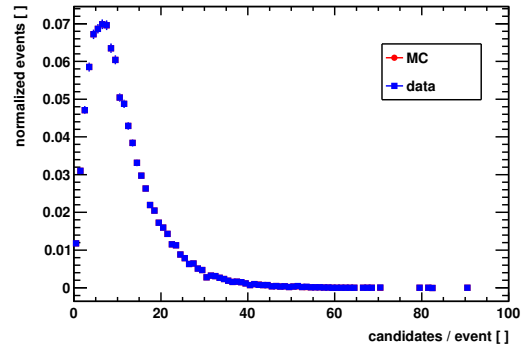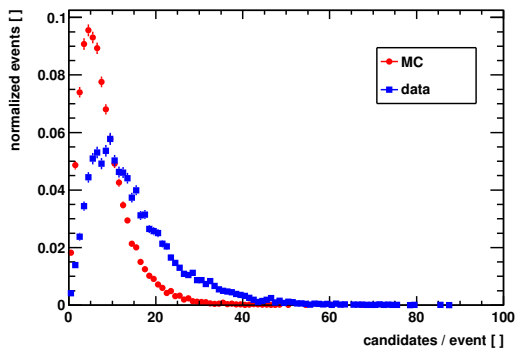
(b) after correction

(c) before correction
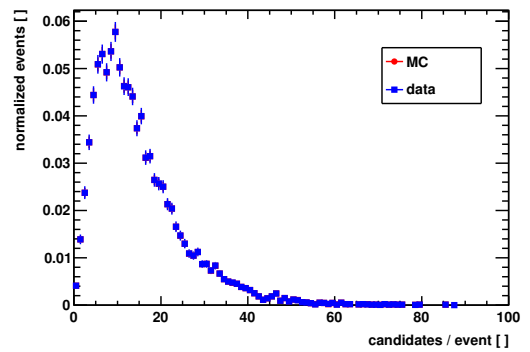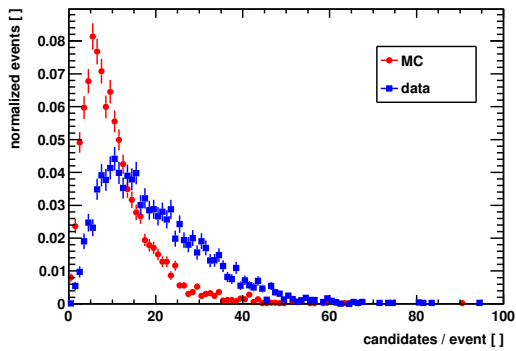
(d) after correction

(e) before correction
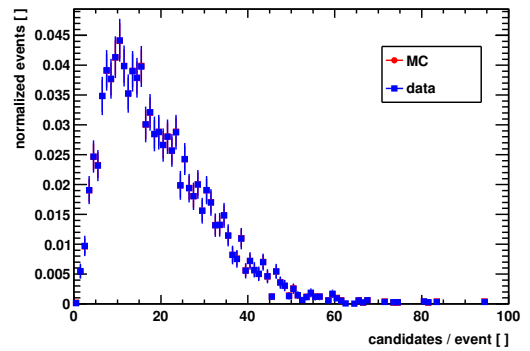
(f) after correction

Figure A.4.: Track multiplicity distribution for events with 3 PVs per event (a) before and (b) after correction, with 4 PVs per event (c) before and (d) after correction and with 5 and more PVs per event (e) before and (f) after correction.
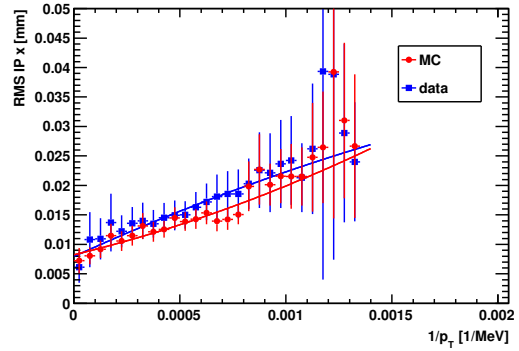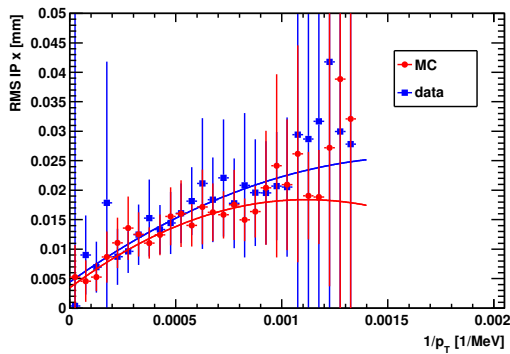
A.3. Impact Parameter Resolution Correction
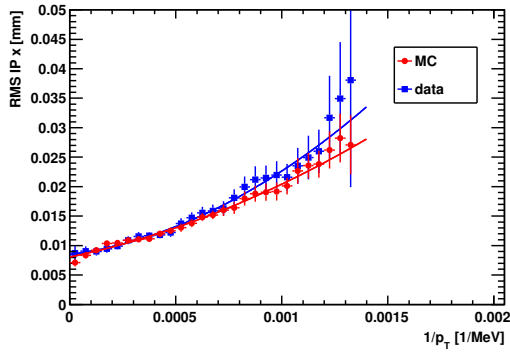
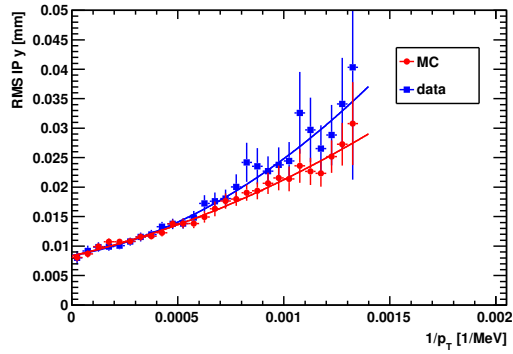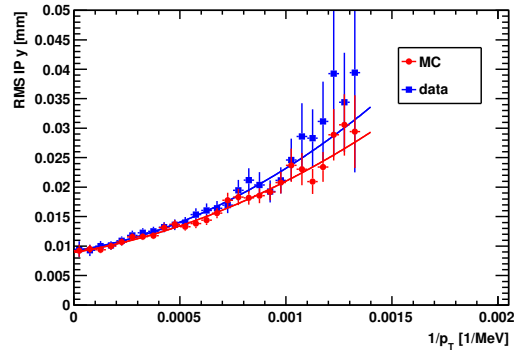(a) 1 PV
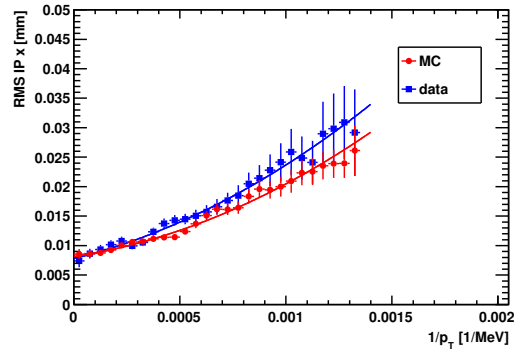
(b) 2 PVs
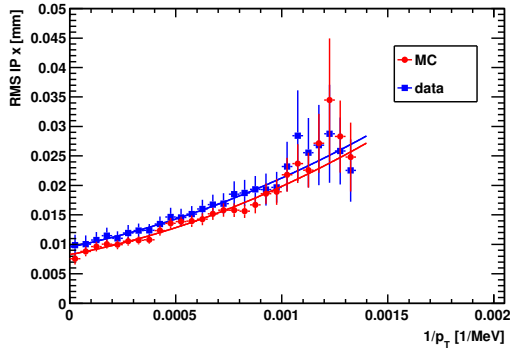
(c) 3 PVs

