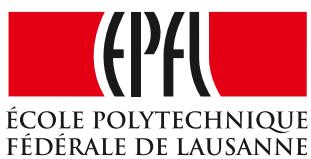


# **Measurement of time-dependent $CP$ violation in $B^0 \rightarrow D^\mp \pi^\mp$ decays and optimisation of Flavour Tagging algorithms at LHCb**

Thèse n. XXXX  
présentée le XX xxxx 2018  
à la Faculté des Sciences de Base  
Laboratoire de physique des hautes énergies  
Programme doctoral en Physique  
École Polytechnique Fédérale de Lausanne



pour l'obtention du grade de Docteur ès Sciences  
par

Vincenzo Battista

acceptée sur proposition du jury:  
Prof. Olivier Schneider, directeur de thèse

Lausanne, EPFL, 2018



*The ability to quote is a serviceable substitute for wit*  
W. S. Maugham



# Abstract

Keywords:  $B$  physics, CKM angle  $\gamma$ ,  $CP$  violation, flavour tagging, mixing, LHCb, LHC.



# Résumé

Mots clefs : Physique du  $B$ , angle CKM  $\gamma$ , violation de  $CP$ , étiquetage du saveur, oscillations, LHCb, LHC.



# Sommario

Parole chiave: fisica del mesone  $B$ , angolo CKM  $\gamma$ , violazione di  $CP$ , etichettatura del sapore, oscillazioni, LHCb, LHC.



# Acknowledgements

*Lausanne, XX XXXX 2018*

V. B.



# Contents

<b>Abstract (English/Français/Italiano)</b>	<b>v</b>
<b>Acknowledgements</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Standard Model of particle physics . . . . .	2
1.2 The Cabibbo-Kobayashi-Maskawa matrix . . . . .	4
1.3 Physics of the neutral $B$ mesons . . . . .	7
1.3.1 Oscillation of neutral mesons . . . . .	7
1.3.2 Decay of neutral mesons . . . . .	9
1.3.3 $CP$ violation in neutral meson systems . . . . .	10
1.4 $B^0 \rightarrow D^\pm \pi^\mp$ analysis motivation and phenomenology . . . . .	14
1.5 Personal contribution . . . . .	17
<b>2 Experimental setup</b>	<b>19</b>
2.1 The Large Hadron Collider . . . . .	20
2.2 The LHCb experiment . . . . .	22
2.2.1 Tracking system . . . . .	22
2.2.2 Particle identification (PID) . . . . .	27
2.2.3 Trigger system . . . . .	29
2.2.4 Event reconstruction and simulation software . . . . .	32
2.2.5 Data collected by LHCb . . . . .	33
<b>3 Flavour tagging</b>	<b>35</b>
3.1 Flavour tagging algorithms . . . . .	36
3.2 Flavour tagging strategy for the $B^0 \rightarrow D^\pm \pi^\mp$ time-dependent analysis . .	42
3.2.1 Calibration of the opposite-side tagger combination . . . . .	44
3.2.2 Calibration of the same-side tagger combination . . . . .	48
3.3 Optimisation of the opposite-side electron tagger . . . . .	49
3.3.1 Sample definition . . . . .	51
3.3.2 Preselection optimisation . . . . .	52
3.3.3 BDT classifier implementation . . . . .	52
3.3.4 Performance evaluation . . . . .	59

## Contents

---

<b>4 Selection of <math>B^0 \rightarrow D^\mp\pi^\pm</math> decays</b>	<b>69</b>
4.1 Data sample and preselection . . . . .	70
4.1.1 Stripping and trigger requirements . . . . .	70
4.1.2 Preselection and sample definitions . . . . .	72
4.1.3 Vetoing against physics backgrounds . . . . .	72
4.1.4 Wrongly associated primary vertices . . . . .	75
4.1.5 Development of an MVA classifier . . . . .	76
4.1.6 BDT selection optimisation . . . . .	77
4.1.7 Multiple candidates . . . . .	79
4.1.8 Selection performance . . . . .	79
4.2 Simulation and expected sample composition . . . . .	82
4.2.1 PIDK correction . . . . .	82
4.2.2 Surviving physics backgrounds . . . . .	83
4.3 Fits to the $B^0$ invariant mass . . . . .	86
4.3.1 Probability density functions . . . . .	87
4.3.2 Fit to data . . . . .	88
4.3.3 <i>sWeight</i> calculation . . . . .	90
4.3.4 Fits of subsamples . . . . .	93
<b>5 Measurement of <math>CP</math> violation in <math>B^0 \rightarrow D^\mp\pi^\pm</math> decays</b>	<b>99</b>
5.1 Decay-time resolution . . . . .	100
5.1.1 Companion track momentum reweighting . . . . .	100
5.1.2 Resolution determination from decay-time error parameterisation .	101
5.2 Time-dependent selection efficiency . . . . .	103
5.3 Decay-time fit to data . . . . .	104
5.4 Fit validation . . . . .	109
5.5 Systematics . . . . .	118
5.5.1 Systematic uncertainties from Gaussian constraints . . . . .	118
5.5.2 Systematic uncertainties estimated with pseudoexperiments . . . . .	119
5.5.3 Systematics related to the background subtraction . . . . .	123
<b>6 Summary and interpretation of the results</b>	<b>127</b>
6.1 Interpretation of the results . . . . .	128
6.2 Summary and perspectives . . . . .	129
<b>Appendices</b>	<b>135</b>
A Opposite-side tagging studies . . . . .	136
A.1 Mass fit of $B^\pm \rightarrow D^0\pi^\pm$ . . . . .	136
A.2 Reweighting of $B^\pm \rightarrow D^0\pi^\pm$ to $B^0 \rightarrow D^\mp\pi^\pm$ . . . . .	137
A.3 GOF tests for OS calibration on $B^\pm \rightarrow D^0\pi^\pm$ data . . . . .	139
B Opposite-side electron optimisation . . . . .	143
C $B^0 \rightarrow D^\mp\pi^-$ selection studies . . . . .	148
C.1 BDT input features . . . . .	148

C.2	Multiple candidates . . . . .	148
D	Particle identification plots . . . . .	152
E	Correlation between $B^0$ mass and decay time . . . . .	155
F	PDF definitions . . . . .	156
G	Signal PDF for the decay-time fit . . . . .	158
H	Correlation matrix of the $B^0 \rightarrow D^\mp\pi^\pm$ decay-time fit . . . . .	160
I	Decay-time fits of $B^0 \rightarrow D^\mp\pi^\pm$ data subsamples . . . . .	161
J	Decay-time fit validation with bootstrapping . . . . .	168
K	Test of the decay-time fit via a toy tagger . . . . .	172
L	Correlation between systematic uncertainties . . . . .	174
L.1	Correlation of $\Delta m$ systematics . . . . .	174
L.2	Correlation of systematics due to fit biases . . . . .	174
L.3	Correlation of systematics due to background subtraction . . . . .	174
M	Inputs for the $B^0 \rightarrow D^\pm\pi^\mp$ simulation . . . . .	176



# 1 Introduction

2

### 1.1 The Standard Model of particle physics

The *Standard Model* (SM) of particle physics [1, 2, 3] is a *non-abelian, Yang-Mills quantum field theory* based on the  $SU(3) \times SU(2) \times U(1)$  gauge symmetry group. This model provides a coherent, unified and experimentally-established picture of electromagnetic, weak and strong interactions, as well as a description of the known elementary particles (quarks, leptons, gauge bosons and Higgs boson, Fig. 1.1).

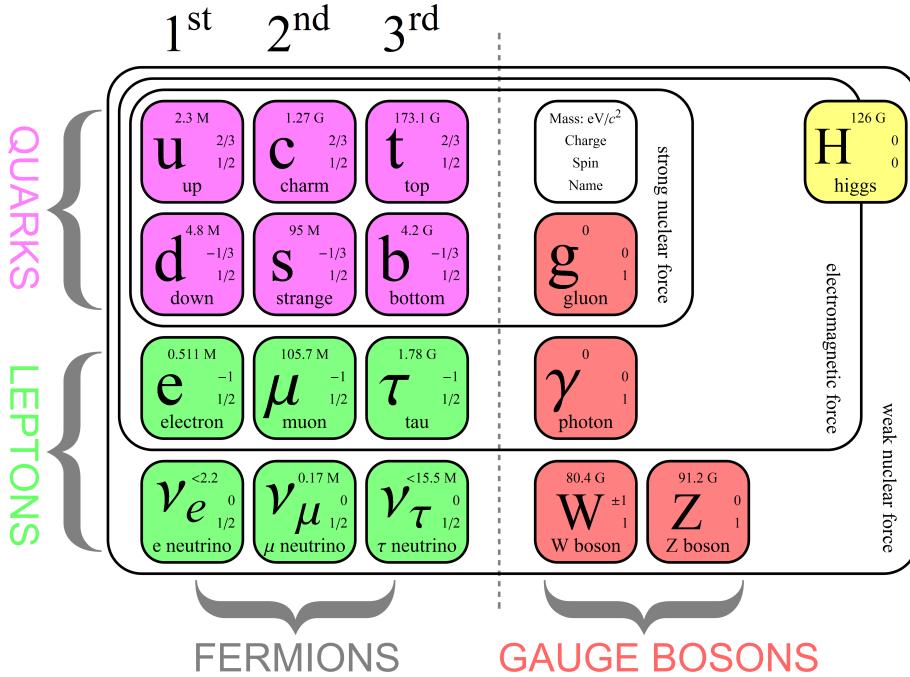


Figure 1.1 – Elementary particles described by the SM [4].

- All particles are either *fermions* or *bosons*. The fermions (leptons, quarks) have half-integer spin and follow the Fermi-Dirac statistics [5, 6], whereas bosons (gauge bosons, Higgs boson) have integer spin and follow the Bose-Einstein statistics [7].
- Leptons (spin- $\frac{1}{2}$ ) include three charged<sup>1</sup>, massive particles (electron  $e^-$ , muon  $\mu^-$  and tau  $\tau^-$ ), which interact via the electromagnetic and weak interactions, and three neutral, (nearly) massless particles, called *neutrinos* ( $\nu_e$ ,  $\nu_\mu$  and  $\nu_\tau$ ), which only experience weak interactions.
- Six different flavours of quarks (spin- $\frac{1}{2}$ ) exist: the *up-type* quarks up ( $u$ ), charm ( $c$ ) and top/truth ( $t$ ), having charge  $+\frac{2}{3}$ , and the *down-type* quarks down ( $d$ ), strange ( $s$ ), and bottom/beauty ( $b$ ), which have charge  $-\frac{1}{3}$ . They can interact via electromagnetic, weak and strong interactions, and they are all massive.

<sup>1</sup>Electric charge is always quoted in units of the fundamental charge, defined as minus the charge of the electron.

## 1.1. The Standard Model of particle physics

---

- 20 The fundamental interactions are *mediated* by gauge bosons (spin-1). The photon ( $\gamma$ )  
21 is responsible for the electromagnetic interaction, whereas the  $Z^0$  and  $W^\pm$  bosons are  
22 the mediators for the weak interaction. These two forces are considered to be different  
23 manifestations of a single *electroweak* interaction, which is responsible for all electric and  
24 magnetic phenomena as well as some radioactive decays. The strong interaction among  
25 quarks is mediated by the gluons  $g$ . Photon and gluons are massless, whereas the weak  
26 force gauge bosons have a non-zero mass.
- 27 Each particle has an *antiparticle* partner, which has the same mass as the corresponding  
28 particle, but opposite quantum numbers (electric charge, lepton numbers, etc...).
- 29 Quarks do not exist in a free state: they can only be bound inside *hadrons* via the  
30 *confinement* mechanism, a feature of the strong interaction. A hadron can be composed  
31 by a quark-antiquark pair (*meson*), or by three quarks or antiquarks (*baryons*). Examples  
32 of mesons include the  $B^0$  ( $\bar{b}d$ ) and  $D^+$  ( $c\bar{d}$ ) mesons, whereas the proton ( $uud$ ) and the  
33 neutron ( $udd$ ) are examples of baryons. Recently more complex states (tetraquarks [8],  
34 pentaquarks [9]) have been evidenced.
- 35 The non-zero mass of leptons, quarks and weak force gauge bosons would require a  
36 gauge symmetry breaking term in the SM Lagrangian density. The *Brout-Englert-Higgs*  
37 *mechanism* [10, 11, 12] introduces a scalar (spin-0) field, called Higgs field, and a potential  
38 that allows the Higgs field to have a non-zero vacuum expectation value. This implies  
39 that the gauge symmetry is broken *dynamically*, and that the masses of the particles arise  
40 from the resulting interaction with the Higgs field. The quantum of the Higgs field is  
41 known as Higgs boson, the last SM particle discovered experimentally [13, 14].
- 42 The fourth fundamental interaction, the gravitational force, is described by another field  
43 theory, the General Relativity (GR), currently not unified with the SM.
- 44 Any experimental signature that is not described by the SM would be a hint for *new*  
45 *physics* (NP). Although the SM is known to be an incomplete theory because of different  
46 unsolved problems, such as dark matter, *naturalness*, matter-antimatter asymmetry, lack  
47 of SM-GR unification, etc..., no evidence for NP has been found so far.

## 48 1.2 The Cabibbo-Kobayashi-Maskawa matrix

The Lagrangian density describing the weak interactions between quarks and  $W^\pm$  (*charged current interaction*) can be written as

$$\mathcal{L}_{cc} = \frac{g}{\sqrt{2}} \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix} V_{\text{CKM}} \gamma^\mu \frac{(1 - \gamma^5)}{2} \begin{pmatrix} d \\ s \\ b \end{pmatrix} W_\mu^+ + h.c., \quad (1.1)$$

where  $g$  is a coupling constant,  $\gamma^\mu$  are Dirac matrices and  $V_{\text{CKM}}$ , known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix [15, 16], couples the *flavour* eigenstates  $d$ ,  $s$  and  $b$  to the *mass* (or *physical*) eigenstates  $d'$ ,  $s'$  and  $b'$ :

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (1.2)$$

- 49 The CKM matrix is unitary ( $V_{\text{CKM}}^\dagger V_{\text{CKM}} \equiv 1$ ), so it can be written in terms of four  
 50 independent parameters, namely three angles and a *complex phase*  $\delta$ . The latter is the  
 51 source of all *CP-violating* phenomena in the SM, i.e. asymmetries between particles and  
 52 anti-particles; in fact, the *complexity* of  $V_{\text{CKM}}$  implies that the SM Lagrangian density is  
 53 non *CP*-invariant, in agreement with the experimentally observed *CP* violation.

A first, standard parameterisation of the CKM matrix [17] gives

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \quad (1.3)$$

- 54 where  $s_{ij} = \sin \theta_{ij}$  and  $c_{ij} = \cos \theta_{ij}$ .

Another, useful parameterisation is given by *Wolfenstein* [18] and points out the order of magnitude of each matrix element. By defining the quantities  $\lambda$ ,  $A$ ,  $\rho$  and  $\eta$  with

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \quad s_{23} = A\lambda^2 = A \left| \frac{V_{cb}}{V_{us}} \right|, \quad (1.4)$$

$$s_{13}e^{i\delta} = V_{ub}^* = A\lambda^3(\rho + i\eta), \quad (1.5)$$

the  $V_{\text{CKM}}$  matrix can be rewritten as a series expansion in powers of  $\lambda$ , given that  $\lambda$  is a small number:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \quad (1.6)$$

55 From Eq. 1.6, one can see that quark transitions within the same family (e.g.  $u \rightarrow d$ )  
 56 are more probable, whereas transitions between different families (e.g.  $b \rightarrow c$ ) are more  
 57 suppressed.  $CP$  violation is a consequence of  $\eta \neq 0$  and  $\eta \neq \pi$ .

The unitarity condition  $V_{\text{CKM}}^\dagger V_{\text{CKM}} = 1$  can be rewritten in terms of six scalar equations. Two of them are particularly relevant for the  $b$ -hadron phenomenology:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \quad (1.7)$$

$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0. \quad (1.8)$$

These two equations can be graphically represented as *triangles* in the  $(\bar{\rho}, \bar{\eta})$  complex plane, where  $\bar{\rho} = \rho(1 - \lambda^2/2 + \dots)$  and  $\bar{\eta} = \eta(1 - \lambda^2/2 + \dots)$ . Having defined the following angles,

$$\alpha = \phi_2 = \arg \left[ -\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*} \right], \quad \beta = \phi_1 = \arg \left[ -\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*} \right], \quad (1.9)$$

$$\gamma = \phi_3 = \arg \left[ -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right], \quad \beta_s = \chi = \arg \left[ -\frac{V_{cb}V_{cs}^*}{V_{tb}V_{ts}^*} \right], \quad (1.10)$$

58 the triangles given by Eqs. 1.7 and 1.8 can be depicted as shown in Fig. 1.2. The first  
 59 triangle, defined by Eq. 1.7, is known as the *Unitarity Triangle* (UT) and its elements  
 60 can be measured from analyses of  $B^0$ ,  $B_s^0$  and  $B^\pm$  decays. The other triangle (Eq. 1.8)  
 61 can be studied from decays of  $B_s^0$  mesons.

The scale of the amount of  $CP$  violation in the SM is given by the Jarlskog invariant  $J$  [19], defined such as

$$\Im[V_{ij}V_{kl}V_{il}^*V_{kj}^*] = J \sum_{m,n} \varepsilon_{ikm}\varepsilon_{jln}, \quad (1.11)$$

62 where  $\varepsilon$  denotes the fully-antisymmetric tensor. The measured value of  $J$  is too small by  
 63 several orders of magnitude to explain the observed matter-antimatter asymmetry in the  
 64 universe, according to the baryogenesis model [20]. So, new sources of  $CP$  violation not  
 65 foreseen by the SM have to exist, and thus measuring the UT with the highest possible  
 66 precision is crucial to constrain these new physics scenarios.

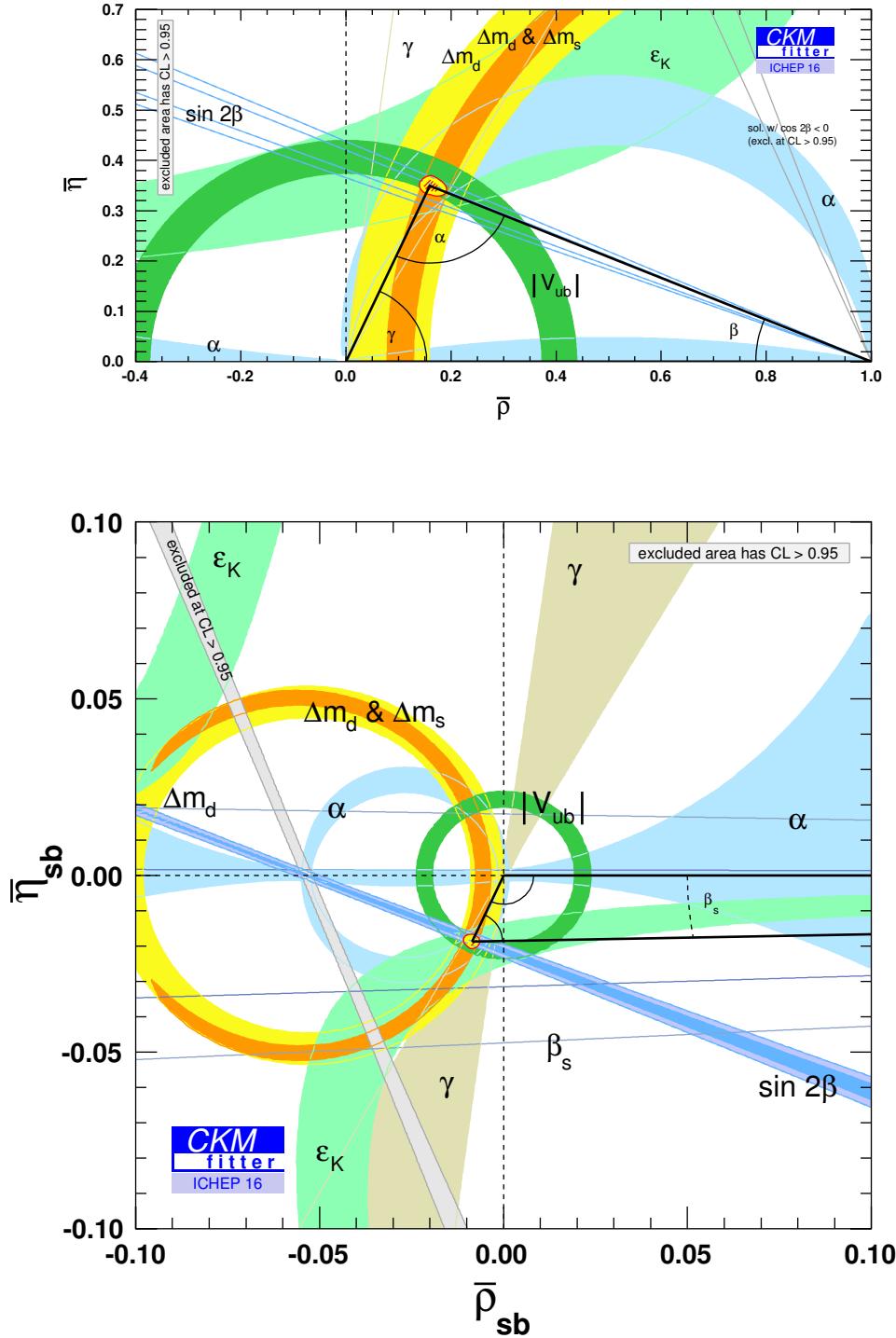


Figure 1.2 – Graphical representation of two of the six unitarity conditions of the CKM matrix, superimposed with the current experimental constraints [21].

<sup>67</sup> **1.3 Physics of the neutral  $B$  mesons**

<sup>68</sup> The theory of neutral  $B$  meson oscillation, decays and  $CP$  violation presented here is  
<sup>69</sup> derived from Refs. [22, 23].

<sup>70</sup> **1.3.1 Oscillation of neutral mesons**

Neutral  $B$  meson states are characterised by the following quark content:

$$|B^0\rangle = |\bar{d}\bar{b}\rangle, \quad |\bar{B}^0\rangle = |\bar{d}\bar{b}\rangle, \quad (1.12)$$

$$|B_s^0\rangle = |\bar{s}\bar{b}\rangle, \quad |\bar{B}_s^0\rangle = |\bar{s}\bar{b}\rangle. \quad (1.13)$$

All neutral mesons will be denoted as  $P^0$  or  $\bar{P}^0$  hereafter. The  $\bar{P}^0$  state is obtained from  $P^0$  via the  $CP$  operator up to an arbitrary phase factor  $e^{i\phi_{CP}}$ . Since charged-currents do not conserve flavour quantum numbers (e.g. strangeness, beauty etc.), a neutral meson can transform itself into its own anti-meson, and viceversa. So, the time evolution of a neutral  $B$  meson system can be generally written as

$$|\Psi(t)\rangle = a(t)|P^0\rangle + b(t)|\bar{P}^0\rangle + \sum_i c_i(t)|f_i\rangle, \quad (1.14)$$

<sup>71</sup> where  $|f_i\rangle$  are all the possible final states with  $c_i(0) = 0$  as initial condition.

Since the typical timescale of weak interactions is much longer than the strong interaction timescale, we can neglect all weak interactions among final states (*Weisskopf-Wigner approximation*). So, we can write the Schrödinger equation for  $|\Psi(t)\rangle$  in terms of an effective, non-hermitian hamiltonian  $\mathcal{H}$ :

$$i\partial_t \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \mathcal{H} \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} a(t) \\ b(t) \end{pmatrix}. \quad (1.15)$$

The  $\mathcal{H}$  matrix can be rewritten as the sum of two hermitian matrices  $M$  and  $\Gamma$ :

$$\mathcal{H} = M - \frac{i}{2}\Gamma = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}. \quad (1.16)$$

Assuming  $CPT$  invariance ( $H_{11} = H_{22} = H_0$ ,  $M_{11} = M_{22} = M_0$ ,  $\Gamma_{11} = \Gamma_{22} = \Gamma_0$ ), the

## Chapter 1. Introduction

---

eigenvalues of  $\mathcal{H}$  are

$$\lambda_L = m_L - \frac{i}{2}\Gamma_L = H_0 + \sqrt{H_{12}H_{21}} = H_0 + \sqrt{\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\left(M_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}, \quad (1.17)$$

$$\lambda_H = m_H - \frac{i}{2}\Gamma_H = H_0 - \sqrt{H_{12}H_{21}} = H_0 - \sqrt{\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\left(M_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}, \quad (1.18)$$

where  $L$  (“light”) and  $H$  (“heavy”) refer to the value of the mass for each eigenstate. The corresponding eigenvectors are

$$|P_L\rangle = p|P^0\rangle + q|\bar{P}^0\rangle, \quad |P_H\rangle = p|P^0\rangle - q|\bar{P}^0\rangle, \quad (1.19)$$

where  $p$  and  $q$  satisfy  $|p|^2 + |q|^2 = 1$  and are given by

$$\frac{q}{p} = \sqrt{\frac{H_{21}}{H_{12}}} = \sqrt{\frac{M_{21} - \frac{i}{2}\Gamma_{21}}{M_{12} - \frac{i}{2}\Gamma_{12}}} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{21}^* - \frac{i}{2}\Gamma_{21}^*}}. \quad (1.20)$$

The relative phase  $\phi_M$  between  $M_{12}$  and  $\Gamma_{12}$  is an observable quantity describing indirect  $CP$  violation (Sec. 1.3.3):

$$M_{12} = M_{12}^* e^{i\phi_{CP}}, \quad \Gamma_{12} = \Gamma_{12}^* e^{i\phi_{CP}} e^{i\phi_M}. \quad (1.21)$$

For the neutral  $B$  meson system, the ratio  $|\Gamma_{12}/M_{12}|$  is expected to be small in the SM; as a consequence, it can be shown that

$$\frac{q}{p} = -e^{-i\phi_M} \left[ 1 - \frac{1}{2} \left| \frac{\Gamma_{12}}{M_{12}} \right| \sin \phi_M + \mathcal{O} \left( \left| \frac{\Gamma_{12}}{M_{12}} \right|^2 \right) \right], \quad (1.22)$$

<sup>72</sup> which implies  $|q/p| \sim 1$ .

The difference and the average of the masses and widths of the two mass eigenstates are defined as

$$\Delta m = m_H - m_L = \Re(\lambda_H - \lambda_L), \quad m = \frac{m_L + m_H}{2} = \frac{\Re(\lambda_H + \lambda_L)}{2}, \quad (1.23)$$

$$\Delta\Gamma = \Gamma_L - \Gamma_H = -2\Im(\lambda_L - \lambda_H), \quad \Gamma = \frac{\Gamma_L + \Gamma_H}{2} = -\frac{\Im(\lambda_H + \lambda_L)}{4}. \quad (1.24)$$

<sup>73</sup> The sign convention for  $\Delta\Gamma$  is chosen to have a positive experimental value for the  $B_s^0$  system (for  $B^0$ , experiments give a result compatible with zero, in agreement with the SM).

The time evolution of the states  $|P^0(t)\rangle$  and  $|\bar{P}^0(t)\rangle$  when they are initially produced as

$|P^0(0)\rangle$  and  $|\bar{P}^0(0)\rangle$  can be obtained from the effective hamiltonian:

$$|P^0(t)\rangle = g_+(t)|P^0(t)\rangle + \frac{q}{p}g_-(t)|\bar{P}^0(t)\rangle, \quad (1.25)$$

$$|\bar{P}^0(t)\rangle = g_+(t)|\bar{P}^0(t)\rangle + \frac{p}{q}g_-(t)|P^0(t)\rangle. \quad (1.26)$$

The functions  $g_{\pm}(t)$  are built in terms of the eigenvalues:

$$g_{\pm} = \frac{1}{2} \left( e^{-im_H t - \frac{1}{2}\Gamma_H t} \pm e^{-im_L t - \frac{1}{2}\Gamma_L t} \right). \quad (1.27)$$

The probabilities that a state initially produced as  $P^0$  or  $\bar{P}^0$  becomes a  $P^0$  or  $\bar{P}^0$  at time  $t$

$$|\langle P^0(0)|P^0(t)\rangle|^2 = |g_+(t)|^2 = \frac{e^{-\Gamma t}}{2} \left( \cosh \frac{\Delta\Gamma t}{2} + \cos \Delta m t \right), \quad (1.28)$$

$$|\langle \bar{P}^0(0)|P^0(t)\rangle|^2 = \left| \frac{q}{p} g_-(t) \right|^2 = \left| \frac{q}{p} \right|^2 \frac{e^{-\Gamma t}}{2} \left( \cosh \frac{\Delta\Gamma t}{2} - \cos \Delta m t \right), \quad (1.29)$$

$$|\langle P^0(0)|\bar{P}^0(t)\rangle|^2 = \left| \frac{p}{q} g_-(t) \right|^2 = \left| \frac{p}{q} \right|^2 \frac{e^{-\Gamma t}}{2} \left( \cosh \frac{\Delta\Gamma t}{2} - \cos \Delta m t \right), \quad (1.30)$$

$$|\langle \bar{P}^0(0)|\bar{P}^0(t)\rangle|^2 = |g_+(t)|^2 = \frac{e^{-\Gamma t}}{2} \left( \cosh \frac{\Delta\Gamma t}{2} + \cos \Delta m t \right). \quad (1.31)$$

76 The equations above describe the *oscillation* of the  $B^0$  and  $B_s^0$  mesons.

### 77 1.3.2 Decay of neutral mesons

The amplitude for the decay of a neutral meson into a final state  $f$  can be obtained from the effective hamiltonian  $\mathcal{H}$ :

$$A_f = \langle f | \mathcal{H} | P^0 \rangle, \quad \bar{A}_f = \langle \bar{f} | \mathcal{H} | \bar{P}^0 \rangle, \quad (1.32)$$

$$A_{\bar{f}} = \langle \bar{f} | \mathcal{H} | P^0 \rangle, \quad \bar{A}_{\bar{f}} = \langle f | \mathcal{H} | \bar{P}^0 \rangle. \quad (1.33)$$

After defining the parameters

$$\lambda_f = \frac{q}{p} \frac{\bar{A}_f}{A_f}, \quad \frac{1}{\bar{\lambda}_{\bar{f}}} = \lambda_{\bar{f}} = \frac{q}{p} \frac{\bar{A}_{\bar{f}}}{A_{\bar{f}}}, \quad (1.34)$$

it is possible to write the *decay rates* for neutral mesons decaying into  $f$  or  $\bar{f}$ :

$$\begin{aligned} \frac{d\Gamma(P^0 \rightarrow f)}{dt}(t) &= N_f |A_f|^2 \frac{1 + |\lambda_f|^2}{2} e^{-\Gamma t} \\ &\quad \left[ \cosh \left( \frac{\Delta\Gamma t}{2} \right) + D_f \sinh \left( \frac{\Delta\Gamma t}{2} \right) + C_f \cos(\Delta m t) - S_f \sin(\Delta m t) \right], \end{aligned}$$

(1.35)

$$\begin{aligned} \frac{d\Gamma(\bar{P}^0 \rightarrow f)}{dt}(t) = N_f |A_f|^2 \left| \frac{p}{q} \right|^2 \frac{1 + |\lambda_f|^2}{2} e^{-\Gamma t} \\ \left[ \cosh \left( \frac{\Delta\Gamma t}{2} \right) + D_f \sinh \left( \frac{\Delta\Gamma t}{2} \right) - C_f \cos(\Delta m t) + S_f \sin(\Delta m t) \right], \end{aligned} \quad (1.36)$$

$$\begin{aligned} \frac{d\Gamma(P^0 \rightarrow \bar{f})}{dt}(t) = N_f |\bar{A}_{\bar{f}}|^2 \left| \frac{q}{p} \right|^2 \frac{1 + |\bar{\lambda}_{\bar{f}}|^2}{2} e^{-\Gamma t} \\ \left[ \cosh \left( \frac{\Delta\Gamma t}{2} \right) + D_{\bar{f}} \sinh \left( \frac{\Delta\Gamma t}{2} \right) + C_{\bar{f}} \cos(\Delta m t) - S_{\bar{f}} \sin(\Delta m t) \right], \end{aligned} \quad (1.37)$$

$$\begin{aligned} \frac{d\Gamma(\bar{P}^0 \rightarrow \bar{f})}{dt}(t) = N_f |\bar{A}_{\bar{f}}|^2 \frac{1 + |\bar{\lambda}_{\bar{f}}|^2}{2} e^{-\Gamma t} \\ \left[ \cosh \left( \frac{\Delta\Gamma t}{2} \right) + D_{\bar{f}} \sinh \left( \frac{\Delta\Gamma t}{2} \right) - C_{\bar{f}} \cos(\Delta m t) + S_{\bar{f}} \sin(\Delta m t) \right], \end{aligned} \quad (1.38)$$

where  $N_f$  is a time-independent normalisation factor and

$$D_f = -\frac{2\Re\lambda_f}{1 + |\lambda_f|^2}, \quad C_f = \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2}, \quad S_f = \frac{2\Im\lambda_f}{1 + |\lambda_f|^2}, \quad (1.39)$$

$$D_{\bar{f}} = -\frac{2\Re\bar{\lambda}_{\bar{f}}}{1 + |\bar{\lambda}_{\bar{f}}|^2}, \quad C_{\bar{f}} = -\frac{1 - |\bar{\lambda}_{\bar{f}}|^2}{1 + |\bar{\lambda}_{\bar{f}}|^2}, \quad S_{\bar{f}} = -\frac{2\Im\bar{\lambda}_{\bar{f}}}{1 + |\bar{\lambda}_{\bar{f}}|^2}, \quad (1.40)$$

<sup>78</sup> are known as *CP coefficients*.

### <sup>79</sup> 1.3.3 CP violation in neutral meson systems

<sup>80</sup> Three types of *CP* violation can occur. They are briefly sketched in Fig. 1.3 and described  
<sup>81</sup> in the following paragraphs.

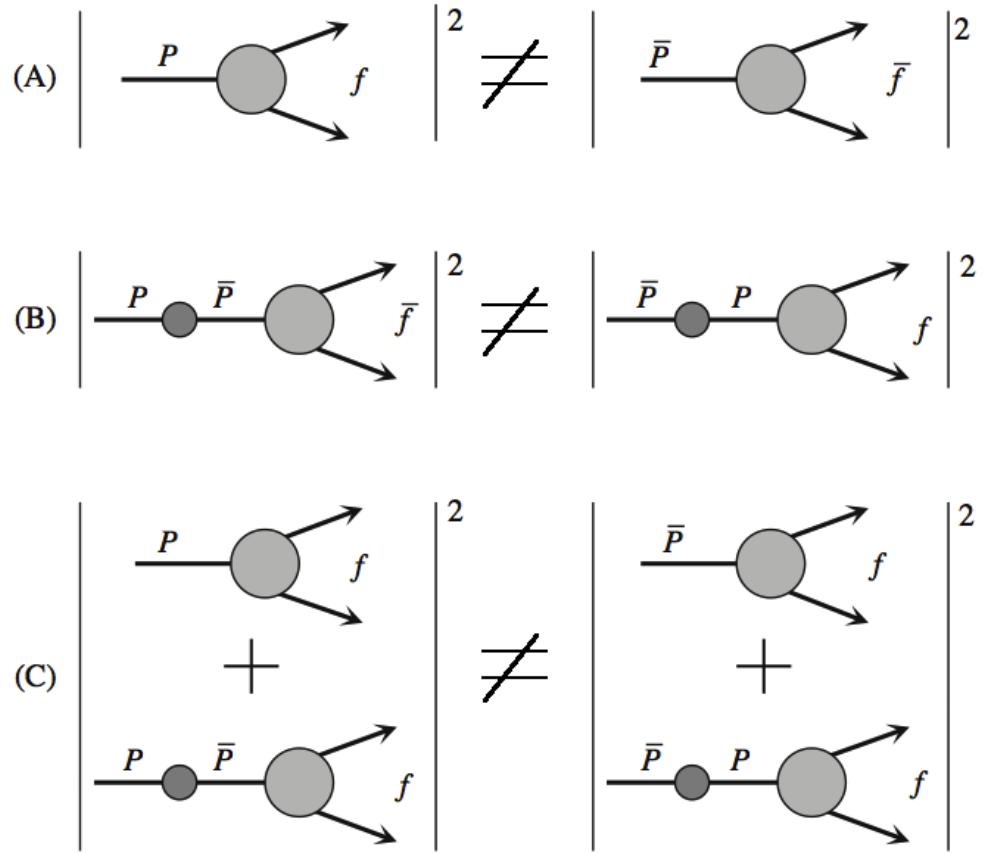


Figure 1.3 – Graphical representation of  $CP$  violation in decay (A), mixing (B) and interference between mixing and decay (C) [23].

### 82 $CP$ violation in decays

$CP$  violation in decays, also known as *direct  $CP$*  violation, happens when the decay rate for  $P \rightarrow f$  is different from that of the  $CP$ -conjugated process  $\bar{P} \rightarrow \bar{f}$ :

$$\left| \frac{\bar{A}_{\bar{f}}}{A_f} \right| \neq 1. \quad (1.41)$$

This kind of  $CP$  violation occurs if, for each decay, at least two amplitudes with different weak ( $\phi_j$ ) and strong ( $\delta_j$ ) phases contribute:

$$A_f = \sum_j |A_j| e^{i(\delta_j + \phi_j)}, \quad \bar{A}_{\bar{f}} = \sum_j |\bar{A}_j| e^{i(\delta_j - \phi_j)}. \quad (1.42)$$

In fact, the strong phases are invariant under  $CP$  conjugation, whereas the weak phases change sign. The following asymmetry between final states can be measured to determine

direct  $CP$  violation experimentally for *charged mesons*, where mixing effects are absent:

$$\mathcal{A}_{f^\pm} = \frac{\Gamma(P^- \rightarrow f^-) - \Gamma(P^+ \rightarrow f^+)}{\Gamma(P^- \rightarrow f^-) + \Gamma(P^+ \rightarrow f^+)} = \frac{\left| \frac{\bar{A}_f}{A_f} \right|^2 - 1}{\left| \frac{\bar{A}_f}{A_f} \right|^2 + 1} \quad (1.43)$$

<sup>83</sup>  **$CP$  violation in mixing**

$CP$  violation in mixing, also called *indirect*  $CP$  violation, occurs when the oscillation rate for  $\bar{P}^0 \rightarrow P^0$  is different from that of the  $CP$ -conjugated process  $P^0 \rightarrow \bar{P}^0$ . These two oscillation probabilities are given by Eqs. 1.29 and 1.30. It turns out that they are identical unless

$$\left| \frac{q}{p} \right| \neq 1. \quad (1.44)$$

From Eq. 1.22, it can be seen that  $CP$  violation in mixing occurs when the relative phase  $\phi_M$  is different from any multiple of  $\pi$ . It is possible to measure the  $|q/p|$  ratio by comparing the oscillation rates in flavour-specific, semileptonic decays of neutral mesons  $P^0 \rightarrow l^+ X$  and  $\bar{P}^0 \rightarrow l^- X$ , where no direct  $CP$  violation occurs. The decays where oscillation occurred are identified by reconstructing “wrong sign” leptons. The so-called semileptonic asymmetry

$$\mathcal{A}_{SL} = \frac{\frac{d\Gamma(\bar{P}^0 \rightarrow l^+ X)}{dt} - \frac{d\Gamma(P^0 \rightarrow l^- X)}{dt}}{\frac{d\Gamma(\bar{P}^0 \rightarrow l^+ X)}{dt} + \frac{d\Gamma(P^0 \rightarrow l^- X)}{dt}} = \frac{1 - |q/p|^4}{1 + |q/p|^4} \quad (1.45)$$

<sup>84</sup> is independent of time.

<sup>85</sup>  **$CP$  violation in the interference between mixing and decay**

This type of decay occurs when a neutral meson can decay directly to a given final state,  $P^0 \rightarrow f$ , or via mixing,  $P^0 \rightarrow \bar{P}^0 \rightarrow f$ . This can happen only if the final state  $f$  is common to both  $P^0$  and  $\bar{P}^0$ . This type of  $CP$  violation can occur also if other sources of  $CP$  violation (mixing or decay) are absent. In general, the interference between mixing and decay can be accessed by studying the following asymmetries:

$$\mathcal{A}_f(t) = \frac{\frac{d\Gamma(P^0 \rightarrow f)}{dt} - \frac{d\Gamma(\bar{P}^0 \rightarrow f)}{dt}}{\frac{d\Gamma(P^0 \rightarrow f)}{dt} + \frac{d\Gamma(\bar{P}^0 \rightarrow f)}{dt}}, \quad \mathcal{A}_{\bar{f}}(t) = \frac{\frac{d\Gamma(P^0 \rightarrow \bar{f})}{dt} - \frac{d\Gamma(\bar{P}^0 \rightarrow \bar{f})}{dt}}{\frac{d\Gamma(P^0 \rightarrow \bar{f})}{dt} + \frac{d\Gamma(\bar{P}^0 \rightarrow \bar{f})}{dt}}. \quad (1.46)$$

### 1.3. Physics of the neutral $B$ mesons

---

A relevant example is the case of neutral  $B$  mesons, where  $|q/p| = 1$ . Using Eqs. 1.35-1.38, the asymmetries of Eq. 1.46 take the following forms:

$$\mathcal{A}_f(t) = \frac{C_f \cos(\Delta mt) - S_f \sin(\Delta mt)}{\cosh\left(\frac{\Delta\Gamma t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma t}{2}\right)}, \quad \mathcal{A}_{\bar{f}}(t) = \frac{-C_{\bar{f}} \cos(\Delta mt) + S_{\bar{f}} \sin(\Delta mt)}{\cosh\left(\frac{\Delta\Gamma t}{2}\right) + D_{\bar{f}} \sinh\left(\frac{\Delta\Gamma t}{2}\right)}. \quad (1.47)$$

<sup>86</sup> The  $CP$  coefficients can be directly measured from a time-dependent analysis of certain  
<sup>87</sup>  $B$  decays.

## 1.4 $B^0 \rightarrow D^\pm \pi^\mp$ analysis motivation and phenomenology

In this thesis, a decay-time dependent analysis of the decay  $B^0 \rightarrow D^\pm \pi^\mp$  is presented, where the  $D^\pm$  meson is reconstructed as  $D^\pm \rightarrow K^\mp \pi^\pm \pi^\pm$ . The pion produced together with the  $D^\pm$  meson will be named *bachelor* or *companion* hereafter. The objective of this study is to perform a measurement of  $CP$  asymmetries, in order to constrain the CKM angle  $\gamma$  [15, 16]. The  $\gamma$  angle is particularly important to test the CKM picture of the  $CP$  violation. In fact,  $\gamma$  is the least-known parameter of the UT, whereas the theoretical predictions for its value are very clean and free from hadronic uncertainties [24, 25, 26]. So, improving the experimental precision on  $\gamma$  is a milestone of the current Flavour Physics programme.  $CP$  violation appears in the interference between the Cabibbo favoured  $b \rightarrow c$  amplitude without mixing,  $A(B^0 \rightarrow D^- \pi^+)$ , and the Cabibbo suppressed  $b \rightarrow u$  amplitude with mixing,  $A(\bar{B}^0 \rightarrow D^- \pi^+)$ . Two of the corresponding Feynman diagrams for these amplitudes are depicted in Fig. 1.4.

The measurement is performed by analysing the four time-dependent decay rates  $\frac{d\Gamma(B^0 \rightarrow D^- \pi^+)}{dt}$ ,  $\frac{d\Gamma(B^0 \rightarrow D^+ \pi^-)}{dt}$ ,  $\frac{d\Gamma(\bar{B}^0 \rightarrow D^- \pi^+)}{dt}$  and  $\frac{d\Gamma(\bar{B}^0 \rightarrow D^+ \pi^-)}{dt}$ . Identifying the final state as  $f = D^- \pi^+$  or  $\bar{f} = D^+ \pi^-$ , and assuming  $CPT$  symmetry, no  $CP$  violation in mixing ( $|q/p| = 1$ ) and decay ( $|A_f|^2 = |\bar{A}_{\bar{f}}|^2$ ,  $|A_{\bar{f}}|^2 = |\bar{A}_f|^2$ ), and  $\Delta\Gamma = 0$ , the time-dependent decay rates for  $B$  mesons initially produced as  $B^0$  or  $\bar{B}^0$  can be written as follows:

$$\frac{d\Gamma(B^0 \rightarrow f)}{dt}(t) = \frac{1}{4\tau} e^{-t/\tau} [1 + C_f \cos(\Delta mt) - S_f \sin(\Delta mt)], \quad (1.48)$$

$$\frac{d\Gamma(\bar{B}^0 \rightarrow f)}{dt}(t) = \frac{1}{4\tau} e^{-t/\tau} [1 - C_f \cos(\Delta mt) + S_f \sin(\Delta mt)], \quad (1.49)$$

$$\frac{d\Gamma(B^0 \rightarrow \bar{f})}{dt}(t) = \frac{1}{4\tau} e^{-t/\tau} [1 + C_{\bar{f}} \cos(\Delta mt) - S_{\bar{f}} \sin(\Delta mt)], \quad (1.50)$$

$$\frac{d\Gamma(\bar{B}^0 \rightarrow \bar{f})}{dt}(t) = \frac{1}{4\tau} e^{-t/\tau} [1 - C_{\bar{f}} \cos(\Delta mt) + S_{\bar{f}} \sin(\Delta mt)], \quad (1.51)$$

where  $\Delta m$  and  $\tau = 1/\Gamma$  are given by Eqs. 1.23 and 1.24, respectively. The  $CP$  coefficients

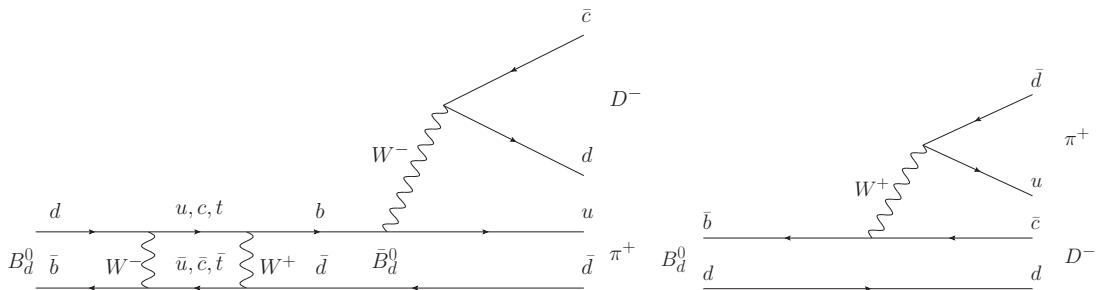


Figure 1.4 – Feynman diagrams contributing to  $B^0 \rightarrow D^- \pi^+$ , with (left) and without (right) mixing.