(d) 4 PVs

(e) 5 PVs

Figure A.5.: Impact parameter resolution distributions in x direction for events with 1 PV per event (a), 2 PVs per event (b), 3 PVs per event (c), 4 PVs per event (d) and 5 or more PVs per event (e).

(a) 1 PV

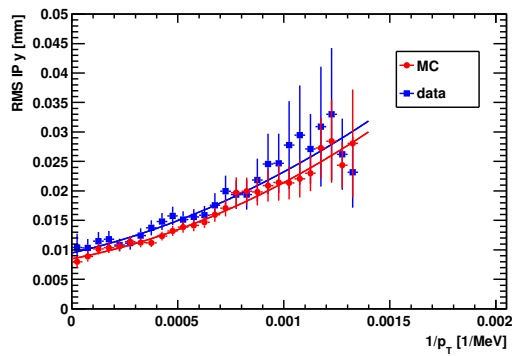(b) 2 PVs

(c) 3 PVs

(d) 4 PVs

(e) 5 PVs

Figure A.6.: Impact parameter resolution distributions in y direction for events with 1 PV per event (a), 2 PVs per event (b), 3 PVs per event (c), 4 PVs per event (d) and 5 or more PVs per event (e).

(a) 1 PV

(b) 2 PVs

(c) 3 PVs

(d) 4 PVs

(e) 5 PVs

Figure A.7.: Corrected impact parameter resolution distributions in x direction for events with 1 PV per event (a), 2 PVs per event (b), 3 PVs per event (c), 4 PVs per event (d) and 5 or more PVs per event (e).
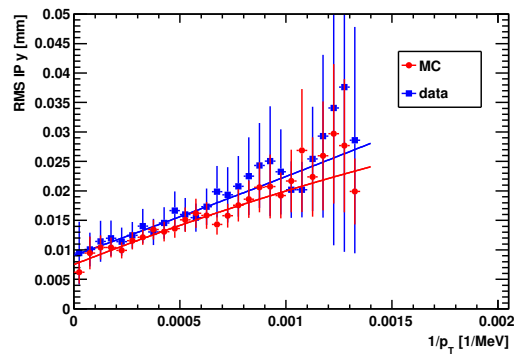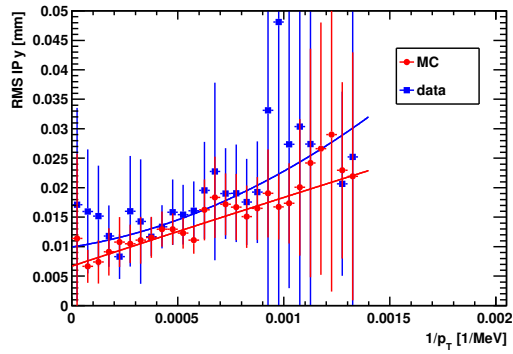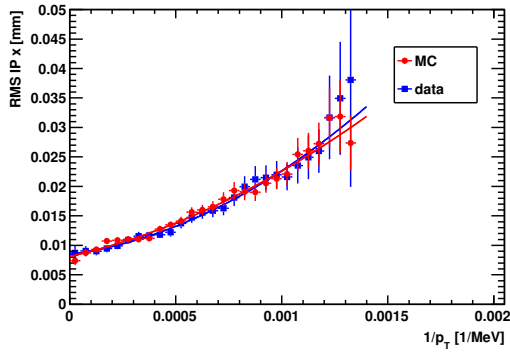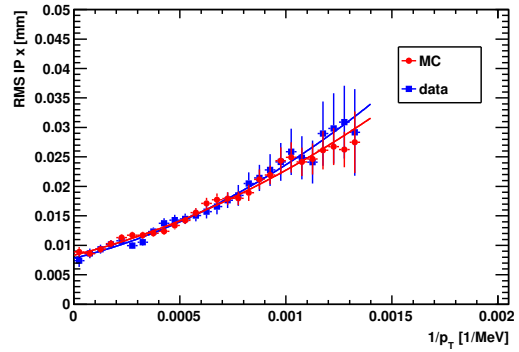
(a) 1 PV