## 1.4. $B^0 \rightarrow D^\pm \pi^\mp$ analysis motivation and phenomenology

---

defined in Eqs. 1.39 and 1.40, can be expressed as

$$S_f = -\frac{2r_{D\pi} \sin[\delta - (\gamma + 2\beta)]}{1 + r_{D\pi}^2}, \quad S_{\bar{f}} = \frac{2r_{D\pi} \sin[\delta + (\gamma + 2\beta)]}{1 + r_{D\pi}^2}, \quad (1.52)$$

$$C_f = -C_{\bar{f}} = C = \frac{1 - r_{D\pi}^2}{1 + r_{D\pi}^2}, \quad (1.53)$$

where  $\beta$  (Eq. 1.9) is related to the  $B^0$  mixing phase,

$$r_{D\pi} = \frac{|\bar{A}_f|}{|A_f|} = \frac{|A_{\bar{f}}|}{|\bar{A}_{\bar{f}}|} \quad (1.54)$$

is the magnitude of the ratio between the doubly Cabibbo suppressed and favoured amplitudes, and  $\delta$  is the strong phase difference between these amplitudes.

A measurement of  $\gamma$  can be obtained by measuring the  $CP$  coefficients and taking external measurements of  $\beta$  and  $r_{D\pi}$  as input. The angle  $\beta$  is known with very high precision, both theoretically and experimentally [27]. An estimation of  $r_{D\pi}$  was performed by the BaBar and Belle collaborations [28, 29], by measuring the branching fraction of  $B^0 \rightarrow D_s^{(*)+} \pi^-$  decays and assuming SU(3) symmetry, yielding an average of about 1.7% with a relative error around 15%, mainly due to SU(3) symmetry breaking. Because of the very small value of  $r_{D\pi}$ , this analysis is not sensitive to  $C_f$  and  $C_{\bar{f}}$ ; for this reason, these coefficients are simply fixed to +1 and -1, respectively.

The small value for the  $r_{D\pi}$  parameter, which reduces the sensitivity on  $S_{f/\bar{f}}$ , makes this measurement challenging as compared to similar analyses like  $B_s^0 \rightarrow D_s^\mp K^\pm$ . However, the  $B_s^0$  production rate is significantly smaller than the  $B^0$  production fraction ( $f_s/f_d = 0.259 \pm 0.015$  [30]). Moreover, the  $B_s^0 \rightarrow D_s^\mp K^\pm$  branching ratio,  $(2.27 \pm 0.19) \times 10^{-4}$  [31], is much smaller than the  $B^0 \rightarrow D^\mp \pi^\pm$  branching ratio,  $(2.52 \pm 0.13) \times 10^{-3}$  [32]. So, the resulting, larger number of  $B^0 \rightarrow D^\mp \pi^\pm$  candidates compensates for this reduced sensitivity.

Measurements of  $\sin(2\beta + \gamma)$  in  $B^0 \rightarrow D^{(*)\pm} \pi^\mp$  were performed previously by the BaBar and Belle collaborations [33, 34, 35, 36]. I'm one of the authors of a measurement of  $2\beta_s - \gamma$  in  $B_s^0 \rightarrow D_s^\mp K^\pm$  decays with  $3 \text{ fb}^{-1}$  of data [37].

The measurement presented in this thesis is performed in terms of a flavour-tagged, decay-time dependent analysis of the Run 1 dataset (Sec. 2.2.5). The dataset includes two sub-samples recorded with opposite directions of the magnetic field (“up” and “down”) in the spectrometer dipole. The selection of the decays, which is explained in detail in Sec. 4.1, includes the use of vetoes to reduce the number of components that must be modelled in the sample, and a boosted decision tree (BDT) to reduce the amount of *combinatorial* background. The expected sample composition after the selection is discussed in Sec. 4.2 based on studies with simulated samples. A fit to the invariant mass

129 distribution of the resulting dataset is performed to extract *sWeights* [38] for the signal  
130 component. The fit is described in detail in Sec. 4.3. The training and calibration of  
131 the flavour-tagging algorithms, which infer the initial flavour of the reconstructed  $B^0$   
132 candidates, is summarised in Sec. 3.1. Finally, the estimation of the  $CP$  coefficients is the  
133 result of an unbinned, *sWeighted* likelihood fit to the distributions of the decay time and  
134 the flavour tagging observables.

### 135 1.5 Personal contribution

136 The work presented in this thesis has been carried out in close cooperation with LHCb  
137 members of the University of Dortmund. My main responsibilities related to the  $B^0 \rightarrow$   
138  $D^\mp\pi^\pm$  analysis during my PhD time were:

- 139     • opposite-side flavour tagging calibration (Sec. 3.2.1);
- 140     • correction of particle identification (Sec. 4.2.1);
- 141     • mass fit for *sWeights* calculation (Sec. 4.3);
- 142     • time-dependent analysis, in particular acceptance parameterisation (Sec. 5.2), decay-  
143       time fit (Sec. 5.3), and estimation of systematic uncertainties (Sec. 5.5).

144 Moreover, I was responsible for:

- 145     • calibration and performance studies of the LHCb Silicon Tracker (Sec. 2.2.1). The  
146       results of these studies are reported in Ref. ??;
- 147     • monitoring and maintenance of the Silicon Tracker as on-call expert during data-  
148       taking operations (*piquet*);
- 149     • opposite-side electron tagger optimisation (Sec. 3.3).



## 2 Experimental setup

151

**2.1 The Large Hadron Collider**

The *Large Hadron Collider* (LHC) is a circular collider with a circumference of 26.66 km. It is located at CERN, near Geneva, between Switzerland and France. The LHC is designed to produce proton-proton ( $pp$ ) collisions with a *luminosity* of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  and a centre-of-mass energy of 14 TeV. In the first data-taking period before the first long shutdown, called Run 1 (2010–2012), the centre-of-mass energy reached 7 TeV (2010–2011) and 8 TeV (2012).

The proton bunches, produced from hydrogen gas, pass through different intermediate accelerating stages (Fig. 2.1):

- LINAC 2 (50 MeV);
- Proton Synchrotron Booster (1.4 GeV);
- Proton Synchrotron (25 GeV);
- Super Proton Synchroton (450 GeV).

Finally, they are injected clockwise and counter-clockwise into the LHC and accelerated to their final energy. Each bunch contains  $\sim 10^{11}$  protons, and the nominal number of bunches per beam is 2808. At LHC, in addition to LHCb, there are two general-purpose detectors (ATLAS and CMS), a detector dedicated to quark matter and quark-gluon plasma physics (ALICE) and other smaller experiments (TOTEM, LHCf, MoEDAL) dedicated to different topics.

The LHC can also accelerate particles other than protons, such as lead or xenon nuclei, in order to collect data samples for specific studies.

## 2.1. The Large Hadron Collider

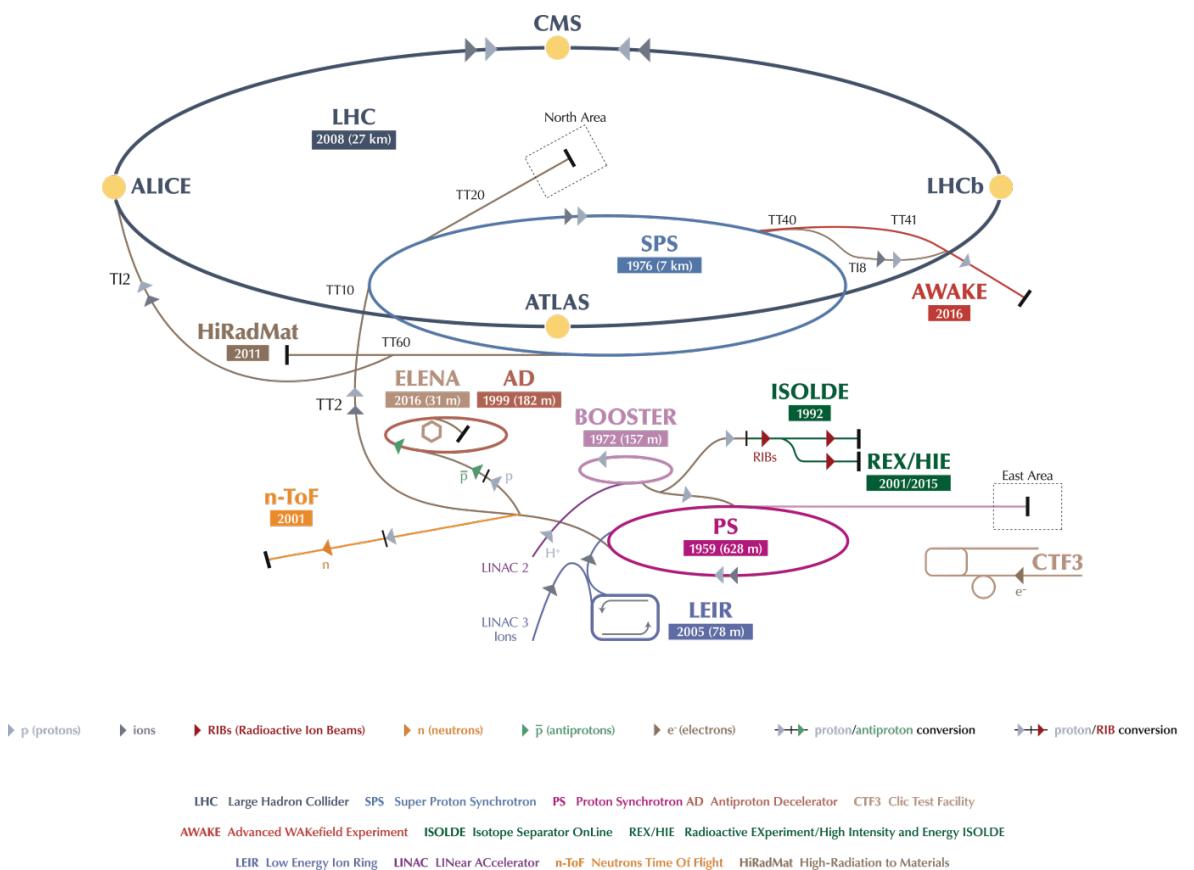


Figure 2.1 – Overview of the CERN accelerators complex.

<sup>173</sup> **2.2 The LHCb experiment**

<sup>174</sup> The *Large Hadron Collider beauty* (LHCb) experiment [39] is a single-arm forward  
<sup>175</sup> spectrometer (see Fig. 2.2) that exploits the forward production of the  $b$ - and  $c$ - quark  
<sup>176</sup> pairs in  $pp$  collisions (Fig. 2.3). The LHCb angular coverage is comprised between 15 mrad  
<sup>177</sup> and 250 (300) mrad in the vertical (horizontal) plane. The LHCb coordinate system  
<sup>178</sup> consists of a right-handed set of axes,  $x$ ,  $y$ ,  $z$ , where the positive  $z$  direction extends into  
<sup>179</sup> the LHCb detector,  $y$  is perpendicular to the LHCb cavern ground and  $x$  is orthogonal to  
<sup>180</sup> the other two.

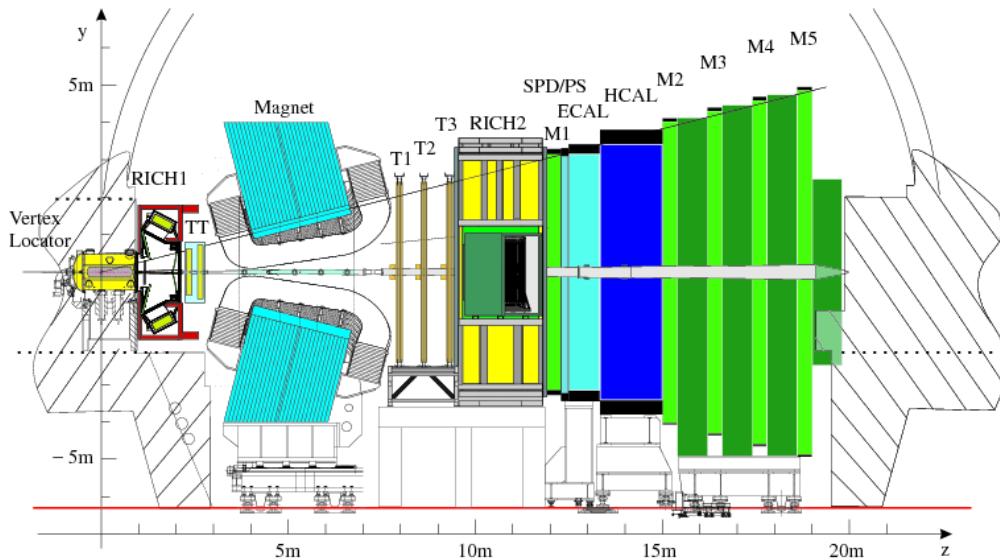


Figure 2.2 – Side view of the LHCb detector.

<sup>181</sup> The LHCb experiment is composed of different sub-detectors. The tracking system  
<sup>182</sup> includes a vertex and tracking detector called *VErtex LOcator* (VELO), the *Tracker*  
<sup>183</sup> *Turicensis* (TT), located upstream a magnetic dipole with an integrated field of 4 Tm, the  
<sup>184</sup> *Inner Tracker* (IT), situated downstream the magnet in three separated stations around  
<sup>185</sup> the beryllium beam pipe, and the *Outer Tracker* (OT), installed in the same stations as  
<sup>186</sup> the IT. The *Particle IDentification* (PID) system comprises two *Ring Imaging CHerenkov*  
<sup>187</sup> detectors (RICH), an *Electromagnetic CALorimeter* (ECAL), which also includes a *Pre-*  
<sup>188</sup> *Showr* (PS) and *Scintillator Pad Detector* (SPD), a *Hadronic CALorimeter* (HCAL)  
<sup>189</sup> and five *muon stations* (M1–M5).

<sup>190</sup> **2.2.1 Tracking system**

<sup>191</sup> The tracking system has a crucial role for the time-dependent analysis of  $B^0 \rightarrow D^\mp \pi^\pm$   
<sup>192</sup> decays at LHCb. The excellent vertex and impact-parameter resolution allows to separate  
<sup>193</sup> true, long-lived  $B^0$ -mesons from random combinations of other random tracks, and to

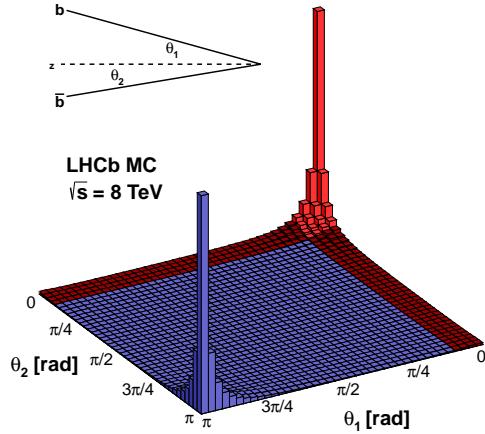


Figure 2.3 – Joint distribution of the  $b$  and  $\bar{b}$  production angles with respect to the beam direction at  $\sqrt{s} = 8$  TeV, as obtained from simulation. The red part shows the LHCb acceptance.

improve the decay-time resolution. Moreover, the optimal momentum resolution implies a low resolution on the reconstructed invariant mass, which is crucial to separate true  $B^0 \rightarrow D^\mp \pi^\pm$  decays from other physical background.

The tracking system is also essential for the flavour tagging algorithms, which rely on the quality of the reconstructed tracks and vertices to discriminate correctly tagged neutral  $B$ -mesons from wrongly tagged ones.

As summary of the performances of the tracking system (decay-time, impact-parameter, and momentum resolution) is shown in Fig. 2.4.

## 202 The VErtex LOcator (VELO)

The VELO [40] (Fig. 2.5) is a silicon micro-strip detector surrounding the interaction point, which detects charged particles, performs the first track reconstruction step and identifies decay vertices. The sensitive region of the VELO is composed of n-on-n silicon micro-strip half-disk sensors with two different read-out strip geometries, called  $r$ -type and  $\phi$ -type, which measure the radial ( $r$ ) and azimuthal ( $\phi$ ) position in polar coordinates. The silicon sensors are 8.4 cm in diameter and have an inner hole with radius 0.8 cm. The strip pitch ranges from 38 to 108  $\mu\text{m}$  (38 to 97  $\mu\text{m}$ ) for  $r$  ( $\phi$ ) sensors, while the sensor thickness is 300  $\mu\text{m}$ . The VELO consists of 21 stations placed perpendicular to the beam axis. Each station has two independent halves that can be moved apart during beam injection and then closed again when the beam orbit is stabilised. Each

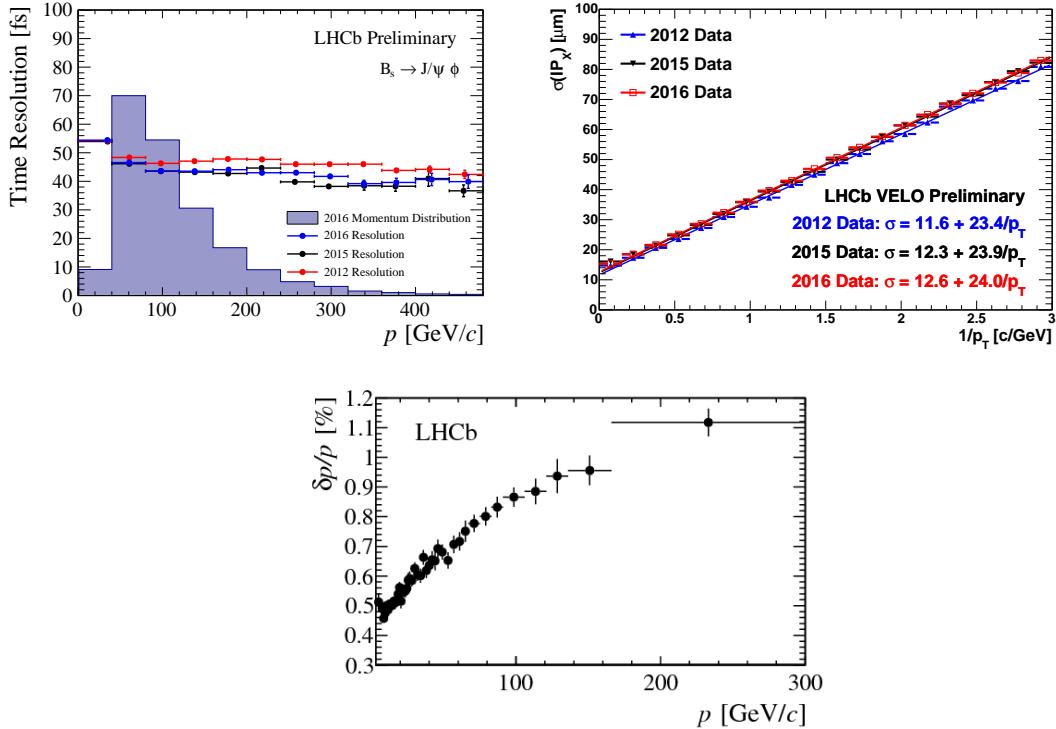


Figure 2.4 – Top left: decay-time resolution as a function of the momentum for reconstructed  $B_s^0 \rightarrow J/\psi\phi$  decays on 2012 (red), 2015 (black) and 2016 (blue) data. The momentum distribution is superimposed as the solid, violet histogram. Top right: resolution of the  $x$ -coordinate of the impact parameter as a function of the inverse of the reconstructed transverse momentum on 2012 (blue), 2015 (black) and 2016 (red) data. The result of a linear fit for each dataset is superimposed. Bottom: relative momentum resolution as a function of the reconstructed momentum.

half-station is composed by one  $r$ -type and one  $\phi$ -type sensor. The total length of the VELO detector is about 1 m. The impact parameter (IP) resolution of a track is measured to be  $\sigma_{IP} = 11.6 \pm 23.4/p_T \mu\text{m}$  in  $x$  and  $\sigma_{IP} = 11.2 \pm 23.2/p_T \mu\text{m}$  in  $y$ , where  $p_T$  is the *transverse momentum* (in  $\text{GeV}/c$ ) of the particle with respect to the beam axis.

### 217 The Tracker Turicensis (TT)

218 The TT [41] (Fig. 2.6) is a silicon micro-strip detector covering a total area of about  
 219  $7.9 \text{ m}^2$  upstream the magnet and divided into two separate stations (TTa, TTb). Each  
 220 station has two layers. The TT helps in improving the track momentum resolution and  
 221 detecting long-lived particles that decay outside the VELO acceptance. TTa is composed  
 222 of a X and an U layer, while TTb includes a V and an X layer. The X layers have read-out  
 223 strips aligned vertically, whereas the U and V *stereo* layers are rotated by  $+5^\circ$  and  $-5^\circ$

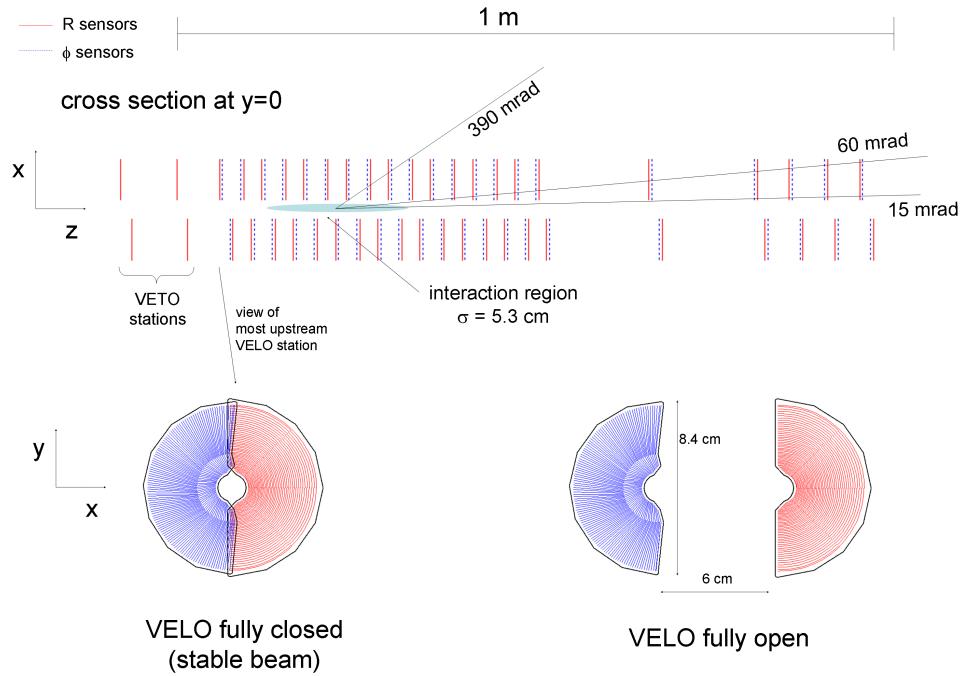


Figure 2.5 – Schematic overview of the VELO detector (top) and its sensors (bottom).

224 with respect to the vertical in the  $xy$  plane.

225 The TT active area is made of p-on-n silicon micro-strip sensors. Since the sensors are  
226 exposed to a significant radiation dose due to the high track multiplicity, they are cooled  
227 to  $0^{\circ}\text{C}$  in order to minimise the damage.

228 A TT read-out module contains from one to four sensors connected in series, resulting in  
229 read-out strips up to 37 cm long. The strip pitch is 183  $\mu\text{m}$  and the sensor thickness is  
230 500  $\mu\text{m}$ . The hit resolution is about 50  $\mu\text{m}$ .

### 231 The Inner Tracker (IT)

232 The IT [42] (Fig. 2.7) is also a silicon micro-strip detector. Together with the TT, it  
233 forms the *Silicon Tracker* (ST). It is dedicated to detect charged particles in the high  
234 track-density region around the beam pipe downstream the magnet. It is separated into  
235 three stations, where each station consists of four boxes. Each box has four layers made  
236 of seven read out modules arranged in a X-U-V-X layout similar to that of the TT. The  
237 total coverage of the IT is about  $4.2 \text{ m}^2$ . The boxes directly above and below the beam  
238 pipe are made of single-sensor modules, called *short modules*, whereas the side boxes are  
239 made of two bonded silicon sensor modules, called *long modules*. The IT strip pitch is

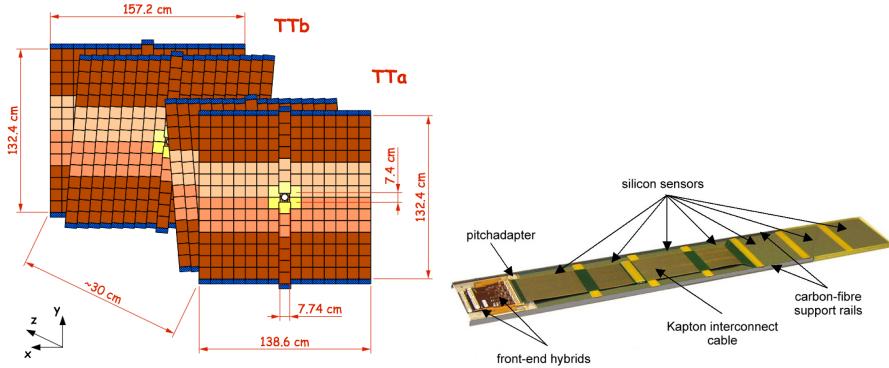


Figure 2.6 – Schematic overview of the TT stations/layers (left) and one of the TT readout modules (right).

<sup>240</sup> 198  $\mu\text{m}$ , while the p-on-n sensor thickness is 320 (410)  $\mu\text{m}$  for the short (long) modules.  
<sup>241</sup> The hit resolution is about 50  $\mu\text{m}$ .

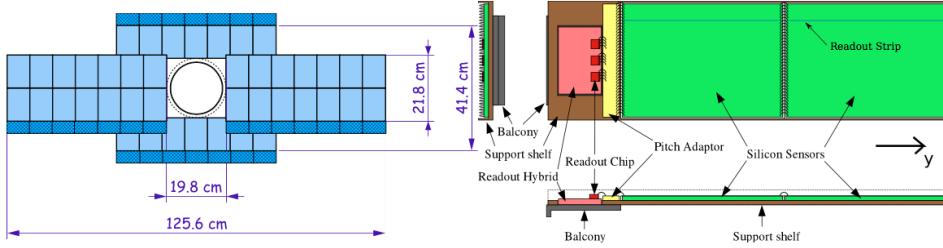


Figure 2.7 – Schematic overview of an IT station (left) and one of the long IT readout modules (right).

## <sup>242</sup> The Outer Tracker (OT)

<sup>243</sup> The OT [43] (Fig. 2.8) is a gaseous straw-tube detector filled with an Ar/CO<sub>2</sub>/O<sub>2</sub>  
<sup>244</sup> (70%/28.5%/1.5%) gas mixture. It is dedicated to the detection of charged particles in  
<sup>245</sup> the low track density region outside the IT acceptance and covers a large area of about  
<sup>246</sup> 340 m<sup>2</sup>. The OT is composed of three stations, where each station has four layers in  
<sup>247</sup> a X-U-V-X configuration. Each station is separated physically in left and right sides  
<sup>248</sup> with respect to the beam pipe mounted in support structures called C frames because of  
<sup>249</sup> their shape. Each layer is divided into two mono-layers. The OT has different types of  
<sup>250</sup> modules, the long F modules and the short S1, S2, S3 modules that are cut in two pieces  
<sup>251</sup> to leave space for the IT. The straw tube and anode wire diameters are 5 mm and 25  $\mu\text{m}$   
<sup>252</sup> respectively. The hit resolution is about 200  $\mu\text{m}$ .

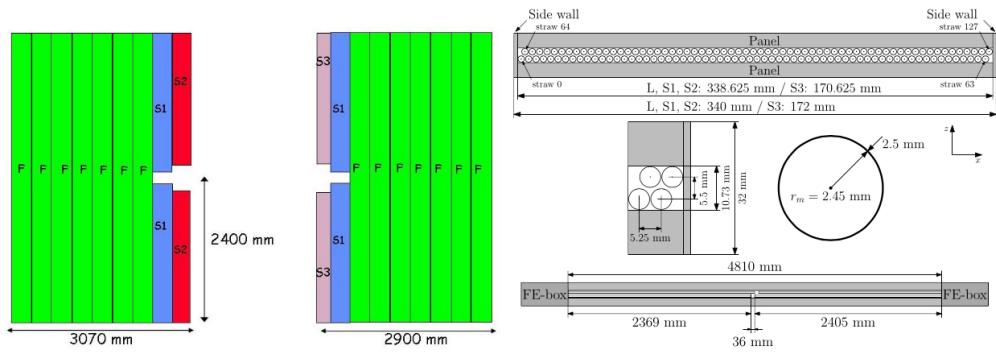


Figure 2.8 – Schematic overview of an OT layer (left) and an OT module layout (right).

### 253 Spillover noise in the Silicon Tracker

254 Starting from 2015, the time spacing between the LHCb proton *bunches* has been 25 ns,  
 255 half of the value adopted before. This had a direct impact on the front-end electronics of  
 256 the silicon detectors (VELO, TT, IT), because the width in time of the analogue signal  
 257 produced by the front-end electronics is of the same order of magnitude. This means that  
 258 it is possible to still have a non negligible amount of signal in the subsequent collision,  
 259 which can be misidentified as coming from particles produced in that event. This source  
 260 of noise is called *spillover*. The starting seeds of tracking algorithms, called *clusters*,  
 261 can be polluted by spillover clusters which may increase the number of fake (or *ghost*)  
 262 reconstructed tracks.

263 In the first part of my PhD activity, I studied the effect of spillover clusters in the ST  
 264 using both simulated events and real collision data. This study showed that this time  
 265 spacing has little impact on the detector *occupancy*, and that the increase of ghost tracks  
 266 is negligible. Moreover, in the same study it was shown that the charge deposited by  
 267 particles in the detector can be exploited as a *feature* in multivariate analyses in order  
 268 to further reduce the ghost track contamination. These results were documented in a  
 269 note [44] internal to the LHCb collaboration.

### 270 2.2.2 Particle identification (PID)

#### 271 The Ring Imaging Cherenkov (RICH) detectors

272 When a charged particle is travelling faster than the speed of light in a medium, Cherenkov  
 273 light is produced at an angle that depends on the velocity of the particle and the refractive  
 274 index of the medium (*radiator*). By knowing the momentum from the tracker and the  
 275 velocity from the RICH detectors, the mass can be determined and therefore provide  
 276 particle identification. Two RICH detectors [45] (Fig. 2.9) are used in order to provide  
 277 PID in different momentum ranges.

## Chapter 2. Experimental setup

---

278 RICH1 is responsible for providing PID in the momentum range from 1 to 60 GeV/c. The  
279 angular acceptance ranges from 25 to 50 (300) mrad in the vertical (horizontal) plane.  
280 The adopted radiator is fluorobutane ( $C_4F_{10}$ ). RICH1 is located between the VELO  
281 and the TT. The Cherenkov photons are guided to Hybrid Photon Detectors (HPD) via  
282 dedicated mirrors.

283 The average kaon identification efficiency in the momentum range from 2 to 100 GeV/c is  
284  $\sim 95\%$ . The average probability that pions are wrongly identified as kaons is  $\sim 5\%$  in  
285 the same momentum range.

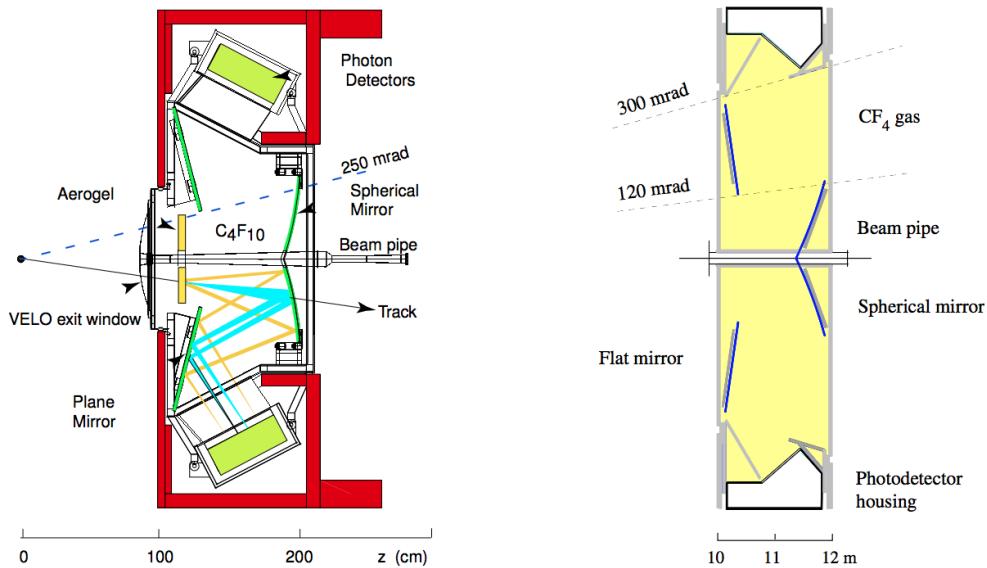


Figure 2.9 – Side view of RICH1 (left) and top view of RICH2 (right).

286 RICH2 is optimised for the momentum range from 15 to 100 GeV/c. The angular accept-  
287 tance ranges from 15 to 100 (120) mrad in the vertical (horizontal) plane where most  
288 of the high-momentum tracks are produced. RICH2 uses tetrafluoromethane ( $CF_4$ ) as  
289 radiator.

### 290 The Electromagnetic CALorimeter (ECAL), Pre-Shower (PS) and Scintillator 291 Pad Detector (SPD)

292 The ECAL [46] is used for the detection and measurement of the energy of electrons and  
293 photons. The ECAL is built as a sandwich of alternating scintillators and lead layers in  
294 the  $xy$  plane. Scintillation light produced by the shower of particles generated by the  
295 lead plates is read out by Wave-Length Shifter (WLS) fibres coupled to PhotoMultiplier  
296 Tubes (PMTs). The SPD is installed upstream the ECAL to separate electrons from  
297 photons. The PS is installed between the SPD and the ECAL. Both SPD and PS use

298 scintillator pads read out by WLS fibres coupled to MultiAnode PhotoMultiplier Tubes  
 299 (MAPMT). The acceptance range of the ECAL is from 25 up to 300 (250) mrad in  
 300 the horizontal (vertical) plane. The relative energy resolution of the ECAL is given by  
 301  $\sigma_E/E = 10\%/\sqrt{E} \oplus 1\%$ , where  $E$  is given in GeV.

### 302 The Hadronic CALorimeter (HCAL)

303 The HCAL [46] is used for the detection and measurement of the energy of hadrons for  
 304 the first level trigger. A HCAL cell is a sampling device made of alternating iron and  
 305 scintillator tiles, where the latter are located along the beam direction. The HCAL has  
 306 the same acceptance coverage of the ECAL. The relative energy resolution of the HCAL  
 307 is given by  $\sigma_E/E = (69 \pm 5\%)/\sqrt{E} \oplus (9 \pm 2)\%$ , where  $E$  is given in GeV.

### 308 Muon detectors

309 The muon system [47] (Fig. 2.10) is a gaseous detector composed of five stations (M1  
 310 to M5) interleaved by 80 cm thick iron filters. The gaseous detectors are Multi-Wire  
 311 Proportional Chambers (MWPC), except for the innermost part of M1, where triple Gas  
 312 Electron Multipliers (GEM) detectors are used to cope with the higher track density. The  
 313 angular acceptance ranges from 20 (16) to 308 (256) mrad in the horizontal (vertical)  
 314 plane. The muon detector has 1380 chambers and covers a total area of 435 m<sup>2</sup>. Each  
 315 muon chamber is composed of four layers of MPWC, except for M1, where two layers are  
 316 used. The hit efficiency of the chambers is higher than 99% enabling a trigger efficiency  
 317 greater than 95% for muons. The adopted gas mixture (40% Ar, 55% CO<sub>2</sub>, 5% CF<sub>4</sub>)  
 318 allows a fast triggering on muons (40 MHz).

#### 319 2.2.3 Trigger system

320 The bunch crossing rate of LHC is very high (40 MHz) because more than 99% of the  
 321  $pp$  collisions do not produce interesting events. It is not possible to record events with  
 322 such a high rate: therefore, a trigger system [48] is required to reduce the rate from  
 323 40 MHz down to a few kHz. The rate reduction is achieved via selection criteria which  
 324 ensure that events containing heavy flavour decays are stored. The signatures of these  
 325 interesting decays include high transverse momentum ( $p_T$ ) and transverse energy ( $E_T$ )  
 326 of the decay products, as well as displaced decay vertices or tracks with large impact  
 327 parameter (IP) with respect to the  $pp$  collision point due to the relative long lifetimes of  $b$   
 328 and  $c$  hadrons. The trigger is divided into two sequential stages: a hardware stage called  
 329 Level-0 (L0) trigger, and a software stage called High Level Trigger (HLT). Different  
 330 trigger decisions are separated into various *lines*, each of which provides information on  
 331 different physics processes (e.g. decay topology, presence of muons etc...). All the trigger  
 332 steps are summarised in Fig. 2.11.

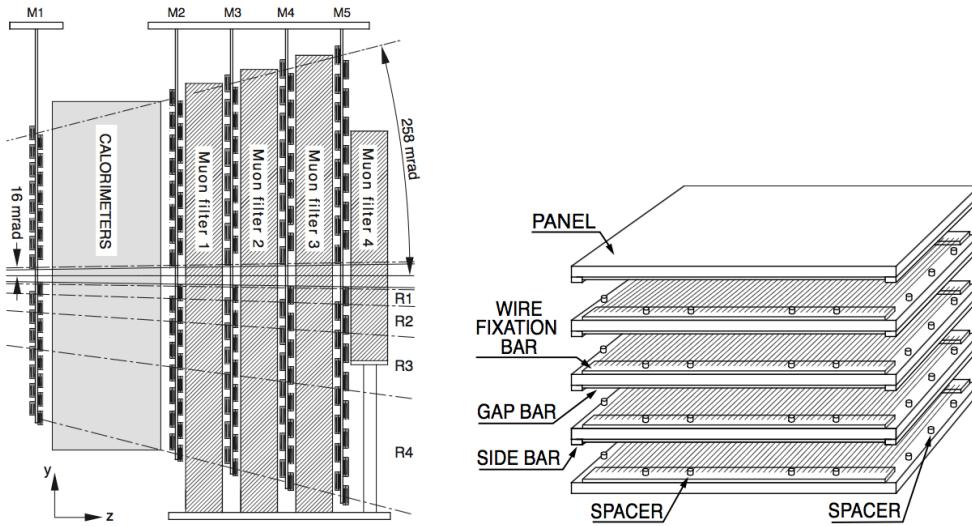


Figure 2.10 – Schematic overview of the muon system (left) and a MWPC (right).

333 Two types of trigger response are assigned offline, when some physics channel is analysed.  
 334 The TOS (*Trigger On Signal*) trigger occurs when the presence of the signal is sufficient  
 335 to have a positive trigger decision. The TIS (*Trigger Independent of Signal*) trigger occurs  
 336 when, after removing signal tracks and hits associated to them, another signature in the  
 337 event is sufficient to have a positive trigger decision.

338 After the trigger stage, the data go through further offline selection steps, where exclusive  
 339 (e.g.  $B \rightarrow D^\mp\pi^\pm$ ,  $B^\pm \rightarrow J/\psi K^\pm$ ) and inclusive (e.g.  $b$  hadron  $\rightarrow J/\psi X$ ,  $J/\psi \rightarrow$   
 340  $\mu^+\mu^-$ ) decays are reconstructed at higher quality than was possible in the strict timing  
 341 requirements of the trigger, and further selections are applied. This offline selection step  
 342 is known as *stripping*, and each set of selection requirements is called *stripping line*.

### 343 Level-0 trigger (L0)

344 The L0 mainly exploits the calorimeters and muon chambers. The idea behind the L0 is  
 345 to select events that contain high  $p_T$  muons and high  $E_T$  hadrons, electrons and photons,  
 346 which very likely come from  $b$ - and  $c$ -hadron decays, by using simple signatures that do  
 347 not require the reconstruction of the event. The L0 reduces the data rate from 40 MHz  
 348 down to 1 MHz, which is the rate limit of the LHCb readout electronics.

### 349 High Level Trigger (HLT)

350 The HLT is separated into two stages, HLT1 and HLT2, and runs on about 29 000  
 351 commercial CPU cores.

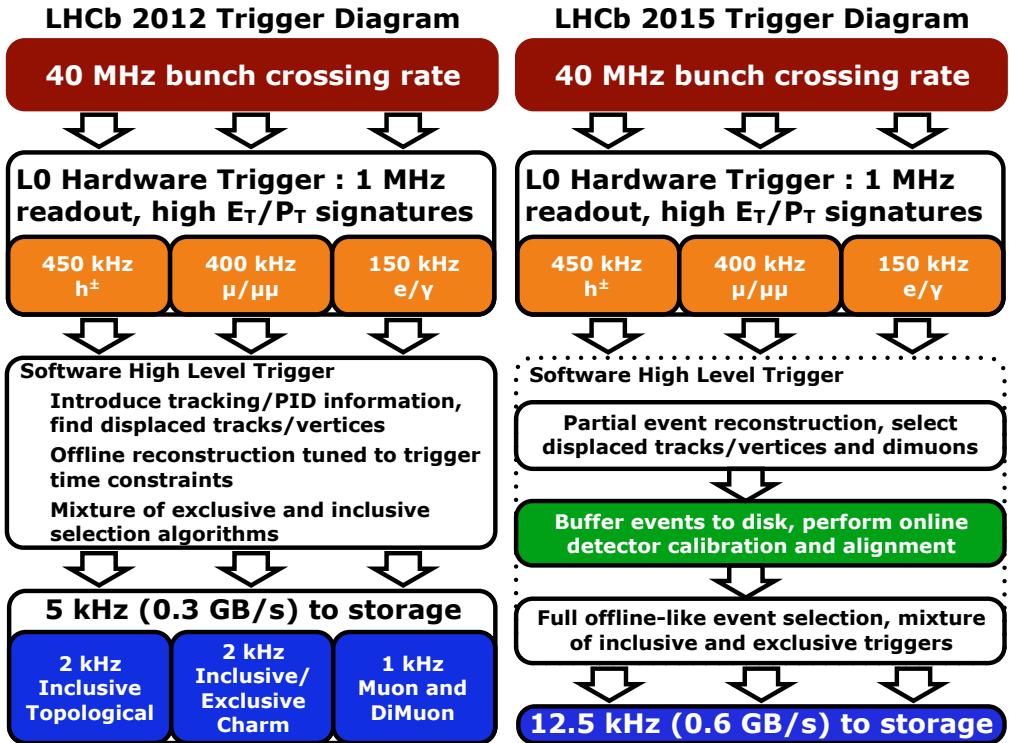


Figure 2.11 – Summary of the trigger strategies followed during the Run 1 (2011–2012, left) and Run 2 (2015–2017, right) data-taking periods. During Run 2, an online detector and calibration alignment was introduced, as well as full event selections (both inclusive and exclusive).

- 352 At the HLT1 level, the full detector information is read out, and then track/vertex  
 353 reconstruction and PID are performed. The exploited signatures are mainly the presence  
 354 of high  $p_T$  tracks, high transverse energy calorimeter clusters (photons and  $\pi^0$ ), high  
 355 invariant mass of muon pairs, and tracks with large IP. All the HLT1 trigger lines are  
 356 *inclusive*, meaning that only decay products common to various decay processes are  
 357 selected rather than specific decays. After the HLT1, the rate goes down to about 70 kHz.
- 358 The HLT2 is a combination of inclusive selections and algorithms that reconstruct entirely  
 359 (*exclusively*) some decay processes. The main lines are topological lines using Multi-  
 360 Variate Analysis (MVA) techniques with different sets of kinematic and position features  
 361 as input, exclusive charm lines and high mass displaced di-hadron/lepton lines. After the  
 362 HLT2, the events are stored on tape for further offline analysis.