(b) 2 PVs

(c) 3 PVs

(d) 4 PVs

(e) 5 PVs

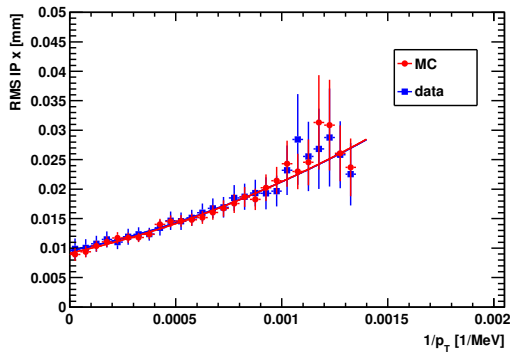Figure A.8.: Corrected impact parameter resolution distributions in y direction for events with 1 PV per event (a), 2 PVs per event (b), 3 PVs per event (c), 4 PVs per event (d) and 5 or more PVs per event (e).
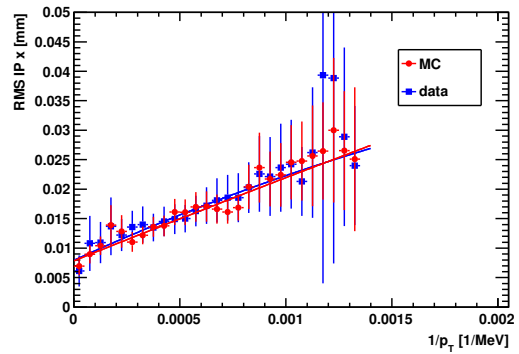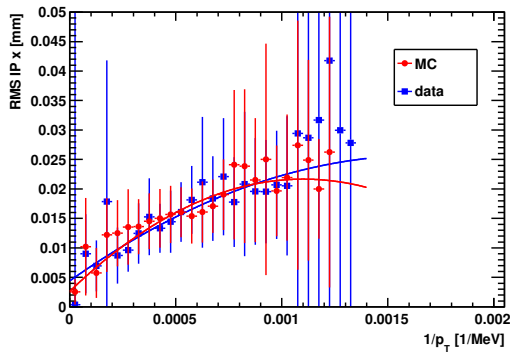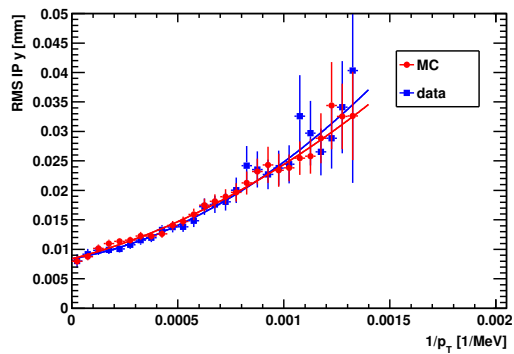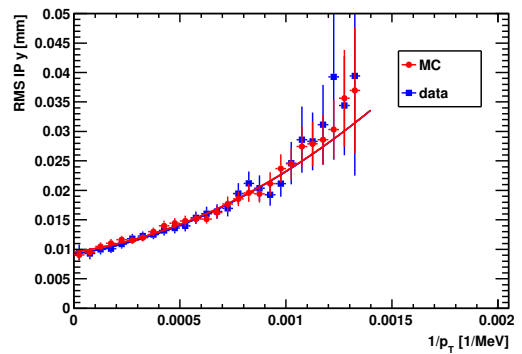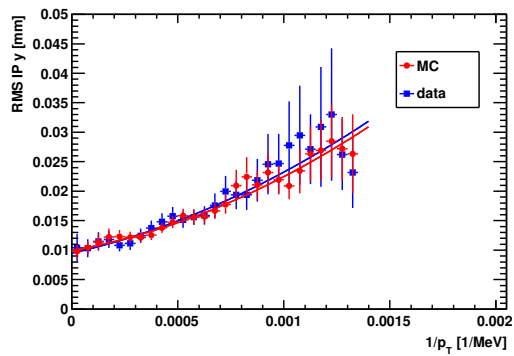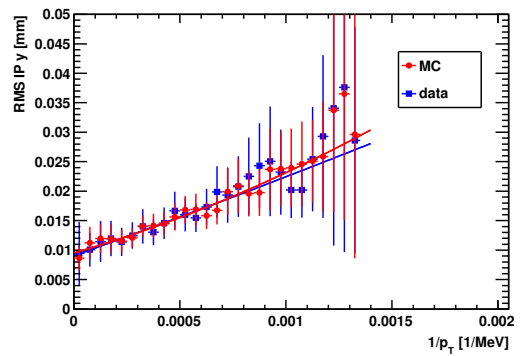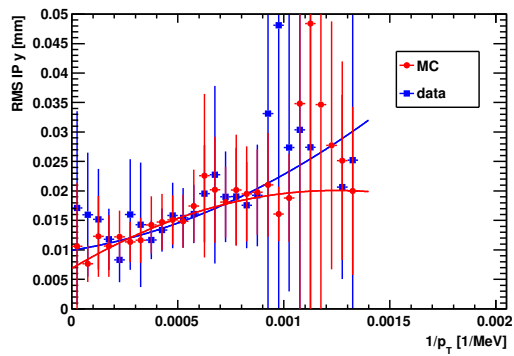
B. Appendix to the Measurement of Δm_d

B.1. Semileptonic Branching Ratios used in MC Generation

Process	Branching fraction (%)
$B^0 \to D^- \mu^+ \nu_\mu$	2.17 ± 0.12
$B^0 \to D^{*-}(\to D^- \pi^0/\gamma)\mu^+ \nu_\mu$	1.62 ± 0.04
$B^0 \to D_0^{*-} \mu^+ \nu_\mu \to D^- \mu^+ \nu_\mu X$	0.14 ± 0.05
$B^0 \to D_1'^- \mu^+ \nu_\mu \to D^- \mu^+ \nu_\mu X$	0.06 ± 0.02
$B^0 \to D_1^- \mu^+ \nu_\mu \to D^- \mu^+ \nu_\mu X$	0.18 ± 0.03
$B^0 \to D_2^{*-} \mu^+ \nu_\mu \to D^- \mu^+ \nu_\mu X$	0.17 ± 0.07
Total $B^0 \to D^{(*(*))-}\mu^+\nu_\mu X$	4.34 ± 0.15
$B^0 \to D^- \pi^0 \mu^+ \nu_\mu$	0.02 ± 0.02
$B^0 \to D^{*-}(\to D^- \pi^0/\gamma)\pi^0 \mu^+ \nu_\mu$	0.01 ± 0.01
$B^0 \to D^{*-}(\to D^- \pi^0/\gamma)\pi^0 \pi^0 \mu^+ \nu_\mu$	0.02 ± 0.02
$B^0 \to D^{*-}(\to D^- \pi^0/\gamma)\pi^+ \pi^- \mu^+ \nu_\mu$	0.09 ± 0.09
$B^0 \to D^- \pi^0 \pi^0 \mu^+ \nu_\mu$	0.03 ± 0.03
$B^0 \to D^- \pi^+ \pi^- \mu^+ \nu_\mu$	0.12 ± 0.12
Total non-resonant B^0	0.30 ± 0.16
$B^0 \to D^- \tau^+ \nu_\tau \to D^- \mu^+ \nu_\mu X$	0.19 ± 0.04
$B^0 \to D^{*-} \tau^+ \nu_\tau \to D^- \mu^+ \nu_\mu X$	0.08 ± 0.03
$B^0 \to D_0^{*-} \tau^+ \nu_\tau \to D^- \mu^+ \nu_\mu X$	0.01 ± 0.02
$B^0 \to D_1'^- \tau^+ \nu_\tau \to D^- \mu^+ \nu_\mu X$	0.01 ± 0.02
$B^0 \to D_1^- \tau^+ \nu_\tau \to D^- \mu^+ \nu_\mu X$	0.01 ± 0.03
$B^0 \to D_2^{*-} \tau^+ \nu_\tau \to D^- \mu^+ \nu_\mu X$	0.01 ± 0.04
Total $B^0 \to D^{(*(*))-}\tau^+\nu_\tau X$	0.31 ± 0.05
Total SL cocktail $B^0 \to D^- \mu^+ X$	4.95 ± 0.23