<sup>363</sup> **2.2.4 Event reconstruction and simulation software**

<sup>364</sup> **Track and vertex reconstruction**

<sup>365</sup> Starting from the *hits* in the tracking detectors, tracks and vertices are reconstructed  
<sup>366</sup> via dedicated algorithms. Different track types are reconstructed, as shown in Fig. 2.12.  
<sup>367</sup> Each track is characterised by hits collected in different sub-detectors. For example,  
<sup>368</sup> downstream tracks, with no hits in the VELO, are typically associated to long-lived  
<sup>369</sup> particles such as  $\Lambda$  and  $K_S^0$ . Because of the vertical magnetic field, tracks are bend in the  
<sup>370</sup>  $xz$  plane. By knowing the reconstructed particle trajectory and the magnetic field map,  
<sup>371</sup> it is possible to measure the momentum of the particle.

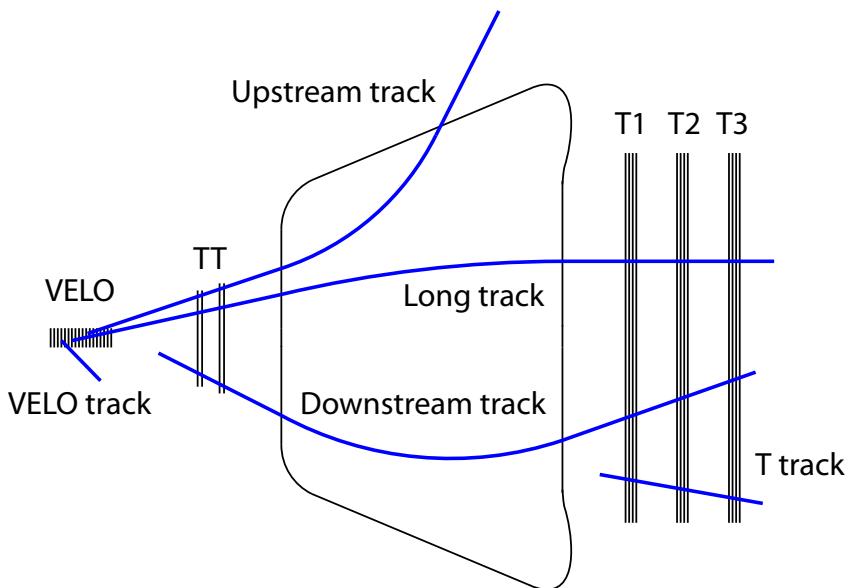


Figure 2.12 – Schematic description of the different track types reconstructed in LHCb.

<sup>372</sup> **PID**

The Cherenkov radiation emitted by charged particles in the RICH radiators produces *rings* in the RICH acceptance, which are reconstructed via pattern recognition algorithms. For each reconstructed pattern, the *likelihood*  $\mathcal{L}_\pi$  for the ring to be produced by a pion (the most common particle in the LHCb environment) is computed. The momentum of the particle is also used in the likelihood computation. Then, likelihood functions for other hypotheses (kaon, proton, electron, muon) are computed and compared with the

pion likelihood. For a given particle X ( $X=e, \mu, p, k$ ), the PIDX observable is defined as

$$\text{PIDX} = \ln \mathcal{L}_X - \ln \mathcal{L}_\pi. \quad (2.1)$$

373 In typical LHCb analyses, requirements on the PID observables are applied to suppress  
 374 physical backgrounds due to wrongly-identified particles. The PID is adopted, together  
 375 with other kinematic and detector-related observables, as input feature for neural networks  
 376 which predict the probability for a particle to be either an electron, a muon, a proton, a  
 377 kaon, or a pion. This probability is denoted as PROBNNX ( $X=e, \mu, p, k, \pi$ ).

378 **Simulation**

379 The Monte Carlo (MC) simulation of  $pp$  collisions, particle decays, and interactions with  
 380 the detector are crucial in the validation of physics analyses. The parton-parton collision  
 381 and hadronisation simulation is performed by PYTHIA [49], interfaced to EvtGen [50] for  
 382 the decay of the hadrons and leptons for standard productions. The QED corrections to  
 383 the decay (i.e. the emission of radiation photons) is generated by the PHOTOS package [51].  
 384 The interactions of particles with detector material and their tracking in the magnetic  
 385 field are simulated by GEANT4 [52, 53].

386 **2.2.5 Data collected by LHCb**

The collision rate  $R$  [ $\text{s}^{-1}$ ] in LHC can be expressed in terms of the *cross-section*  $\sigma$  [ $\text{cm}^2$ ] and the *luminosity*  $\mathcal{L}$  [ $\text{s}^{-1}\text{cm}^{-2}$ ] as

$$R = \mathcal{L}\sigma. \quad (2.2)$$

387 For a given data-taking period, the *integrated luminosity*  $L$  [ $\text{cm}^{-2}$ ] is a measure of the  
 388 amount of recorded data. The typical unit for luminosity is the inverse *barn*, which  
 389 corresponds to  $10^{24} \text{ cm}^{-2}$ . The LHCb integrated luminosity is of the order of an inverse  
 390 *femtobarn* ( $\text{fb}^{-1}$ ); one inverse femtobarn corresponds to the production of about  $10^{11} b\bar{b}$   
 391 quark pairs.

392 The LHCb detector collected data produced mainly from  $pp$  collisions in the 2010–2017  
 393 period, so far. During the 2011–2012 data-taking period, called *Run 1*, about  $3 \text{ fb}^{-1}$  of  
 394 data were collected. The centre-of-mass energy  $\sqrt{s}$  of the  $pp$  interactions was 7 TeV and  
 395 8 TeV in 2011 and 2012, respectively, and the time spacing between *bunches* of protons  
 396 in the LHC was 50 ns. The 2013–2014 period, known as *Long Shutdown 1* (LS1), was  
 397 dedicated to work on the LHC machine to bring the energy up. The *Run 2* data-taking  
 398 period started in 2015, and is planned to last until the end of 2018. The centre-of-mass  
 399 energy of the  $pp$  collisions during Run 2 is 13 TeV, whereas the time spacing between  
 400 proton bunches is 25 ns. In the 2015–2017 data taking period, about  $3.7 \text{ fb}^{-1}$  of data

## Chapter 2. Experimental setup

---

401 were collected. The data collected during each year is summarised in Fig. 2.13.

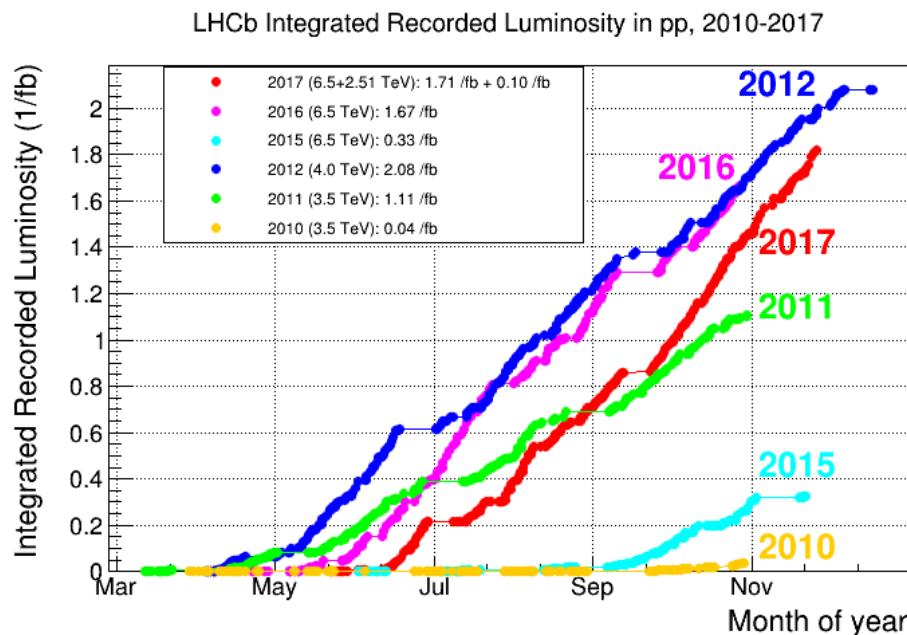


Figure 2.13 – Integrated luminosity of  $pp$  collisions collected each year by LHCb.

# 3 Flavour tagging

403

<sup>404</sup> **3.1 Flavour tagging algorithms**

<sup>405</sup> In this chapter, a description of the *flavour tagging* techniques at LHCb is reported. After  
<sup>406</sup> a brief introduction to the methods, the calibration of the *Opposite Side* (OS) and *Same*  
<sup>407</sup> *Side* (SS) algorithms for the  $B^0 \rightarrow D^\pm \pi^\mp$  time-dependent analysis are described. Finally,  
<sup>408</sup> a reoptimisation of the OS *electron* (OSe) tagger on both Run 1 and Run 2 data is  
<sup>409</sup> reported. This work was made in collaboration with the University of Dortmund. During  
<sup>410</sup> my PhD work activity, I focused mainly on the OS calibration and OSe reoptimisation.

<sup>411</sup> The identification of the flavour at production time of a neutral  $B$  meson, *i.e.* whether it  
<sup>412</sup> contained a  $b$  or a  $\bar{b}$  quark at production, is a key element for any time-dependent analysis  
<sup>413</sup> aiming at the measurement of oscillations and  $CP$  asymmetries, and in particular for the  
<sup>414</sup>  $B^0 \rightarrow D^\pm \pi^\mp$  analysis reported in this thesis. In fact, this information is needed in order  
<sup>415</sup> to measure the decay rates or asymmetries introduced in Sec. 1.3. Complications arise  
<sup>416</sup> from two facts:

- <sup>417</sup> • neutral  $B$  meson oscillate, so the flavour at the production time might differ from  
<sup>418</sup> the flavour at the decay time;
- <sup>419</sup> • many final states, such as  $D^\pm \pi^\mp$ , are not  $CP$  eigenstates. This means that both  $B$   
<sup>420</sup> and  $\bar{B}$  mesons can decay to the same final state particles, and so the flavour cannot  
<sup>421</sup> be obtained from the charge of these particles, even if there were no oscillations.

<sup>422</sup> For these reasons, the flavour has to be reconstructed by exploiting particles not produced  
<sup>423</sup> in the decay of the neutral  $B$  meson being analysed.

<sup>424</sup> Techniques to infer the initial flavour of a reconstructed candidate are called flavour  
<sup>425</sup> tagging algorithms. Several flavour tagging algorithms exist in LHCb; they can be  
<sup>426</sup> categorised into same-side taggers (SS taggers) and opposite-side taggers (OS taggers). A  
<sup>427</sup> schematic representation of the taggers that can be used for tagging  $B^0$  mesons is shown  
<sup>428</sup> in Fig. 3.1.

<sup>429</sup> The SS taggers infer the production flavour of the signal  $B$  meson by selecting charged  
<sup>430</sup> particle candidates that have a high chance of being remnants of the hadronisation process  
<sup>431</sup> of the  $B$  candidate [54, 55]. For  $B^0$  mesons, the same-side pion tagger (SS $\pi$ ), which  
<sup>432</sup> exploits charged  $\pi$  mesons produced in the hadronisation of the  $B^0$  meson, and the  
<sup>433</sup> same-side proton tagger (SS $p$ ), which looks for co-produced protons, have been developed.  
<sup>434</sup> For both taggers, the charge of the pion or proton is correlated with the production flavour  
<sup>435</sup> of the signal  $B^0$  meson. The response of the two taggers is combined into a common SS  
<sup>436</sup> response.

<sup>437</sup> In contrast, the OS taggers exploit the predominant production process of  $B$  mesons  
<sup>438</sup> via  $b\bar{b}$  quark pair production [56]. They partially reconstruct the decay of the *other*

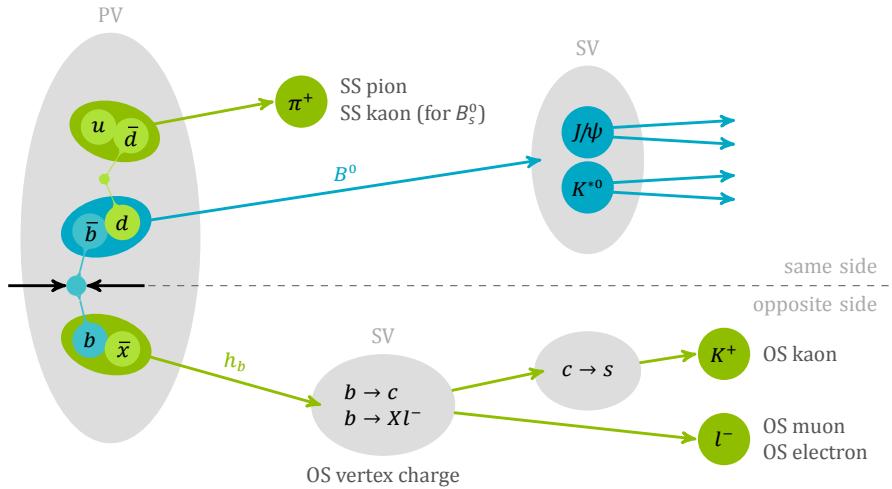


Figure 3.1 – Flavour tagging algorithms used in LHCb. In this cartoon, the signal channel is considered to be  $B^0 \rightarrow J/\psi K^{*0}$ .

439  $b$  hadron produced along with each reconstructed signal  $B$  meson, and infer its initial  
 440 flavour. In fact, the flavour of the signal  $B$  meson and the other  $b$  hadron produced  
 441 in the same collision are opposite. Several OS taggers have been developed in LHCb,  
 442 where the combination of the OS kaon (OSK), muon (OS $\mu$ ), electron (OSe), and vertex  
 443 charge (OSVtx) tagging algorithms represents the current standard OS combination. An  
 444 additional OS tagger, the OS charm tagger (OSCharm) [57], can be exploited, and can  
 445 be combined with the OS standard combination.

446 Given a reconstructed candidate, each flavour tagging algorithm provides a flavour tag  $d$   
 447 and a prediction  $\eta$  for the probability of the tag to be wrong. This mistag probability  $\eta$  is  
 448 defined in the range  $[0, 0.5]$  and is based on the output of multivariate classifiers, which  
 449 are trained on datasets of flavour-specific decays, and combine several kinematic and  
 450 geometric information on the tagging particle(s) and the event. The flavour tag takes the  
 451 values  $d = +1$  for an initial  $B^0$ ,  $d = -1$  for an initial  $\bar{B}^0$ , and  $d = 0$  when no tag could  
 452 be assigned; this happens, for example, if the tagging particle fails the selection criteria  
 453 of a given tagging algorithm, or if its trajectory lies outside the detector acceptance.

454 More details on flavour tagging at LHCb can be found in Refs. [58, 59, 60].

#### 455 Performance characteristics

The performance of flavour tagging algorithms can be characterised by different quantities. If  $N_U$  is the number of untagged candidates and  $N_W$  ( $N_R$ ) is the number of wrongly (rightly) tagged candidates, the *tagging efficiency* (*i.e.* the fraction of tagged candidates)

### Chapter 3. Flavour tagging

---

can be defined as

$$\varepsilon_{\text{tag}} = \frac{N_R + N_W}{N_R + N_W + N_U}. \quad (3.1)$$

The fraction of wrongly tagged candidates, or *mistag fraction*, is given by

$$\omega = \frac{N_W}{N_R + N_W}. \quad (3.2)$$

A non-zero mistag fraction dilutes the time-dependent asymmetries, reducing the experimental sensitivity to them. For instance, the measured decay rates for a  $B \rightarrow f$  decay and its  $CP$ -conjugate decay are

$$\frac{d\Gamma^{\text{meas}}}{dt} = (1 - \omega) \frac{d\Gamma}{dt} + \omega \frac{d\bar{\Gamma}}{dt}, \quad (3.3)$$

$$\frac{d\bar{\Gamma}^{\text{meas}}}{dt} = \omega \frac{d\Gamma}{dt} + (1 - \omega) \frac{d\bar{\Gamma}}{dt}. \quad (3.4)$$

As a consequence, the measured  $CP$  asymmetry is

$$A^{\text{meas}}(t) = \frac{\frac{d\bar{\Gamma}^{\text{meas}}}{dt} - \frac{d\Gamma^{\text{meas}}}{dt}}{\frac{d\bar{\Gamma}^{\text{meas}}}{dt} + \frac{d\Gamma^{\text{meas}}}{dt}} = (1 - 2\omega) \frac{\frac{d\bar{\Gamma}}{dt} - \frac{d\Gamma}{dt}}{\frac{d\bar{\Gamma}}{dt} + \frac{d\Gamma}{dt}} = D A^{\text{phys}}(t), \quad (3.5)$$

where  $A^{\text{phys}}$  is the physical (true)  $CP$  asymmetry. The quantity  $D = 1 - 2\omega$  is known as average *dilution*. If  $\omega = 0$  (perfect tagger), then  $D = 1$  and no asymmetry dilution occurs. If  $\omega = 0.5$  (random tagger), then  $D = 0$ , and it is not possible to measure the asymmetry anymore.

The quantity that can be interpreted as the figure of merit to optimise a tagging algorithm is the *effective tagging efficiency*, also called *tagging power*:

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}} (1 - 2\omega)^2 = \varepsilon_{\text{tag}} D^2. \quad (3.6)$$

Assuming that  $\varepsilon_{\text{eff}}$  is known without uncertainty, it can be shown that the statistical uncertainty on the physical asymmetry is given by

$$\sigma_{A^{\text{phys}}} = \frac{\sqrt{1 - A^{\text{meas}}^2}}{\sqrt{N\varepsilon_{\text{tag}}}(1 - 2\omega)} = \frac{\sqrt{1 - A^{\text{meas}}^2}}{\sqrt{N\varepsilon_{\text{eff}}}}, \quad (3.7)$$

where  $N$  is the total number of candidates. So, according to Eq. 3.7, the greater the tagging power, the smaller the resulting statistical uncertainty on the  $CP$  asymmetry. Instead of using an average mistag fraction or *probability*  $\omega$ , it is possible to exploit the mistag probability  $\eta$  estimated by the tagging algorithm. This probability  $\eta$  is evaluated for each  $B$  candidate individually, rather than being a global quantity. Usually,  $\eta$  needs to be *calibrated* via a function  $\omega(\eta)$  in order to return the true mistag probability (details

### 3.1. Flavour tagging algorithms

in Sec. 3.1). So, the tagging power can be rewritten as

$$\varepsilon_{\text{eff}} = \frac{1}{N} \sum_{i=1}^N D_i^2 = \frac{1}{N} \sum_{i=1}^N (1 - 2\omega(\eta_i))^2, \quad (3.8)$$

where  $\omega(\eta_i) = 0.5$  ( $D_i = 0$ ) for untagged candidates.

An example of dilution effect can be seen in Fig. 3.2, which shows how the measured amplitude of an asymmetry gets smaller for increasing values of  $\eta$ .

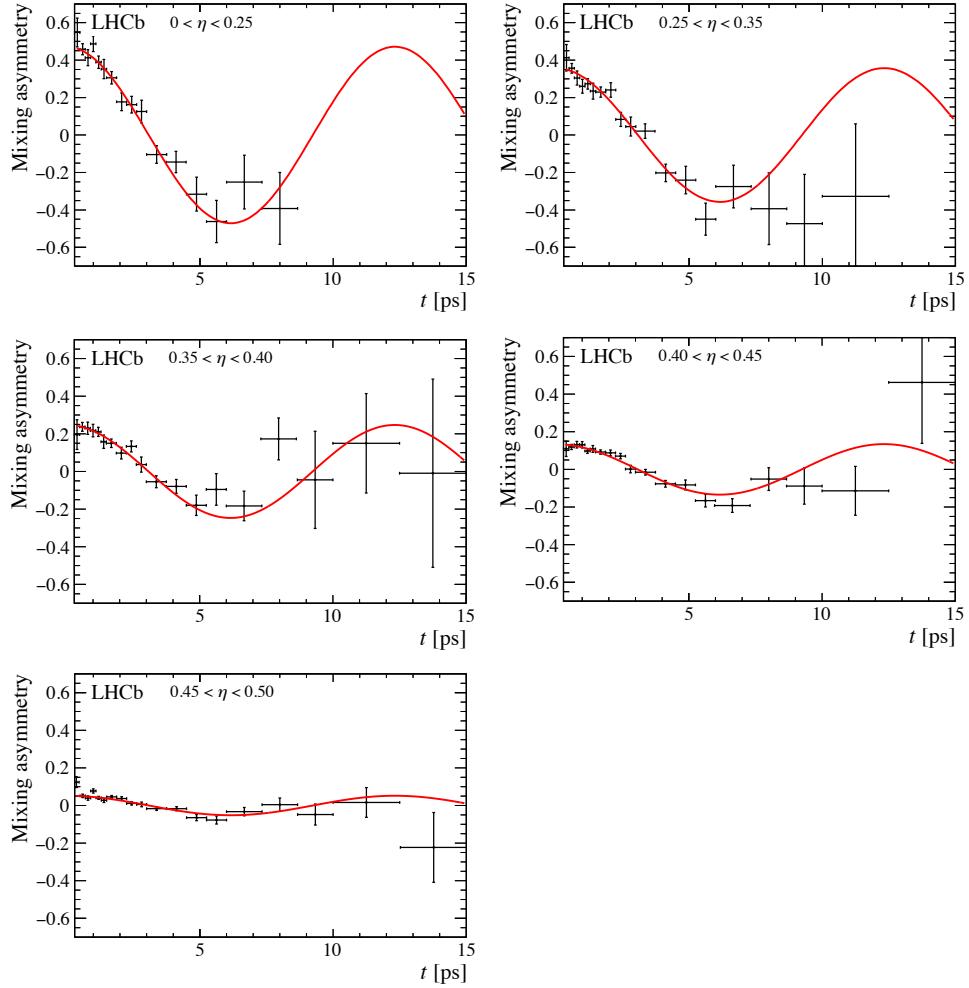


Figure 3.2 – Mixing asymmetry for SS-pion-tagged  $B_s^0 \rightarrow D_s^\mp \pi^\pm$  candidates in bins of increasing estimated mistag  $\eta$  [55].

<sup>463</sup> **Calibration of the tagging output**

<sup>464</sup> The output of the flavour tagging algorithms is the result of training multivariate classifiers  
<sup>465</sup> (MVA) using datasets of flavour-specific  $B$  decays, and transforming the classifier output  
<sup>466</sup> into mistag estimates  $\eta$  through regression. However, as the training and validation  
<sup>467</sup> samples are different from the signal sample used in the  $CP$  measurement (*e.g.* in terms  
<sup>468</sup> of trigger and selection criteria that affect the distribution of the MVA input features),  
<sup>469</sup> the output needs to be calibrated. Again, using control samples of flavour-specific decays,  
<sup>470</sup> calibration functions  $\omega(\eta)$  are obtained to transform the mistag estimate  $\eta$  of the algorithm  
<sup>471</sup> to the mistag probability  $\omega$  measured in the control sample.

A common choice for the calibration function is a linear function:

$$\omega(\eta) = p_0 + p_1 (\eta - \langle \eta \rangle). \quad (3.9)$$

<sup>472</sup> The use of the arithmetic mean  $\langle \eta \rangle$  of the  $\eta$  distribution aims at a decorrelation of  $p_0$   
<sup>473</sup> and  $p_1$ , hence a perfect calibration of the taggers would result in  $p_0 = \langle \eta \rangle$  and  $p_1 = 1$ .

The performance of the flavour taggers is not necessarily independent of the initial flavour of the  $B^0$ . The charged decay products, like the  $K^\pm$  mesons that are used by the OS kaon tagger, can have significantly different interaction rates with the detector material and therefore different reconstruction efficiencies. This can result in different tagging efficiencies  $\varepsilon_{\text{tag}}$  and mistag probabilities  $\omega$  for  $B^0$  and  $\bar{B}^0$ . These tagging asymmetries can dilute or enhance the observed raw asymmetry and need to be corrected for. The asymmetries of the mistag probability, *i.e.* the difference of the tagging calibration parameters  $p_0$  and  $p_1$  for initial  $B^0$  and  $\bar{B}^0$ , can be parameterised with two independent calibration functions:

$$\begin{aligned} \omega^{B^0}(\eta) &= p_0^{B^0} + p_1^{B^0} (\eta - \langle \eta \rangle), \\ \omega^{\bar{B}^0}(\eta) &= p_0^{\bar{B}^0} + p_1^{\bar{B}^0} (\eta - \langle \eta \rangle). \end{aligned} \quad (3.10)$$

Equivalently, we can parameterise the calibration parameters  $p_i$  (with  $i = 0, 1$ ) as

$$p_i^{B^0} = p_i + \frac{\Delta p_i}{2}, \quad p_i^{\bar{B}^0} = p_i - \frac{\Delta p_i}{2}. \quad (3.11)$$

The difference between the mistag of  $B^0$  and  $\bar{B}^0$  can be written as

$$\Delta\omega(\eta) = \omega^{B^0}(\eta) - \omega^{\bar{B}^0}(\eta) = \Delta p_0 + \Delta p_1 (\eta - \langle \eta \rangle). \quad (3.12)$$

<sup>474</sup> In this thesis, new models for the calibration functions are adopted instead of the standard  
<sup>475</sup> linear calibrations. These different parameterisations are called *Generalised Linear Models*  
<sup>476</sup> (GLM), and are implemented in the EPM (*Espresso Performance Monitor*) package [61].  
<sup>477</sup> As it will be explained in this section, these models allow a great flexibility to cope with

478 non-linearities, and solve technical issues that may occur in fits that make use of flavour  
 479 tagging. During my PhD, I worked to refine some of these models and to include them in  
 480 the fitting routines used for decay-time fits.

In general, a GLM of order  $N$  that relates the predicted mistag probability  $\eta$  to the calibrated probability  $\omega$  can be written as follows:

$$\omega(\eta) = g(h(\eta)) = g \left( g^{-1}(\eta) + \sum_{i=1}^N \left( p_i + \frac{d\Delta p_i}{2} \right) f_i(\eta) \right). \quad (3.13)$$

481 The functions  $f_i(\eta)$  are called *basis functions*, and they can be chosen as polynomials or  
 482 spline functions. The set on basis functions is automatically orthogonalised by the EPM  
 483 by using the Gram-Schmidt method [62]; this ensures that the corresponding calibration  
 484 parameters  $p_i$  and  $\Delta p_i$  are correlated as little as possible.

485 The parameter  $d$  is the tagging decision, which is incorporated into the model in order to  
 486 parameterise  $\omega(\eta)$  for the two possible flavours.

487 The function  $g$  is known as *link function*. Usually, this is chosen as the inverse of a  
 488 cumulative distribution function in order to map input values into the interval  $[0, 1]$ , such  
 489 that the output can be naturally interpreted as a probability.

For the  $B^0 \rightarrow D^\pm \pi^\mp$  analysis presented in this thesis, the adopted link function  $g$  is a *modified logistic function*, defined as

$$g(h) = \frac{1}{2(1 + e^h)}, \quad (3.14)$$

490 where  $h$  is defined in Eq. 3.13. This link function is built such that the calibrated mistag  
 491 probability is defined in the interval  $(0, 0.5)$ . This choice solves a numerical issue that  
 492 often occurs when standard link functions (*e.g.* identity or logistic) are adopted. In fact,  
 493 if  $\omega > 0.5$ , then an arbitrary prescription has to be taken (*e.g.*, label the candidate as  
 494 untagged, or flip the tagging decision and take  $1 - \omega$  as new calibrated mistag). If the  
 495 calibration parameters are free in a time-dependent fit, this choice has to be made during  
 496 the minimisation process, according to the values  $\omega$  takes at each iteration. This means  
 497 that the relative number of  $B$  and  $\bar{B}$ , or the relative number of tagged and untagged  
 498 candidates, may change during the fit, which leads to numerical instabilities due to  
 499 discontinuous changes in the likelihood function.

500 The EPM estimates the calibration parameters  $p_i$  and  $\Delta p_i$  via an unbinned maximum  
 501 likelihood fit called *binomial regression*; this is an improvement over traditional, binned  
 502 least-squares fits, which are affected by a systematic uncertainty due to the binning  
 503 choice.

504 **Combination of multiple taggers**

When more than one tagger is available per event, the tagging decisions and mistag probabilities provided by each tagger can be combined into a single decision and a single probability using the equations

$$p(b) = \prod_i \left( \frac{1}{2} - d_i \left( \frac{1}{2} - \eta_i \right) \right), \quad p(\bar{b}) = \prod_i \left( \frac{1}{2} + d_i \left( \frac{1}{2} - \eta_i \right) \right), \quad (3.15)$$

where  $p(\bar{b}/b)$  is the probability that the signal  $B^0$  contains a  $\bar{b}/b$ ,  $d_i$  is the decision taken by the  $i$ -th tagger and  $\eta_i$  is the predicted mistag probability of the  $i$ -th tagger. These probabilities are normalised as

$$P(\bar{b}) = \frac{p(\bar{b})}{p(\bar{b}) + p(b)}, \quad P(b) = 1 - P(\bar{b}). \quad (3.16)$$

505 If  $P(\bar{b}) > P(b)$  the combined tagging decision is  $d = +1$  and the final mistag probability  
506 is  $\eta = 1 - P(\bar{b})$ . Otherwise if  $P(b) > P(\bar{b})$  the combined tagging decision and the mistag  
507 probability are  $d = -1$  and  $\eta = 1 - P(b)$ .

508 Equation 3.15 is valid under the assumption that all taggers in the combination are  
509 independent. In the  $B^0 \rightarrow D^\pm \pi^\mp$  analysis presented in this thesis, the OS taggers are  
510 combined in a single OS combination, and the same is done for the SS taggers. Effects  
511 due to correlations among taggers within a combination are corrected for by calibrating  
512 the combined predicted mistag.

513 **3.2 Flavour tagging strategy for the  $B^0 \rightarrow D^\pm \pi^\mp$  time-  
514 dependent analysis**

515 In the  $B^0 \rightarrow D^\pm \pi^\mp$  analysis presented in this thesis, the OS combination (including the OS  
516 charm tagger) and the SS combination are used. The implementation of the OS algorithms  
517 used in the combination are the same as described in Refs. [56, 57]; the OS algorithms  
518 other than the OS charm tagger were built as neural networks trained on  $B^+ \rightarrow J/\psi K^+$   
519 Run 1 data, whereas the OS charm tagger was implemented with a BDT trained on a  
520 cocktail of simulated  $B^+ \rightarrow J/\psi K^+$ ,  $B^0 \rightarrow J/\psi K^{*0}$  and  $B_s^0 \rightarrow J/\psi \phi$  decays. The SS  
521 taggers have been reimplemented for this specific analysis by exploiting  $B^0 \rightarrow J/\psi K^{*0}$   
522 decays. The functional form of the tagging calibrations is studied in control samples of  
523 flavour-specific decays properly corrected to resemble the signal decay. The calibration  
524 parameters are determined directly in the decay-time-dependent fit of the signal described  
525 in Sec. 5.3; they are nuisance parameters of the likelihood function. Determining the  
526 calibration parameters from the data along with the  $CP$  observables is possible because  
527 the  $CP$  coefficients  $C_f$  and  $C_{\bar{f}}$  of Eqs. 1.35–1.38 are fixed in this analysis (to 1 and

### 3.2. Flavour tagging strategy for the $B^0 \rightarrow D^\pm \pi^\mp$ time-dependent analysis

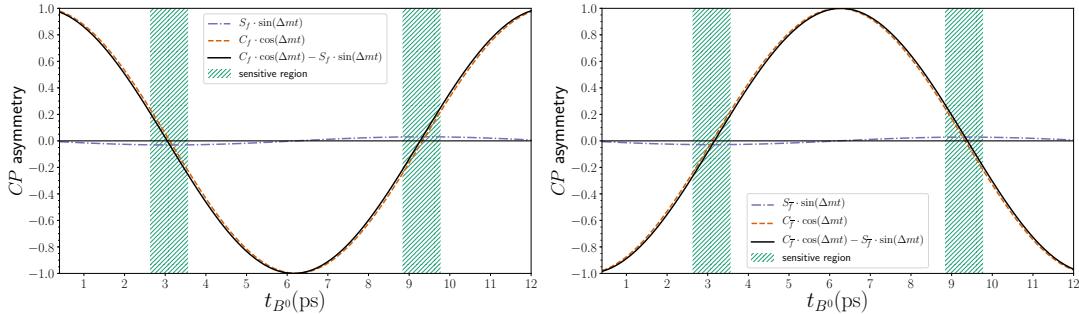


Figure 3.3 –  $\bar{B}^0$  versus  $B^0$  time-dependent asymmetries for the  $D^- \pi^+$  (left) and  $D^+ \pi^-$  (right) final states. The values of  $C_f$ ,  $C_{\bar{f}}$ ,  $S_f$  and  $S_{\bar{f}}$  are the ones used in simulation (see Appendix M). The sensitivity to  $S_f$  and  $S_{\bar{f}}$  is maximised in the intervals called “sensitive regions”, since the  $\sin(\Delta m)$  amplitude becomes of the same order of the  $\cos(\Delta m)$  amplitude, which is close to zero. In the outer regions, since  $C_f$  ( $C_{\bar{f}}$ ) is fixed to 1 (-1) in the fit, the mistag dilution (which depends on the flavour tagging calibration parameters) adapts to fit the  $\cos(\Delta m)$  amplitude, giving sensitivity to the calibration parameters.

528 –1 respectively). Hence, the cosine terms give sensitivity to the calibration parameters  
 529 independently of the sine terms, which are proportional to the  $S_f$  and  $S_{\bar{f}}$  coefficients. A  
 530 heuristic explanation is presented in Fig. 3.3. This strategy avoids any assumption on the  
 531 portability to the signal sample of the calibration parameters determined from the control  
 532 data. Such a strategy was studied extensively on simulation: the increase of the statistical  
 533 uncertainty of the  $S_f$  and  $S_{\bar{f}}$  coefficients given by the additional degrees of freedom of  
 534 the calibration parameters is smaller than the systematic uncertainties associated with  
 535 the calibration portability. Moreover, the use of the calibration parameters from the  
 536 control sample causes biases on  $S_f$  and  $S_{\bar{f}}$  of the order of their statistical uncertainty;  
 537 when letting the calibration parameters float in the fit, such biases are suppressed or  
 538 disappear, at the cost of a moderate increase of the statistical uncertainty. In addition,  
 539 while the precision of the OS tagger calibration from the control sample is similar to  
 540 the one from the signal sample, the calibration of the SS tagger derived from the signal  
 541 sample (Tab. 5.4) is much more precise than that from the control sample (Tab. 3.4).

542 In what follows, the study of the tagging calibration from the control sample is presented.  
 543 For all reasons discussed above, these studies are not meant for determining the calibration  
 544 parameters to use in the time fit to the signal data (usual strategy adopted in all flavour-  
 545 tagged time-dependent analyses), but they serve the purpose of: i) determining the best  
 546 functional form of the calibration functions to be used in the fit to the signal; ii) having  
 547 some reference values for the calibration parameters for a comparison with those extracted  
 548 from the signal.

549 The calibration for the OS combination are determined using  $B^+ \rightarrow D^0 \pi^+$  decays, as  
 550 described in Sec. 3.2.1. The SS pion and the SS proton taggers were developed using

551  $B^0 \rightarrow D^\mp\pi^\pm$  data and assuming negligible  $CP$  violation. The use of these algorithms in  
552 this analysis could bias the measurement. Therefore, the SS taggers are retrained using  
553  $B^0 \rightarrow J/\psi K^{*0}$  decays. The calibration of the SS combination is described in Sec. 3.2.2.

#### 554 3.2.1 Calibration of the opposite-side tagger combination

##### 555 Data sample selection

556 The calibration parameters of the OS tagger combination (namely the combination of  
557 the OS electron, muon, kaon, vertex charge, and charm algorithms) are determined using  
558  $B^+ \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$  candidates reconstructed in  $3\text{ fb}^{-1}$  of data. Such a control decay  
559 mode provides very high statistics (more than 300k OS-tagged signal candidates) and is  
560 very similar to the signal decay  $B^0 \rightarrow D^-\pi^+$ .

561 Candidate  $B^\pm \rightarrow D^0\pi^\pm$  decays are selected through the `B2D0PiD2HHBeauty2CharmLine`  
562 stripping line, versions `S21r1` (2011 data) and `S21` (2012 data), of the `BhadronCompleteEvent`  
563 stream. The  $B^+$  candidates are required to be TOS, *i.e.* to trigger on `Hlt1TrackAllL0Decision`  
564 at the HLT1 stage, and at least one among `Hlt2Topo2BodyBBDTDecision`, `Hlt2Topo3BodyBBDTDecision`,  
565 and `Hlt2Topo4BodyBBDTDecision` at HLT2. The additional requirements listed in Ta-  
566 ble 3.1 are applied to further suppress backgrounds and enhance the signal purity.

567 A fit to the mass distribution of  $B^+$  candidates is done to calculate `sWeights`, used in  
568 the subsequent steps of the analysis to subtract the backgrounds surviving the selection.  
569 This fit is described in details in Appendix A.1.

570 Event-by-event weights are calculated to equalise the  $B^\pm \rightarrow D^0\pi^\pm$  and  $B^0 \rightarrow D^\pm\pi^\mp$   
571 distributions of the variables on which the tagging calibration can depend. The procedure  
572 and the results of this reweighting are reported in Appendix A.2. Additionally, the  
573 number of  $B^+$  and  $B^-$  candidates are made equal in the sample to avoid any spoil of the  
574 calibration parameters due to a  $B^+/B^-$  production asymmetry or a detection asymmetry.  
575 All these weights, along with `sWeights` [38], are applied during the calibration procedure.

##### 576 Calibration

577 The calibration of the estimated mistag  $\eta$  is performed on the fully reweighted  $B^+ \rightarrow D^0\pi^+$   
578 dataset. A GLM model with NSpline basis function [61] is adopted. The projection of the  
579 fitted calibration function over the  $B^+ \rightarrow D^0\pi^+$  dataset is shown in Fig. 3.4, whereas the  
580 fitted calibration parameters are listed in Table 3.2. The number of free parameters in  
581 the adopted GLM model (10) has been chosen in order to have satisfactory goodness-of-fit  
582 (GOF) metrics (details in Appendix A.3).

### 3.2. Flavour tagging strategy for the $B^0 \rightarrow D^\pm \pi^\mp$ time-dependent analysis

Table 3.1 – Selection requirements for the  $B^\pm \rightarrow D^0 \pi^\pm$  candidates.

Description	Variable	Requirement
Bachelor track		
muon identification criteria	IsMuon	= 0
ghost probability	$p_{\text{ghost}}$	< 0.1
quality of track	$\chi^2_{\text{track}}/\text{ndof}$	< 2
$D^0$ daughter tracks		
$\ln L_K - \ln L_\pi$	PIDK	> -2 (kaon), < 8 (pion)
ghost probability	$p_{\text{ghost}}$	< 0.1
quality of track	$\chi^2_{\text{track}}/\text{ndof}$	< 2.5
$D^0$ candidate		
invariant mass	$m_{K\pi}$	$m_{K\pi} \in [1830, 1904] \text{ MeV}/c^2$
$B^+$ candidate		
decay time	$\tau_{B^+}$	$\tau_{B^+} \in [0.2, 15] \text{ rmps}$
minimum IP $\chi^2$ w.r.t. PV	MIN $\chi^2_{\text{IPPV}}$	< 15

Table 3.2 – Fitted OS calibration parameters on the  $B^+ \rightarrow D^0 \pi^+$  reweighted dataset.

Parameter	Fitted value
$p_0$	$-0.136 \pm 0.019$
$p_1$	$-0.006 \pm 0.022$
$p_2$	$-0.0107 \pm 0.0083$
$p_3$	$-0.5 \pm 0.10$
$p_4$	$-0.85 \pm 0.46$
$\Delta p_0$	$-0.129 \pm 0.038$
$\Delta p_1$	$0.042 \pm 0.045$
$\Delta p_2$	$-0.020 \pm 0.017$
$\Delta p_3$	$0.42 \pm 0.21$
$\Delta p_4$	$1.91 \pm 0.92$

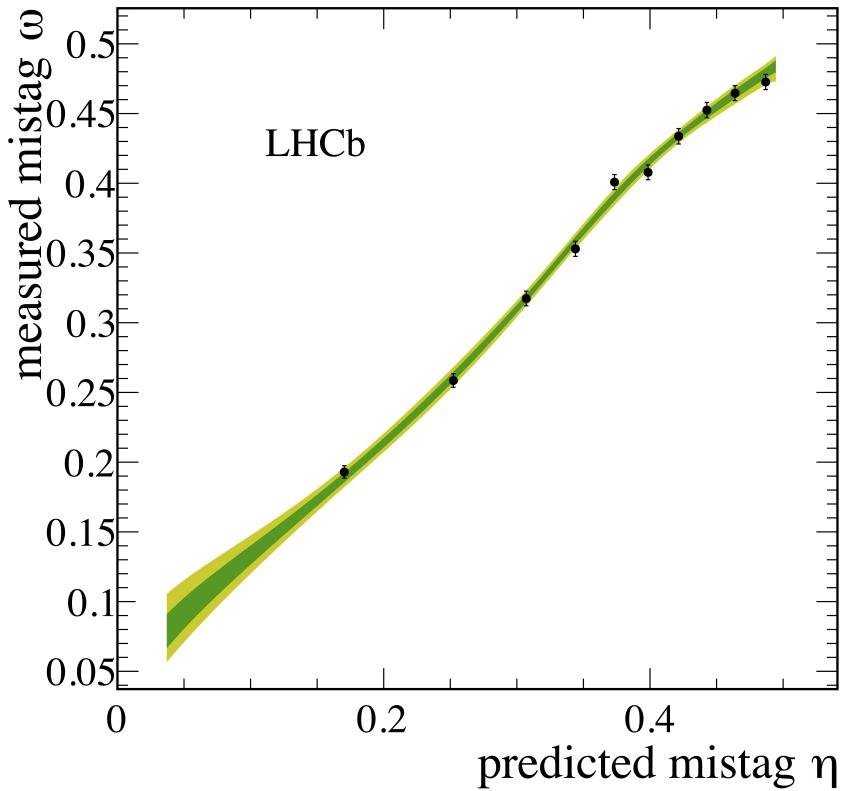


Figure 3.4 – Mistag  $\omega$  measured in bins of predicted mistag  $\eta$  for reweighted  $B^+ \rightarrow D^0\pi^+$  candidates (data points) and fitted calibration function. The green (yellow) band indicate the 68% (95%) confidence interval on the calibration function.

**583 Calibration portability**

584 The aim of the calibration is to return a mistag  $\omega$  as close as possible to the *true* mistag,  
 585 which would be given by a *true calibration*. The latter is not defined for  $B^0 \rightarrow D^-\pi^+$   
 586 decays in data, but it is possible to estimate it for  $B^0 \rightarrow D^-\pi^+$  decays on MC. In fact,  
 587 since the true flavour of the  $B^0$  meson is known in MC, this true MC calibration can be  
 588 done in the same way as  $B^+ \rightarrow D^0\pi^+$ , where the true flavour is given by the  $B$  charge.

589 This  $B^0 \rightarrow D^-\pi^+$  calibration is performed after equalising the number of  $B^0$  and  $\bar{B}^0$  in  
 590 the sample, in order to disentangle tagging asymmetries from  $CP$  violation and production  
 591 asymmetries.

592 The  $B^+ \rightarrow D^0\pi^+$  MC calibration is performed in exactly the same way as described in  
 593 Sec. 3.2.1, except that no *sWeights* are considered, since only true MC signal decays are  
 594 used.

### 3.2. Flavour tagging strategy for the $B^0 \rightarrow D^\pm \pi^\mp$ time-dependent analysis

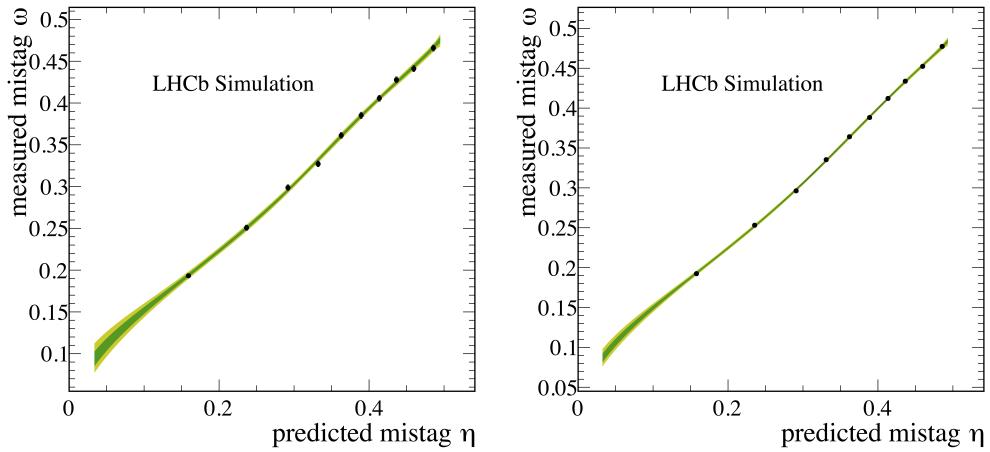


Figure 3.5 – Mistag  $\omega$  measured in bins of predicted mistag  $\eta$  for reweighted  $B^+ \rightarrow D^0\pi^+$  (left) and  $B^0 \rightarrow D^-\pi^+$  (right) candidates (data points) and fitted calibration functions. The green (yellow) band indicates the 68% (95%) confidence interval on the calibration functions.

Table 3.3 – Comparison between the fitted OS tagging calibration parameters using truth-matched  $B^+ \rightarrow D^0\pi^+$  and  $B^0 \rightarrow D^-\pi^+$  MC decays. The discrepancy in each parameter is computed assuming independent datasets.

Parameter	$B^+ \rightarrow D^0\pi^+$	$B^0 \rightarrow D^-\pi^+$	Discrepancy ( $\sigma$ )
$p_0$	$-0.065 \pm 0.011$	$-0.0996 \pm 0.0066$	2.70
$p_1$	$-0.190 \pm 0.012$	$-0.1492 \pm 0.0077$	-2.84
$p_2$	$-0.0105 \pm 0.0044$	$-0.0191 \pm 0.0029$	1.63
$p_3$	$-0.295 \pm 0.054$	$-0.234 \pm 0.036$	-0.93
$p_4$	$-0.42 \pm 0.26$	$-0.14 \pm 0.20$	-0.85
$\Delta p_0$	$-0.059 \pm 0.022$	$-0.058 \pm 0.013$	-0.03
$\Delta p_1$	$0.044 \pm 0.024$	$0.030 \pm 0.015$	0.46
$\Delta p_2$	$-0.0012 \pm 0.0088$	$-0.0126 \pm 0.0058$	1.08
$\Delta p_3$	$-0.08 \pm 0.11$	$-0.046 \pm 0.073$	-0.25
$\Delta p_4$	$-0.34 \pm 0.53$	$-0.29 \pm 0.39$	-0.08

595 The two calibrations using the  $B^0 \rightarrow D^-\pi^+$  and  $B^+ \rightarrow D^0\pi^+$  MC samples are shown in  
 596 Fig. 3.5 and compared in Table 3.3. A more robust comparison is obtained from a  $\chi^2$   
 597 function describing the discrepancy between the two calibrations by taking the covariance  
 598 matrices into account. The overall discrepancy (corresponding to the  $\chi^2$  minimum) is  
 599 around 2  $\sigma$ .

Table 3.4 – Fitted SS calibration parameters obtained on the  $B^0 \rightarrow J/\psi K^{*0}$  data sample (calibration subsample).

$p_0$	$p_1$	$\Delta p_0$	$\Delta p_1$
$-0.091 \pm 0.059$	$-0.027 \pm 0.065$	$0.034 \pm 0.084$	$0.032 \pm 0.094$

### 600 3.2.2 Calibration of the same-side tagger combination

601 As described in Ref. [55], the SS pion and proton taggers were both trained on the  
 602 2012 data sample of  $B^0 \rightarrow D^\mp \pi^\pm$  decays. As the effect of  $CP$  violation was neglected  
 603 during the training the algorithms and the underlying MVAs cannot be blindly used  
 604 when measuring  $CP$  violation in the same decay channel. Thus,  $B^0 \rightarrow J/\psi K^{*0}$  decays  
 605 are chosen instead, as they represent a flavour-specific  $B^0$  decay with a large signal yield  
 606 of about 350000 candidates in 2012 data.

607 Once the SS pion and proton taggers are implemented, they are combined into a single  
 608 SS combination as described in Sec. 3.1.

### 609 Calibration

610 The calibration is performed on a  $B^0 \rightarrow J/\psi K^{*0}$  data subsample that is not used for the  
 611 SS pion and proton training. A GLM model having a first order polynomial is chosen as  
 612 basis function and a modified logistic function (Eq. 3.14) is used as link. The number  
 613 of free parameters in this model (4) is tuned in order to have satisfactory goodness-  
 614 of-fit (GOF) metrics. Together with the  $sWeights$ , additional weights to correct the  
 615  $B^0 \rightarrow J/\psi K^{*0}$  data to resemble the  $B^0 \rightarrow D^\pm \pi^\mp$  data are applied during the calibration.  
 616 The resulting calibration parameters are listed in Table 3.4 and a graphical representation  
 617 of the calibration is presented in Fig. 3.6.

### 618 Calibration portability

619 In the same way as for the OS taggers (Sec. 3.2.1), the portability of the SS tagging  
 620 calibration is checked on Monte Carlo. For  $B^0 \rightarrow D^- \pi^+$  the calibration is performed  
 621 using the true flavour of the  $B^0$  meson after equalising the number of  $B^0$  and  $\bar{B}^0$  in the  
 622 sample, in order to disentangle tagging asymmetries from  $CP$  violation and production  
 623 asymmetries. Also on  $B^0 \rightarrow J/\psi K^{*0}$  the true flavour of the  $B^0$  meson is used for the  
 624 calibration, and no  $sWeights$  are needed, since only the true MC signal decays are used.

625 The two calibrations using the  $B^0 \rightarrow D^- \pi^+$  and  $B^0 \rightarrow J/\psi K^{*0}$  Monte Carlo samples  
 626 are shown in Fig. 3.7 and compared in Table 3.5. A full comparison that takes into  
 627 account the correlation between the parameters is obtained from a  $\chi^2$  test similar to  
 628 the one described in Sec. 3.2.1. The agreement is around  $0.1\sigma$ . Even though this test

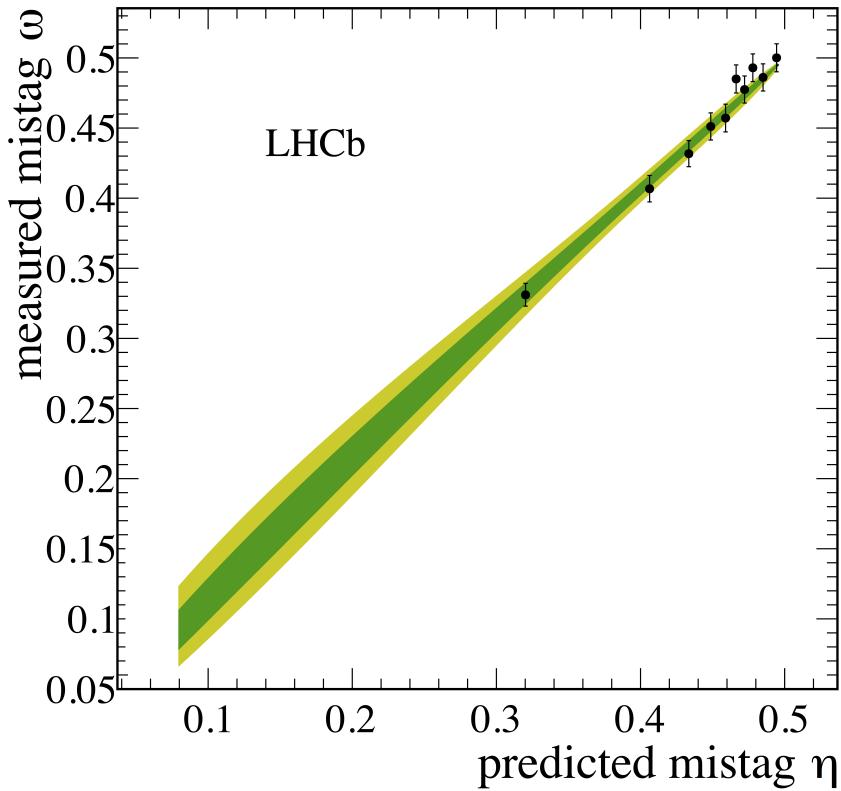


Figure 3.6 – Mistag  $\omega$  measured in bins of predicted mistag  $\eta$  for reweighted  $B^0 \rightarrow J/\psi K^{*0}$  candidates (data points) and fitted calibration function. The green (yellow) band indicate the 68% (95%) confidence interval on the calibration function.

629 doesn't hint to issues of portability between the decay modes, the same strategy used  
 630 for the OS calibrations is followed, i.e fitting the parameter directly in data with the  
 631  $CP$  asymmetries. This is motivated by the fact that the  $B^0 \rightarrow D^- \pi^+$  signal sample has  
 632 much more sensitivity to determine the parameters than the  $B^0 \rightarrow J/\psi K^{*0}$  sample. In  
 633 addition, with this approach no systematic related to calibration portability is necessary,  
 634 consistent with the OS tagger treatment.

### 635 3.3 Optimisation of the opposite-side electron tagger

636 The performance of the flavour tagging algorithms depends on the data taking conditions,  
 637 in particular the centre-of-mass energy of the  $pp$  collision.

638 On one hand, the tagging power of the SS taggers shows an increase on Run 2 data  
 639 as compared to Run 1, thanks to either a higher tagging efficiency (SS $\pi$  and SS $p$ ) or a

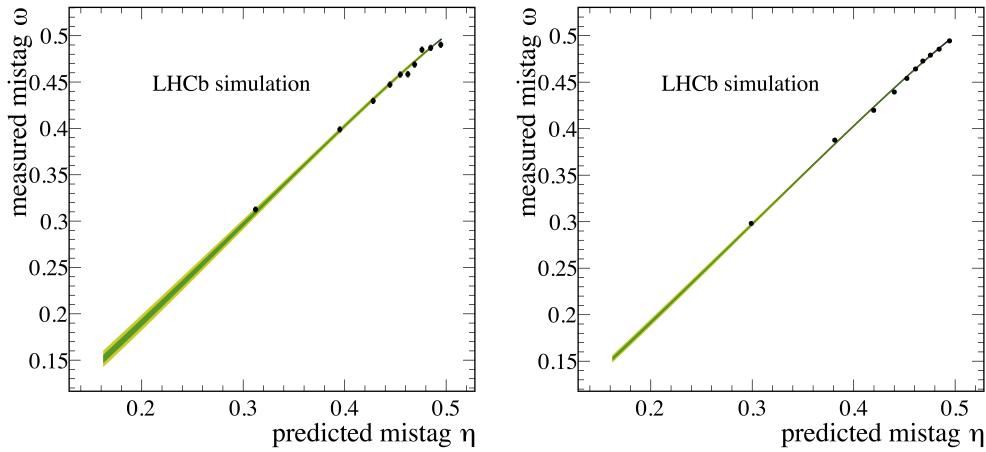


Figure 3.7 – Mistag  $\omega$  measured in bins of predicted mistag  $\eta$  for reweighted  $B^0 \rightarrow J/\psi K^{*0}$  (left) and  $B^0 \rightarrow D^- \pi^+$  (right) candidates (data points) and fitted calibration functions. The green (yellow) band indicate the 68% (95%) confidence interval on the calibration functions.

Table 3.5 – Comparison between the fitted SS tagging calibration parameters using truth-matched  $B^0 \rightarrow J/\psi K^{*0}$  and  $B^0 \rightarrow D^- \pi^+$  MC decays. The discrepancy in each parameter is computed assuming independent datasets.

Parameter	$B^0 \rightarrow J/\psi K^{*0}$	$B^0 \rightarrow D^- \pi^+$	Discrepancy ( $\sigma$ )
$p_0$	$-0.016 \pm 0.017$	$-0.019 \pm 0.008$	-0.19
$p_1$	$0.063 \pm 0.021$	$0.060 \pm 0.010$	-0.14
$\Delta p_0$	$-0.029 \pm 0.033$	$-0.027 \pm 0.015$	0.04
$\Delta p_1$	$-0.026 \pm 0.041$	$0.015 \pm 0.019$	0.90

640 lower mistag rate (SSK). This is due to the higher boost of the  $b\bar{b}$  quark pair at 13 TeV,  
 641 which makes the momentum spectrum of  $B$  mesons and fragmentation tracks harder, and  
 642 increases the acceptance of the fragmentation tracks.

643 On the other hand, the tagging power of the existing OSe, OS $\mu$ , and OS $K$  taggers  
 644 decreases on Run 2 data. The reason for this degradation is mainly due to the higher  
 645 track multiplicity, which increases the probability to have a wrong tag decision. Moreover,  
 646 because of the different Run 2 kinematics, the criteria to select the tagging particles are  
 647 no longer optimal, thus giving a lower tagging efficiency.

648 The performance of the OSCharm and OSVtx algorithms is, on average, compatible or  
 649 better on Run 2 as compared to Run 1.

650 In this section, the reoptimisation of the OSe tagger is presented. This reoptimisation is  
 651 performed both on Run 2 data, in order to recover the observed loss in tagging power, and

### 3.3. Optimisation of the opposite-side electron tagger

---

on Run 1 data, to further improve the already existing algorithm. This reoptimisation consists of two main steps. First, selection criteria are applied to select electron-like particles, yielding a sample of  $B$  signal candidates with a low average tagging power. Then, a BDT classifier is applied to discriminate between  $B$  candidates with right and wrong tag decisions for each selected track. Finally, for each  $B$  candidate, the BDT output for the track with the highest transverse momentum is converted into a predicted mistag probability.

A similar approach is followed for the development of the OS $K$  and OS $\mu$  taggers. In this case, the reoptimisation on Run 1 data does not show any gain in performance, whereas some significant gain is found on Run 2 data.

#### 3.3.1 Sample definition

The OSe algorithm is developed in a data-driven fashion by using *sWeighted* samples of  $B^\pm \rightarrow J/\psi K^\pm$  decays. The full Run 1 dataset (2011+2012) is used to optimise the algorithm on Run 1 conditions, whereas the 2016 dataset is exploited to optimise the tagger on Run 2 conditions. An alternative optimisation on *sWeighted* 2016  $B^\pm \rightarrow D^0\pi^\pm$  data is performed in parallel. The motivation for this is to cross-check the Run 2 implementation on an independent decay mode, which is characterised by a different kinematics than that of  $B^\pm \rightarrow J/\psi K^\pm$ . Hereafter, the OSe tagger optimised on Run 1  $B^\pm \rightarrow J/\psi K^\pm$  data will be indicated as “Run 1 new” version, in order to distinguish it from the previous “Run 1 old” version introduced in Ref. [56], which was based on simple selection criteria and a neural network for the mistag estimation. The OSe tagger optimised on Run 2  $B^\pm \rightarrow J/\psi K^\pm$  data, and Run 2  $B^\pm \rightarrow D^0\pi^\pm$  data will be denoted as “Run 2 B2CC” and “Run 2 B2OC” versions, respectively. Moreover, it is understood that the tunings of all the PROBNN features mentioned in this section are MC12TuneV2 for Run 1 data and MC15TuneV1 for Run 2 data.

Each dataset is divided in four subsamples:

- the first subset, including  $\sim 25\%$  of the total data, is used for the optimisation of the electron preselection (Sec. 3.3.2);
- the second subset (*training sample*), including  $\sim 50\%$  of the total data, is adopted as training set for the BDT classifier used for the predicted mistag estimation (Sec. 3.3.3). This sample is also used for tuning some *hyperparameters* (number of trees, maximum depth) which define the BDT classifier.
- the third and the fourth subsets (*evaluation set 1* and *2*), each including  $\sim 12.5\%$  of the total data, are adopted together as test sets to check for overtraining (Sec. 3.3.4). The evaluation set 1 is also used to calibrate the obtained tagger, which is then applied to the second evaluation set in order to measure the performance; the

688 procedure is then repeated by swapping the two samples (*two-fold validation*).

### 689 3.3.2 Preselection optimisation

690 Electron-like particles are selected by means of a set of requirements. The reconstructed  
691 tracks must not be associated to the  $B$  signal decay tree, and must not have hits in  
692 the muon detector in order to exclude muons (already exploited by the OS $\mu$  algorithm).  
693 Moreover, these tracks have to be of type long, lie in the ECAL acceptance, and have  
694 sufficient reconstruction quality ( $\chi^2/\text{ndof} < 3$ ). Also, the inverse of the rigidity  $e/p$  has  
695 to be comprised between 0.85 and 2, and the charge deposited in the VELO detector  
696 must be smaller than 1.4. Finally, tracks where the fit for the track IP with respect to  
697 the primary vertex did not converge are excluded.

A further selection is applied in order to enhance the average tagging power of the resulting sample, which is defined as

$$\langle \varepsilon_{\text{eff}} \rangle = \varepsilon_{\text{tag}} \left[ 1 - 2 \frac{\sum_{i=1}^N w_i f_R^i}{\sum_{i=1}^N w_i (f_R^i + f_W^i)} \right]^2, \quad (3.17)$$

698 where  $w_i$  are the *sWeights*, while  $f_R^i$  ( $f_W^i$ ) is the fraction of particles giving the right  
699 (wrong) flavour for the  $i$ th  $B$  candidate, with  $f_R^i + f_W^i = 1$  for every candidate.

700 The expression of Eq. 3.17 is taken as the figure of merit to maximise during the selection  
701 optimisation. This maximisation is performed numerically by using gradient boosted  
702 regression trees to model  $\langle \varepsilon_{\text{eff}} \rangle$  as a function of the applied cuts [63]. The cuts are  
703 optimised separately for the Run 1 new, Run 2 B2CC and Run 2 B2OC algorithms; in  
704 all cases, about 25% of the available data for each sample is used.

705 The resulting, optimised requirements are reported in Table 3.6, while the convergence  
706 plots of the minimisation are shown in Fig. 3.8.

707 After the optimisation, the performance of the selection (including the average tagging  
708 power defined in by Eq. 3.17) is evaluated on the remaining 75% of data for each sample,  
709 yielding the results shown in Table 3.7.

### 710 3.3.3 BDT classifier implementation

711 The selection described in Sec. 3.3.2 is applied on the remaining part of the data ( $\sim 75\%$ )  
712 used by each OSe implementation. The BDT classifier is trained to identify  $B$  candidates  
713 as correctly or incorrectly tagged. The list of the features considered to build the BDTs  
714 are reported in Table 3.8. The distributions of the input features for the two possible

### 3.3. Optimisation of the opposite-side electron tagger

Table 3.6 – Optimised requirements for the preselection of tracks used by the OSe algorithm.  $p_{\text{ghost}}$  is the probability for the track to be a fake combination of hits. IPPU denotes the impact parameter with respect to the pile-up vertex in the event, which might be reconstructed in the event in addition to the nominal PV.  $\Delta\phi$  is the difference in azimuthal angle between the track and the signal  $B$  candidate.

Requirement	Run 1 new	Run 2 B2CC	Run 2 B2OC
$p_{\text{ghost}} <$	0.861	0.843	0.348
PROBNN $\pi <$	0.934	0.983	0.980
PROBNN $p <$	0.719	0.271	0.732
PROBNN $k <$	0.765	0.695	0.954
PROBNN $e >$	0.0610	0.243	0.040
PROBNN $\mu <$	0.938	0.158	0.263
PID $e >$	4.555	4.333	-0.691
$p_T > (\text{MeV}/c)$	1132	1403	1263
$p > (\text{MeV})$	3114	5035	2246
$\sigma_{\text{IP}}/\text{IP} >$	0.0200	0.0420	1.410
$\sigma_{\text{IPPU}}/\text{IPPU} >$	12.101	9.335	2.758
min $\Delta\phi >$	0.00803	0.0167	0.0299

Table 3.7 – Performance of the preselection (OSe algorithm) applied on the data not used for the preselection optimisation ( $\sim 75\%$  of the total dataset for each sample). The average tagging power  $\langle \varepsilon_{\text{eff}} \rangle$  is defined in Eq. 3.17.

Algorithm	$\varepsilon_{\text{tag}} (\%)$	$\langle \omega \rangle (\%)$	$\langle \varepsilon_{\text{eff}} \rangle (\%)$
Run 1 new	$3.440 \pm 0.019$	$33.31 \pm 0.27$	$0.383 \pm 0.007$
Run 2 B2CC	$2.514 \pm 0.017$	$33.50 \pm 0.32$	$0.274 \pm 0.006$
Run 2 B2OC	$3.664 \pm 0.024$	$34.32 \pm 0.32$	$0.360 \pm 0.006$

values of the target are shown in Figs. 3.9, 3.10 and 3.11 for the Run 1 new, Run 2 B2CC and Run 2 B2OC samples, respectively. The Pearson correlation coefficients between the input features are reported in Fig. 3.12.

The BDT classifier consists of an ensemble of 300 gradient-boosted decision trees [64], where each tree can have a maximum depth of 3. The objective of the classifier is a binary logistic loss function plus a quadratic regularisation term to control model complexity (with regularisation parameter  $\lambda = 1$ ). Some hyperparameters were tested by means of a cross-validation+bootstrapping method on the training set, as described in Appendix B. The importance (or F score) of each feature, defined as the total number of times a feature is chosen as split node by any tree in the BDT ensemble, is presented in Fig. 3.13, while the *partial dependence* of the predicted mistag  $\eta$  (on the training set) as a function of each input feature is shown in Appendix B. The receiver operating characteristic (ROC) curves, which report the rate of true right-tagged candidates as a function of the rate

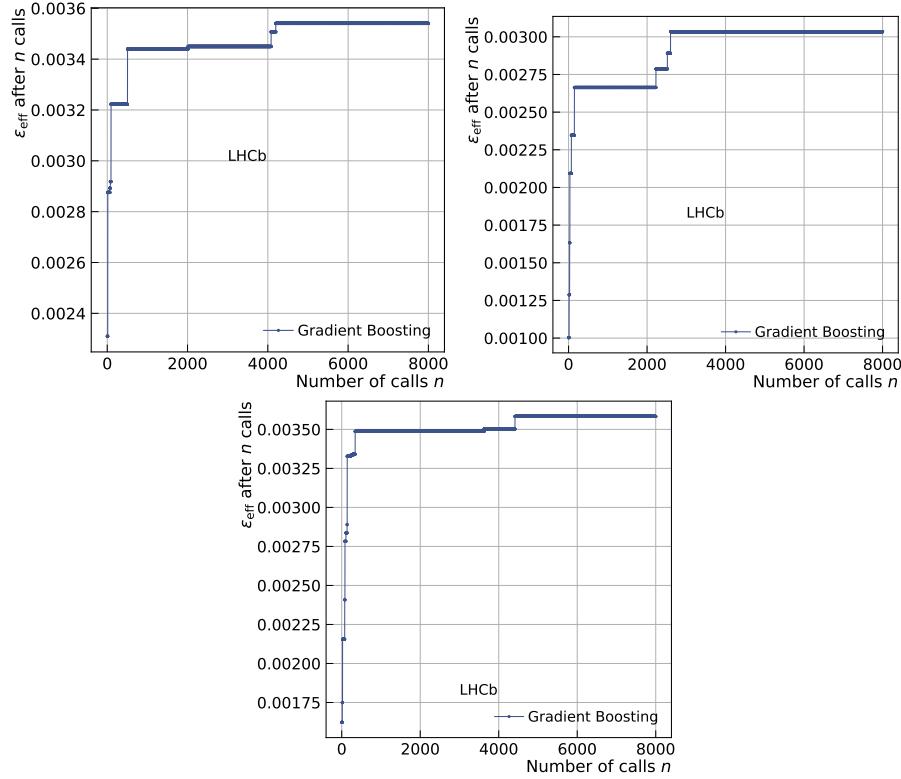


Figure 3.8 – Maximised value of the average tagging power as a function of the gradient boosted regression tree algorithm iteration for the Run 1 new (top left), Run 2 B2CC (top right) and Run 2 B2OC (bottom) implementations of the OSe tagger.

728 of false right-tagged candidates, are reported in Fig. 3.14. The feature selection, BDT  
 729 training and feature importance evaluation chain has been repeated iteratively in order  
 730 to exclude highly-correlated and poorly-important features, until the BDT performance  
 731 started to degrade significantly.

For each candidate, the BDT predicts the probability  $P$  that such candidates is correctly tagged. In order to obtain a mistag probability  $\eta$ , the following transformation is applied on both  $P$  and tagging decision  $d$ :

$$(\eta, d) = \begin{cases} (P, -d) & \text{if } P \leq 0.5 \\ (1 - P, d) & \text{otherwise} \end{cases} \quad (3.18)$$

732 The distributions of  $\eta$  for training and test samples, splitted per target value, are shown  
 733 in Fig. 3.15.

### 3.3. Optimisation of the opposite-side electron tagger

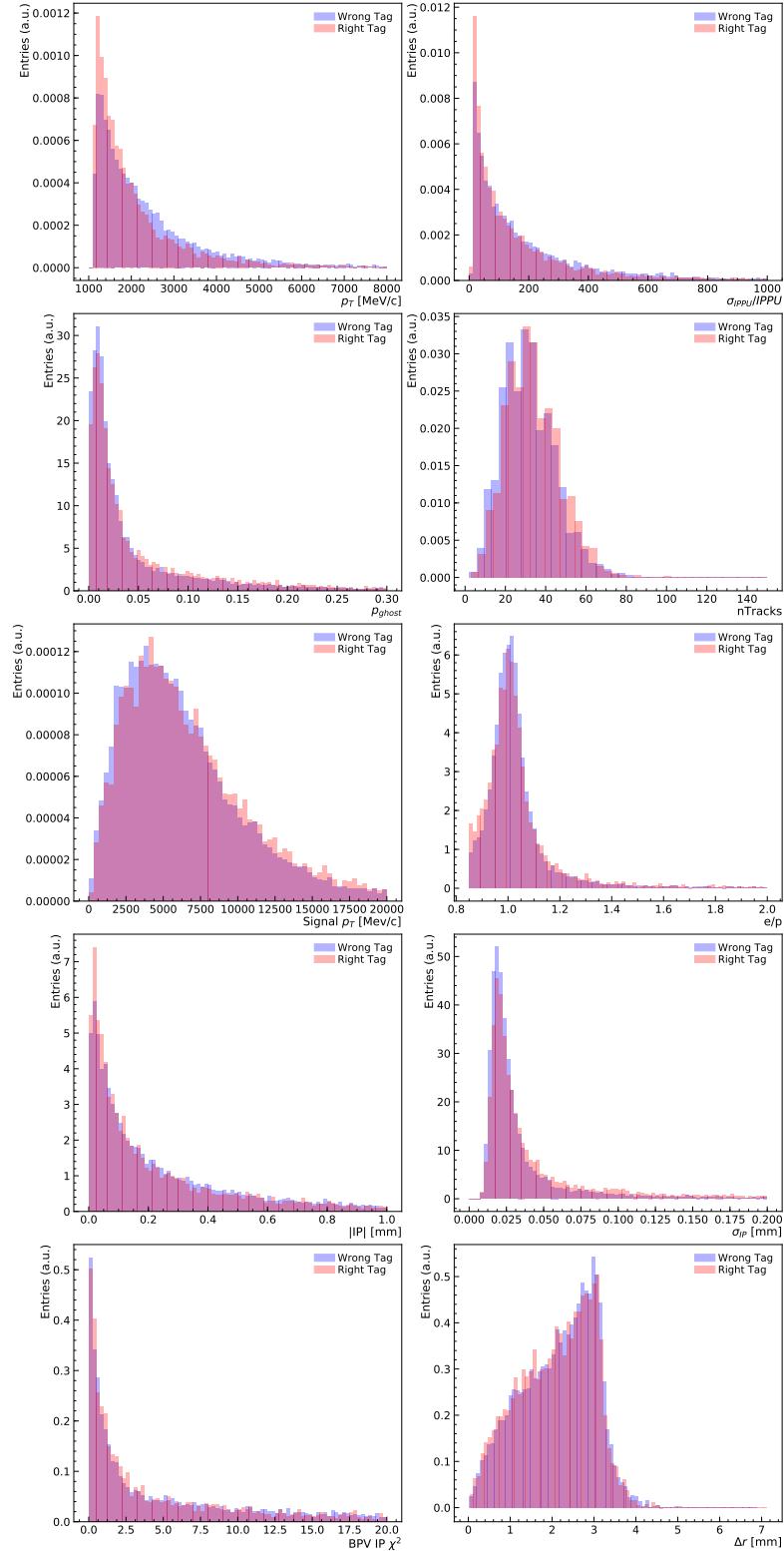


Figure 3.9 – Distributions (for the *sWeighted*, Run 1  $B^\pm \rightarrow J/\psi K^\pm$  sample) of the input features of the BDT classifier, for candidates with a right (red) and wrong (blue) decision from the OSe tagger.

### Chapter 3. Flavour tagging

---

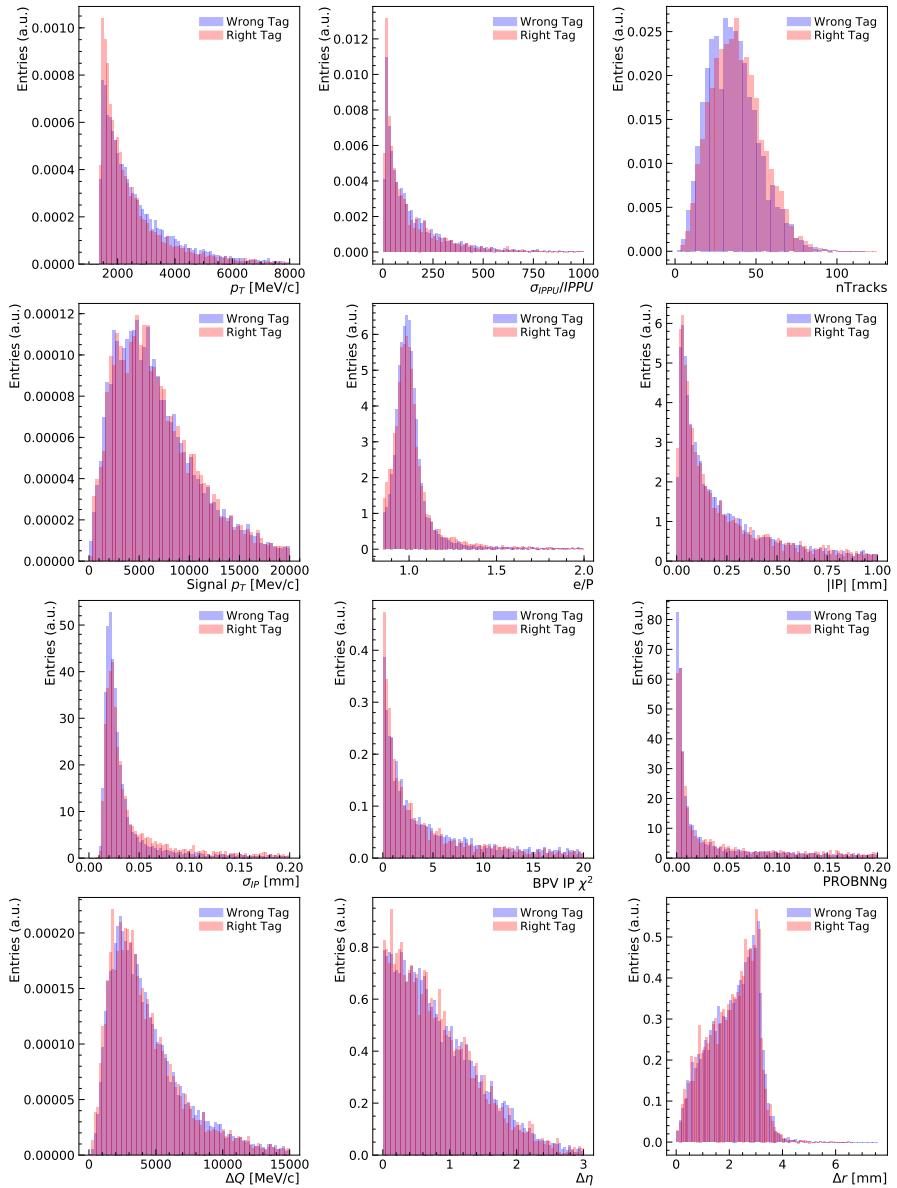


Figure 3.10 – Distributions (for the *sWeighted*, Run 2  $B^\pm \rightarrow J/\psi K^\pm$  sample) of the input features of the BDT classifier, for candidates with a right (red) and wrong (blue) decision from the OSe tagger.

### 3.3. Optimisation of the opposite-side electron tagger

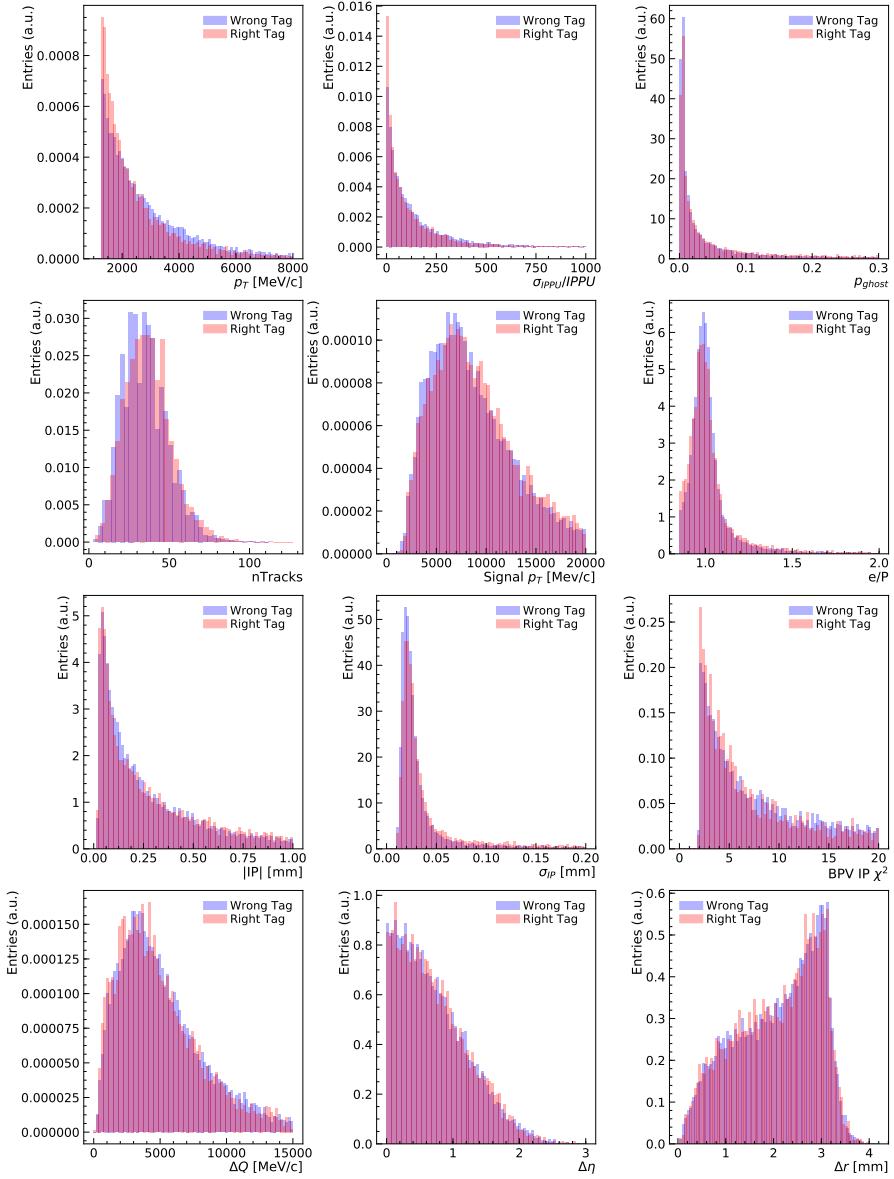


Figure 3.11 – Distributions (for the *sWeighted*, Run 2  $B^\pm \rightarrow D^0\pi^\pm$  sample) of the input features of the BDT classifier, for candidates with a right (red) and wrong (blue) decision from the OSe tagger.

### Chapter 3. Flavour tagging

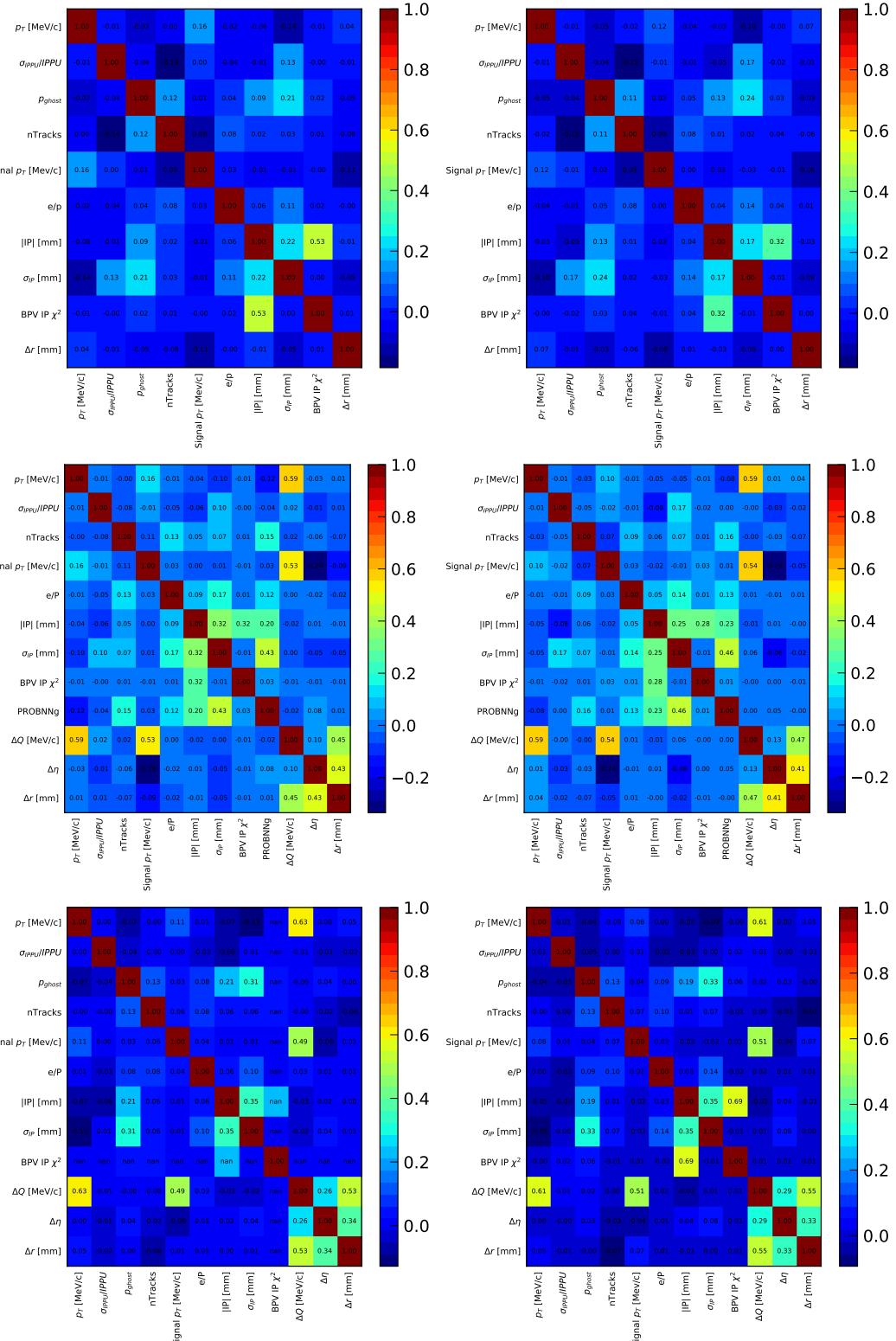


Figure 3.12 – Pearson correlation coefficients between the input features of the Run 1 new (top), Run 2 B2CC (middle) and Run 2 B2OC (bottom) BDT classifiers, for candidates with a correct (left) and wrong (right) decision from the OSe tagger.

### 3.3. Optimisation of the opposite-side electron tagger

Table 3.8 – Features considered for the BDT used to evaluate the predicted mistag of the OSe tagger. For each tuning (Run 1 new, Run 2 B2CC and Run 2 B2OC), the symbol ✓(✗) indicates if a given feature is included (discarded).

Feature	Description	Run 1 new	Run 2 B2CC	Run 2 B2OC
nTracks	Number of reconstructed tracks	✓	✓	✓
$p_T$	Transverse momentum of tagging track	✓	✓	✓
$\sigma_{IP}$	IP uncertainty of tagging track	✓	✓	✓
Signal $p_T$	Transverse momentum of $B$ candidate	✓	✓	✓
BPV IP $\chi^2$	IP $\chi^2$ of tagging track w.r.t $B$ vertex	✓	✓	✓
$p_{ghost}$	Ghost probability	✓	✗	✓
$e/p$	Inverse rigidity	✓	✓	✓
$\Delta r$	difference in $r$ coordinate between $B$ and tagging track	✓	✓	✓
$ IP $	Absolute value of tagging track IP	✓	✓	✓
$\sigma_{IPPU/IPPU}$	Significance of the IP w.r.t pile-up vertex for tagging track	✓	✓	✓
PROBNNg	Ghost probability from neural networks	✗	✓	✗
$\Delta\eta$	Difference in pseudorapidity between $B$ and tagging track	✗	✓	✓
$\Delta Q$	Magnitude of difference in momenta between $B$ and tagging track	✗	✓	✓

#### 3.3.4 Performance evaluation

##### Performance on $B^\pm \rightarrow J/\psi K^\pm$ and $B^\pm \rightarrow D^0 \pi^\pm$ data

Once the BDT is trained on the training sample, the mistag  $\eta$  is predicted for each candidate using the evaluation samples. Then, a two-fold evaluation is applied:

- the mistag calibration is determined on the first evaluation sample. The obtained calibration is then applied to the second evaluation sample, and a calibrated per-event tagging power is computed on the latter;
- same as above, but with the two evaluation samples swapped.

The calibrated per-event tagging power is computed by considering, for each tagged  $B$  candidate, only the tagging particle with the highest transverse momentum. The calibration model consists of a first order natural spline with a logistic link function. The result of these calibrations are shown in Fig. 3.16. The calibrated per-event tagging power is reported in Table 3.9.

### Chapter 3. Flavour tagging

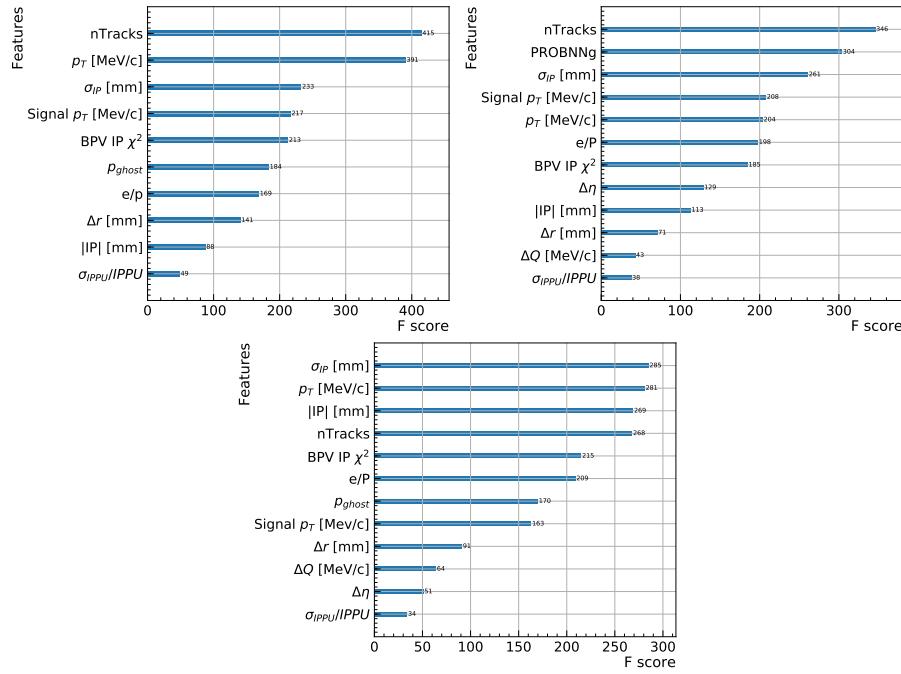


Figure 3.13 – Feature importance for the BDT classifiers of the Run 1 new (top left), Run 2 B2CC (top right) and Run 2 B2OC (bottom) implementations of the OSe tagger.

Table 3.9 – Calibrated, per-event tagging power  $\varepsilon_{\text{eff}}$  (in %) of the OSe algorithms obtained on the evaluation sets of each OSe implementation. The errors include both statistical uncertainty and uncertainties from the calibration procedure. The average is computed by assuming uncorrelated measurements.

Algorithm	set 1	set 2	average
Run 1 new	$0.513 \pm 0.040$	$0.496 \pm 0.038$	$0.504 \pm 0.028$
Run 2 B2CC	$0.324 \pm 0.031$	$0.364 \pm 0.033$	$0.343 \pm 0.023$
Run 2 B2OC	$0.455 \pm 0.043$	$0.434 \pm 0.041$	$0.444 \pm 0.030$

#### 747 Performance on $B^0 \rightarrow D^\mp \pi^\pm$ data

748 The performance (tagging efficiency, mistag probability, tagging power) of the calibrated  
 749 OSe tagger is evaluated on Run 1 (2012) and Run 2 (2016) *sWeighted* data samples of  
 750  $B^0 \rightarrow D^\mp \pi^\pm$  decays. These decays ensure a robust estimation of the performance thanks  
 751 to the high statistics collected at LHCb. Moreover, this channel was not exploited in the  
 752 development of the OSe tagger, so that it constitutes an independent validation of these  
 753 algorithms. The performance of the other OS taggers (OS $\mu$ , OS $K$ , OSCharm, OSVtx,  
 754 and their combination) is presented as well in this section in order to provide a complete  
 755 overview.