Table B.1.: Exclusive semileptonic branching ratios used to generate the signal simulation sample in the $B^0 \to D^- \mu^+ \nu_\mu X$. The branching ratios for D^{**} resonances decaying into D^- and at least one pion are already included. For modes decaying via τ leptons, the branching fraction $\tau \to \mu \nu_\mu \nu_\tau$ is included as well. The D^- branching ratio in $K^+ \pi^- \pi^-$ is not included, taken from [3]

Process	Branching fraction (%)
$B^0 \to D^{*-}\mu^+\nu_\mu$	3.39 ± 0.074
$B^0 \to D_0^{*-}\mu^+\nu_\mu \to D^{*-}\mu^+\nu_\mu X$	0.040 ± 0.014
$B^0 \to D_1^{'-}\mu^+\nu_\mu \to D^{*-}\mu^+\nu_\mu X$	0.040 ± 0.015
$B^0 \to D_1^-\mu^+\nu_\mu \to D^{*-}\mu^+\nu_\mu X$	0.12 ± 0.018
$B^0 \to D_2^{*-}\mu^+\nu_\mu \to D^{*-}\mu^+\nu_\mu X$	0.040 ± 0.018
Total $B^0 \to D^{*(*)-}\mu^+\nu_\mu$	3.63 ± 0.081
$B^0 \to D^{*-}\pi^0\mu^+\nu_\mu$	0.030 ± 0.031
$B^0 \to D^{*-}\pi^0\pi^0\mu^+\nu_\mu$	0.044 ± 0.044
$B^0 \to D^{*-}\pi^+\pi^-\mu^+\nu_\mu$	0.170 ± 0.166
Total non-resonant B^0	0.24 ± 0.17
$B^0 \to D^{*-}\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.176 ± 0.059
$B^0 \to D_0^{*-}\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.002 ± 0.001
$B^0 \to D_1^{'-}\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.004 ± 0.001
$B^0 \to D_1^-\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.006 ± 0.002
$B^0 \to D_2^{*-}\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.003 ± 0.001
Total $B^0 \to D^{*(*)-}\tau^+\nu_\tau X$	0.190 ± 0.059
Total SL cocktail $B^0 \to D^{*-}\mu^+ X$	4.07 ± 0.20

Table B.2.: Exclusive semileptonic branching ratios used to generate the signal simulation sample in the $B^0 \to D^{*-}\mu^+\nu_\mu X$. The branching ratios for D^{**} resonances decaying into D^* and at least one pion are already included, as well as the branching fraction of D^{*+} decaying into $D^0\pi^+$. For modes decaying via τ leptons, the branching fraction $\tau \to \mu\nu_\mu\nu_\tau$ is included as well. The $\bar{D}^0$ branching ratio in $K^+\pi^-$ is not included, taken from [3]

Process	Branching fraction (%)
$B^+ \to D_0^{*0}\mu^+\nu_\mu \to D^-\mu^+\nu_\mu X$	0.26 ± 0.01
$B^+ \to D_1^{'0}\mu^+\nu_\mu \to D^-\mu^+\nu_\mu X$	0.07 ± 0.02
$B^+ \to D_1^0\mu^+\nu_\mu \to D^-\mu^+\nu_\mu X$	0.21 ± 0.03
$B^+ \to D_2^{*0}\mu^+\nu_\mu \to D^-\mu^+\nu_\mu X$	0.23 ± 0.10
Total $B+ \to D^{**}\mu^+\nu_\mu X$	0.77 ± 0.14
$B^+ \to D^-\pi^+\mu^+\nu_\mu$	0.04 ± 0.04
$B^+ \to D^-\pi^0\pi^+\mu^+\nu_\mu$	0.05 ± 0.05
$B^+ \to D^{*-}(\to D^-\pi^0/\gamma)\pi^+\mu^+\nu_\mu$	0.03 ± 0.03
$B^+ \to D^{*-}(\to D^-\pi^0/\gamma)\pi^0\pi^+\mu^+\nu_\mu$	0.04 ± 0.04
Total non-resonant B^+	0.16 ± 0.08
$B^+ \to D_0^{*0}\tau^+\nu_\tau \to D^-\mu^+\nu_\mu X$	0.010 ± 0.003
$B^+ \to D_1^{'0}\tau^+\nu_\tau \to D^-\mu^+\nu_\mu X$	0.013 ± 0.004
$B^+ \to D_1^0\tau^+\nu_\tau \to D^-\mu^+\nu_\mu X$	0.006 ± 0.002
$B^+ \to D_2^{*0}\tau^+\nu_\tau \to D^-\mu^+\nu_\mu X$	0.015 ± 0.005
Total $B^+ \to D^{**}\tau^+\nu_\tau \to D^-\mu^+\nu_\mu X$	0.04 ± 0.01
Total SL cocktail $B^+ \to D^-\mu^+ X$	0.97 ± 0.16

Table B.3.: Exclusive semileptonic branching ratios used to generate the B^+ background sample in the $B^0 \to D^-\mu^+\nu_\mu X$. The branching ratios for D^{**} resonances decaying into D^- and at least one pion are already included, as well as the branching fraction of D^{*+} decaying into $D^0\pi^+$. For modes decaying via τ leptons, the branching fraction $\tau \to \mu\nu_\mu\nu_\tau$ is included as well. The $D^- \to K^+\pi^-\pi^-$ branching ratio is not included, taken from [3]

Process	Branching fraction (%)
$B^+ \to D_0^{*0}\mu^+\nu_\mu \to D^{*-}\mu^+\nu_\mu X$	0.03 ± 0.01
$B^+ \to D_1^{'0}\mu^+\nu_\mu \to D^{*-}\mu^+\nu_\mu X$	0.09 ± 0.03
$B^+ \to D_1^0\mu^+\nu_\mu \to D^{*-}\mu^+\nu_\mu X$	0.22 ± 0.03
$B^+ \to D_2^{*0}\mu^+\nu_\mu \to D^{*-}\mu^+\nu_\mu X$	0.07 ± 0.03
Total $B^+ \to D^{**}\mu^+\nu_\mu X$	0.41 ± 0.05
$B^+ \to D^{*-}\pi^+\mu^+\nu_\mu$	0.06 ± 0.06
$B^+ \to D^{*-}\pi^+\pi^0\mu^+\nu_\mu$	0.08 ± 0.08
Total non-resonant B^+	0.15 ± 0.11
$B^+ \to D_0^{*0}\mu^+\nu_\mu\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.0012 ± 0.0002
$B^+ \to D_1^{'0}\mu^+\nu_\mu\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.0076 ± 0.0011
$B^+ \to D_1^0\mu^+\nu_\mu\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.0104 ± 0.0015
$B^+ \to D_2^{*0}\mu^+\nu_\mu\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.0048 ± 0.0007
Total $B^+ \to D^{**}\tau^+\nu_\tau \to D^{*-}\mu^+\nu_\mu X$	0.0241 ± 0.0020
Total SL cocktail $B^+ \to D^-\mu^+ X$	0.58 ± 0.12

Table B.4.: Exclusive semileptonic branching ratios used to generate the B^+ background sample in the $B^0 \to D^{*-}\mu^+\nu_\mu X$. The branching ratios for D^{**} resonances decaying into D^* and at least one pion are already included, as well as the branching fraction of D^{*+} decaying into $D^0\pi^+$. For modes decaying via τ leptons, the branching fraction $\tau \to \mu\nu_\mu\nu_\tau$ is included as well. The $\overline{D}^0 \to K^+\pi^-$ branching ratio is not included, taken from [3]

Sample	Branching fraction	
Signal $B^0 \to D^{*-}\mu^+\nu_\mu X$ cocktail	$(4.07 \pm 0.20 \pm 0.08)\%$	$\times \mathcal{B}(D^0 \to K^-\pi^+)$
Signal $B^0 \to D^-\mu^+\nu_\mu X$ cocktail	$(4.95 \pm 0.23 \pm 0.08)\%$	$\times \mathcal{B}(D^- \to K^+\pi^-\pi^-)$
Bkg $B^+ \to D^{*-}\mu^+\nu_\mu X$ cocktail	$(0.58 \pm 0.12 \pm 0.06)\%$	$\times \mathcal{B}(D^0 \to K^-\pi^+)$
Bkg $B^+ \to D^-\mu^+\nu_\mu X$ cocktail	$(0.97 \pm 0.16 \pm 0.08)\%$	$\times \mathcal{B}(D^- \to K^+\pi^-\pi^-)$

Table B.5.: Total branching ratios corresponding to the generated MC cocktails. The first uncertainty comes from the measured uncertainties on the branching fractions. The second uncertainty on the semileptonic B cocktails comes from varying the branching fractions for the different D^{**} decay models, taken from [3]

B.2. Parameters of Fits to the D Mass in Categories

Parameter	Cat. 0	Cat. 1	Cat. 2	Cat. 3
$\mu_m[\mathrm{MeV}]$	1871.06 ± 0.04	1871.03 ± 0.03	1871.05 ± 0.02	1871.10 ± 0.01
$\sigma_{m,G,1}\,[\mathrm{MeV}]$	6.9 ± 0.2	7.3 ± 0.2	7.0 ± 0.2	6.51 ± 0.07
$\sigma_{m,G,2}\,[\mathrm{MeV}]$	12.7 ± 0.9	12.7 ± 0.9	11.4 ± 0.6	11.3 ± 0.2
$\sigma_{m,CB}\,[\mathrm{MeV}]$	4.2 ± 0.4	4.5 ± 0.2	4.7 ± 0.1	3.0 ± 0.1
$\alpha_{m,CB}$	2.1 ± 0.2	2.46 ± 0.09	2.35 ± 0.07	1.73 ± 0.07
$n_{m,CB}$	0.00001	0.00001	0.00001	0.00001
$f_{m,G}$	0.71 ± 0.04	0.69 ± 0.04	0.60 ± 0.04	0.66 ± 0.02
$f_{m,CB}$	0.20 ± 0.04	0.19 ± 0.05	0.25 ± 0.05	0.28 ± 0.02
$m_0[\mathrm{MeV}^{-1}]$	-0.00276 ± 0.0005	-0.00324 ± 0.0004	-0.00223 ± 0.0003	-0.00211 ± 0.0002
$\#nSig$	50525 ± 537	83662 ± 673	170472 ± 991	467393 ± 1667
$\#nBkg$	15781 ± 504	27494 ± 630	55527 ± 932	150530 ± 1569

Table B.6.: $B^0 \to D^- \mu^+ \nu_\mu X$ mode in 2011 data: fit parameters obtained from fits to the reconstructed D^- mass separately in four tagging categories.

Parameter	Cat. 0	Cat. 1	Cat. 2	Cat. 3
$\mu_m[\mathrm{MeV}]$	1871.09 ± 0.03	1871.12 ± 0.02	1871.12 ± 0.01	1871.126 ± 0.008
$\sigma_{m,G,1}\,[\mathrm{MeV}]$	7.2 ± 0.2	7.1 ± 0.1	6.77 ± 0.10	6.63 ± 0.05
$\sigma_{m,G,2}\,[\mathrm{MeV}]$	11.9 ± 0.6	12.2 ± 0.5	11.5 ± 0.3	11.4 ± 0.1
$\sigma_{m,CB}\,[\mathrm{MeV}]$	4.7 ± 0.1	4.09 ± 0.10	4.3 ± 0.1	3.71 ± 0.07
$\alpha_{m,CB}$	2.37 ± 0.07	2.25 ± 0.05	2.12 ± 0.05	1.94 ± 0.04
$n_{m,CB}$	0.00001	0.00001	0.00001	0.00001
$f_{m,G}$	0.63 ± 0.04	0.68 ± 0.03	0.63 ± 0.02	0.66 ± 0.01
$f_{m,CB}$	0.22 ± 0.05	0.21 ± 0.03	0.27 ± 0.02	0.27 ± 0.01
$m_0[\mathrm{MeV}^{-1}]$	-0.00247 ± 0.0003	-0.00272 ± 0.0002	-0.00238 ± 0.0002	-0.002228 ± 0.00010
$\#nSig$	133404 ± 872	226153 ± 1126	482720 ± 1669	1300091 ± 2798
$\#nBkg$	40607 ± 816	71260 ± 1055	155597 ± 1567	424489 ± 2637

Table B.7.: $B^0 \to D^- \mu^+ \nu_\mu X$ mode in 2012 data: fit parameters obtained from fits to the reconstructed D^- mass separately in four tagging categories.