### 3.3. Optimisation of the opposite-side electron tagger

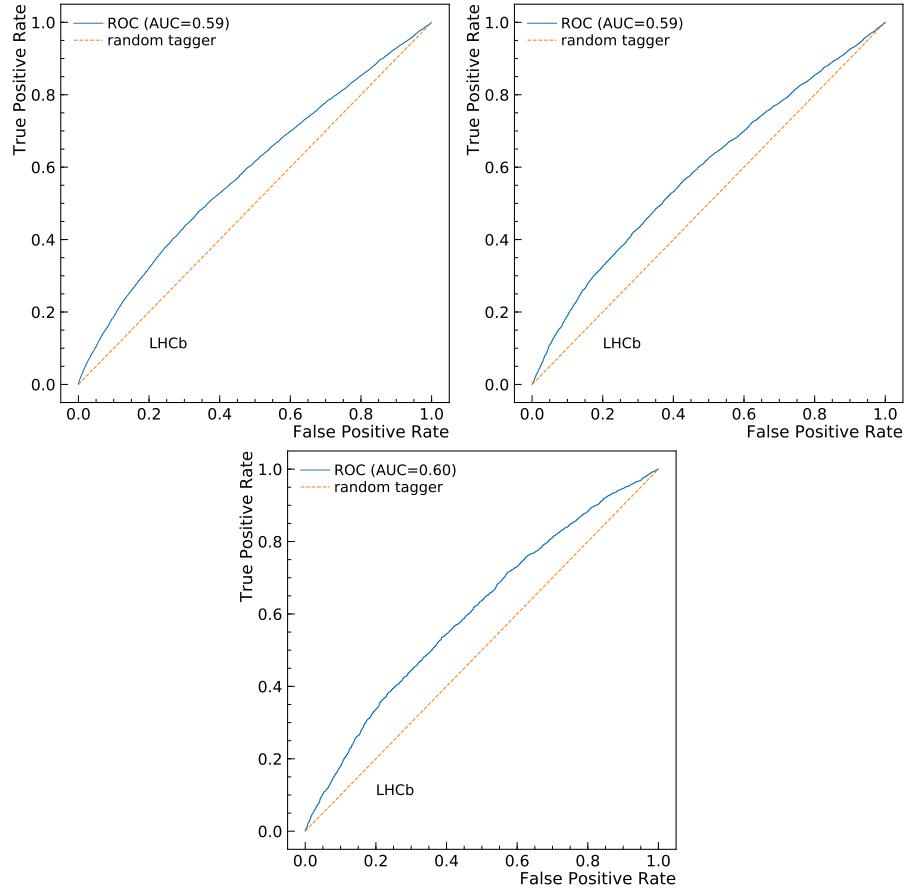


Figure 3.14 – Rate of true right-tagged candidates as a function of the rate of false right-tagged candidates (ROC curves) for the BDT classifiers of the Run 1 new (top left), Run 2 B2CC (top right) and Run 2 B2OC (bottom) implementations of OSe. The obtained ROC curves are represented in blue, while the expected ROC curve in case of random tag decision is shown as a dashed orange line. For each BDT, the *Area Under the ROC Curve* (AUC score) is reported as well.

756 The calibration and the performance evaluation are done as follows:

- 757     • each sample (Run 1 and Run 2) is split randomly in two subsamples;
- 758     • the calibrations are found on one subsample for all OS taggers;
- 759     • the obtained calibrations are applied to the other subsample, and the calibrated  
760        performance is evaluated.
- 761     • the calibrated OS taggers are combined, the combination is calibrated in order to  
762        correct for effects due to correlations among taggers, and the performance of the  
763        calibrated combination is evaluated.

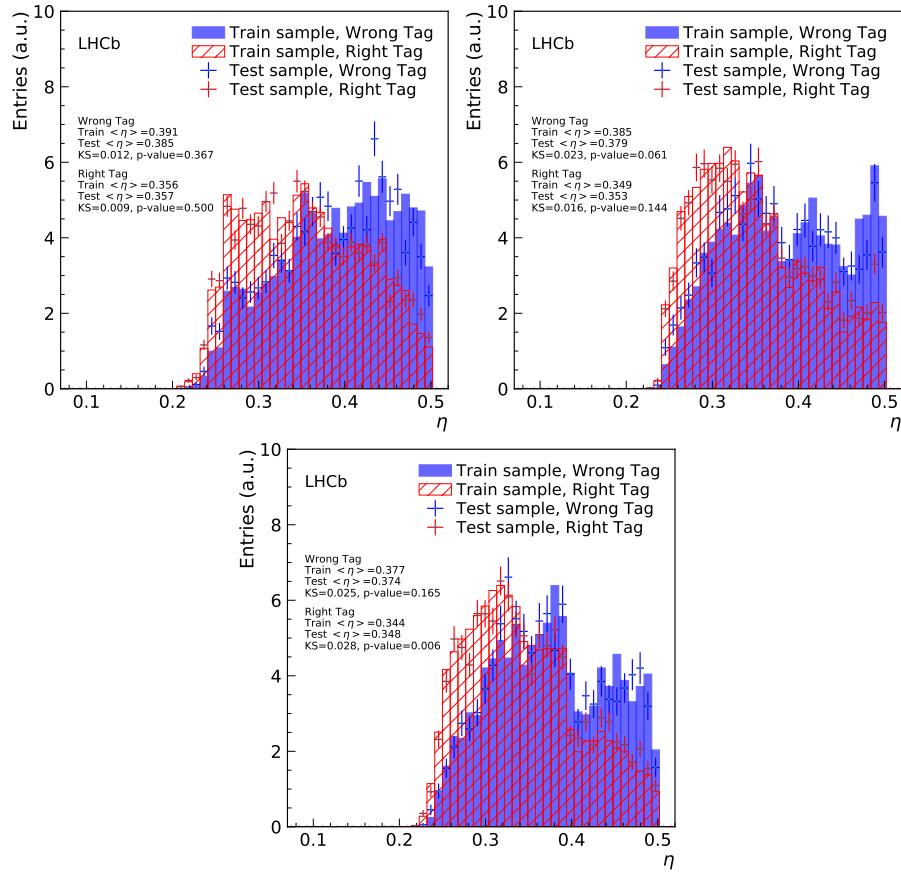


Figure 3.15 – *s*Weighted distributions of the mistag probability  $\eta$  predicted by the BDT classifiers of the Run 1 new (top left), Run 2 B2CC (top right) and Run 2 B2OC (bottom) versions of the OSe tagger. The blue-solid (red-hatched) histogram represents the training data for candidates having the wrong (right) tag decision. The blue (red) points indicate the test data for candidates with wrong (right) tag decision. The overtraining is checked, separately for candidates with wrong and right tag decision, by means of a Kologorov-Smirnov (KS) test to measure the compatibility between training data and test data. The conventional value of 0.05 is chosen as significance level to reject the hypothesis of compatibility.

764 The calibrations are obtained via a time-dependent analysis of the  $B^0 \rightarrow D^\mp \pi^\pm$  decays,  
 765 where acceptance and resolution effects are neglected as described in Ref. [61]. The chosen  
 766 model  $\omega(\eta)$  for each tagger is a GLM model with a logistic link function, and a first order  
 767 spline as basis function. The results of the calibration and the mistag distribution of each  
 768 OSe implementation are shown in Figs. 3.17 and 3.18; the calibration and the mistag of  
 769 the corresponding OS combinations are also reported in Figs. 3.19 and 3.20.

770 The performance is reported in Tables 3.10 and 3.11. The Run 1 new tuning allows to gain a  
 771 relative 9% in tagging power for the OSe tagger on Run 1 data; the corresponding, relative

### 3.3. Optimisation of the opposite-side electron tagger

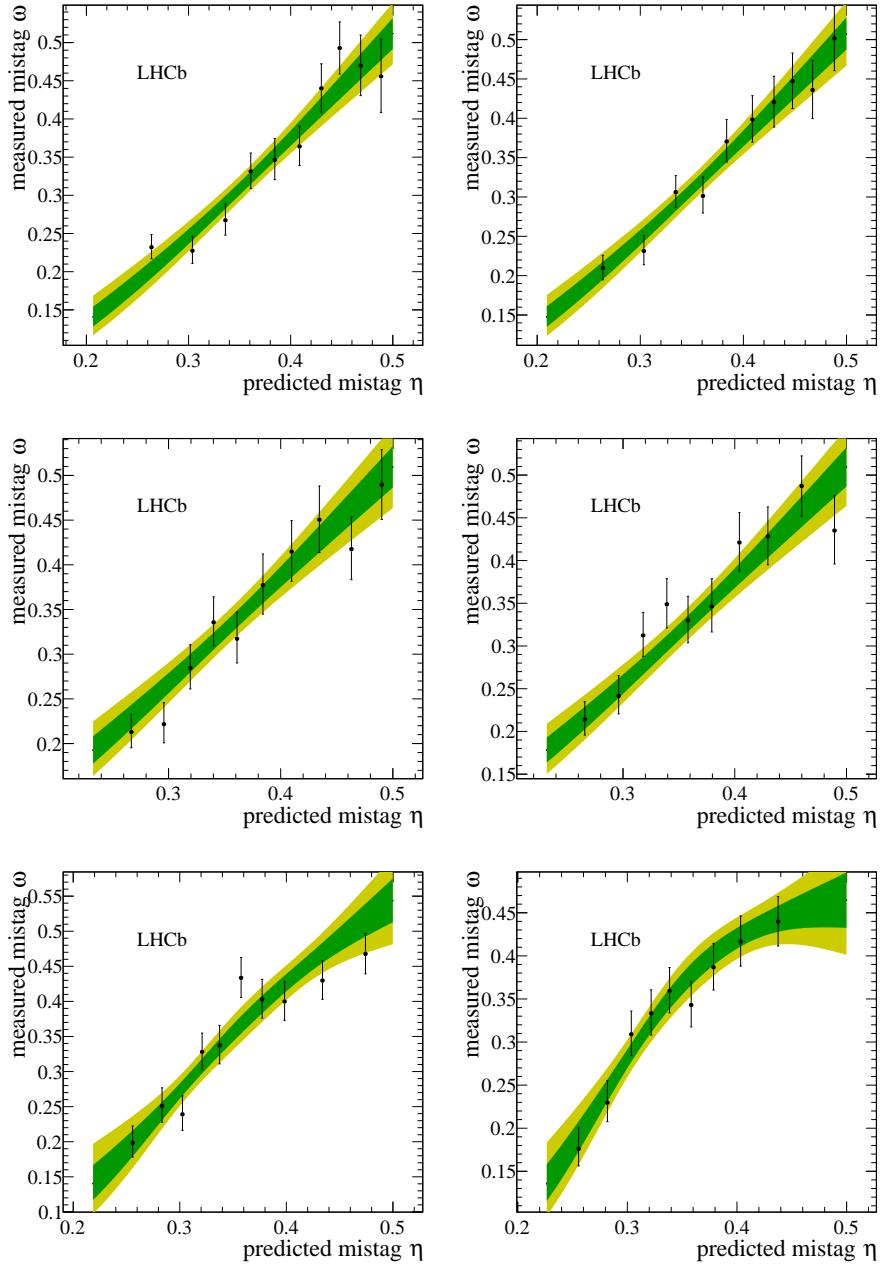


Figure 3.16 – OSe mistag calibration results for the (top) Run 1 new, (middle) Run 2 B2CC and (bottom) Run 2 B2OC optimisations. Left: calibration obtained on the second evaluation sample plotted together with the first evaluation sample. Right: calibration obtained on the first evaluation sample plotted together with the second evaluation sample. The *sWeighted* data sample is shown as black points. The green (yellow) band indicates the 68% (95%) C.L. interval for the fitted calibration functions.

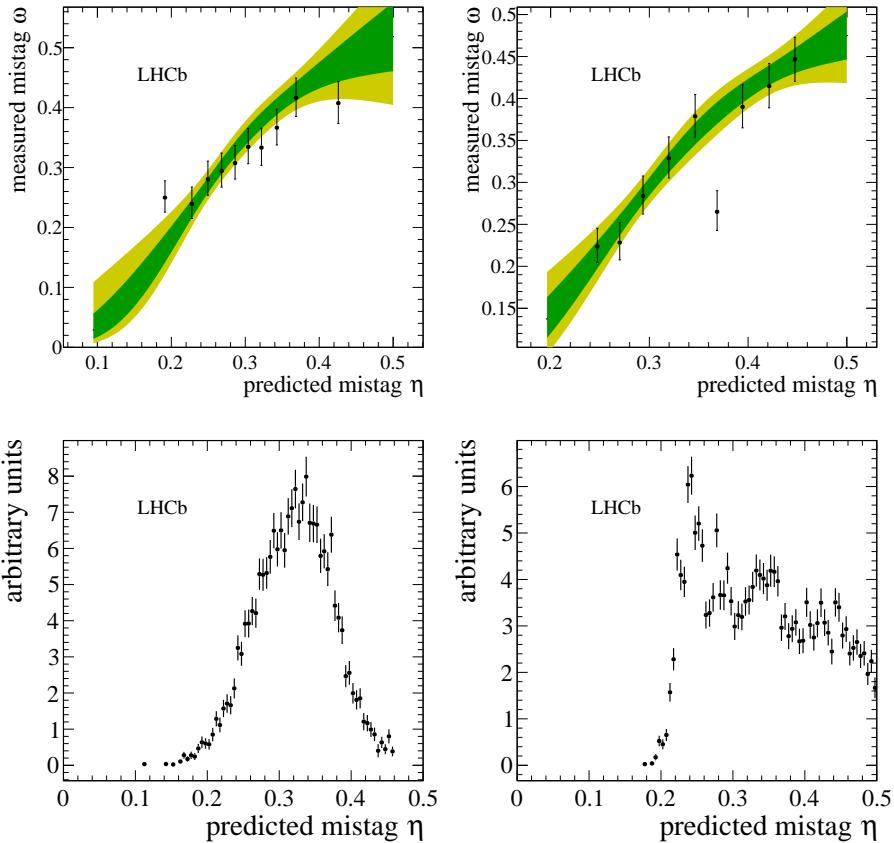


Figure 3.17 – Top: mistag calibration results on *sWeighted* Run 1  $B^0 \rightarrow D^\mp \pi^\pm$  data for the Run 1 old (left) and Run 1 new (right) versions of the OSe tagger. The *sWeighted* data sample is shown as black points. The green (yellow) band indicates the 68% (95%) C.L. interval for the fitted calibration functions. Bottom: distributions of the uncalibrated mistag  $\eta$ .

772 gain of the OS combination is 3%. The tagging power of OSVtx and OSCharm increases  
 773 on Run 2 data compared to Run 1; for this reason, no optimisations are required. The  
 774 tagging power of OSe, OS $\mu$ , and OSK with the Run 1 tunings is lower on Run 2 data  
 775 compared to Run 1. However, compared to the Run 1 tunings, the Run 2 tunings show a  
 776 relative improvement in tagging power of about  $\sim 160\%$  for OSe, and  $\sim 6\%$  for OS $\mu$  and  
 777 OSK on Run 2 data. This allows to recover similar performances as the ones obtained on  
 778 Run 1 data with the Run 1 tunings, both for the individual taggers and their combination.  
 779 Moreover, the Run 2 B2CC and B2OC tunings show consistent tagging powers on Run  
 780 2 data, meaning that the optimisation is robust against the different kinematics of the  
 781 adopted decays.

### 3.3. Optimisation of the opposite-side electron tagger

Table 3.10 – Performance (tagging efficiency, average mistag and tagging power in %) of the OS taggers on *sWeighted* Run 1  $B^0 \rightarrow D^\mp\pi^\pm$  data. The numbers for OSe and the OS combination are shown separately for the Run 1 old and Run 1 new tunings. The first uncertainty is statistical and the second comes from the calibration.

Tagger	$\varepsilon_{\text{tag}}$	$\langle \omega \rangle$	$\varepsilon_{\text{eff}}$
OSVtx	$22.026 \pm 0.100$	$37.295 \pm 0.030 \pm 0.376$	$1.422 \pm 0.009 \pm 0.084$
OSCharm	$4.632 \pm 0.050$	$34.026 \pm 0.049 \pm 0.824$	$0.473 \pm 0.006 \pm 0.049$
OSe Run 1 old	$3.028 \pm 0.041$	$30.570 \pm 0.113 \pm 0.963$	$0.457 \pm 0.008 \pm 0.045$
OSe Run 1 new	$4.337 \pm 0.049$	$33.089 \pm 0.085 \pm 0.777$	$0.496 \pm 0.007 \pm 0.046$
OS $\mu$ Run 1	$8.539 \pm 0.067$	$28.756 \pm 0.071 \pm 0.582$	$1.541 \pm 0.016 \pm 0.085$
OSK Run 1	$18.800 \pm 0.094$	$36.724 \pm 0.031 \pm 0.417$	$1.325 \pm 0.009 \pm 0.083$
OS combination Run 1 old	$39.004 \pm 0.117$	$34.679 \pm 0.035 \pm 0.273$	$3.662 \pm 0.020 \pm 0.131$
OS combination Run 1 new	$39.733 \pm 0.118$	$34.576 \pm 0.035 \pm 0.270$	$3.781 \pm 0.021 \pm 0.133$

Table 3.11 – Performance (tagging efficiency, average mistag and tagging power in %) of the OS taggers on *sWeighted* Run 2  $B^0 \rightarrow D^\mp\pi^\pm$  data. The numbers for OSe, OS $\mu$ , OSK, and the OS Combination are shown separately for the Run 1, Run 2 B2CC, and Run 2 B2OC tunings. The first uncertainty is statistical and the second comes from the calibration.

Tagger	$\varepsilon_{\text{tag}}$	$\langle \omega \rangle$	$\varepsilon_{\text{eff}}$
OSVtx	$20.834 \pm 0.075$	$36.139 \pm 0.029 \pm 0.301$	$1.601 \pm 0.009 \pm 0.070$
OSCharm	$5.025 \pm 0.040$	$33.875 \pm 0.041 \pm 0.624$	$0.523 \pm 0.005 \pm 0.040$
OSe Run 1 old	$1.868 \pm 0.025$	$34.300 \pm 0.096 \pm 0.941$	$0.184 \pm 0.003 \pm 0.022$
OSe Run 2 B2CC	$4.451 \pm 0.038$	$33.352 \pm 0.081 \pm 0.608$	$0.493 \pm 0.006 \pm 0.036$
OSe Run 2 B2OC	$3.333 \pm 0.033$	$30.917 \pm 0.075 \pm 0.702$	$0.486 \pm 0.006 \pm 0.036$
OS $\mu$ Run 1	$8.343 \pm 0.051$	$30.357 \pm 0.042 \pm 0.466$	$1.288 \pm 0.010 \pm 0.061$
OS $\mu$ Run 2 B2CC	$9.151 \pm 0.053$	$30.837 \pm 0.041 \pm 0.432$	$1.344 \pm 0.010 \pm 0.061$
OS $\mu$ Run 2 B2OC	$8.040 \pm 0.050$	$29.174 \pm 0.043 \pm 0.463$	$1.395 \pm 0.010 \pm 0.062$
OSK Run 1	$15.737 \pm 0.067$	$35.902 \pm 0.030 \pm 0.357$	$1.251 \pm 0.008 \pm 0.063$
OSK Run 2 B2CC	$19.516 \pm 0.073$	$36.889 \pm 0.026 \pm 0.310$	$1.342 \pm 0.007 \pm 0.064$
OSK Run 2 B2OC	$15.793 \pm 0.067$	$35.565 \pm 0.030 \pm 0.348$	$1.316 \pm 0.008 \pm 0.063$
OS combination Run 1 old	$36.239 \pm 0.088$	$35.285 \pm 0.024 \pm 0.227$	$3.139 \pm 0.013 \pm 0.097$
OS combination Run 2 B2CC	$40.154 \pm 0.090$	$35.123 \pm 0.025 \pm 0.210$	$3.555 \pm 0.014 \pm 0.100$
OS combination Run 2 B2OC	$36.555 \pm 0.089$	$34.225 \pm 0.026 \pm 0.220$	$3.638 \pm 0.015 \pm 0.102$

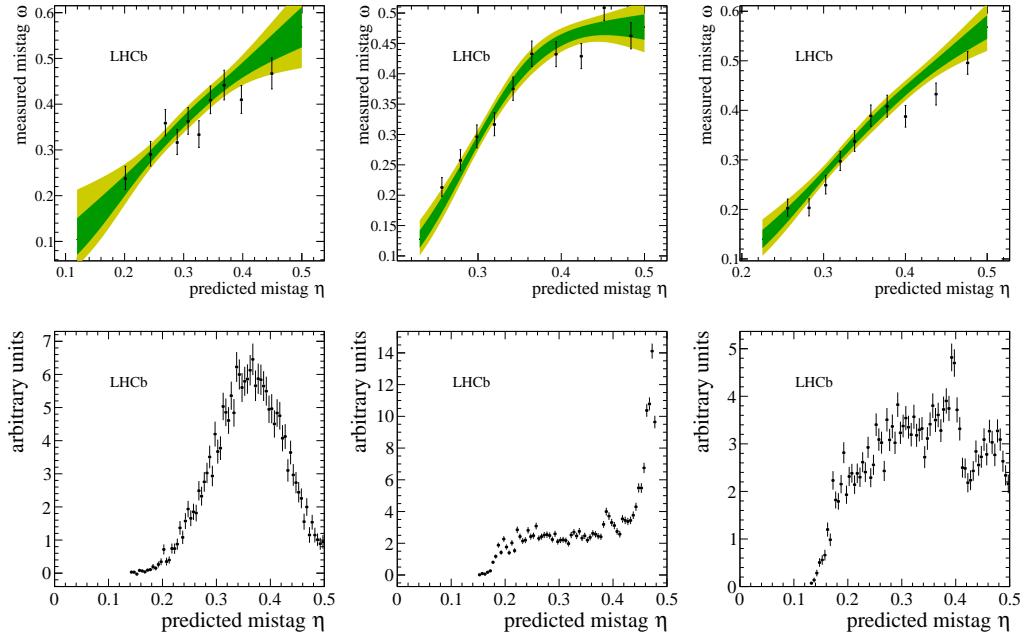


Figure 3.18 – Mistag calibration results on  $s\text{Weighted}$  Run 2  $B^0 \rightarrow D^\mp \pi^\pm$  data for the OSe taggers. The results obtained with the Run 1 old (left), Run 2 B2CC (center), and Run 2 B2OC (right) tunings are shown. The  $s\text{Weighted}$  data sample is shown as black points. The green (yellow) band indicates the 68% (95%) C.L. interval for the fitted calibration functions. Bottom: distributions of the uncalibrated mistag  $\eta$ .

### 3.3. Optimisation of the opposite-side electron tagger

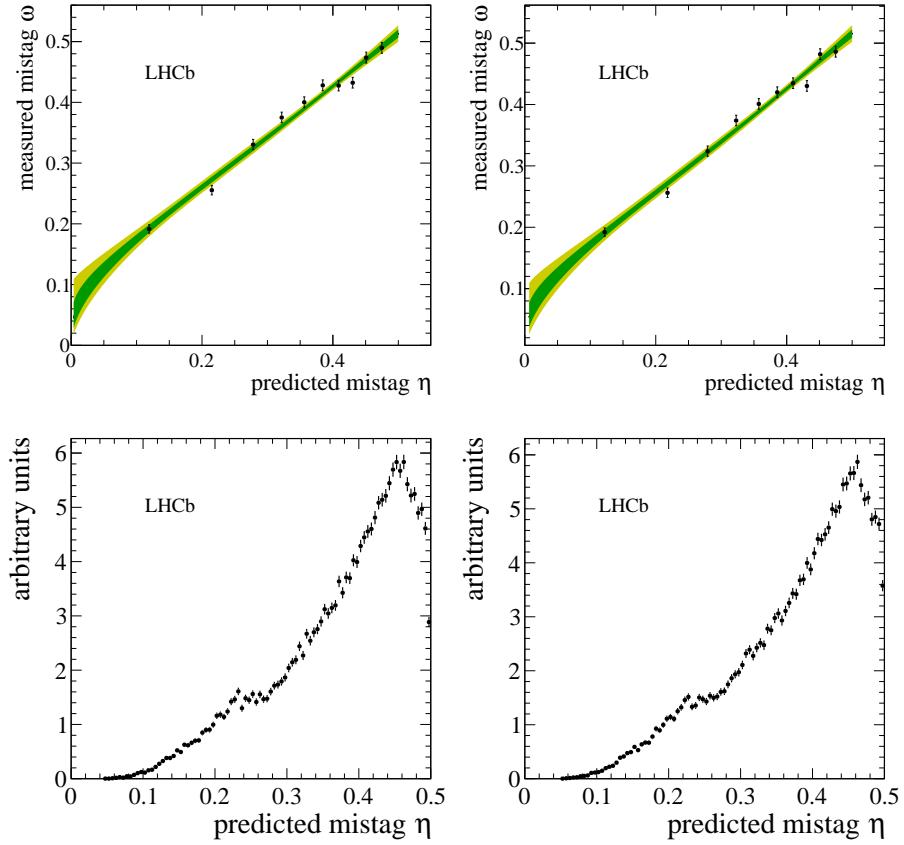


Figure 3.19 – Top: mistag calibration results on *sWeighted* Run 1  $B^0 \rightarrow D^\mp \pi^\pm$  data for the combination of the OS taggers. The results obtained with the Run 1 old (left) and Run 1 new (right) tunings of OSe are shown. The *sWeighted* data sample is shown as black points. The green (yellow) band indicates the 68% (95%) C.L. interval for the fitted calibration functions. Bottom: distributions of the uncalibrated mistag  $\eta$ .

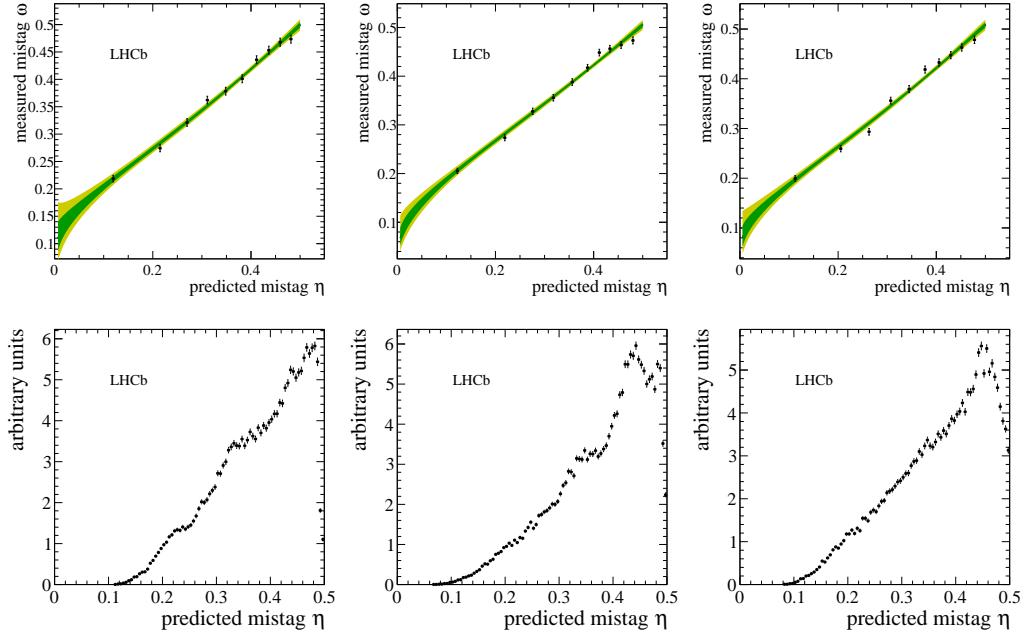


Figure 3.20 – Mistag calibration results on *sWeighted* Run 2  $B^0 \rightarrow D^\mp\pi^\pm$  data for the combination of the OS taggers. The results obtained with the Run 1 old (left), Run 2 B2CC (center), and Run 2 B2OC (right) tunings of OSe, OS $\mu$ , and OSK are shown. The *sWeighted* data sample is shown as black points. The green (yellow) band indicates the 68% (95%) C.L. interval for the fitted calibration functions. Bottom: distributions of the uncalibrated mistag  $\eta$ .

## 4 Selection of $B^0 \rightarrow D^\mp \pi^\pm$ decays

783

## 784 4.1 Data sample and preselection

785 The sample of data is passed through the following selection steps:

- 786 1. stripping and trigger requirements;
- 787 2. a *cut-based* preselection;
- 788 3. vetoes for misidentified backgrounds and wrongly associated *primary vertices* (PVs);
- 789 4. a multivariate classification (MVA);
- 790 5. a final randomised multiple candidate selection.

791 In what follows, the details of each step are provided.

### 792 4.1.1 Stripping and trigger requirements

793 Signal  $B^0 \rightarrow D^\mp\pi^\pm$  candidates are reconstructed using a dedicated stripping line (called  
 794 `B02DPiD2HHHBeauty2CharmLine`). Each event is required to have less than 500 long  
 795 tracks. The criteria that the charged tracks have to fulfill are listed in Table 4.1. Three of  
 796 these hadrons have to form a common vertex to build a  $D^\mp$  meson. Further requirements  
 797 on the  $D^\mp$  combination are given in Table 4.2. The  $B^0$  candidates are built by combining  
 798 a  $D^\mp$  candidate and a bachelor particle if the requirements listed in Table 4.3 are fulfilled.  
 799 Finally, a bagged boosted decision tree (BDT) classifier [65], which is trained on simulated  
 800 data, is applied. A minimum value of 0.05 is required for the output value of the BDT.  
 801 Stripped candidates are then filtered according to how they were selected at the trigger  
 802 level: no specific requirements are made at L0; at HLT1,  $B$  candidates are required  
 803 to be TOS from the `Hlt1TrackAllL0Decision` trigger line; at HLT2,  $B$  candidates  
 804 are required to be TOS from one of the following lines: `Hlt2Topo2BodyBBDTDecision`,  
 805 `Hlt2Topo3BodyBBDTDecision` or `Hlt2Topo4BodyBBDTDecision`. These trigger lines are  
 806 described in detail in Refs. [66, 67].

Table 4.1 – Stripping requirements applied in the selection of charged tracks. The more stringent requirements given in brackets are for the bachelor track. The  $\text{IP}\chi^2$  is the vertex-fit  $\chi^2$  difference for the PV reconstructed with and without the  $B^0$  candidate.

track $\chi^2/\text{ndof}$	$< 3.0$ ( $< 2.5$ )
momentum $p$	$> 1 \text{ GeV}/c$ ( $> 5 \text{ GeV}/c$ )
transverse momentum $p_T$	$> 100 \text{ MeV}/c$ ( $> 500 \text{ MeV}/c$ )
$\text{IP}\chi^2$ w.r.t. any PV	$> 4.0$
ghost probability	$< 0.4$

## 4.1. Data sample and preselection

Table 4.2 – Stripping requirements on the three-track combinations forming  $D^\mp$  candidates. DOCA is the Distance Of Closest Approach of the daughter particles w.r.t. each other, and DIRA indicates the cosine of the angle between the momentum of the  $D^\mp$  meson and the direction from the best PV to the decay vertex. The best PV is defined as the vertex with the lowest IP $\chi^2$ .

$\sum p_T(hhh)$	$> 1800 \text{ MeV}/c$
DOCA	$< 0.5 \text{ mm}$
$m(hhh)$	$\in [1769.62, 2068.49] \text{ MeV}/c^2$
$D^\mp$ vertex $\chi^2/\text{ndof}$	$< 10.0$
$D^\mp$ vertex separation $\chi^2$ to any PV	$> 36$
$D^\mp$ DIRA	$> 0.0$

Table 4.3 – Stripping requirements on the  $D^\mp\pi^\pm$  combination (before the BDT requirement mentioned in the text).

$B^0$ vertex $\chi^2/\text{ndof}$	$< 10.0$
reconstructed $B^0$ proper decay time $t$	$> 0.2 \text{ ps}$
IP $\chi^2$ w.r.t. the best PV	$< 25.0$
$B^0$ DIRA	$> 0.999$

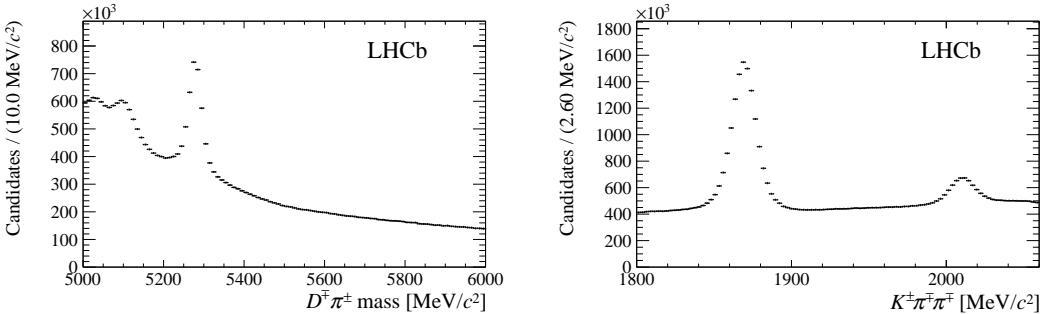


Figure 4.1 –  $D^\mp\pi^\pm$  and  $K^\pm\pi^\mp\pi^\mp$  mass distributions of the reconstructed  $B^0 \rightarrow D^\mp\pi^\pm$ ,  $D^\mp \rightarrow K^\pm\pi^\mp\pi^\mp$  candidates after the stripping and trigger selection.

807 Figure 4.1 shows the  $D^\mp\pi^\pm$  and  $K^\pm\pi^\mp\pi^\mp$  mass distributions of the reconstructed can-  
 808 didates after the stripping and trigger selections. In the  $D^\mp\pi^\pm$  mass distribution the  
 809  $B^0 \rightarrow D^\mp\pi^\pm$  signal peak is already visible at around  $5280 \text{ MeV}/c^2$ . The structure  
 810 at masses lower than the  $B$  peak originates from partially reconstructed  $B^0 \rightarrow D\rho$   
 811 and  $B^0 \rightarrow D^*\pi$  decays. The  $K^\pm\pi^\mp\pi^\mp$  mass distribution features a clearly visible  
 812  $D^\mp \rightarrow K^\pm\pi^\mp\pi^\mp$  peak at  $1870 \text{ MeV}/c^2$  and a  $D^{*\mp} \rightarrow D^0(K^\pm\pi^\mp)\pi^\mp$  peak around  
 813  $2010 \text{ MeV}/c^2$ .

Table 4.4 – Offline preselection requirements.

$B^0$ candidate decay time	$> 0.2$ ps
$ m(K^\pm\pi^\mp\pi^\mp) - m_{D^\mp}^{\text{PDG}} $	$< 35$ MeV/ $c^2$
PID $K$ for pions	$< +8$ from $D^\mp$
PID $K$ for kaon	$> -2$ from $D^\mp$

### 4.1.2 Preselection and sample definitions

Additional preselection criteria (shown in Table 4.4) are applied offline. In order to obtain the correct correlations between the uncertainties on vertex positions, particle momenta, flight distances, decay times, and invariant masses, a Kalman filter, known as `DecayTreeFitter` (DTF) [68], is used. The decay-time related observables are derived from a DTF fit where the position of the primary vertex has been used to constrain the production vertex of the  $B^0$  meson. To determine the momentum and the invariant mass of the  $B^0$  meson, the invariant mass of the  $D^\mp$  meson is constrained to the central value of the PDG ( $m_{D^\mp}^{\text{PDG}} = 1869.61$  MeV/ $c^2$  [22]) in a separate DTF fit. The PID $K$  variable of Eq. 2.1 is used to identify the kaon and the pions from the  $D^\mp$  decays, and to identify the bachelor pion from the  $B^0$  decay. The requirement on the PID $K$  of the bachelor pion defines two samples of candidates: the so-called *pion sample* ( $\text{PID}K \leq 5$ ) and the so-called *kaon sample* ( $\text{PID}K > 5$ ). This distinction will be useful in the fit to the  $B^0$  mass distribution described in Sec. 4.3 for determining the sample composition. All the following selection steps are applied to both the pion and kaon samples.

### 4.1.3 Veto against physics backgrounds

Misidentification of muons, kaons and protons as pions leads to exclusive backgrounds. These are suppressed by means of explicit *veto*s. In order to reduce contributions from semileptonic decays such as  $B^0 \rightarrow D^- (\rightarrow K^+\pi^-\pi^-)\mu^+\nu_\mu$ , the bachelor pion is required to have no hits in the muon chambers. A  $p \rightarrow \pi$  mis-identification can lead to background contributions from  $\Lambda_b^0 \rightarrow \Lambda_c^+ (\rightarrow K^-\pi^+p)\pi^-$ . To reduce these contributions, the proton mass hypothesis is applied separately to both pion candidates from the  $D^\pm$  final state. The invariant mass of the three hadrons is recalculated and if the candidate is inside a  $\pm 30$  MeV/ $c^2$  ( $\pm 50$  MeV/ $c^2$ ) window around the  $\Lambda_c^+$  mass,  $m_{\Lambda_c^+} = 2286.46$  MeV/ $c^2$  [22], it is required to have  $\text{PID}p < -8.0$  ( $\text{PID}p < -5.0$ ). A plot showing the distributions before and after applying the veto is given in Fig. 4.2. This requirement shows a signal efficiency of  $(93.48 \pm 0.06)\%$ . The rejection of  $\Lambda_b^0 \rightarrow \Lambda_c^+\pi^-$  is checked with simulation. After stripping and preselection alone,  $(99.720 \pm 0.004)\%$  of the  $\Lambda_b^0 \rightarrow \Lambda_c^+\pi^-$  decays are rejected, and this veto rejects  $(76.6 \pm 0.6)\%$  of the remaining  $\Lambda_b^0 \rightarrow \Lambda_c^+\pi^-$  decays.

In the same way as protons may be misidentified as pions, it is possible for kaons to be misidentified as pions. Such a mis-identification would lead to background contributions

## 4.1. Data sample and preselection

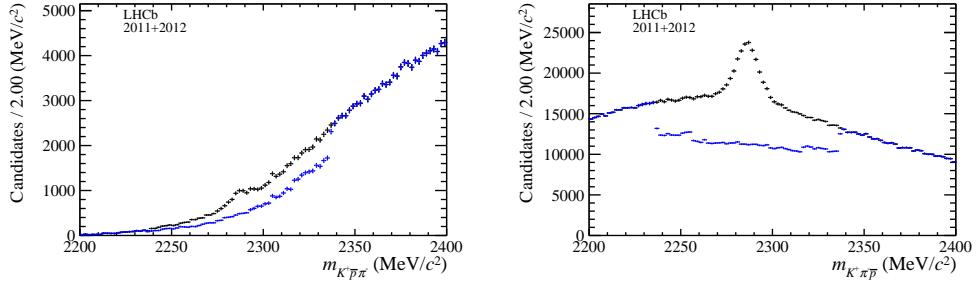


Figure 4.2 – Distributions of the invariant mass of the  $K^\pm\pi^\mp p^\mp$  combinations where each of the two daughter pions of the  $D^\mp$  meson candidate is assigned in turn the proton mass. The distribution is given without (black) and with (blue) the  $\Lambda_c^\mp$  veto described in the text. On the left (right) the proton mass hypothesis is applied to the pion with the lower (higher)  $p_T$ .

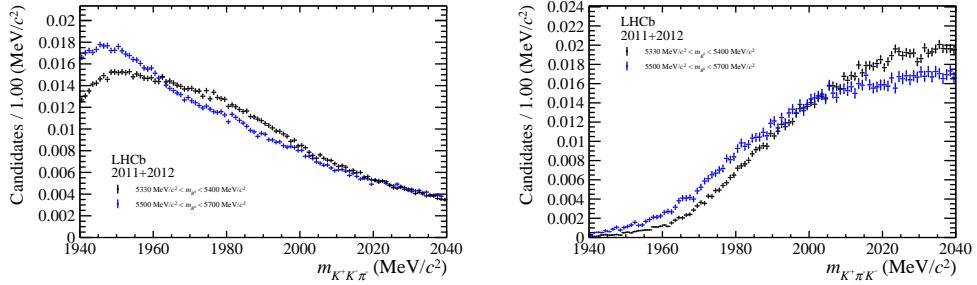


Figure 4.3 – Distributions of the invariant mass of the  $K^\pm K^\mp \pi^\mp$  combinations where each of the two daughter pion of the  $D^\mp$  meson candidate is assigned in turn the kaon mass. The distributions are given in the  $B_s^0$  meson mass range from 5330 to 5400  $\text{MeV}/c^2$  (black) and in the  $B^0$  meson mass range from 5500 to 5700  $\text{MeV}/c^2$  (blue) after applying the  $\Lambda_c^\mp$  veto. On the left (right) the kaon mass hypothesis is applied to the pion with the lower (higher)  $p_T$ .

from  $B_s^0 \rightarrow D_s^\mp (\rightarrow K^\pm K^\mp \pi^\mp) \pi^\pm$ . To check for these contributions, the kaon mass hypothesis is applied in turn to each of the two pions from the  $D^\pm$  final state. The invariant mass of the resulting  $K^\pm K^\mp \pi^\mp$  system is recalculated and plotted for two different ranges of the  $B_s^0$  mass: the first range, from 5330 to 5400  $\text{MeV}/c^2$ , covers the signal region of the  $B_s^0$  meson as possible background contribution. The second range, from 5500 to 5700  $\text{MeV}/c^2$  is the upper mass *sideband* for this possible background contamination. As can be seen in Fig. 4.3 the distribution in the  $B_s^0$  signal region does not show any significant peaking structure compared to that in the upper  $B_s^0$  mass sideband region. The visible differences are expected as the distributions arise from different kinematic ranges. To double check for possible resonant contributions from a kaon mis-identification, the decay of the  $D^\mp$  meson after applying the kaon mass hypothesis is investigated. Possible resonant decays of the  $D$  meson can take place via a

## Chapter 4. Selection of $B^0 \rightarrow D^\mp \pi^\pm$ decays

857  $K^*$  or  $\phi$  resonance. These resonances would be visible in the  $K\pi$  and  $KK$  invariant mass  
 858 distributions, which are plotted for the same two ranges in Fig. 4.4. As the distributions  
 859 in the signal and background ranges look compatible, the  $D_s^\mp$  contamination is negligible  
 and no veto is applied.

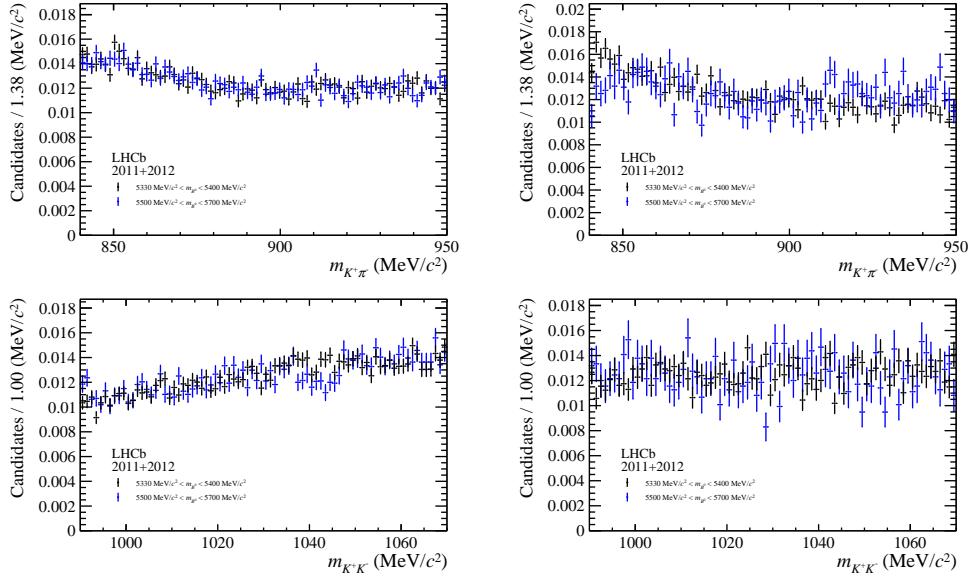


Figure 4.4 – Distributions of the invariant mass of the  $K^\mp\pi^\pm$  combination (top) and the  $K^\mp K^\pm$  combination (bottom), where each of the two daughter pion of the  $D^\mp$  meson candidate is assigned in turn the kaon mass. The distributions are shown in the  $B_s^0$  meson mass range of [5330, 5400] MeV/ $c^2$  (black) and [5500, 5700] MeV/ $c^2$  (blue) as described in the text. Additionally, the  $D_s^\mp$  mass is required to be in [1940, 2040] MeV/ $c^2$ . On the left (right) the kaon mass hypothesis is applied to the pion with the lower (higher)  $p_T$ .

860

861 In the same way as for the  $D^\mp$  meson daughters, it is also possible that the bache-  
 862 lor pion candidate is actually a misidentified kaon. This mis-identification could lead  
 863 to background contributions of  $B^0 \rightarrow D^0 (\rightarrow K^\pm \pi^\mp) \pi^\mp K^\pm$ . A similar background,  
 864  $B^0 \rightarrow D^0 (\rightarrow K^\pm \pi^\mp) K^\mp \pi^\pm$ , could arise if one  $D^\mp$  meson daughter pion is misidenti-  
 865 fied as a kaon and combined with the bachelor pion. To check for this contribution, the  
 866 kaon mass hypothesis is applied to the bachelor pion and the  $D$  meson daughter pions,  
 867 and the invariant mass distributions for the four possible  $K\pi$  systems are plotted after  
 868 applying the MVA classifier described in Sec. 4.1.5 (Fig. 4.5). As the distributions show  
 869 no significant peaking structures, this contribution is neglected and no specific cuts are  
 870 applied.