Parameter	Cat. 0	Cat. 1	Cat. 2	Cat. 3
$\mu_m\,[\,\mathrm{MeV}\,]$	1866.36 ± 0.05	1866.25 ± 0.03	1866.28 ± 0.03	1866.26 ± 0.01
$\sigma_{m,G1}\,[\,\mathrm{MeV}\,]$	8.0 ± 0.2	8.18 ± 0.05	8.2 ± 0.1	8.21 ± 0.02
$\sigma_{m,G2}\,[\,\mathrm{MeV}\,]$	15.5 ± 0.8	16.1 ± 0.2	16.0 ± 0.5	16.25 ± 0.09
$f_{m,G}$	0.75 ± 0.02	0.764 ± 0.006	0.73 ± 0.01	0.717 ± 0.003
$\sigma_{m,CB}\,[\,\mathrm{MeV}\,]$	4.3 ± 0.2	4.6 ± 0.1	4.8 ± 0.1	5.09 ± 0.04
$\alpha_{m,CB}$	2.27 ± 0.09	2.35 ± 0.04	2.46 ± 0.05	2.54 ± 0.01
$n_{m,CB}$	0.00001	0.00001	0.00001	0.00001
$f_{m,CB}$	0.17 ± 0.03	0.153 ± 0.005	0.16 ± 0.02	0.152 ± 0.002
$m_0\,[\,\mathrm{MeV}^{-1}\,]$	-0.0064 ± 0.001	-0.00847 ± 0.0009	-0.00823 ± 0.0007	-0.00882 ± 0.0003
$\mu_{\delta m}\,[\,\mathrm{MeV}\,]$	145.33 ± 0.02	145.392 ± 0.005	145.36 ± 0.01	145.374 ± 0.002
$\sigma_{\delta m,G1}\,[\,\mathrm{MeV}\,]$	0.60 ± 0.05	0.46 ± 0.01	0.55 ± 0.07	0.554 ± 0.005
$\sigma_{\delta m,G2}\,[\,\mathrm{MeV}\,]$	1.07 ± 0.06	0.928 ± 0.010	0.96 ± 0.04	0.964 ± 0.006
$f_{\delta m,G}$	0.33 ± 0.06	0.21 ± 0.01	0.19 ± 0.05	0.216 ± 0.008
$\sigma_{\delta m,RJ}\,[\,\mathrm{MeV}\,]$	0.63 ± 0.04	0.67 ± 0.02	0.61 ± 0.03	0.638 ± 0.006
$\delta_{\delta m}$	1	1	1	1
$\gamma_{\delta m}$	-0.44 ± 0.1	-0.234 ± 0.02	-0.260 ± 0.04	-0.2078 ± 0.006
$f_{\delta m,RJ}$	0.31 ± 0.07	0.47 ± 0.02	0.36 ± 0.06	0.341 ± 0.007
$m_{\delta m}^{th}\,[\,\mathrm{MeV}\,]$	139	139	139	139
$a_{\delta m}$	4.0 ± 0.3	3.7 ± 0.9	3.7 ± 0.1	3.7 ± 0.3
$b_{\delta m}$	-5.0 ± 1	-5.0 ± 4	-5.00 ± 0.2	-5.0 ± 6
$c_{\delta m}$	1.5 ± 0.2	1.8 ± 0.4	1.79 ± 0.09	1.9 ± 0.2
$\#nSig$	27892 ± 202	45659 ± 333	93517 ± 372	256789 ± 807
$\#nBkg$	1444 ± 77	2468 ± 96	5979 ± 154	17005 ± 237
$\#nBkg_{D^0 fB}$	7267 ± 146	11292 ± 265	27066 ± 279	81426 ± 674

Table B.8.: $B^0 \rightarrow D^{*-}\mu^+\nu_\mu X$ mode in 2011 data: fit parameters obtained from fits to the reconstructed D^0 mass and the mass difference δm of the reconstructed D^* and D^0 masses separately in four tagging categories.

Parameter	Cat. 0	Cat. 1	Cat. 2	Cat. 3
$\mu_m[\mathrm{MeV}]$	1866.24 ± 0.03	1866.25 ± 0.03	1866.31 ± 0.02	1866.311 ± 0.009
$\sigma_{m,G1}[\mathrm{MeV}]$	8.4 ± 0.1	8.27 ± 0.10	8.19 ± 0.08	8.31 ± 0.01
$\sigma_{m,G2}[\mathrm{MeV}]$	17.2 ± 0.7	16.3 ± 0.4	16.2 ± 0.3	16.42 ± 0.07
$f_{m,G}$	0.75 ± 0.01	0.73 ± 0.01	0.716 ± 0.008	0.723 ± 0.002
$\sigma_{m,CB}[\mathrm{MeV}]$	5.1 ± 0.1	4.69 ± 0.09	5.14 ± 0.07	5.07 ± 0.03
$\alpha_{m,CB}$	2.54 ± 0.05	2.52 ± 0.04	2.53 ± 0.03	2.568 ± 0.009
$n_{m,CB}$	0.00001	0.00001	0.00001	0.00001
$f_{m,CB}$	0.13 ± 0.01	0.17 ± 0.01	0.167 ± 0.010	0.148 ± 0.001
$m_0[\mathrm{MeV}^{-1}]$	-0.01000 ± 0.0010	-0.00951 ± 0.0007	-0.00848 ± 0.0005	-0.00865 ± 0.0002
$\mu_{\delta m}[\mathrm{MeV}]$	145.37 ± 0.02	145.37 ± 0.02	145.385 ± 0.007	145.371 ± 0.002
$\sigma_{\delta m,G1}[\mathrm{MeV}]$	0.51 ± 0.08	0.50 ± 0.06	0.52 ± 0.03	0.548 ± 0.004
$\sigma_{\delta m,G2}[\mathrm{MeV}]$	0.92 ± 0.04	0.95 ± 0.03	0.96 ± 0.02	0.953 ± 0.004
$f_{\delta m,G}$	0.18 ± 0.06	0.21 ± 0.04	0.22 ± 0.03	0.216 ± 0.006
$\sigma_{\delta m,RJ}[\mathrm{MeV}]$	0.70 ± 0.08	0.68 ± 0.07	0.68 ± 0.03	0.649 ± 0.005
$\delta_{\delta m}$	1	1	1	1
$\gamma_{\delta m}$	-0.225 ± 0.03	-0.252 ± 0.03	-0.196 ± 0.02	-0.2242 ± 0.005
$f_{\delta m,RJ}$	0.38 ± 0.08	0.40 ± 0.06	0.38 ± 0.03	0.356 ± 0.006
$m_{\delta m}^{th}[\mathrm{MeV}]$	139	139	139	139
$a_{\delta m}$	0.6 ± 0.8	3.8 ± 0.2	3.71 ± 0.08	3.7 ± 0.2
$b_{\delta m}$	-0.08 ± 0.7	-5.0 ± 1	-4.999 ± 0.06	-4.95 ± 0.6
$c_{\delta m}$	1.4 ± 0.1	1.8 ± 0.1	1.79 ± 0.06	1.74 ± 0.10
$\#nSig$	62958 ± 311	104965 ± 389	223564 ± 583	604909 ± 1563
$\#nBkg$	3427 ± 118	5785 ± 153	14969 ± 246	43312 ± 430
$\#nBkg_{D^0 fB}$	17932 ± 232	27536 ± 282	68982 ± 450	205630 ± 1312

Table B.9.: $B^0 \rightarrow D^{*-}\mu^+\nu_\mu X$ mode in 2012 data: fit parameters obtained from fits to the reconstructed D^0 mass and the mass difference δm of the reconstructed D^* and D^0 masses separately in four tagging categories.

B.3. Parameters of Fits to the Flight Distance Resolution

$B^0 \rightarrow D^- \mu^+ \nu_\mu X$	2011	2012
μ_t [ps]	0.000346 ± 0.000217	$1.7 \cdot 10^{-5} \pm 0.000161$
$\sigma_{t,1}$ [ps]	0.033 ± 0.001	0.035 ± 0.001
$\sigma_{t,2}$ [ps]	0.067 ± 0.002	0.071 ± 0.001
$\sigma_{t,3}$ [ps]	0.181 ± 0.008	0.204 ± 0.006
$f_{t,1}$	0.67 ± 0.024	0.685 ± 0.019
$f_{t,2}$	0.96 ± 0.004	0.956 ± 0.003

| $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$	2011	2012
μ_t [ps]	0.000957 ± 0.000108	0.001288 ± 0.000143
$\sigma_{t,1}$ [ps]	0.035 ± 0.0	0.034 ± 0.001
$\sigma_{t,2}$ [ps]	0.074 ± 0.001	0.072 ± 0.001
$\sigma_{t,3}$ [ps]	0.206 ± 0.003	0.189 ± 0.004
$f_{t,1}$	0.646 ± 0.011	0.658 ± 0.014
$f_{t,2}$	0.937 ± 0.002	0.94 ± 0.003

Table B.10.: Parameters of the flight distance resolution fit for tagged events in $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ mode (upper row) and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ mode (lower row), 2011 MC (left) and 2012 MC (right).

$B^+ \to D^- \mu^+ \nu_\mu X$	2011	2012
μ_t [ps]	0.001455 ± 0.000507	0.001266 ± 0.000313
$\sigma_{t,1}$ [ps]	0.038 ± 0.002	0.038 ± 0.001
$\sigma_{t,2}$ [ps]	0.076 ± 0.004	0.077 ± 0.003
$\sigma_{t,3}$ [ps]	0.279 ± 0.062	0.189 ± 0.017
$f_{t,1}$	0.721 ± 0.046	0.695 ± 0.036
$f_{t,2}$	0.979 ± 0.006	0.954 ± 0.01

$B^+ \to D^{*-} \mu^+ \nu_\mu X$	2011	2012
μ_t [ps]	0.003549 ± 0.00012	0.001263 ± 0.000462
$\sigma_{t,1}$ [ps]	0.043 ± 0.001	0.04 ± 0.002
$\sigma_{t,2}$ [ps]	0.104 ± 0.003	0.083 ± 0.004
$\sigma_{t,3}$ [ps]	0.27 ± 0.058	0.24 ± 0.036
$f_{t,1}$	0.315 ± 0.015	0.688 ± 0.042
$f_{t,2}$	0.092 ± 0.005	0.968 ± 0.009