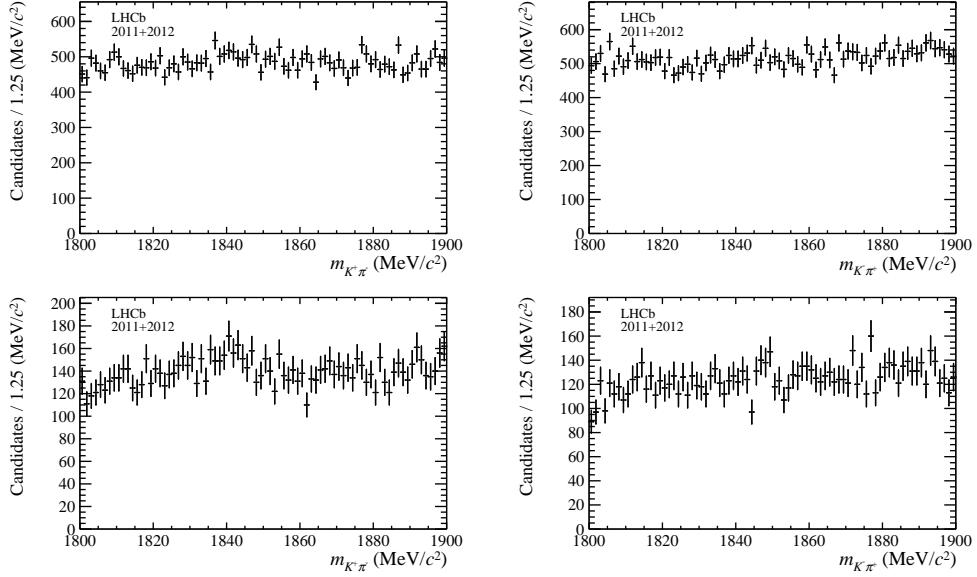


Figure 4.5 – Distributions of the invariant mass of the four possible  $K^\mp\pi^\pm$  combinations. In the top (bottom) plots the bachelor pion is combined with the pion from the  $D^\mp$  meson with lower (higher)  $p_T$ . In the left (right) plots the kaon mass hypothesis is applied to the bachelor pion (pions from the  $D^\mp$  meson).

#### 871 4.1.4 Wrongly associated primary vertices

872 Given an average number of total  $pp$  interactions per bunch crossing of  $\nu = 2.5$ , a  
 873 large fraction of events have more than one reconstructed PV. The PV to which the  $B^0$   
 874 candidate has the smallest  $IP\chi^2$  (*best* PV) is chosen as the  $B^0$  production vertex.

875 In events where the association of the  $B^0$  candidate to its best PV is wrong, the re-  
 876 constructed decay time of this candidate will be incorrect. These wrongly associated  
 877 candidates cause a large tail in the decay time distribution, which can be clearly observed  
 878 in signal MC where the true decay time is known: giving each candidate a weight equal  
 879 to  $e^{t/\tau}$ , where  $\tau$  is the true lifetime, leads to an excess of candidates at high decay times.  
 880 To remove these incorrect associations in MC, one can compare the  $z$  position of the  
 881 associated PV with the  $z$  position of the true PV and reject the candidate if the distance  
 882 between those positions exceeds 5 times its uncertainty (Fig. 4.6). In real data, the  
 883 true PV is unknown, so a selection cut on the smallest  $B^0$  impact parameter  $\chi^2_{DTF,PV}$   
 884 (called  $MinIP\chi^2$ ) with respect to any other PV is applied (Fig. 4.7). Events with only  
 885 a single PV are not affected. The main advantage of this  $MinIP\chi^2$  variable is that all  
 886 PVs are treated equally, without any biasing choice. The cut on  $MinIP\chi^2$  is optimised  
 887 to retain 98 % of the truth-matched signal candidates in MC. The optimal requirement  
 888 is then found at  $MinIP\chi^2 > 16.5$ . A plot showing the signal MC weighted decay time  
 889 distribution after applying this cut is given in Fig. 4.7.

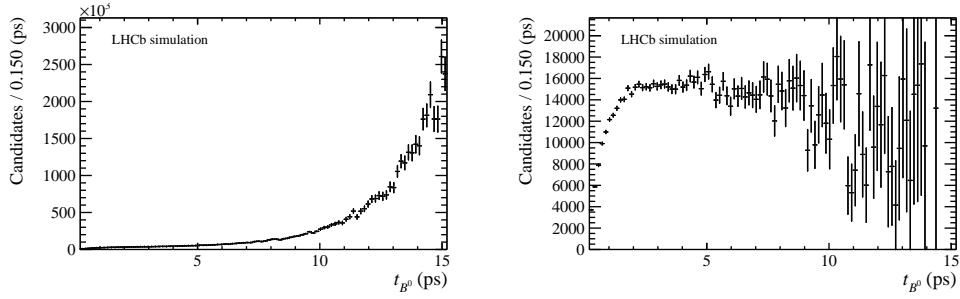


Figure 4.6 – Left: decay time distribution of signal MC events weighted with  $e^{t/\tau}$ , where  $\tau$  is the true lifetime. At high decay times an excess of candidates can be observed. Right: same distribution after requiring that the absolute difference between the best PV  $z$  position and the true PV  $z$  position is within 5 times the best PV  $z$  position uncertainty. As the excess of candidates at high decay times vanishes (from left to right), it is concluded that this excess is due to candidates wrongly associated to their PV.

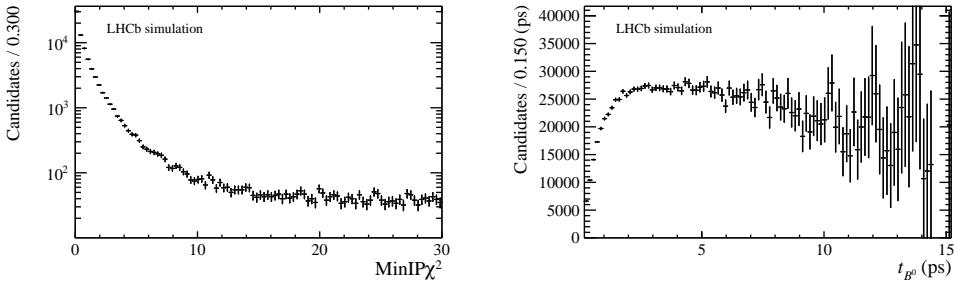


Figure 4.7 – Left: distribution of the smallest  $B^0$  impact parameter  $\chi^2$ ,  $\text{MinIP}\chi^2$ , with respect to any other PV, for signal MC events. Right: decay time distribution of signal MC events weighted with  $e^{t/\tau}$ , where  $\tau$  is the true lifetime, after requiring that  $\text{MinIP}\chi^2 > 16.5$ .

#### 890 4.1.5 Development of an MVA classifier

891 The combinatorial background, consisting of candidates created from random combinations  
 892 of tracks, is rejected by using a Boosted Decision Tree (BDT) classifier [69, 70]. The  
 893 signal input to the training stage consists of signal MC candidates simulated under 2012  
 894 data-taking conditions, while the upper mass sideband above  $5500 \text{ MeV}/c^2$  from the 2012  
 895 data sample is used as template for the combinatorial background. The BDT is trained  
 896 on one half of these samples, the other half being used to test its performance. Before the  
 897 BDT training, all previous selection steps (the cut-based preselection, the mass vetoes and  
 898 the wrongly associated PV veto) are applied. To reduce the number of input features, the  
 899 ones with a correlation larger than 97% with any other feature are removed. The 16 final  
 900 input features are listed in Table 4.5. The correlation matrices between the input features  
 901 in the signal and the background samples are shown in Fig. 4.8, while the distributions of

Table 4.5 – List of input features used in the training of the BDT.

$B^0$ candidate	$\cos \propto [\overrightarrow{\text{Vertex}}(B^0) - \overrightarrow{\text{PV}}, \vec{p}(B^0)]$ vertex $\chi^2$ DTF $\chi^2$ with best PV constraint
$D^\mp$ candidate	IP $\chi^2$ w.r.t. $B^0$ vertex IP $\chi^2$ w.r.t. best PV radial flight distance flight distance $\chi^2$ w.r.t. $B^0$ vertex vertex $\chi^2/\text{ndof}$ transverse momentum $\cos \propto [\overrightarrow{\text{Vertex}}(D^\mp) - \overrightarrow{\text{Vertex}}(B^0), \vec{p}(D^\mp)]$
bachelor $\pi^\pm$	IP $\chi^2$ w.r.t. the best PV transverse momentum track $\chi^2/\text{ndof}$
$D^\mp$ daughters	IP $\chi^2$ w.r.t. best PV

902 the input features can be found in Appendix C.1.

903 The BDT implementation of TMVA [71] is used. The BDT is built out of 1700 trees,  
 904 with a depth limited to four. For each node, at least 2.5 % of the training events have to  
 905 be present. The chosen boosting method is the AdaBoost [72] algorithm with a boost  
 906 factor  $\beta = 0.5$ . The number of trees and the maximal depth of trees have been increased  
 907 iteratively until no significant increase of the performance without overtraining was  
 908 observed. The BDT is tested on the events that are not used in the training. The plot of  
 909 this overtraining check is given in Fig. 4.9.

#### 910 4.1.6 BDT selection optimisation

911 To estimate the best requirement on the output of the BDT classifier, the statistical  
 912 uncertainty of the  $CP$  coefficients derived from the analysis of simulated samples is used as  
 913 the figure of merit (FoM). To determine the sensitivity, the preselection, the mass vetoes  
 914 and the wrongly associated PV veto are applied and the BDT classifier is calculated for  
 915 every candidate. The BDT cut point is then scanned with a step size of 0.01 from  $-0.15$   
 916 to  $0.10$  and a step size of 0.05 in the outer regions. For each cut point, a simulated (*toy*)  
 917 sample is generated. This sample contains the same signal and combinatorial background  
 918 yields as determined from the real dataset via a maximum likelihood fit of the  $B^0$  mass  
 919 distribution. Finally, a time-dependent analysis of each toy dataset is performed in order  
 920 to estimate the statistical uncertainty on  $S_f$  and  $S_{\bar{f}}$ . These statistical uncertainties as a  
 921 function of the BDT cut are shown in Fig. 4.10. Based on these distributions, the BDT  
 922 cut point is chosen to be at 0.05.

## Chapter 4. Selection of $B^0 \rightarrow D^\mp \pi^\pm$ decays



Figure 4.8 – Correlation matrices of the input features used in the training of the BDT for signal (top) and background (bottom).

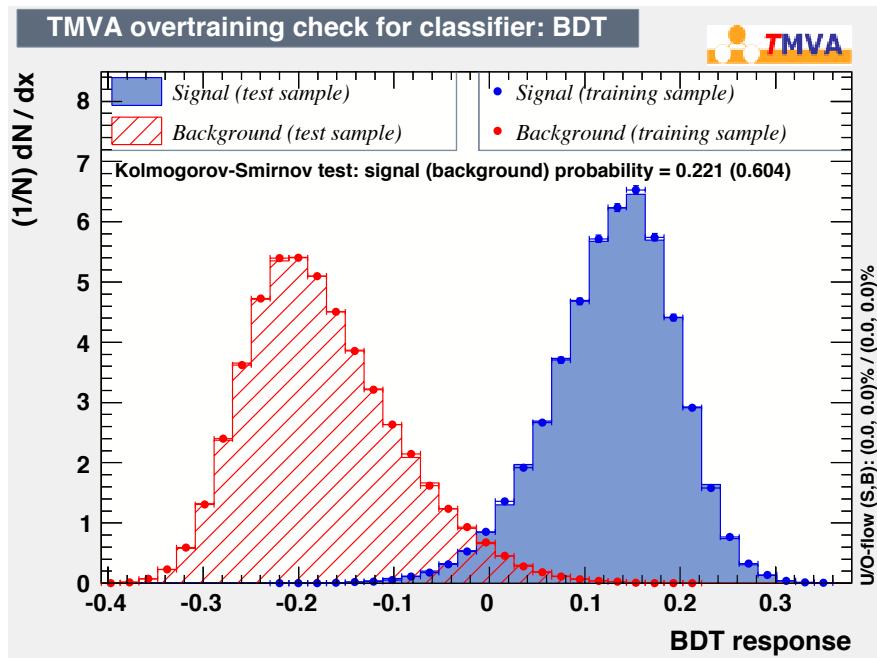


Figure 4.9 – Distributions of the BDT response on training and test samples.

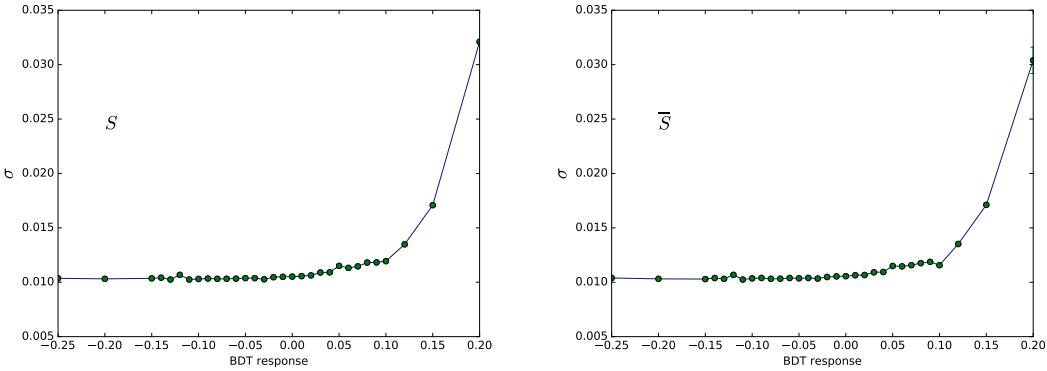


Figure 4.10 – Expected statistical uncertainty on of  $S_f$  (left) and  $S_{\bar{f}}$  (right) as a function of the cut on the output of the BDT classifier, as obtained from simulated samples.

### 923 4.1.7 Multiple candidates

924 After the stripping selection and trigger requirements, approximately 9% of the events  
 925 contain at least two  $B^0$  candidates, and 18 – 20% of all  $B^0$  candidates share an event.  
 926 If the offline selection is also applied, around 0.4% of the events contain multiple  $B^0$   
 927 candidates, and 0.8% of all  $B^0$  candidates share an event. More details are given in  
 928 Appendix C.2. In order to be consistent with the prescription used in the stripping and  
 929 trigger requirements, only the best PV is chosen; all events in which the best PV is no  
 930 longer present after the offline selection are removed. Finally, since the remaining  $B^0$   
 931 candidates are considered to be equally likely signal candidates, a single  $B^0$  candidate  
 932 per event is chosen randomly following the prescription of Ref. [73], which prevents any  
 933 unexpected bias due to a more specific choice.

### 934 4.1.8 Selection performance

935 The offline selection performances are listed in Table 4.6. They are determined by using  
 936 data candidates of the 2012 sample with an invariant  $B^0$  mass above  $5500 \text{ MeV}/c^2$  to  
 937 represent combinatorial background, and signal MC candidates (see Sec. 4.2) to represent  
 938 the signal. In Fig. 4.11 a comparison of the mass distribution of the selected 2011 and  
 939 2012 data is shown. Additionally, the BDT performances are quoted in Table 4.7  
 940 split by both magnet polarity and year of data taking. Finally, in order to check the  
 941 contribution of non-resonant  $B^0 \rightarrow K^+ \pi^- \pi^- \pi^+$  decays, the  $B^0$  and  $D^\mp$  invariant mass  
 942 distributions are analysed after applying the full offline selection (with the exception of  
 943 the  $B^0$  and  $D^\mp$  mass cuts) in two ways. First, the  $D^\mp$  mass distribution is plotted for  
 944 candidates falling in a  $B^0$  mass signal window (Fig. 4.12). From this plot, the maximal  
 945 contamination from non-resonant decays can be estimated to be roughly 1%. Then,  
 946 the  $B^0$  distribution after excluding the  $D^\mp$  signal window is plotted. To quantify the

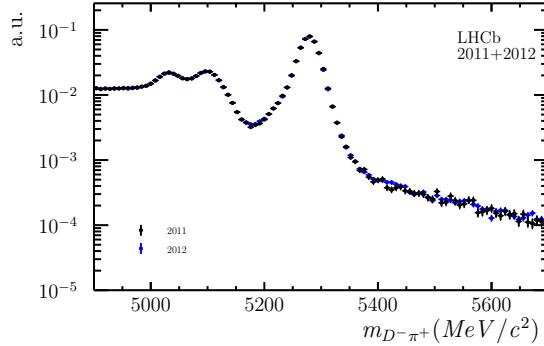


Figure 4.11 – Mass distributions of the  $B^0 \rightarrow D^\mp\pi^\pm$  candidates passing the full offline selection for the 2011 (blue) and 2012 (black) data samples.

Table 4.6 – Signal efficiencies and background rejections of the different selection steps given with respect to the previous selection step. The preselection efficiency is computed w.r.t. the number of candidates passing trigger and stripping requirements. The last row shows the overall selection performance.

Selection step	$\varepsilon_{\text{sig}}$	$1 - \varepsilon_{\text{bkg}}$
preselection	$(93.61 \pm 0.06) \%$	$(85.20 \pm 0.02) \%$
$\Lambda_c^\mp$ veto	$(93.48 \pm 0.06) \%$	$(9.85 \pm 0.03) \%$
semileptonic veto	$(98.96 \pm 0.03) \%$	$(7.66 \pm 0.03) \%$
wrongly associated PV veto	$(97.75 \pm 0.04) \%$	$(15.81 \pm 0.04) \%$
BDT selection	$(83.63 \pm 0.10) \%$	$(97.18 \pm 0.01) \%$
total	$(70.7 \pm 0.1) \%$	$(99.911 \pm 0.002) \%$

947 non-resonant  $B^0 \rightarrow K^+\pi^-\pi^-\pi^+$  decays, the sum of an exponential and a Gaussian with  
 948 a fixed shape is used to fit the resulting  $B^0$  mass distribution, as shown in Fig. 4.13. As  
 949 the fitted  $B^0$  yield is  $645 \pm 242$ , the non-resonant contribution is assumed to be negligible.  
 950

## 4.1. Data sample and preselection

Table 4.7 – BDT performance for each magnet polarity (up and down) and year of data taking. The quoted efficiencies contain signal and background.

	2011, up	2011, down	2012, up	2012, down
# cand. before BDT	398 357	569 853	1 301 800	1 316 597
# cand. after BDT	210 844	285 137	601 345	609 880
$\varepsilon_{\text{sig+bkg}}$	$(50.67 \pm 0.08)\%$	$(50.04 \pm 0.07)\%$	$(46.19 \pm 0.04)\%$	$(46.32 \pm 0.04)\%$

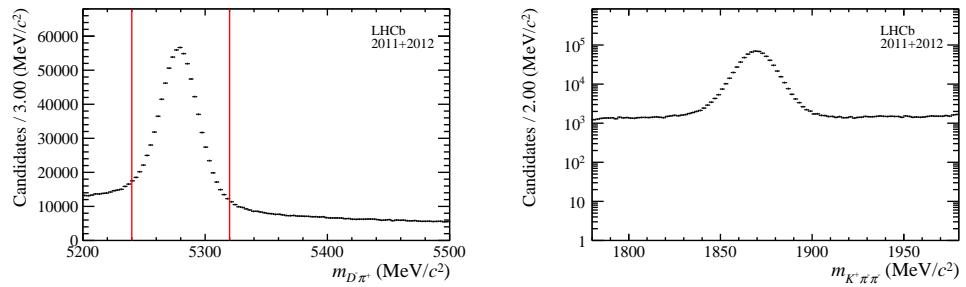


Figure 4.12 – Left:  $B^0$  mass distribution with red vertical lines indicating the selected signal window. Right: resulting  $D^\mp$  mass distribution in the  $B^0$  signal window.

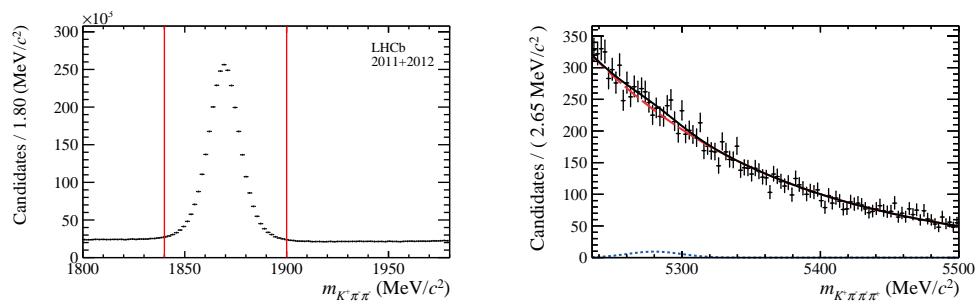


Figure 4.13 – Left:  $D^\mp$  mass distribution with red vertical lines indicating the excluded range. Right:  $B^0$  mass distribution outside the  $D^\mp$  mass window with the fitting function overlaid.

Sample	Event type	$N_{\text{sel}}$	Efficiency [%]
$B^0 \rightarrow D^\mp\pi^\pm$	11164003	101096	$1.966 \pm 0.006$
$B^0 \rightarrow D^\mp K^\pm$	11264011	19300	$1.833 \pm 0.013$
$B^0 \rightarrow D^\mp\rho^\pm$	11164401	2408	$0.1178 \pm 0.0024$
$B^0 \rightarrow D^{*\mp}\pi^\pm$	11164404	14901	$0.721 \pm 0.006$
$B_s^0 \rightarrow D_s^-\pi^+$	13264021	7942	$0.1531 \pm 0.0017$
$\Lambda_b^0 \rightarrow \Lambda_c^+\pi^-$	15164001	325	$0.0155 \pm 0.0009$
$B^0 \rightarrow D^\mp K^{*\pm}$	11164470	361	$0.0358 \pm 0.0019$

Table 4.8 – Samples of simulated data used in the analysis, with the numbers of candidates  $N_{\text{sel}}$  after the final selection, and the selection efficiencies. Efficiencies include generator level, trigger, stripping, offline selection and tagging efficiencies. The  $B^0 \rightarrow D^\mp\pi^\pm$  signal sample is generated with the parameters given in Appendix M.

## 951 4.2 Simulation and expected sample composition

952 Simulated samples are used to (i) gain a detailed overview of all sources of  $b$ -hadron  
 953 decays that contribute to the sample and (ii) model the relevant distributions studied in  
 954 the analysis. Simulated data undergoes the same reconstruction and selection as real data.  
 955 Each sample is split into four subsamples according to magnet polarity (up or down) and  
 956 year of data taking (2011 or 2012), in proportions similar to those present in real data.

957 The simulated samples used are listed in Table 4.8, together with the number of true  
 958 signal events passing the final selection and the corresponding total efficiencies. The PID  
 959 requirements on the bachelor pion are not applied in order to compute these efficiencies.

960

### 961 4.2.1 PIDK correction

962 The PIDK distributions in data and MC differ. To correct for that, the PIDK distributions  
 963 in MC are resampled using the binned PIDK probability density functions of dedicated  
 964 calibration samples. These calibration samples consist of kinematically clean  $D^{*+} \rightarrow$   
 965  $D^0(\rightarrow K^-\pi^+)\pi^+$  decays, for which no requirement on RICH information is applied in  
 966 the reconstruction.

967 The need for this resampling is due to the fact that, if the same cut is applied on data and  
 968 MC, the resulting distributions in other observables may differ if the PIDK distributions  
 969 in data and MC are different. Moreover, a correct PIDK distribution in MC allows the  
 970 proper estimation (on MC) of the efficiency or misidentification rate for a given PIDK  
 971 cut, which is an essential ingredient in the fit to the  $B^0$  invariant mass distribution (as  
 972 described in Sec. 4.3).

973 The following strategy is adopted. A two-dimensional binning in momentum,  $p$ , and

## 4.2. Simulation and expected sample composition

---

974 pseudorapidity,  $\eta$ , is defined. For each bin, the corresponding PID $K$  distribution in the  
975 calibration sample is built and for each event in the MC sample, a random PID $K$  value  
976 is sampled from the PID $K$  distribution associated with the corresponding bin in the  
977 calibration sample. More details are given in Appendix D.

978 Because of the  $\Lambda_c^\mp$  veto described in Sec. 4.1.3, the PID $p$  variable for the  $D^\mp$  daughter  
979 particles is resampled as well in a similar manner using  $\Lambda^0 \rightarrow p\pi^-$  decays as calibration  
980 channel.

981 The nominal binning used for the PID resampling is the following:

- 982     • momentum: 100 uniform bins between 2000 MeV/ $c$  and 200000 MeV/ $c$ , two equal  
983       bins between 200000 MeV/ $c$  and 300000 MeV/ $c$ ;  
984     • pseudorapidity: one bin between 1.5 and 1.55, 69 uniform bins between 1.55 and  
985       5.0

986 In order to check the robustness of the method and evaluate a systematic uncertainties,  
987 two alternative binning schemes are defined:

- 988     • *narrow* binning: the number of bins in the uniform binning parts for  $p$  and  $\eta$  are  
989       increased to 140 and 80, respectively;  
990     • *wide* binning: the number of bins in the uniform binning parts for  $p$  and  $\eta$  are  
991       decreased to 60 for both.

992 The PID $K$  variable is resampled for the  $B^0 \rightarrow D^\mp\pi^\pm$  and  $B^0 \rightarrow D^\mp K^\pm$  Monte Carlo  
993 samples using these two alternative schemes as well. The result of this resampling is  
994 shown in Fig. 4.14 for  $B^0 \rightarrow D^\mp\pi^\pm$  decays. The effect on all the other physics background  
995 is expected to be very small and thus neglected, since they are all located in different  
996 regions of the  $D^\mp\pi^\pm$  invariant mass. More details on this will be given in Sections 4.3  
997 and 5.5.

### 998 4.2.2 Surviving physics backgrounds

999 Some physics background candidates that survive the selection chain described in the  
1000 previous section are expected. In the pion sample, these are:

- 1001     •  $B^0 \rightarrow D^\mp K^\pm$ : Peaking background due to the bachelor kaon being wrongly  
1002       identified as a pion.  
1003     •  $B^0 \rightarrow D^\mp\rho^\pm(\rightarrow \pi^\pm\pi^0)$ : Low mass background due to a missing neutral pion in  
1004       the reconstruction.

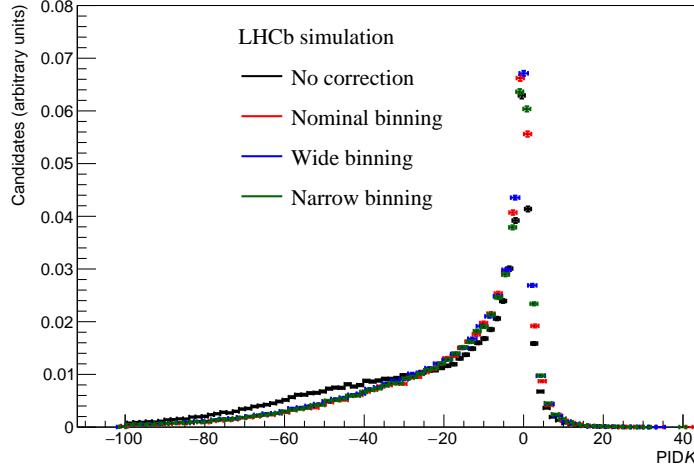


Figure 4.14 –  $\text{PID}_K$  distribution for simulated  $B^0 \rightarrow D^\mp\pi^\pm$  decays without resampling (black), after the nominal resampling (red), after the resampling with the wide binning scheme (blue), and after the resampling with the narrow binning scheme (green).

- $B^0 \rightarrow D^\mp*(\rightarrow D^\mp\gamma/\pi^0)\pi^\pm$ : Low mass background due to a missing neutral particle in the reconstruction.

In the kaon sample, the following backgrounds are expected:

- $B^0 \rightarrow D^\mp\pi^\pm$ : Signal candidates having the bachelor pion wrongly identified as a kaon.
- $B^0 \rightarrow D^\mp\rho^\pm(\rightarrow \pi^\pm\pi^0)$ : Low mass background where, in addition to the missing pion in the final state, a reconstructed pion is wrongly identified as a kaon.
- $B^0 \rightarrow D^\mp K^{\pm*}(\rightarrow \pi^0 K^\pm)$ : Low mass background where the neutral pion is missing in the reconstruction.

The background fractions expected in the pion sample with respect to the  $B^0 \rightarrow D^\mp\pi^\pm$  signal are reported in Table 4.9. These fractions are computed using the branching fractions of the expected decay as inputs and from the ratio of efficiencies estimated from MC and corrected as described in Sec. 4.2.1. Where relevant we consider also the ratio of the fragmentation probabilities of  $b$  quarks to different  $b$  hadrons, which are  $33.9 \pm 9$ ,  $21.2 \pm 6.9$  and  $11.1 \pm 1.4$  for  $B^0$ ,  $B_s^0$  and  $\Lambda_b^0$ , respectively [32]. These expectations will be compared with the results from the fit to data described in the next section.

The  $B_s^0 \rightarrow D_s^- \pi^+$  and  $\Lambda_b^0 \rightarrow \Lambda_c^+(\rightarrow K^- \pi^+ p) \pi^-$  backgrounds are suppressed to a negligible fraction by the offline selection described in Sec. 4.1, and are thus ignored in the description

## 4.2. Simulation and expected sample composition

---

Decay	$\mathcal{B}$ from Ref. [32] [%]	$\epsilon_{\text{bkg}}$ [%]	$f_{\text{bkg}}$ [%]
$B^0 \rightarrow D^\mp K^\pm$	$0.00186 \pm 0.00020$	$0.684 \pm 0.008$	$2.61 \pm 0.31$
$B^0 \rightarrow D^\mp \rho^\pm (\rightarrow \pi^\pm \pi^0)$	$0.071 \pm 0.011$	$0.1149 \pm 0.0024$	$16.7 \pm 2.8$
$B^0 \rightarrow D^{\mp*} (\rightarrow D^\mp \gamma / \pi^0) \pi^\pm$	$0.0080 \pm 0.0004$	$0.705 \pm 0.006$	$11.6 \pm 0.8$
$\Lambda_b^0 \rightarrow \Lambda_c^+ (K^- \pi^+ p) \pi^-$	$0.032 \pm 0.004$	$0.0150 \pm 0.0008$	$0.62 \pm 0.24$
$B_s^0 \rightarrow D_s^- \pi^+$	$0.0164 \pm 0.0014$	$0.1493 \pm 0.0017$	$1.64 \pm 0.32$

Table 4.9 – Background contributions expected in the pion sample. Each fraction  $f_{\text{bkg}}$  is relative to the  $B^0 \rightarrow D^\mp \pi^\pm$  yield. The  $B^0 \rightarrow D^\mp \pi^\pm$  branching ratio and total selection efficiency in the pion sample are  $(0.254 \pm 0.014)\%$  [32] and  $(1.924 \pm 0.006)\%$ , respectively.

1023 of the sample composition. Moreover, in the kaon sample, the  $B^0 \rightarrow D^{\mp*} \pi^\pm$  and  
 1024  $B^0 \rightarrow D^{\mp*} K^\pm$  components, which are expected to be negligible, are ignored as well.  
 1025 More precisely, these components are taken into account by the PDF describing  $B^0 \rightarrow$   
 1026  $D^\mp K^{\pm*} (\rightarrow \pi^0 K^\pm)$ , since they are expected to sit in the same mass region.

### 1027 4.3 Fits to the $B^0$ invariant mass

1028 The *sPlot* technique [74] is applied in order to statistically isolate the signal contribution  
 1029 for the subsequent decay time fit. The  $D^\mp\pi^\pm$  invariant mass, where the  $D^\mp$  mass is  
 1030 constrained to its known value in order to improve the mass resolution, is adopted as  
 1031 discriminating observable thanks to its small correlation with the  $B^0$  decay time (see  
 1032 Appendix E).

In a first step, a binned extended maximum likelihood fit (“Fit A”) is performed in order to define the PDFs describing the signal and background components. The choice of a binned fit is justified by the very high statistics of the data sample. The invariant mass range of the fit is [5090, 6000] MeV/ $c^2$ . Only tagged candidates are considered, *i.e.* candidates with at least one nonzero tagging decision from the OS or SS taggers. The reason for this is that untagged candidates do not contribute to the sensitivity on the  $CP$  coefficients. The fit is performed simultaneously on the pion and kaon samples (see Sec. 4.1.2). This approach is adopted in order to control the contamination from the  $B^0 \rightarrow D^\mp K^\pm$  background in the pion sample. The number of  $B^0 \rightarrow D^\mp X^\pm$  candidates in the  $Y$  sample (with  $X, Y = \pi, K$ ),  $N_{B^0 \rightarrow DX}^Y$ , can be defined via the following relations:

$$N_{B^0 \rightarrow D\pi}^K = \frac{\epsilon_{\text{PID}}(B^0 \rightarrow D\pi)_K}{\epsilon_{\text{PID}}(B^0 \rightarrow D\pi)_\pi} \times N_{B^0 \rightarrow D\pi}^\pi = \frac{1 - \epsilon_{\text{PID}}(B^0 \rightarrow D\pi)_\pi}{\epsilon_{\text{PID}}(B^0 \rightarrow D\pi)_\pi} \times N_{B^0 \rightarrow D\pi}^\pi, \quad (4.1)$$

$$N_{B^0 \rightarrow DK}^\pi = \frac{\epsilon_{\text{PID}}(B^0 \rightarrow DK)_\pi}{\epsilon_{\text{PID}}(B^0 \rightarrow DK)_K} \times N_{B^0 \rightarrow DK}^K = \frac{1 - \epsilon_{\text{PID}}(B^0 \rightarrow DK)_K}{\epsilon_{\text{PID}}(B^0 \rightarrow DK)_K} \times N_{B^0 \rightarrow DK}^K. \quad (4.2)$$

1033 The quantities  $\epsilon_{\text{PID}}(B^0 \rightarrow DX)_Y$  are the fractions of true  $B^0 \rightarrow D^\mp X^\pm$  decays that are  
 1034 selected in the  $Y$  sample by applying the corresponding PID $K$  cut. These fractions (or  
 1035 efficiencies) are estimated on  $B^0 \rightarrow D^\mp\pi^\pm$  and  $B^0 \rightarrow D^\mp K^\pm$  MC samples where the  
 1036 PID $K$  distributions are resampled from calibration data, as described in Sec. 4.2.1. A  
 1037 systematic uncertainty for these efficiencies is estimated by taking the largest discrepancy  
 1038 between the nominal value and the result obtained with the narrow and wide binning  
 1039 schemes introduced in Sec. 4.2.1. The results of these estimations are reported in  
 1040 Table 4.10.

1041 Finally, an unbinned extended maximum likelihood fit (“Fit B”) is performed on data  
 1042 using the reduced mass interval [5220, 5600] MeV/ $c^2$  in order to extract *sWeights*. In this  
 1043 second fit, all the parameters are fixed to the values found in Fit A, except for the signal  
 1044 and total background yields. The reduced mass window avoids diluting the *sWeights*  
 1045 with background candidates having an invariant mass falling outside this window. This  
 1046 has the added advantage of reducing the dataset size used in the decay time fit.

<sup>1047</sup> 4.3.1 Probability density functions

<sup>1048</sup> The PDFs used to describe both the pion and kaon sample components in Fit A are first  
<sup>1049</sup> estimated on MC samples. The parameters of the combinatorial background PDFs are  
<sup>1050</sup> instead determined directly on data. The PDFs used for the pion sample are:

- <sup>1051</sup> •  $B^0 \rightarrow D^\mp \pi^\pm$ : sum of a double-sided Hypatia [75] and a Johnson SU [76] function  
<sup>1052</sup> ( $\text{PDF}_{B^0 \rightarrow D\pi}^\pi$ ).
- <sup>1053</sup> •  $B^0 \rightarrow D^\mp K^\pm$ : double-sided Hypatia function ( $\text{PDF}_{B^0 \rightarrow DK}^\pi$ ).
- <sup>1054</sup> •  $B^0 \rightarrow D^\mp \rho^\pm$ : Johnson SU function ( $\text{PDF}_{B^0 \rightarrow D\rho}^\pi$ ).
- <sup>1055</sup> •  $B^0 \rightarrow D^\mp \pi^\pm$ : sum of a single-sided Crystal Ball function [77] and a Gaussian  
<sup>1056</sup> function ( $\text{PDF}_{B^0 \rightarrow D^*\pi}^\pi$ ).
- <sup>1057</sup> • Background: sum of an exponential function and a constant offset.

<sup>1058</sup> For the kaon sample they are:

- <sup>1059</sup> •  $B^0 \rightarrow D^\mp \pi^\pm$ : double-sided Hypatia function ( $\text{PDF}_{B^0 \rightarrow D\pi}^K$ ).
- <sup>1060</sup> •  $B^0 \rightarrow D^\mp K^\pm$ : single-sided Hypatia function ( $\text{PDF}_{B^0 \rightarrow DK}^K$ ).
- <sup>1061</sup> •  $B^0 \rightarrow D^\mp \rho^\pm$ : double Gaussian function ( $\text{PDF}_{B^0 \rightarrow D\rho}^K$ ).
- <sup>1062</sup> •  $B^0 \rightarrow D^\mp K^{\pm*}$ : Gaussian function ( $\text{PDF}_{B^0 \rightarrow DK^*}^K$ ).
- <sup>1063</sup> • Background: sum of an exponential function and a constant offset.

<sup>1064</sup> The definitions of all the PDFs listed above are reported in Appendix F. The fits to the  
<sup>1065</sup> MC samples are shown in Figs. 4.15 and 4.16. The parameters obtained from these fits  
<sup>1066</sup> that are then fixed in the data fits are listed in Tables 4.11 and 4.12.

Table 4.10 – Fractions of true  $B^0 \rightarrow D^\mp \pi^\pm$  and  $B^0 \rightarrow D^\mp K^\pm$  decays that are selected in the  $\pi$  or  $K$  sample.

Decay	PID $K$ requirement	fraction
$B^0 \rightarrow D^\mp \pi^\pm$	< 5 ( $\pi$ sample)	$0.9790 \pm 0.0040(\text{stat}) \pm 0.0004(\text{syst})$
$B^0 \rightarrow D^\mp \pi^\pm$	> 5 ( $K$ sample)	$0.0211 \pm 0.0005(\text{stat}) \pm 0.0004(\text{syst})$
$B^0 \rightarrow D^\mp K^\pm$	< 5 ( $\pi$ sample)	$0.373 \pm 0.005(\text{stat}) \pm 0.008(\text{syst})$
$B^0 \rightarrow D^\mp K^\pm$	> 5 ( $K$ sample)	$0.627 \pm 0.007(\text{stat}) \pm 0.010(\text{syst})$

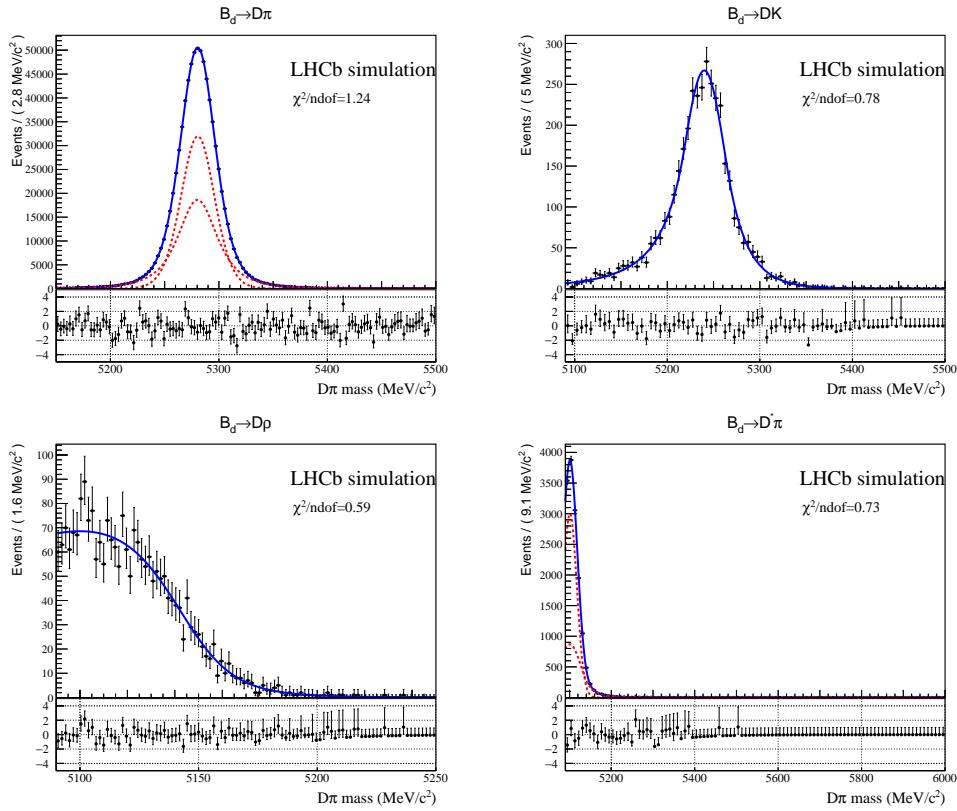


Figure 4.15 – Fits to MC samples of the PDFs describing the physical composition of the  $\pi$  sample.

### <sup>1067</sup> 4.3.2 Fit to data

In order to perform Fit A, two fitting functions are defined as follows:

$$\begin{aligned} f_\pi(m) &= N_{B^0 \rightarrow D\pi}^\pi \text{PDF}_{B^0 \rightarrow D\pi}^\pi + N_{B^0 \rightarrow DK}^\pi \text{PDF}_{B^0 \rightarrow DK}^\pi \\ &+ N_{B^0 \rightarrow D^*\pi}^\pi \text{PDF}_{B^0 \rightarrow D^*\pi}^\pi + N_{B^0 \rightarrow D\rho}^\pi \text{PDF}_{B^0 \rightarrow D\rho}^\pi \\ &+ N_{\text{comb}}^\pi \text{PDF}_{\text{comb}}^\pi. \end{aligned} \quad (4.3)$$

$$\begin{aligned} f_K(m) &= N_{B^0 \rightarrow DK}^K \text{PDF}_{B^0 \rightarrow DK}^K + N_{B^0 \rightarrow D\pi}^K \text{PDF}_{B^0 \rightarrow D\pi}^K \\ &+ N_{B^0 \rightarrow D^*K}^K \text{PDF}_{B^0 \rightarrow D^*K}^K + N_{B^0 \rightarrow D\rho}^K \text{PDF}_{B^0 \rightarrow D\rho}^K \\ &+ N_{\text{comb}}^K \text{PDF}_{\text{comb}}^K. \end{aligned} \quad (4.4)$$

### 4.3. Fits to the $B^0$ invariant mass

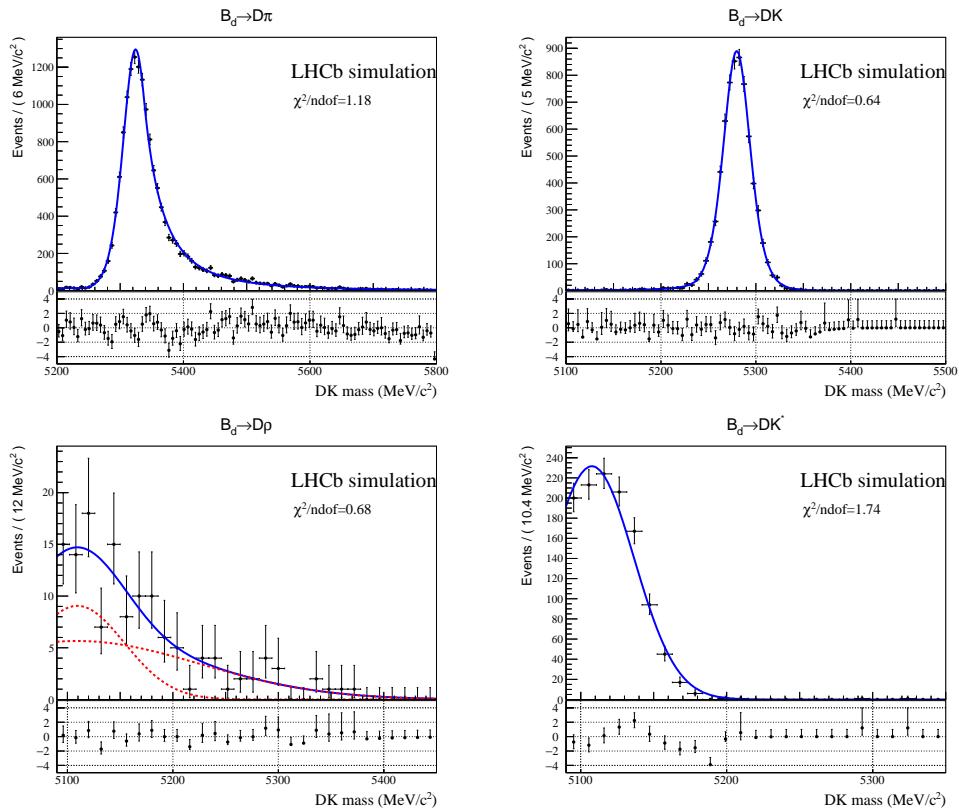


Figure 4.16 – Fits to MC samples of the PDFs describing the physical composition of the  $K$  sample.