Table B.11.: Parameters of the flight distance resolution fit for all events in $B^+ \to D^- \mu^+ \nu_\mu X$ mode (upper row) and $B^+ \to D^{*-} \mu^+ \nu_\mu X$ mode (lower row), 2011 B^+ MC (left) and 2012 B^+ MC (right).
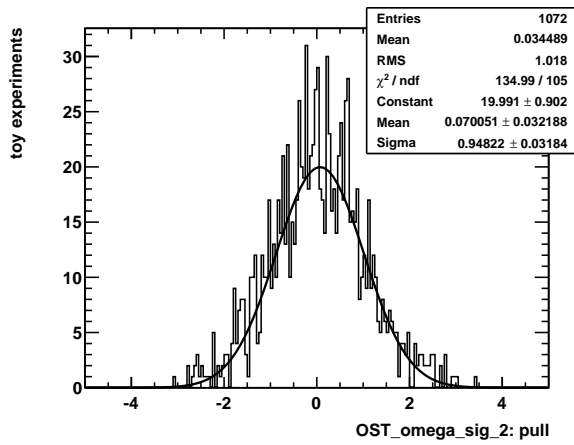
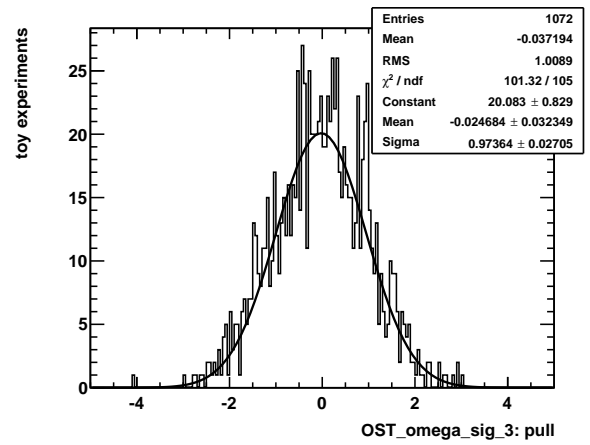# B.4. Additional Toy Experiment Pull Distributions
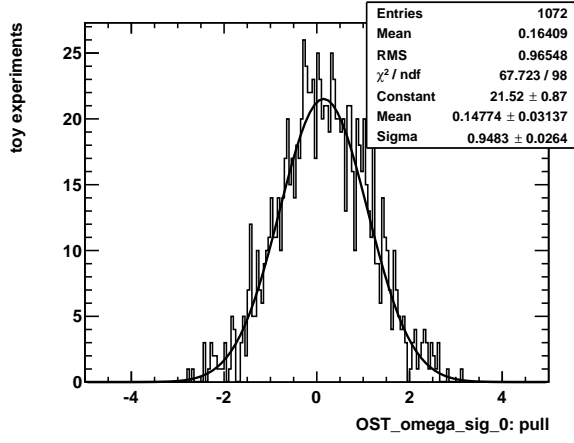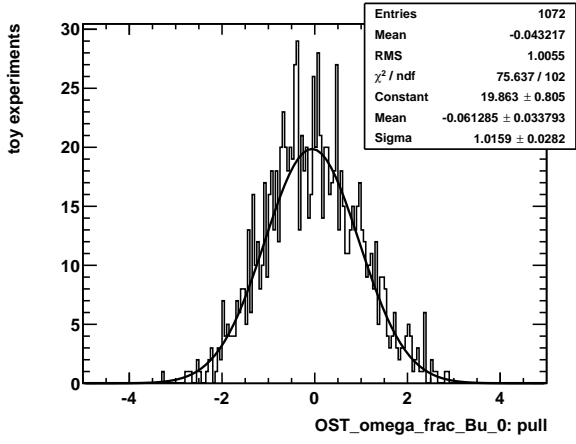

(a) Category 0
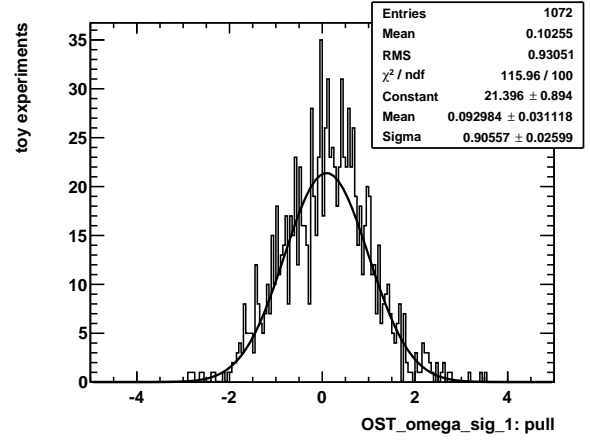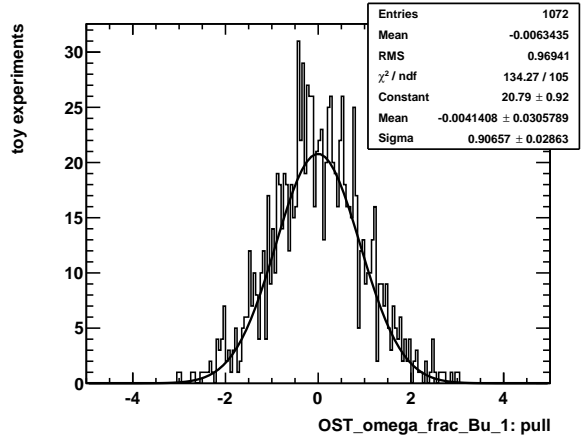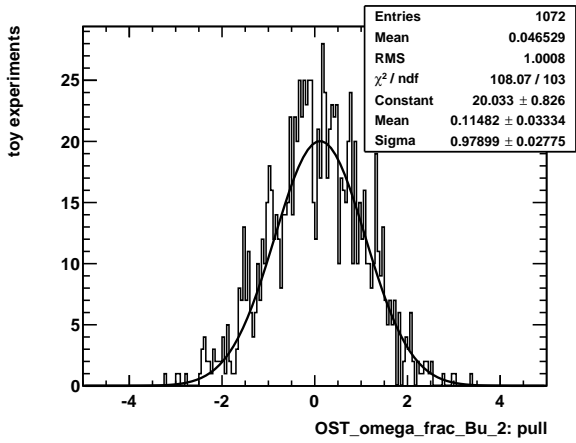

(b) Category 1

(c) Category 2

(d) Category 3

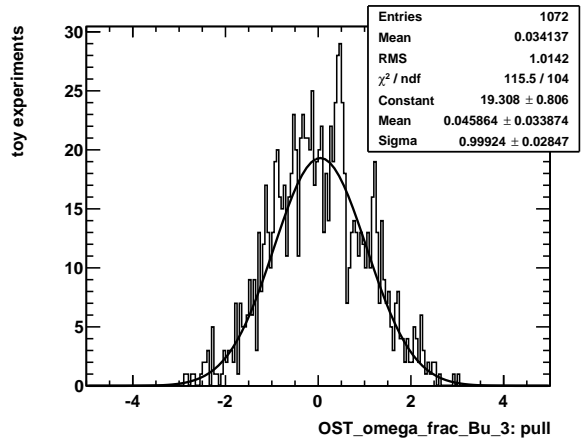Figure B.1.: Pull distributions of the B^0 ω parameters in the four tagging categories.

(a) Category 0

(b) Category 1

(c) Category 2

(d) Category 3

Figure B.2.: Pull distributions of the B^+ ω parameters in the four tagging categories.

Acknowledgements (Danksagungen)

Ich möchte mich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben. Dies sind insbesondere:

Die Heidelberger LHCb Arbeitsgruppe,

meine Familie,

mein liebster Eric

und alle meine Freunde, die mich von der Arbeit ablenken konnten.

Statement of Originality (Erklärung):

I certify that this thesis, and the research to which it refers, are the product of my own work. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Ich versichere, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, 2015-03-30

....
(signature)