Two extended likelihood functions are defined using data and PDFs related to both samples:

$$\mathcal{L}_x = \frac{e^{-N_{x,\text{exp}}} (N_{x,\text{exp}})^{N_{x,\text{obs}}}}{N_{x,\text{obs}}!} \prod_{i=1}^{N_{x,\text{obs}}} \frac{f_x(m_i)}{N_{x,\text{exp}}}, \quad x = \pi, K, \quad (4.5)$$

1068 where  $N_{\pi,\text{exp}} = N_{B^0 \rightarrow D\pi}^\pi + N_{B^0 \rightarrow DK}^\pi + N_{B^0 \rightarrow D^*\pi}^\pi + N_{B^0 \rightarrow D\rho}^\pi + N_{\text{comb}}^\pi$ ,  $N_{K,\text{exp}} = N_{B^0 \rightarrow DK}^K +$   
1069  $N_{B^0 \rightarrow D\pi}^K + N_{B^0 \rightarrow DK^*}^K + N_{B^0 \rightarrow D\rho}^K + N_{\text{comb}}^K$ , and  $N_{x,\text{obs}}$  is the number of observed candidates  
1070 in the  $x$  sample. The product  $\mathcal{L}_\pi \mathcal{L}_K$  is maximised during the fit.

1071 The following strategy is adopted to perform Fit A:

- 1072 • The mean and width parameters ( $\mu_{B^0 \rightarrow D\pi}^\pi, \sigma H_{B^0 \rightarrow D\pi}^\pi, \sigma J_{B^0 \rightarrow D\pi}^\pi, \mu_{B^0 \rightarrow DK}^K, \sigma_{B^0 \rightarrow DK}^K$ )  
1073 of  $\text{PDF}_{B^0 \rightarrow D\pi}^\pi$  and  $\text{PDF}_{B^0 \rightarrow DK}^K$  are floated in the fit.
- 1074 • The tail parameters ( $a1_{B^0 \rightarrow D\pi}^\pi, a2_{B^0 \rightarrow D\pi}^\pi, n1_{B^0 \rightarrow D\pi}^\pi, n2_{B^0 \rightarrow D\pi}^\pi$ ) of  $\text{PDF}_{B^0 \rightarrow D\pi}^\pi$  are  
1075 constrained in the following way:  $a1_{B^0 \rightarrow D\pi}^\pi, a2_{B^0 \rightarrow D\pi}^\pi$  are set to the values found on  
1076 MC and both multiplied by a floating scale factor  $sa_{B^0 \rightarrow D\pi}^\pi$ ; the same constraint is  
1077 applied to  $n1_{B^0 \rightarrow D\pi}^\pi$  and  $n2_{B^0 \rightarrow D\pi}^\pi$ , where the scale factor is labelled as  $sn_{B^0 \rightarrow D\pi}^\pi$ .

- The yield parameters  $N_{B^0 \rightarrow D\pi}^K$  and  $N_{B^0 \rightarrow DK}^\pi$  are constrained according to Eqs. 4.1 and 4.2. The efficiencies  $\epsilon_{\text{PID}}(B^0 \rightarrow D\pi)_{D\pi}$  and  $\epsilon_{\text{PID}}(B^0 \rightarrow DK)_{DK}$  are Gaussian-constrained independently in the fit, using the values reported in Table 4.10. The yield  $N_{B^0 \rightarrow D\rho}^K$  is fixed to be 0.92 times the yield  $N_{B^0 \rightarrow DK^*}^K$ , the latter being floated in the fit. This is done according to the expected  $B^0 \rightarrow D^\mp\rho^\pm$  to  $B^0 \rightarrow D^\mp K^\pm$  ratio in the kaon sample, which is  $0.92 \pm 0.21$ . All the other yields appearing in Eqs. 4.3 and 4.4 are floated in the fit.
- The mean parameters  $(\mu_{B^0 \rightarrow D^*\pi}^\pi, \mu_{B^0 \rightarrow D\rho}^{K/\pi})$  of  $\text{PDF}_{B^0 \rightarrow D^*\pi}^{\pi/K}$ ,  $\text{PDF}_{B^0 \rightarrow D\rho}^{\pi/K}$ , are constrained to be shifted from  $\mu_{B^0 \rightarrow D\pi}^\pi$  (in the  $\pi$  sample) and  $\mu_{B^0 \rightarrow DK}^K$  (in the  $K$  sample) by the same amount found in MC. The shift of the component with respect to the  $B^0 \rightarrow D^\mp\pi^\pm$  ( $B^0 \rightarrow D^\mp K^\pm$ ) peak in the  $\pi$  ( $K$ ) sample is denoted as  $\Delta\mu_{\text{comb}}^{K/\pi}$ . The mean parameters  $(\mu_{B^0 \rightarrow DK^*}^K, \mu_{B^0 \rightarrow DK}^\pi, \mu_{B^0 \rightarrow D\pi}^K)$  of  $\text{PDF}_{B^0 \rightarrow DK^*}^K$ ,  $\text{PDF}_{B^0 \rightarrow DK}^\pi$ ,  $\text{PDF}_{B^0 \rightarrow D\pi}^K$  are floated in the fit.
- The exponent parameters  $(c_{\text{comb}}^{\pi/K})$  and fractions  $(f_{\text{comb}}^{\pi/K})$  of  $\text{PDF}_{\text{comb}}^{\pi/K}$  are floated in the fit.

The projections of the fitted  $f_\pi$  and  $f_K$  in the  $D^\mp\pi^\pm$  and  $D^\mp K^\pm$  invariant mass observables (Fit A) are shown in Fig. 4.17, for the  $\pi$  and  $K$  data samples, respectively. A list of all the parameters fixed in Fit A is given in Tables 4.11 and 4.12. The fitted parameters (including yields and PID efficiencies) are listed in Table 4.13.

As cross-check, the fitted yields of  $B^0 \rightarrow D^\mp K^\pm$ ,  $B^0 \rightarrow D^\mp\rho^\pm$  and  $B^0 \rightarrow D^*\pi^\pm$  in the pion sample are compared with the expected yields, which are obtained, for each background, by multiplying the fitted  $B^0 \rightarrow D^\mp\pi^\pm$  yield in the pion sample by the  $f_{\text{bkg}}$  fractions given in Table 4.9. These yields are reported in Table 4.14. There is a full agreement between expected and observed yields for the  $B^0 \rightarrow D^\mp\rho^\pm$  and  $B^0 \rightarrow D^*\pi^\pm$  components, while the agreement for the  $B^0 \rightarrow D^\mp K^\pm$  is at the level of  $2.5\sigma$ .

### 4.3.3 *s*Weight calculation

After that Fit A is performed, all the floating shape parameters in  $f_\pi(m)$  are fixed, all the background components in the  $\pi$  sample are combined into a single PDF, and the  $B$  mass range is restricted to  $[5220, 5600]$  MeV/c<sup>2</sup>. Concretely,  $f_\pi(m)$  is redefined as

$$f_\pi(m) = N_{B^0 \rightarrow D\pi}^\pi \text{PDF}_{B^0 \rightarrow D\pi}^\pi + N_{\text{bkg}}^\pi \text{PDF}_{\text{bkg}}^\pi. \quad (4.6)$$

The  $N_{\text{bkg}}^\pi$  parameter describes the total number of background events in the new range. The  $\text{PDF}_{\text{bkg}}^\pi$  function is defined as:

$$\begin{aligned} \text{PDF}_{\text{bkg}}^\pi = & f_{\text{comb}}^\pi \text{PDF}_{\text{comb}}^\pi + f_{B^0 \rightarrow DK}^\pi \text{PDF}_{B^0 \rightarrow DK}^\pi + f_{B^0 \rightarrow D\rho}^\pi \text{PDF}_{B^0 \rightarrow D\rho}^\pi \\ & + (1 - f_{\text{comb}}^\pi - f_{B^0 \rightarrow DK}^\pi - f_{B^0 \rightarrow D\rho}^\pi) \text{PDF}_{B^0 \rightarrow D^*\pi}^\pi. \end{aligned} \quad (4.7)$$

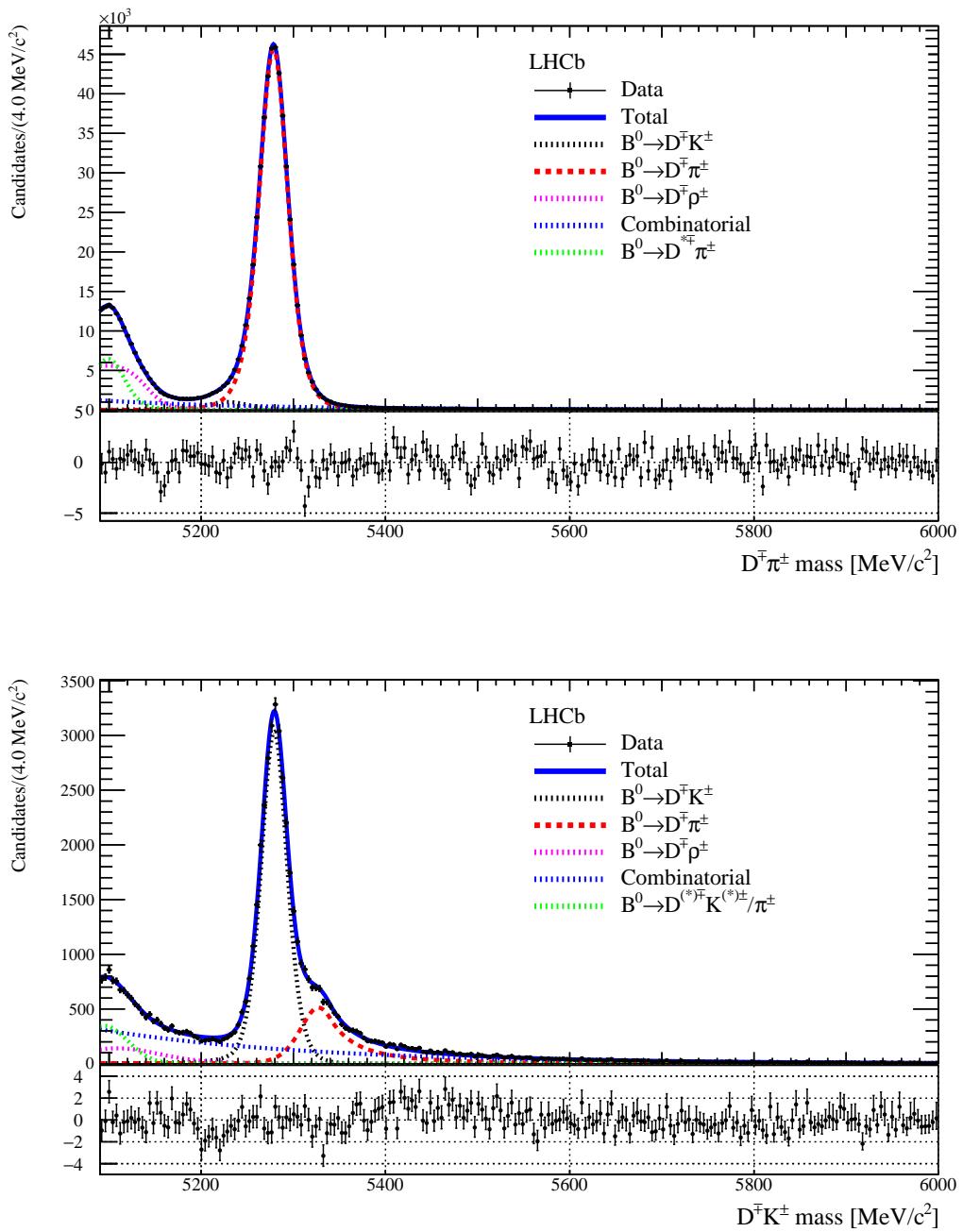


Figure 4.17 – Top:  $D^\mp\pi^\pm$  mass distribution of the  $\pi$  sample. Bottom:  $D^\mp K^\pm$  mass distribution of the  $K$  sample. The result of the simultaneous fit (Fit A) to both samples is superimposed. The plot below each histogram shows the normalised fit residuals (data minus fit divided by fit error).

## Chapter 4. Selection of $B^0 \rightarrow D^\mp \pi^\pm$ decays

---

Table 4.11 – Parameters of  $f_\pi(m)$  fixed or constrained in Fit A. All non-zero values are obtained from the fits to MC samples described in Sec. 4.3.1.

Parameter	Value	Status in Fit A
$a1_{B^0 \rightarrow D\pi}^\pi$	$0.722 \pm 0.091$	constrained
$a2_{B^0 \rightarrow D\pi}^\pi$	$0.96 \pm 0.12$	constrained
$n1_{B^0 \rightarrow D\pi}^\pi$	$5.92 \pm 0.92$	constrained
$n2_{B^0 \rightarrow D\pi}^\pi$	$5.83 \pm 0.38$	constrained
$\beta_{B^0 \rightarrow D\pi}^\pi$	0.0	fixed
$\lambda_{B^0 \rightarrow D\pi}^\pi$	$-1.240 \pm 0.060$	fixed
$\zeta_{B^0 \rightarrow D\pi}^\pi$	0.0	fixed
$f_{B^0 \rightarrow D\pi}^\pi$	$0.436 \pm 0.060$	fixed
$\sigma_{B^0 \rightarrow DK}^\pi$	$23.43 \pm 0.42 \text{ MeV}/c^2$	fixed
$a1_{B^0 \rightarrow DK}^\pi$	$0.898 \pm 0.025$	fixed
$a2_{B^0 \rightarrow DK}^\pi$	$1.092 \pm 0.033$	fixed
$n1_{B^0 \rightarrow DK}^\pi$	$3.83 \pm 0.40$	fixed
$n2_{B^0 \rightarrow DK}^\pi$	$22.0 \pm 7.6$	fixed
$\beta_{B^0 \rightarrow DK}^\pi$	0.0	fixed
$\lambda_{B^0 \rightarrow DK}^\pi$	$-24 \pm 10$	fixed
$\zeta_{B^0 \rightarrow DK}^\pi$	0.0	fixed
$\nu_{B^0 \rightarrow D\rho}^\pi$	$-2.01 \pm 0.15$	fixed
$\mu_{B^0 \rightarrow D\rho}^\pi$	$4.828 \pm 80 \text{ MeV}/c^2$	constrained into $\Delta\mu_{B^0 \rightarrow D\rho}^\pi$
$\sigma_{B^0 \rightarrow D\rho}^\pi$	$550 \pm 190 \text{ MeV}/c^2$	fixed
$\tau_{B^0 \rightarrow D\rho}^\pi$	$1.163 \pm 0.090$	fixed
$\alpha_{B^0 \rightarrow D^*\pi}^\pi$	$-1.443 \pm 0.031$	fixed
$n_{B^0 \rightarrow D^*\pi}^\pi$	$4.65 \pm 0.30$	fixed
$\mu_{B^0 \rightarrow D^*\pi}^\pi$	$5.100.93 \pm 0.23 \text{ MeV}/c^2$	constrained into $\Delta\mu_{B^0 \rightarrow D^*\pi}^\pi$
$\sigma G_{B^0 \rightarrow D^*\pi}^\pi$	$16.52 \pm 0.20 \text{ MeV}/c^2$	fixed
$\sigma C B_{B^0 \rightarrow D^*\pi}^\pi$	$25.84 \pm 0.48 \text{ MeV}/c^2$	fixed
$f_{B^0 \rightarrow D^*\pi}^\pi$	$0.302 \pm 0.011$	fixed

For each background component in the  $\pi$  sample, the fraction  $f_j^\pi$  is determined by the following expression:

$$f_j^\pi = \frac{N_j^\pi \int_{5220 \text{ MeV}/c^2}^{5600 \text{ MeV}/c^2} \text{PDF}_j^\pi dm}{\sum_i N_i^\pi \int_{5220 \text{ MeV}/c^2}^{5600 \text{ MeV}/c^2} \text{PDF}_i^\pi dm}, \quad (4.8)$$

where the indices  $i$  and  $j$  run over the combinatorial,  $B^0 \rightarrow D^\mp K^\pm$  and  $B^0 \rightarrow D^\mp \rho^\pm$  background components in the  $\pi$  sample.

An unbinned extended maximum likelihood fit (Fit B) is then performed to the  $\pi$  sample only. The only floating parameters are the yields  $N_{B^0 \rightarrow D\pi}^\pi$  and  $N_{\text{bkg}}^\pi$ . The result of the fit is reported in Table 4.15. The yields of the signal component and the total background as

Table 4.12 – Parameters of  $f_K(m)$  fixed or constrained in Fit A. All non-zero values are obtained from the fits to MC samples described in Sec. 4.3.1.

Parameter	Value	Status in Fit A
$\sigma_{B^0 \rightarrow D\pi}^K$	$23.97 \pm 0.46 \text{ MeV}/c^2$	fixed
$a1_{B^0 \rightarrow D\pi}^K$	$3.14 \pm 0.14$	fixed
$a2_{B^0 \rightarrow D\pi}^K$	$0.569 \pm 0.039$	fixed
$n1_{B^0 \rightarrow D\pi}^K$	$0.05 \pm 0.11$	fixed
$n2_{B^0 \rightarrow D\pi}^K$	$2.81 \pm 0.12$	fixed
$\beta_{B^0 \rightarrow D\pi}^K$	0.0	fixed
$\lambda_{B^0 \rightarrow D\pi}^K$	$-3.77 \pm 0.57$	fixed
$\zeta_{B^0 \rightarrow D\pi}^K$	0.0	fixed
$\sigma_{B^0 \rightarrow DK}^K$	$17.32 \pm 0.26 \text{ MeV}/c^2$	fixed
$a_{B^0 \rightarrow DK}^K$	$2.34 \pm 0.19$	fixed
$n_{B^0 \rightarrow DK}^K$	$1.56 \pm 0.33$	fixed
$\beta_{B^0 \rightarrow DK}^K$	0.0	fixed
$\lambda_{B^0 \rightarrow DK}^K$	$-3.45 \pm 0.34$	fixed
$\zeta_{B^0 \rightarrow DK}^K$	0.0	fixed
$f_{B^0 \rightarrow D\rho}^K$	$0.58 \pm 0.17$	fixed
$\mu_{B^0 \rightarrow D\rho}^K$	$5109 \pm 24 \text{ MeV}/c^2$	constrained into $\Delta\mu_{B^0 \rightarrow D\rho}^K$
$\sigma1_{B^0 \rightarrow D\rho}^K$	$117 \pm 18 \text{ MeV}/c^2$	fixed
$\sigma2_{B^0 \rightarrow D\rho}^K$	$45 \pm 16 \text{ MeV}/c^2$	fixed

1109 obtained in Fit B are compatible with the yields of the signal and the sum of the yields of  
 1110 each background as obtained in Fit A after rescaling the PDFs in the Fit B mass range.

1111 Fit B is used as starting point to apply the *sPlot* technique and extract *sWeights* used  
 1112 to subtract the total background component from the  $\pi$  sample. The projection of the  
 1113 fitted  $f_\pi(m)$  in Fit B, and a comparison between the weighted and unweighted datasets  
 1114 projected over the  $B^0$  decay time and  $D^\mp$  invariant mass observables, are reported in  
 1115 Fig. 4.18.

#### 1116 4.3.4 Fits of subsamples

1117 In order to validate the data sample, selection and fit procedure, Fit A is repeated in  
 1118 smaller subsamples. These subsamples are divided per year of data taking (2011, 2012),  
 1119 magnet polarity (up, down) and final state ( $D^+\pi^-$ ,  $D^-\pi^+$ ).

1120 In order to cope with the reduced statistics in the 2011 subsample, the combinatorial  
 1121 background PDF of the  $K$  sample ( $\text{PDF}_{\text{comb}}^K$ ) is taken as a simple exponential (instead of  
 1122 an exponential plus a constant function).

1123 The projections of the fitted PDFs describing the pion and kaon samples for each data

Table 4.13 – Results of Fit A.

Parameter	Fitted value	Reference
$\mu_{B^0 \rightarrow DK}^\pi$	$5\,228.62 \pm 0.92$	$5\,239.96 \pm 0.52$
$\sigma_{B^0 \rightarrow DK}^K$	$17.17 \pm 0.15$	$17.32 \pm 0.26$
$\mu_{B^0 \rightarrow DK^*}^K$	$5\,094.8 \pm 3.9$	$5\,107.7 \pm 2.2$
$\sigma_{B^0 \rightarrow DK^*}^K$	$25.5 \pm 2.6$	$28.9 \pm 1.2$
$c1_{\text{comb}}^\pi$	$-0.00576 \pm 0.00017$	/
$c2_{\text{comb}}^\pi$	$-0.0010 \pm 0.0010$	/
$f_{\text{comb}}^\pi$	$0.899 \pm 0.025$	/
$c_{\text{comb}}^K$	$-0.004397 \pm 0.000066$	/
$\mu_{B^0 \rightarrow DK}^K$	$5\,279.19 \pm 0.14$	$5\,279.84 \pm 0.20$
$\mu_{B^0 \rightarrow D\pi}^\pi$	$5\,278.360 \pm 0.032$	$5\,280.291 \pm 0.025$
$sa_{B^0 \rightarrow D\pi}^\pi$	$0.684 \pm 0.022$	/
$sn_{B^0 \rightarrow D\pi}^\pi$	$2.71 \pm 0.80$	/
$\sigma H_{B^0 \rightarrow D\pi}^\pi$	$37.69 \pm 0.69$	$43.9 \pm 4.9$
$\sigma J_{B^0 \rightarrow D\pi}^\pi$	$17.01 \pm 0.17$	$16.82 \pm 0.15$
$\mu_{B^0 \rightarrow D\pi}^K$	$5\,327.32 \pm 0.78$	$5\,324.37 \pm 0.44$
$\epsilon_{\text{PID}}(B^0 \rightarrow DK)_K$	$0.6197 \pm 0.0079$	$0.627 \pm 0.013$
$\epsilon_{\text{PID}}(B^0 \rightarrow D\pi)_\pi$	$0.98048 \pm 0.00041$	$0.979 \pm 0.004$
$N_{B^0 \rightarrow DK}^K$	$28\,820 \pm 242$	/
$N_{B^0 \rightarrow DK^*}^K$	$3\,164 \pm 110$	/
$N_{B^0 \rightarrow D\rho}^\pi$	$73\,766 \pm 1239$	/
$N_{B^0 \rightarrow D^*\pi}^\pi$	$52\,494 \pm 819$	/
$N_{\text{comb}}^K$	$17\,469 \pm 341$	/
$N_{\text{comb}}^\pi$	$56\,230 \pm 1336$	/
$N_{B^0 \rightarrow D\pi}^\pi$	$483\,398 \pm 1040$	/

Table 4.14 – Expected and fitted yields for the physical background components in the pion sample.

Decay	Expected yield [ $10^4$ ]	Fitted yield [ $10^4$ ]
$B^0 \rightarrow D^\mp K^\pm$	$1.26 \pm 0.15$	$1.65 \pm 0.05$
$B^0 \rightarrow D^\mp \rho^\pm$	$8.1 \pm 1.4$	$7.38 \pm 0.12$
$B^0 \rightarrow D^{\mp*} \pi^\pm$	$5.6 \pm 0.4$	$5.25 \pm 0.08$

Table 4.15 – Results of Fit B shown in the second column. The third column shows the yields calculated by integrating the PDFs fitted in Fit A in the mass range used for Fit B.

Parameter	Fitted value (Fit B)	Fitted value (Fit A)
$N_{B^0 \rightarrow D\pi}^\pi$	$479\,045 \pm 732$	$483\,398 \pm 1040$
$N_{\text{bkg}}^\pi$	$34\,381 \pm 300$	$34\,615 \pm 664$

### 4.3. Fits to the $B^0$ invariant mass

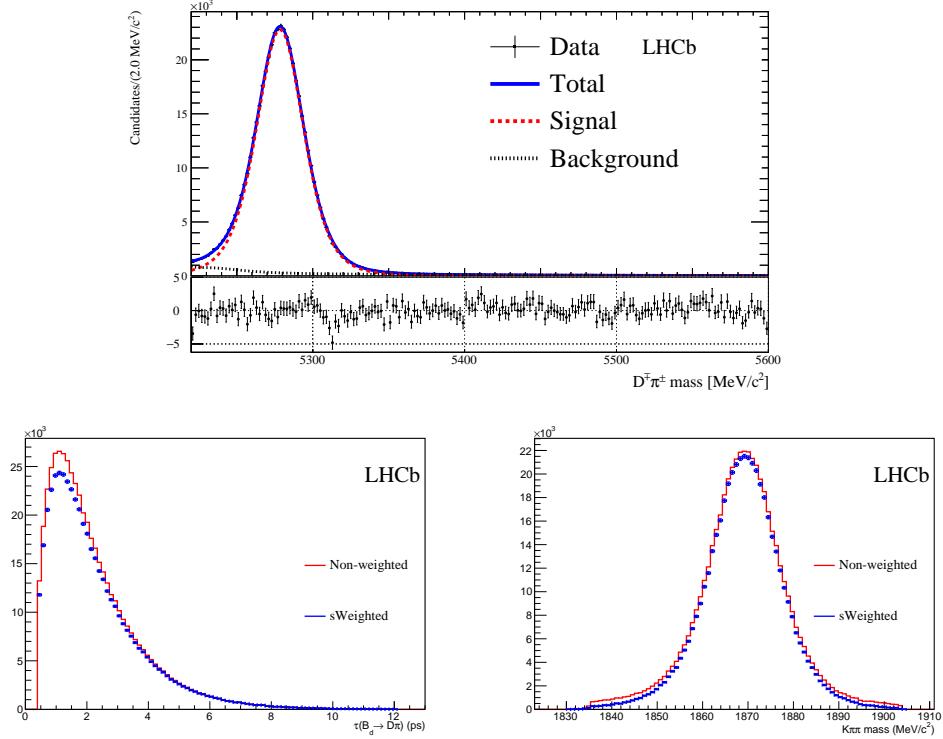


Figure 4.18 – Top:  $D^\mp\pi^\pm$  mass distribution of the  $\pi$  sample with the results of Fit B superimposed. The plot below the histogram shows the normalised fit residuals (data minus fit divided by fit error). Bottom:  $B^0$  decay time (left) and  $K^\mp\pi^\pm\pi^\pm$  mass (right) distributions of the  $\pi$  sample in the  $B^0$  mass region  $[5220, 5600]$   $\text{MeV}/c^2$ , with (blue) and without (red) *sWeights* from Fit B.

1124 subsample are shown in Figs. 4.19 and 4.20, respectively.

1125 Fit B strategy is also repeated exactly as before for each subsample. The corresponding  
 1126 signal and background fitted yields are listed in Table 4.16. The sum of the yields for  
 1127 each subsample is compatible with the total signal yield in the total sample (reported in  
 1128 Table 4.15), which is  $(4.790 \pm 0.007) \times 10^5$ . The asymmetry between the yields of the  
 1129  $D^- \pi^+$  and  $D^+ \pi^-$  samples is  $0.0100 \pm 0.0015$ , which is in agreement with the detection  
 1130 asymmetry between  $\pi^+$  and  $\pi^-$  obtained in this analysis and by previous measurements  
 1131 (more details given in Sec. 5.3). The ratio between the fitted yields on the 2011 and  
 1132 2012 samples is compatible with the different collected luminosities and data taking  
 1133 conditions between the two years (twice as much more luminosity is collected in 2012  
 1134 compared to 2011, and the  $b$ -production cross section is increased by a factor  $\sim 8/7$  in  
 1135 2012 because of the increase of the centre-of-mass energy). Moreover, the ratio between  
 1136 the yields obtained with the magnet up and down samples is in agreement with the ratio  
 1137 of luminosities collected with the two magnet polarities (same luminosity in 2012, +30 %  
 1138 more magnet down data in 2011).

## Chapter 4. Selection of $B^0 \rightarrow D^\mp \pi^\pm$ decays

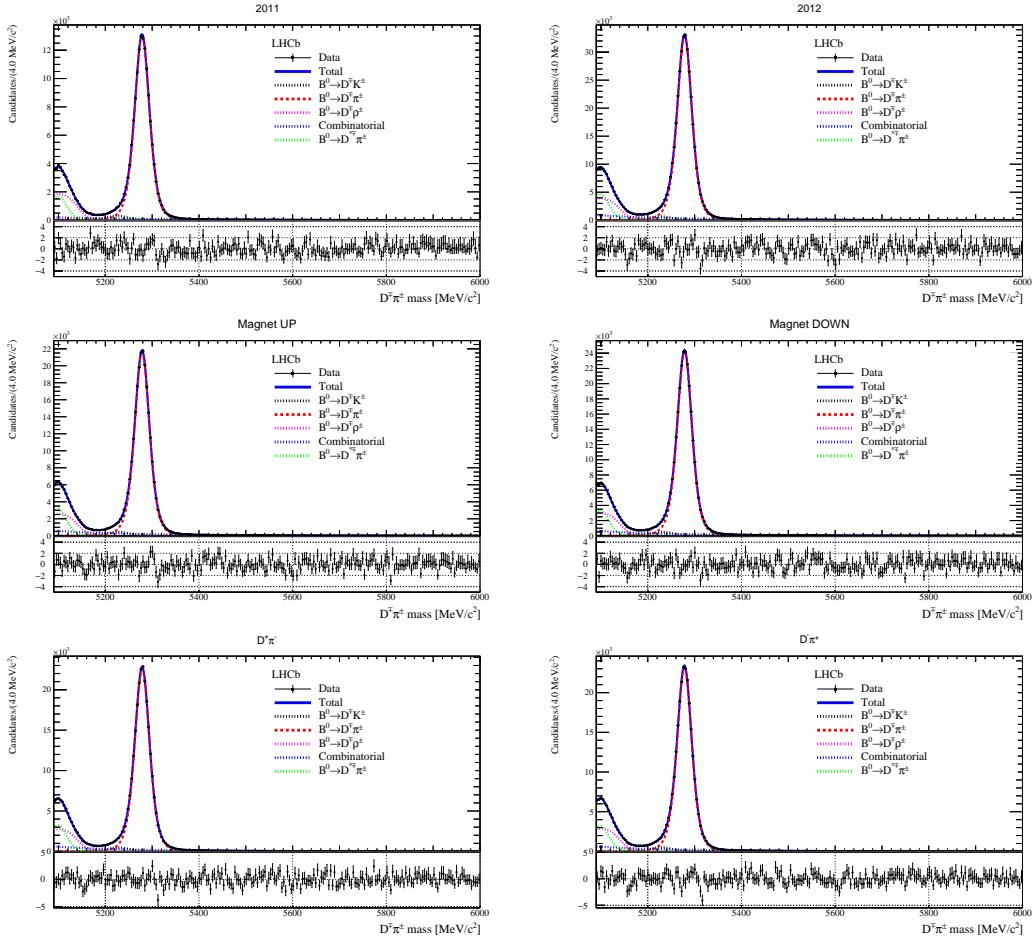


Figure 4.19 – Projections of the fitted PDFs describing the pion sample for each data subsample.

Table 4.16 – Signal yields (in units of  $10^5$ ) in the pion sample for each subsample, obtained from Fit B.

2011	2012	Sum
$1.383 \pm 0.004$	$3.424 \pm 0.006$	$4.807 \pm 0.007$
Magnet Up	Magnet Down	Sum
$2.263 \pm 0.005$	$2.523 \pm 0.005$	$4.786 \pm 0.007$
$D^- \pi^+$	$D^+ \pi^-$	Sum
$2.421 \pm 0.005$	$2.373 \pm 0.005$	$4.794 \pm 0.007$

### 4.3. Fits to the $B^0$ invariant mass

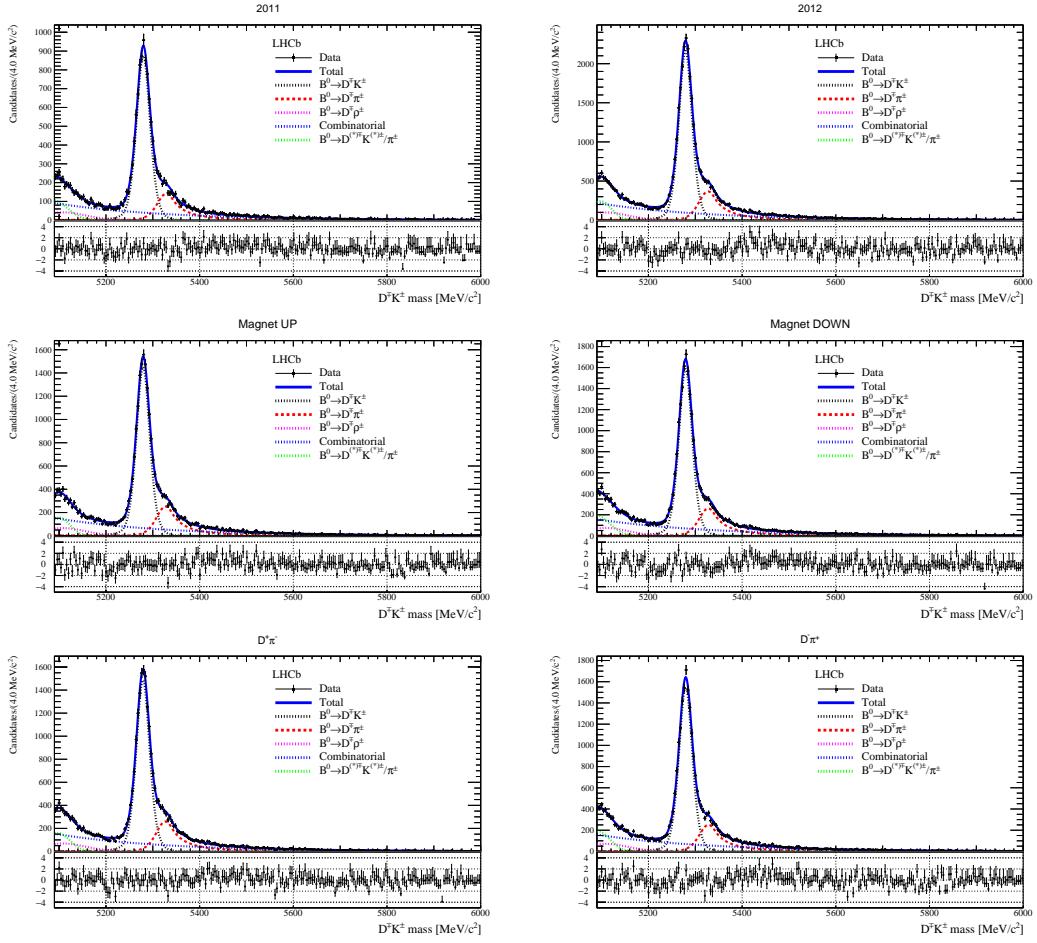


Figure 4.20 – Projections of the fitted PDFs describing the kaon sample for each data subsample.



## 5 Measurement of $CP$ violation in $^{1140} B^0 \rightarrow D^\mp \pi^\pm$ decays

## 1141 5.1 Decay-time resolution

1142 The decay-time resolution is determined from a sample of fake  $B^0$  candidates formed  
1143 from a prompt  $D^\pm$  candidate combined with a track originating from the PV. This is  
1144 sample is referred to as “ $D^\pm$ +track”. The candidates are taken from the B02DKLTUBD2HHH  
1145 stripping line. They candidates are subjected to the same offline selection as that of  
1146 the signal sample without a BDT cut and with two additional requirements: a single  
1147 reconstructed PV per event is required in order to reduce wrong PV associations, and  
1148 the  $D^\pm$  IP  $\chi^2$  with respect to the PV is required to be less than 9 to reduce the  $D^\mp$   
1149 contribution from  $B^0$  decays. The combined stripping and offline selection yields 51053  
1150 candidates. True  $D^\pm$ +track candidates are unfolded from combinatorial background and  
1151 nonresonant decays by means of *sWeights* computed via a fit to the  $K^\pm\pi^\mp\pi^\mp$  invariant  
1152 mass distribution.

### 1153 5.1.1 Companion track momentum reweighting

1154 The decay time resolution is found to be dependent upon the companion track  $p_T$  which  
1155 is considerably lower on average for the  $D^\pm$ +track candidates than it is for genuine  
1156  $B^0 \rightarrow D^\pm\pi^\mp$  signal. This is corrected for by reweighting the prompt sample to have the  
1157 same  $\ln(p_T)$  distribution as that of the  $B^0 \rightarrow D^\mp\pi^\pm$  signal. The logarithm is taken to  
1158 compress the high- $p_T$  tails, and a binning scheme is chosen to have an equal number  
1159 of signal events per bin. The  $\ln(p_T)$  spectra for signal, prompt, and reweighted prompt  
1160 candidates is shown in Fig. 5.1. By following the steps described in Sec. 5.1.2, prior to  
1161 reweighting the average proper time resolution is determined to be  $\sim 71$  fs, and after  
1162 reweighting the resolution is found to be consistent with the value of 54 fs that was  
 obtained in other  $B$  meson time-dependent analyses [78].

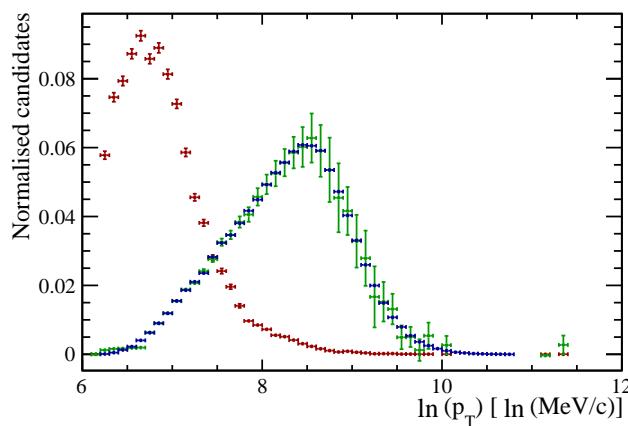


Figure 5.1 – Normalised  $\log(p_T)$  distributions of *sWeighted*  $B^0 \rightarrow D^\pm\pi^\mp$  signal (blue), prompt  $D^\pm$ + track before reweighting (red) and after reweighting (green).

1163 **5.1.2 Resolution determination from decay-time error parameterisa-**  
1164 **tion**

In order to study potential second order corrections to the decay-time error distribution, fits to the decay-time distribution of the  $D^\pm + \text{track}$  sample in bins of the per-event decay time error are performed. The decay-time error is obtained from DTF by propagating the uncertainty on the  $B^0$  four-momentum. The binning scheme is chosen such that the sum of the *sWeights* associated to  $D^\pm + \text{track}$  candidates in each bin is equal. The fit is similar to that used to determine the resolution in Ref. [79], consisting of three components: a delta function convolved with a Gaussian resolution function accounts for the genuine prompt  $D^\pm + \text{track}$  component; a pair of exponential functions convolved with the same Gaussian function accounts for signal candidates coming from  $b$ -hadron decays, and a Gaussian function with a large width accounts for wrong-PV associated backgrounds. The time constant of the exponentials and the mean of the wrong-PV component are fixed from a global fit to the sample, while the mean and width of the resolution, the width of the wrong PV component and the relative fractions of the prompt, wrong-PV and from- $b$  components are all free parameters in the fits to each decay-time error bin. A likelihood fit is performed in 20 bins of the decay-time error from which the measured resolution  $\langle\sigma\rangle_i$  is determined. The results of these fits are presented in Table 5.1, and a representative fit is shown in Fig. 5.2. A  $\chi^2$  fit is performed to the obtained values of the per-bin average error and resolution of the form:

$$\langle\sigma\rangle_i = \langle\sigma\rangle + p_1 (\langle\delta\rangle_i - \langle\delta\rangle) + p_2 (\langle\delta\rangle_i - \langle\delta\rangle)^2 \quad (5.1)$$

1165 where  $\langle\delta\rangle$  is the average per-event decay time error of the whole (unbinned) sample, while  
1166  $\langle\delta\rangle_i$  is the average per-event decay time error in each bin. In the prompt  $D^\pm + \text{track}$   
1167 sample  $\langle\delta\rangle$  is determined to be  $0.0307 \pm 0.0097$  ps, in good agreement with the *sWeighted*  
1168  $B^0 \rightarrow D^\mp \pi^\pm$  signal sample value of  $0.034 \pm 0.011$  ps. The fit determines the average  
1169 resolution,  $\langle\sigma\rangle$ , in addition to the trend. This fit is shown in Fig. 5.2, the result of which  
1170 is presented in Table 5.2. The global average resolution is determined from this fit to be  
1171  $\langle\sigma\rangle = 0.05491 \pm 0.00038$  ps. The procedure is found to be stable and yields compatible  
1172 results with fits to 10 bins ( $0.05523 \pm 0.00041$  ps) and 30 bins ( $0.05464 \pm 0.00037$  ps).

1173 This method, which accounts for second-order corrections to the decay time error, is used to  
1174 define the width of a single Gaussian in the decay time fit to data,  $\mathcal{R}(t-t') = G(t-t', \langle\sigma\rangle)$ ,  
1175 with  $\langle\sigma\rangle = 0.05491 \pm 0.00038$  ps. The uncertainty stated here is statistical. Systematic  
1176 uncertainties will be considered in Sec. 5.5.2.

Table 5.1 – Measured resolution  $\langle\sigma\rangle_i$  obtained from a fit to the  $p_T$  corrected *sPlot* of the decay-time distribution in bins of per-event decay-time error,  $\delta$ , for prompt  $D^\pm +$  track signal. The average per-event decay time error  $\langle\delta\rangle_i$  in each bin is also reported.

Bin $i$	lower edge	$\langle\delta\rangle_i$	$\langle\sigma\rangle_i$
0	0.01	$0.0142 \pm 0.0016$	$0.01731 \pm 0.00053$
1	0.0165376	$0.01801 \pm 0.00075$	$0.02439 \pm 0.00089$
2	0.0192247	$0.02038 \pm 0.00063$	$0.0286 \pm 0.0011$
3	0.0214493	$0.02248 \pm 0.00052$	$0.0347 \pm 0.0011$
4	0.0232264	$0.02388 \pm 0.00036$	$0.0384 \pm 0.0013$
5	0.0245968	$0.02528 \pm 0.00033$	$0.0422 \pm 0.0014$
6	0.0257605	$0.02641 \pm 0.00034$	$0.0449 \pm 0.0014$
7	0.0269093	$0.02753 \pm 0.00033$	$0.0489 \pm 0.0015$
8	0.0280345	$0.02857 \pm 0.00028$	$0.0489 \pm 0.0015$
9	0.0290414	$0.02955 \pm 0.00030$	$0.0525 \pm 0.0018$
10	0.0301189	$0.03054 \pm 0.00024$	$0.0552 \pm 0.0019$
11	0.0309259	$0.03138 \pm 0.00027$	$0.0582 \pm 0.0017$
12	0.0318409	$0.03229 \pm 0.00032$	$0.0594 \pm 0.0016$
13	0.0328907	$0.03347 \pm 0.00036$	$0.0641 \pm 0.0015$
14	0.0341106	$0.03482 \pm 0.00039$	$0.0643 \pm 0.0014$
15	0.0354999	$0.03638 \pm 0.00052$	$0.0658 \pm 0.0014$
16	0.0372226	$0.03830 \pm 0.00063$	$0.0719 \pm 0.0012$
17	0.0395386	$0.04096 \pm 0.00086$	$0.0736 \pm 0.0012$
18	0.0424521	$0.0447 \pm 0.0014$	$0.0786 \pm 0.0011$
19	0.0473915	$0.0561 \pm 0.0095$	$0.0933 \pm 0.0010$

Table 5.2 – Average per-event decay-time error  $\langle\delta\rangle$  and resolution parameters  $p_1$ ,  $p_2$ ,  $\langle\sigma\rangle$  obtained from a fit to the per-bin decay time error.

Parameter	Result			
$\langle\delta\rangle$	0.0307	$\pm$	0.0097	ps
$p_1$	2.031	$\pm$	0.022	
$p_2$	-19.30	$\pm$	1.6	$\text{ps}^{-1}$
$\langle\sigma\rangle$	0.05491	$\pm$	0.00038	ps

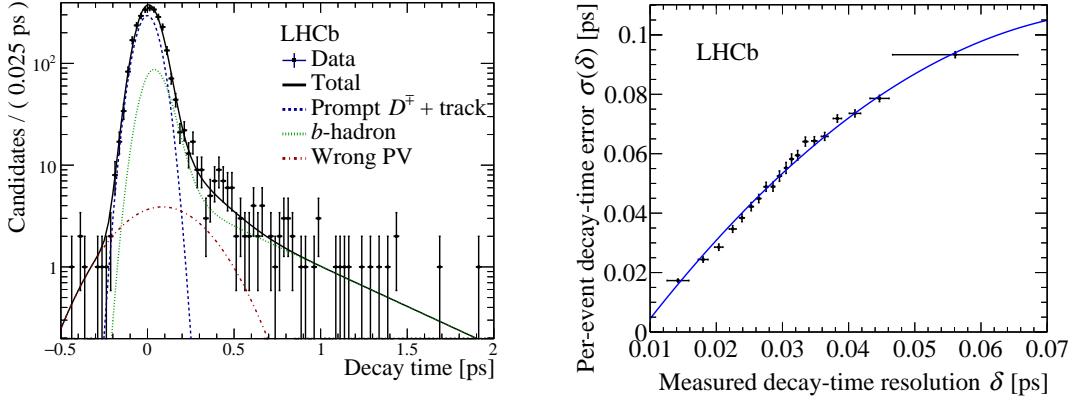


Figure 5.2 – Left:  $p_T$ -corrected and background-subtracted decay-time distribution of the  $D^\pm + \text{track}$  sample for the 15th bin ( $[0.0341, 0.0355]$  ps) in per-event decay time error. The fit result is overlaid as the black solid curve: the wrong-PV, from- $b$ , and prompt components are shown as the red dot-dashed, green dotted, and blue dashed curves, respectively. The numerical results are presented in Table 5.1. Right: measured resolution as a function of the average per-event decay time error determined from fits to the decay time in bins of decay time error. The horizontal bars are the standard deviation of the average per-event decay time error in each bin. The overlaid fit is described in the text.

## 1177 5.2 Time-dependent selection efficiency

1178 Because of some of the selection criteria described in Sec. 4.1, the  $B^0$  decay time distri-  
 1179 bution is biased, *i.e.* different from the shape it would have with a constant selection  
 1180 efficiency. This efficiency, called here and after “acceptance”, is a function of the recon-  
 1181 structed proper time. In particular, it goes very rapidly to zero at low decay times due  
 1182 to the impact parameter requirements which exclude short-lived  $B^0$  candidates; then, it  
 1183 reaches a “plateau” at intermediate decay times; finally, it drops at high decay times due  
 1184 to the acceptance of the VELO reconstruction.

The acceptance function  $a(t)$  is parameterised using splines defined analytically as de-  
 scribed in Ref. [80]. These splines are cubic polynomials defined in sub-ranges of the  
 decay time. The boundaries of each sub-range, called “knots”, are located at 0.4, 0.5,  
 1.0, 1.5, 2.0, 2.3, 2.6, 3.0, 4.0, 10.0, and 12.0 ps. The location of the 11 knots and the  
 higher density of knots at low decay times, where the acceptance is a strongly-varying  
 function of  $t$ , ensure that the resulting acceptance is sufficiently smooth. For each knot  
 $t_i$ , a coefficient  $v_i$  is defined, which is the actual value of the acceptance function  $a(t_i)$ .  
 In order to fix the overall scale of the acceptance function, the  $v_{10}$  coefficient is set to 1.0.  
 Moreover, since statistical fluctuations at high decay times may strongly affect  $v_{11}$ , the

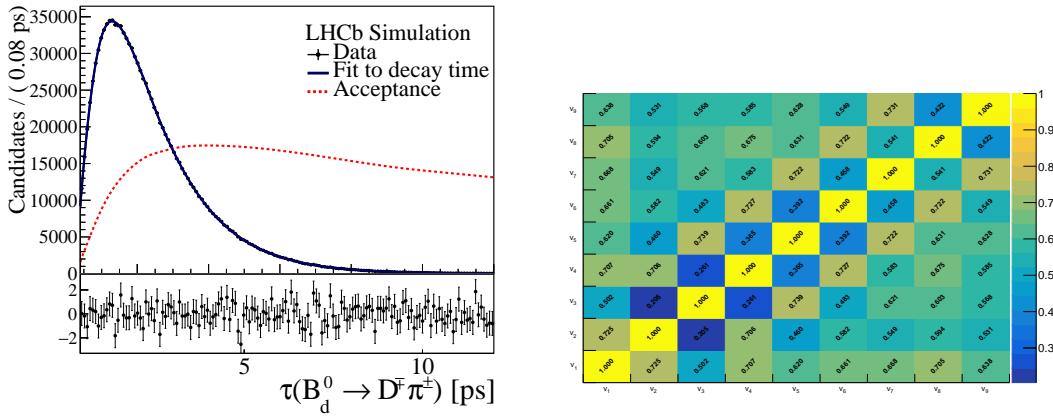


Figure 5.3 – Left: distribution of the reconstructed decay time of simulated and selected  $B^0 \rightarrow D^\mp\pi^\pm$  decays (data points), with fit model superimposed (blue curve), and fitted acceptance function (red dotted curve). Right: correlation matrix of the nine fitted acceptance parameters.

latter is constrained to be the linear extrapolation from the previous two coefficients:

$$v_{11} = v_{10} + \frac{v_{10} - v_9}{t_{10} - t_9} \times (t_{11} - t_{10}). \quad (5.2)$$

The knot positions and the number of knots are optimized in order to fit the  $B^0 \rightarrow D^\mp\pi^\pm$  Monte Carlo decay time distribution with sufficient fit quality. The PDF adopted in this fit is proportional to:

$$a(t) \int dt' \mathcal{R}(t - t') e^{-t'/\tau_d}, \quad (5.3)$$

where  $\mathcal{R}(t - t')$  is the average resolution model discussed in Sec. 5.1 and  $\tau_d$  is the  $B^0$  lifetime value used in the Monte Carlo generation. All acceptance coefficients are floating in the fit, while resolution and lifetime are fixed.

The fit projection is shown if Fig. 5.3 together with the correlation matrix obtained from the fit, whereas the fitted coefficients are listed in Table 5.3.

### 5.3 Decay-time fit to data

The  $CP$  coefficients  $S_f$  and  $S_{\bar{f}}$  are determined from an unbinned maximum likelihood fit where each candidate is weighted with the *sWeights* extracted from the mass fit described in Sec. 4.3.3. Hence, the total PDF is given solely by the PDF  $f(t, dos, dss, \eta os, \eta ss, q)$

### 5.3. Decay-time fit to data

Table 5.3 – Acceptance parameters fitted on the signal Monte Carlo sample.

Parameter name	Fitted value
$v_1$	$0.1961 \pm 0.0016$
$v_2$	$0.3348 \pm 0.0032$
$v_3$	$0.6159 \pm 0.0057$
$v_4$	$0.8667 \pm 0.0073$
$v_5$	$0.9982 \pm 0.0086$
$v_6$	$1.0747 \pm 0.0091$
$v_7$	$1.1051 \pm 0.0094$
$v_8$	$1.1590 \pm 0.0086$
$v_9$	$1.188 \pm 0.014$

describing the signal distribution. This is proportional to

$$a(t) \int dt' R(t-t') P(t'|d_{\text{OS}}, d_{\text{SS}}, \eta_{\text{OS}}, \eta_{\text{SS}}, q) E^{\text{OS}}(\eta_{\text{OS}}) E^{\text{SS}}(\eta_{\text{SS}}) D^{\text{OS}}(d_{\text{OS}}) D^{\text{SS}}(d_{\text{SS}}) Q(q), \quad (5.4)$$

where  $R(t-t')$  is the Gaussian resolution function,  $a(t)$  is the acceptance function,  $E^{\text{OS}}(\eta_{\text{OS}})$ ,  $E^{\text{SS}}(\eta_{\text{SS}})$  are the PDFs of the predicted mistag probability of the taggers,  $D^{\text{OS}}(d_{\text{OS}})$ ,  $D^{\text{SS}}(d_{\text{SS}})$  are the PDFs of the decision of the taggers, and  $Q(q)$  is the PDF of the final states. The term  $P^{\text{time}}(t|d_{\text{OS}}, d_{\text{SS}}, \eta_{\text{OS}}, \eta_{\text{SS}}, q)$  represents the expected  $B$  decay-time distribution for a  $B^0$  or a  $\bar{B}^0$  decaying into a  $D^-\pi^+$  or  $D^+\pi^-$  final state. This is conditional on the tagging decision  $d_i$ , the mistag probability  $\eta_i$  and the final state  $q$ , and it contains the decay rates of Eqs. 1.48–1.51. A detailed description of the time PDF including the tagging parameters, as well as the detection and production asymmetries, is given in Appendix G. The function maximised during the fit is the logarithm of the likelihood obtained from the PDF given in Eq. 5.4,

$$\ln \mathcal{L} = s \sum_{i=1}^N s_W^i \ln f(t_i, d_{\text{OS}}^i, d_{\text{SS}}^i, \eta_{\text{OS}}^i, \eta_{\text{SS}}^i, q^i). \quad (5.5)$$

where  $s_W^i$  are the *sWeights*,  $N$  is the number of candidates in the fitted sample, and  $s$  is a correction factor given by

$$s = \frac{\sum_{i=1}^N s_W^i}{\sum_{i=1}^N (s_W^i)^2}. \quad (5.6)$$

<sup>1191</sup> This factor  $s$  allows to take into account the dilution due to the background subtraction  
<sup>1192</sup> with the *sWeights*, so that correctly-estimated uncertainties from the fit are obtained [81].

1193 In the PDF,  $\Delta\Gamma$  is fixed to zero. Moreover, the  $C_f$  ( $C_{\bar{f}}$ ) coefficient is fixed to 1 ( $-1$ )  
 1194 because the value of  $r_{D\pi}^2$  is such that the sensitivity to  $C_f$  ( $C_{\bar{f}}$ ) is negligible. The  
 1195 tagging efficiency differences  $\varepsilon^i$  ( $i = \text{OS}, \text{SS}$ ) are found to be consistent with zero in the  
 1196  $B^0 \rightarrow D^- \pi^+$  Monte Carlo sample: for this reason, these coefficients are fixed to zero in  
 1197 the fit. Systematic uncertainties will be considered in Sec. 5.5 for all these assumptions.

The following physics parameters are Gaussian-constrained to their measured values,

$$\tau = 1/\Gamma = 1.518 \pm 0.004 \text{ ps}, \quad (5.7)$$

$$\Delta m = 0.5050 \pm 0.0023 \text{ ps}^{-1}, \quad (5.8)$$

1198 where  $\tau$  is taken as the world average value [22], and  $\Delta m$  is the LHCb measurement from  
 1199 semileptonic  $B^0$  decays [82].

1200 The free parameters of the fit are:

- 1201 • the  $S_f$  and  $S_{\bar{f}}$  coefficients;  
 1202 • the detection asymmetry<sup>1</sup>

$$A_D \frac{\varepsilon(f) - \varepsilon(\bar{f})}{\varepsilon(f) + \varepsilon(\bar{f})}, \quad (5.9)$$

where  $\varepsilon(f)$  ( $\varepsilon(\bar{f})$ ) is the detection efficiency of the final state  $f$  ( $\bar{f}$ ), and the production asymmetry

$$A_P = \frac{\sigma(\bar{B}^0) - \sigma(B^0)}{\sigma(\bar{B}^0) + \sigma(B^0)}, \quad (5.10)$$

1202 where  $\sigma(\bar{B}^0)$  ( $\sigma(B^0)$ ) is the inclusive  $\bar{B}^0$  ( $B^0$ ) production cross-section;

- 1203 • the calibration parameters for both OS and SS taggers;  
 1204 • the tagging efficiencies  $\varepsilon_{\text{tag}}^{\text{OS}}$  and  $\varepsilon_{\text{tag}}^{\text{SS}}$ ;  
 1205 • the time acceptance coefficients.

1206 The value of the parameters obtained from the fit to data are listed in Table 5.4. The  
 1207 correlation matrix of the parameters is reported in Appendix H. The projection of the  
 1208 PDF on the decay-time distribution is shown in Fig 5.4, while Fig 5.5 shows the  $B^0$ – $\bar{B}^0$   
 1209 asymmetries of Eq. 1.46 (distorted by experimental effects) for the two final states.

It is possible to define the  $CP$  asymmetries between Cabibbo-favoured (CF) and Cabibbo-

---

<sup>1</sup>This definition has the opposite sign compared to the one in Ref. [83].

### 5.3. Decay-time fit to data

Table 5.4 – Results of the decay time fit. The last column shows existing measurements/independent estimations for a direct comparison, or the Gaussian constraint that is applied to the parameter.

Parameter	Value	Comment
$S_f$	$0.058 \pm 0.021$	Statistical uncertainty when fitting w/o Gauss-const. and PIDK syst., 0.0198
$S_{\bar{f}}$	$0.038 \pm 0.021$	Statistical uncertainty when fitting w/o Gauss-const. and PIDK syst., 0.0199
$A_P$	$-0.0064 \pm 0.0028$	Compare with $-0.0100 \pm 0.0047$ (Eq. 5.14)
$A_D$	$0.0086 \pm 0.0019$	Compare with [83] (Sec. 5.4)
$\Gamma$	$0.6587 \pm 0.0017 \text{ ps}^{-1}$	Gaussian-constrained to $0.6588 \pm 0.0017 \text{ ps}^{-1}$
$\Delta m$	$0.5054 \pm 0.0022 \text{ ps}^{-1}$	Gaussian-constrained to $0.5050 \pm 0.0023 \text{ ps}^{-1}$
$p_0^{\text{OS}}$	$-0.152 \pm 0.021$	OS tagger calibration parameters
$p_1^{\text{OS}}$	$-0.035 \pm 0.024$	
$p_2^{\text{OS}}$	$-0.0070 \pm 0.0089$	
$p_3^{\text{OS}}$	$-0.32 \pm 0.11$	
$p_4^{\text{OS}}$	$-0.47 \pm 0.49$	
$\Delta p_0^{\text{OS}}$	$-0.079 \pm 0.049$	
$\Delta p_1^{\text{OS}}$	$0.141 \pm 0.036$	
$\Delta p_2^{\text{OS}}$	$-0.024 \pm 0.013$	
$\Delta p_3^{\text{OS}}$	$-0.26 \pm 0.16$	
$\Delta p_4^{\text{OS}}$	$-0.52 \pm 0.71$	
$p_0^{\text{SS}}$	$-0.041 \pm 0.021$	SS tagger calibration parameters
$p_1^{\text{SS}}$	$-0.012 \pm 0.022$	
$\Delta p_0^{\text{SS}}$	$-0.085 \pm 0.044$	
$\Delta p_1^{\text{SS}}$	$0.043 \pm 0.033$	
$\varepsilon_{\text{tag}}^{\text{OS}}$	$0.43237 \pm 0.00077$	fraction of OS tagged candidates (relative to tagged candidates only)
$\varepsilon_{\text{tag}}^{\text{SS}}$	$0.93046 \pm 0.00040$	fraction of SS tagged candidates (relative to tagged candidates only)
$v_1$	$0.3192 \pm 0.0062$	time acceptance coefficients
$v_2$	$0.494 \pm 0.010$	
$v_3$	$0.793 \pm 0.016$	
$v_4$	$0.994 \pm 0.019$	
$v_5$	$1.093 \pm 0.021$	
$v_6$	$1.117 \pm 0.021$	
$v_7$	$1.140 \pm 0.021$	
$v_8$	$1.175 \pm 0.019$	
$v_9$	$1.154 \pm 0.026$	

suppressed (CS) rates as follows:

$$A_{\text{CF}} = \frac{\Gamma_{B^0 \rightarrow f}(t) - \Gamma_{\bar{B}^0 \rightarrow \bar{f}}(t)}{\Gamma_{B^0 \rightarrow f}(t) + \Gamma_{\bar{B}^0 \rightarrow \bar{f}}(t)}, \quad (5.11)$$

$$A_{\text{CS}} = \frac{\Gamma_{\bar{B}^0 \rightarrow f}(t) - \Gamma_{B^0 \rightarrow \bar{f}}(t)}{\Gamma_{\bar{B}^0 \rightarrow f}(t) + \Gamma_{B^0 \rightarrow \bar{f}}(t)}.. \quad (5.12)$$

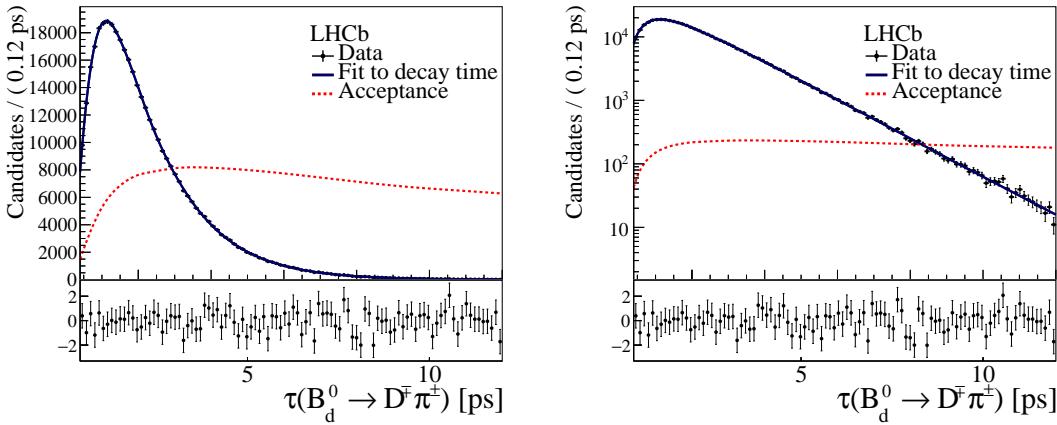


Figure 5.4 – Distribution of the reconstructed decay time of *sWeighted*  $B^0 \rightarrow D^\mp\pi^\pm$  decays (data points), with fit model superimposed (blue curve), and fitted acceptance function (red dotted curve), in linear (left) and logarithmic (right) scale.

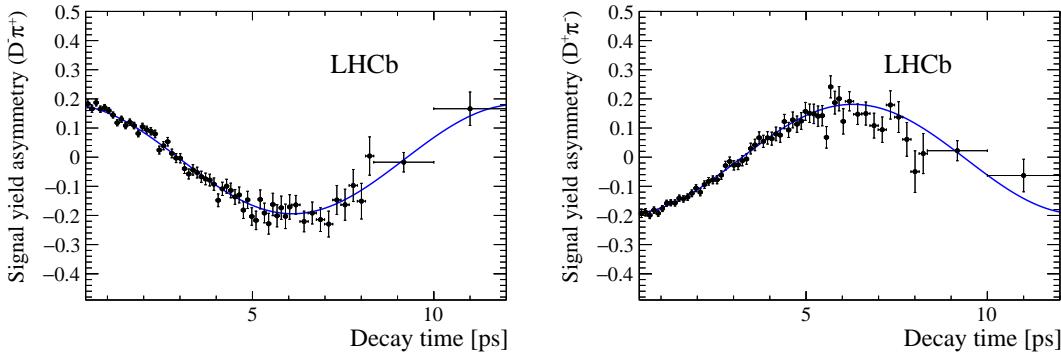


Figure 5.5 – Time-dependent asymmetry between  $B^0$  and  $\bar{B}^0$  decays (data points) for the  $D^-\pi^+$  (left) and  $D^+\pi^-$  (right) final states, with fit model superimposed (blue curve).

- 1210 These asymmetries above (distorted by experimental effects) are plotted together with  
1211 data in Fig. 5.6, separately for the  $f = D^-\pi^+$  and  $\bar{f} = D^+\pi^-$  final states.  
1212 The projection of the PDF considering the decay-time distribution of the four independent  
1213 decays rates,  $B^0 \rightarrow D^-\pi^+$ ,  $\bar{B}^0 \rightarrow D^-\pi^+$ ,  $B^0 \rightarrow D^+\pi^-$  and  $\bar{B}^0 \rightarrow D^+\pi^-$ , for OS and SS  
1214 tagged candidates, are reported in Figs. 5.7 and 5.8, respectively.  
1215 The 2D contour plots for the  $CP$  coefficients  $S_f$  and  $S_{\bar{f}}$  and for the detection and  
1216 production asymmetry are shown in Fig. 5.9.  
1217 Using the flavour tagging calibrations obtained from the fit, the tagging performance of the  
1218 signal sample is computed. The average of the total squared dilution is  $(6.554 \pm 0.017)\%$ .  
1219 Taking into account also untagged candidates, *i.e.* considering the tagging efficiency of

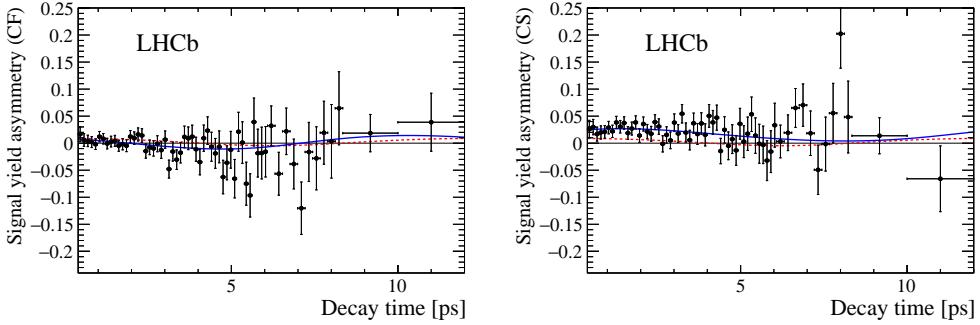


Figure 5.6 – Decay-time-dependent signal-yield asymmetries for (left) Cabibbo-favoured and (right) Cabibbo-suppressed decay topologies, defined in Eqs. 5.11 and 5.12. The points with error bars are the data, the blue solid curve is the fit model, and the red dotted curve indicates the fit model when  $S_{\bar{f}} \equiv -S_f$  (*i.e.* no CP violation in the interference between mixing and decay) is required.

<sub>1220</sub>  $(85.23 \pm 0.05)\%$ , the tagging power is  $(5.59 \pm 0.01)\%$ .

## <sub>1221</sub> 5.4 Fit validation

### <sub>1222</sub> Check of nuisance parameters

<sub>1223</sub> The values of the nuisance parameters obtained in the fit (production/detection asymmetries, flavour tagging calibrations) are compared with available external measurements.

The production asymmetry  $A_P$  is compared with the LHCb measurement of Ref. [83]. The production asymmetry is computed by weighting the production asymmetry measured from this paper in bins of  $p_T$  and  $\eta$ ,  $A_{P,i}$ , with the signal fractions  $\varepsilon_i = \frac{s_i}{\sum_j s_j}$  in each bin  $i$  of  $B^0 \rightarrow D^\mp \pi^\pm$  data:

$$A_P = \sum_i \varepsilon_i A_{P,i}, \quad (5.13)$$

where  $s_i$  is the sum of the *sWeights* in bin  $i$ . This yields

$$A_P = -0.0100 \pm 0.0047 \text{ (stat)} \pm 0.0004 \text{ (syst)}. \quad (5.14)$$

<sub>1225</sub> This value is compatible within  $0.65 \sigma$  with that obtained from the  $B^0 \rightarrow D^\mp \pi^\pm$  decay-time fit.

<sub>1227</sub> The detection asymmetry  $A_D$  is also obtained from Ref. [83], and is measured using <sub>1228</sub>  $B^0 \rightarrow J/\psi K^{*0}$  to be  $0.0098 \pm 0.0046$  and  $0.0056 \pm 0.0030$  for 2011 and 2012, respectively, <sub>1229</sub> and using  $B_s^0 \rightarrow D_s^- \pi^+$  to be  $0.0143 \pm 0.0086$  and  $0.0103 \pm 0.0058$  for 2011 and 2012, <sub>1230</sub> respectively. The central values obtained from the fit to  $B^0 \rightarrow D^\mp \pi^\pm$  are in agreement

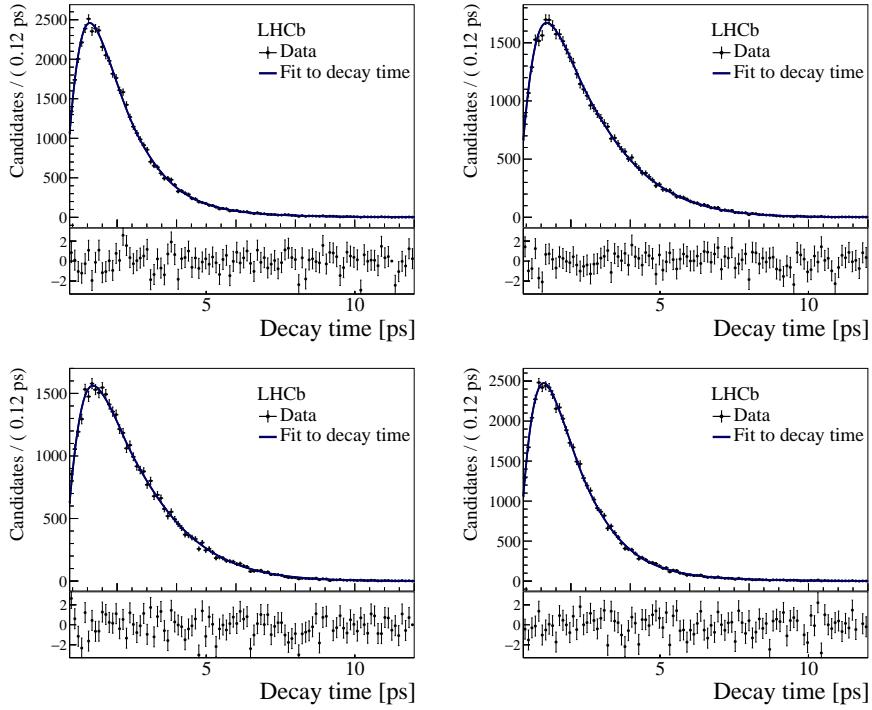


Figure 5.7 – Decay-time distributions of the *sWeighted* data samples for (top left)  $B^0 \rightarrow D^- \pi^+$ , (top right)  $\bar{B}^0 \rightarrow D^- \pi^+$ , (bottom left)  $B^0 \rightarrow D^+ \pi^-$  and (bottom right)  $\bar{B}^0 \rightarrow D^+ \pi^-$  for OS inclusively tagged candidates. The fit result is superimposed as the blue curves.

1231 with this set of results.

1232 The values of the parameters of the tagging calibrations are compared with those found  
 1233 in the control samples as described in Secs. 3.2.1 and 3.2.2. The strategy presented in  
 1234 Sec. 3.2 is followed, and no perfect portability of the calibrations is assumed *a priori*, as  
 1235 Sec. 3.2.1 has shown this not to be the case. However, the values of the parameter values  
 1236 found in the signal fit are expected to be in the same interval of those from the control  
 1237 channels given in Tables 3.2 and 3.4. A full comparison that takes into account the  
 1238 correlation between the parameters with a  $\chi^2$  test which uses the covariance matrices of  
 1239 the fit to data (reduced to the tagging parameters only), and the fit of the calibration from  
 1240 the control channel, is performed. The discrepancy (corresponding to the  $\chi^2$  minimum)  
 1241 is around  $0.91\sigma$  for the OS tagger, and  $0.29\sigma$  for the SS tagger. The parameters which  
 1242 present the largest disagreement are  $\Delta p_3^{\text{OS}}$  for the OS tagger and the  $\Delta p_0^{\text{SS}}$  for the SS  
 1243 tagger.

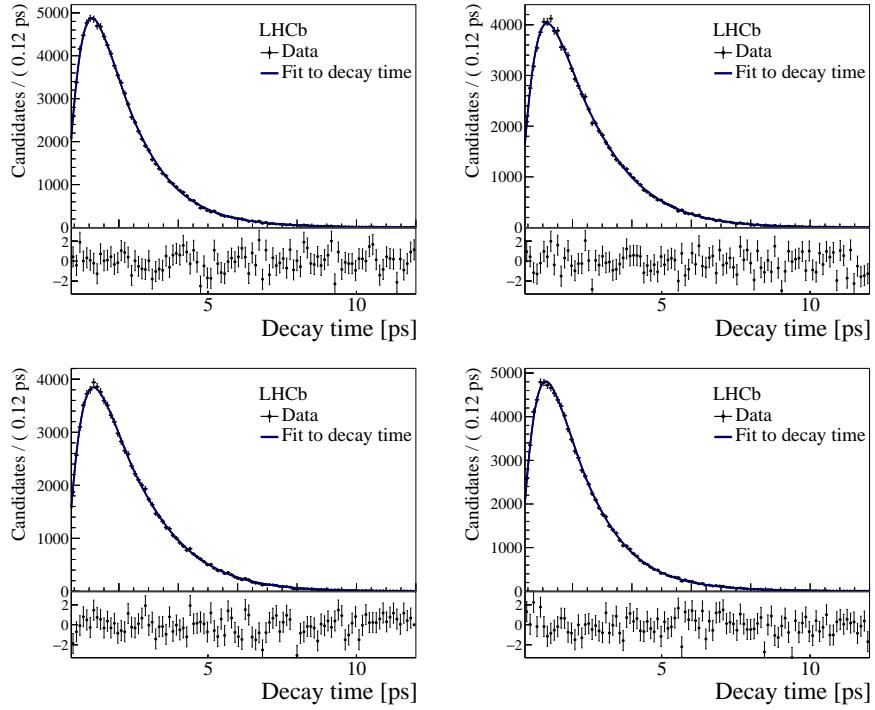


Figure 5.8 – Decay-time distributions of the *sWeighted* data sample for (top left)  $B^0 \rightarrow D^- \pi^+$ , (top right)  $\bar{B}^0 \rightarrow D^- \pi^+$ , (bottom left)  $B^0 \rightarrow D^+ \pi^-$  and (bottom right)  $\bar{B}^0 \rightarrow D^+ \pi^-$  for SS inclusively tagged candidates. The fit result is superimposed as the blue curves.

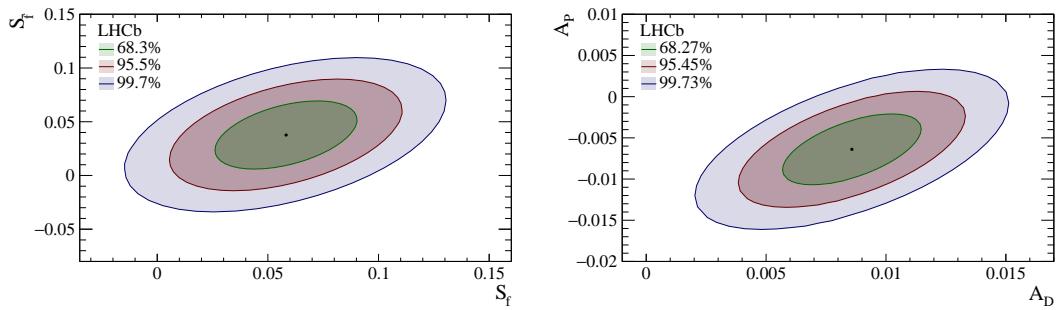


Figure 5.9 – Contour plots for  $(S_f, S_{\bar{f}})$  (left) and  $(A_P, A_D)$  (right) showing the one, two and three sigma contours. The shown uncertainties include the full statistical uncertainty and the systematic uncertainty due to the Gaussian constraints on the mixing frequency  $\Delta m$  and the  $B^0$  decay width  $\Gamma$ .

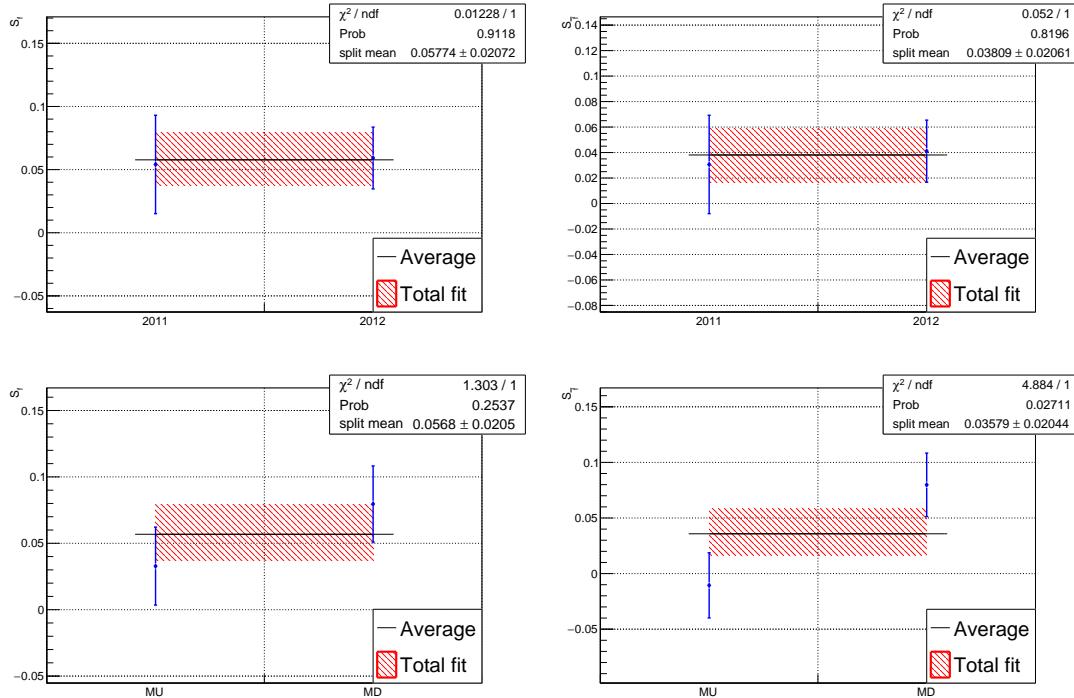


Figure 5.10 – Fitted values of  $S_f$  (left) and  $S_{\bar{f}}$  (right) as a function of the data-taking year (top) and magnet polarity (bottom). The red hatched band shows the values obtained from the nominal fit of the full sample. The horizontal black line is the result of a  $\chi^2$  fit to obtain the weighted average of the results of each subsample.

#### 1244 Fits in data subsamples

1245 A check of the stability of the results against the different data taking conditions is  
 1246 performed by repeating the fit in four subsamples of the data, namely data taken with  
 1247 magnet “up” and “down” polarities, and data taken in 2011 and 2012. The *sWeights* for  
 1248 each subsample are obtained via the mass fits described in Sec. 4.3.4. The detailed results  
 1249 of these time fits are reported in Appendix I. A comparison between the fitted values for  
 1250  $S_f$  and  $S_{\bar{f}}$  obtained in each subsample is shown in Fig. 5.10. In all cases, the parameters  
 1251 are in agreement and have good *p*-values, the smallest being the one between the values  
 1252 of  $S_{\bar{f}}$  from the magnet polarity splits (2.7%). In addition, the average between the fitted  
 1253 values in each split (black line in Fig. 5.10) is always very close to the central value from  
 1254 the nominal fit (red hatched band).

1255 The stability of the results against the tagging algorithm adopted in the fit are also  
 1256 checked. In this case, the data sample with *sWeights* obtained from the nominal mass fit  
 1257 (Sec. 4.3) is split in three independent subsamples according to the tagging decision:

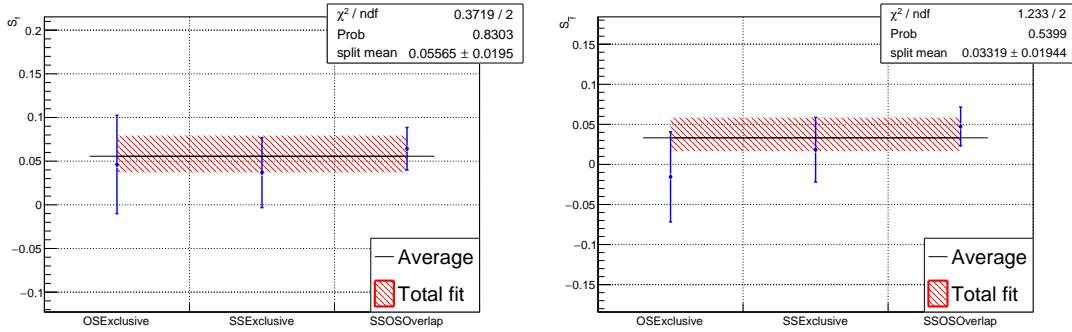


Figure 5.11 – Fitted values of  $S_f$  (left) and  $S_{\bar{f}}$  (right) when candidates tagged exclusively by OS or SS, or both simultaneously are considered. The red hatched band shows the values obtained from the nominal fit of the full sample. The horizontal black line is the result of a  $\chi^2$  fit to obtain the weighted average of the results of each subsample.

- 1258     • candidates tagged exclusively by the OS tagger;
- 1259     • candidates tagged exclusively by the SS tagger;
- 1260     • candidates tagged by both the OS and SS taggers.

1261     The values of  $S_f$  and  $S_{\bar{f}}$  obtained in these subsamples are compared in Fig. 5.11. All  
 1262     values are compatible. More details are given in Appendix I. Given the difference of the  
 1263     tagging algorithms and their calibrations, the stability of the results in this test provides  
 1264     additional confidence on the strategy adopted of floating the calibration parameters in  
 1265     the fit.

1266     The stability of the results against the  $B^0$  kinematics and global properties of the event is  
 1267     tested. More specifically, the decay time fit is repeated in bins of the following variables:

- 1268     • transverse momentum of the  $B^0$  (4 bins);
- 1269     • number of reconstructed primary vertices (3 bins);
- 1270     • total number of reconstructed tracks (3 bins);
- 1271     • difference in pseudorapidity ( $\Delta\eta$ ) between  $D$  meson and bachelor pion (4 bins).

1272     The motivation for these tests is that flavour tagging calibration parameters depend on  
 1273     the above observables; as a consequence, the fitted values for the  $S_f$  and  $S_{\bar{f}}$  coefficients  
 1274     might also show a significant trend in these variables because of the correlation with  
 1275     the flavour tagging calibrations. Moreover, the difference in pseudorapidity is sensitive  
 1276     to potential misalignments in the detectors which might affect the measured value of

1277  $CP$  asymmetries. The values of  $S_f$  and  $S_{\bar{f}}$  obtained in these subsamples are compared in  
1278 Fig. 5.12, whereas more details are given in Appendix I. All values are compatible, and  
1279 no significant dependence of  $S_f$  and  $S_{\bar{f}}$  on the studied variables is observed.

1280 Finally, the time fit is repeated separately for TOS candidates on L0Hadron and all the  
1281 other candidates. The values of  $S_f$  and  $S_{\bar{f}}$  obtained in these subsamples are compared  
1282 in Fig. 5.13. All values are compatible, and no significant dependence of  $S_f$  and  $S_{\bar{f}}$  is  
1283 observed. More details can be found in Appendix I.

#### 1284 Time fits to bootstrapped Monte Carlo samples

1285 The fit is also validated using Monte Carlo simulation. The  $B^0 \rightarrow D^\mp\pi^\pm$  simulated  
1286 sample is *bootstrapped* [84] (*i.e.*, resampled allowing repetition of the same candidates)  
1287 1000 times. Each bootstrapped sample contains the same signal yield as obtained from  
1288 the nominal mass fit (Table 4.15), corrected for the *sWeights* dilution factor of Eq. 5.6 to  
1289 have the same effective yield as in the data time fit.

1290 Each sample is then fitted using exactly the same strategy as described in Sec. 5.3, except  
1291 that the central value of the Gaussian constraints on the  $\Gamma$  and  $\Delta m$  parameters is randomly  
1292 drawn from a Gaussian distribution centred on the Monte Carlo generation value given  
1293 in Appendix M ( $1/\Gamma_{\text{MC}} = 1.519$  ps and  $\Delta m = 0.510$  ps $^{-1}$ ) with a standard deviation  
1294 equal to the width of the constraint ( $\pm 0.004$  ps for  $1/\Gamma$  and  $\pm 0.0023$  ps $^{-1}$  for  $\Delta m$ ). This  
1295 allows fluctuations of the  $\Gamma$  and  $\Delta m$  measurements, and avoids underestimation of the  
1296 fitted uncertainties.

1297 The distributions of the fitted value, uncertainty, pull and residual<sup>2</sup> of  $S_f$  and  $S_{\bar{f}}$  are  
1298 shown in Fig. 5.14. Other fitted parameters are reported in Appendix J. Each of these  
1299 distribution is fitted with a Gaussian function. The width of the fitted pull distributions  
1300 are close to unity, meaning that the uncertainty coming from the likelihood fit is correctly  
1301 estimated. The mean value of the distribution of the uncertainties of each parameter  
1302 is close to the value of the uncertainty found in the fit to data. The on-average better  
1303 precision found in the fit to MC is due to the higher tagging performance of the simulation.

1304 The distribution of the residuals of the  $S_f$  parameter shows a mean of  $0.0071 \pm 0.0006$ ,  
1305 corresponding to one third of the statistical uncertainty of the fit to data; for  $S_{\bar{f}}$ , the  
1306 mean is  $-0.0013 \pm 0.0006$ , which corresponds to about 6% of the statistical uncertainty  
1307 of the fit to data.

1308 Several configurations are implemented to test the bootstrap study and its results, and  
1309 to try to address the origin of these biases. The fits to the bootstrapped samples are  
1310 repeated in the following different configurations:

---

<sup>2</sup>The residual is defined as fitted value minus generated value, whereas the pull is the residual divided by the fitted uncertainty.

## 5.4. Fit validation

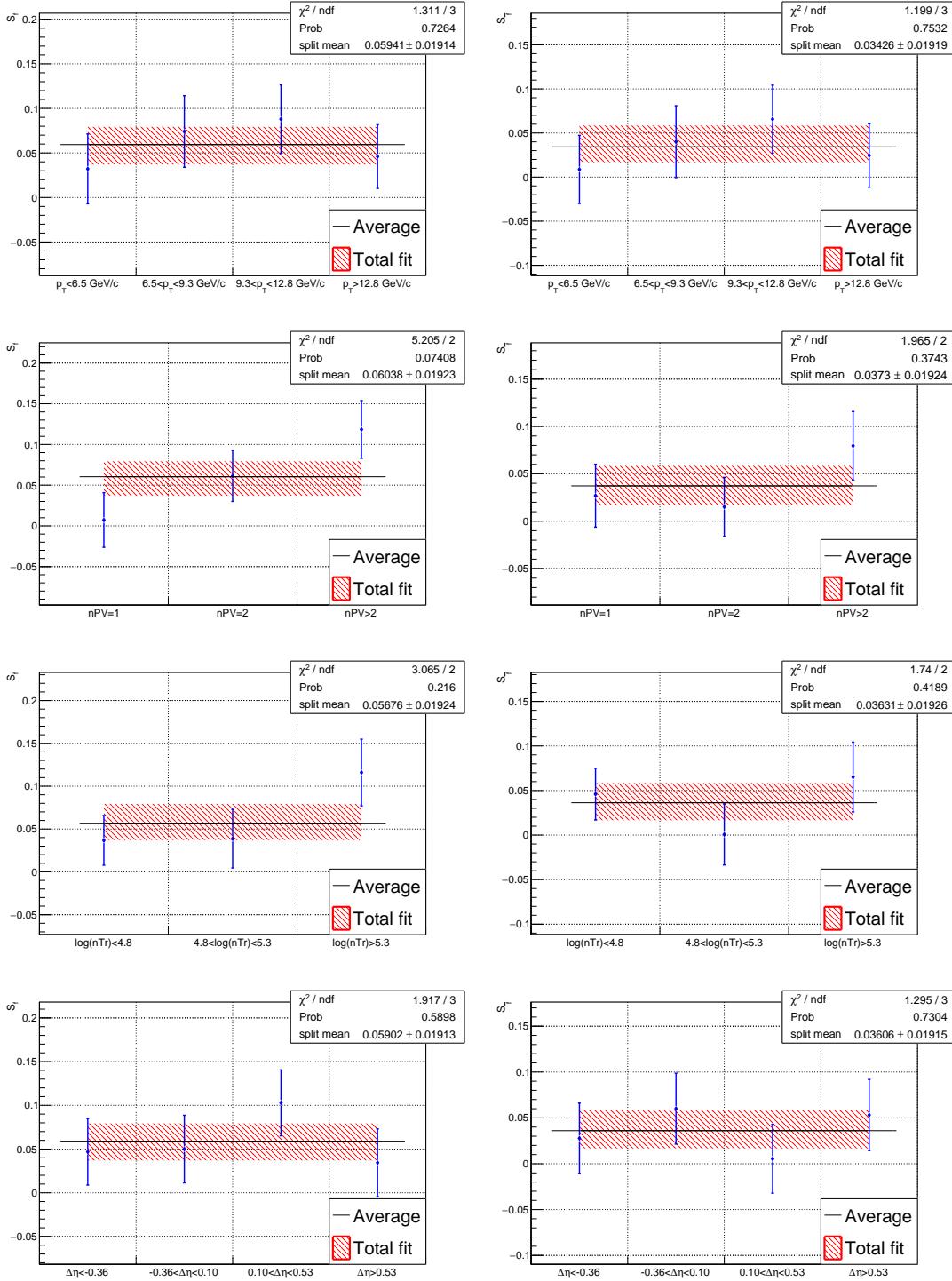


Figure 5.12 – Fitted values of  $S_f$  (left) and  $S_{\bar{f}}$  (right) when the decay time fit is performed in bins of (from top to bottom) the transverse momentum of the  $B^0$ , number of primary vertices, number of tracks and difference in pseudorapidity between the  $D$  meson and the bachelor pion. The red hatched band shows the values obtained from the nominal fit of the full sample. The horizontal black line is the result of a  $\chi^2$  fit to obtain the weighted average of the results of each subsample.

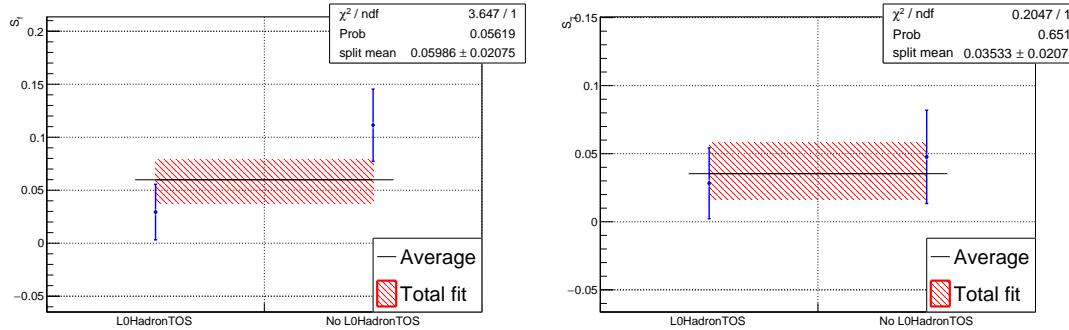


Figure 5.13 – Fitted values of  $S_f$  (left) and  $S_{\bar{f}}$  (right) when the decay time fit is performed separately for TOS candidates on L0Hadron and all the other candidates. The red hatched band shows the values obtained from the nominal fit of the full sample. The horizontal black line is the result of a  $\chi^2$  fit to obtain the weighted average of the results of each subsample.

- 1311 • using the true flavour of the  $B^0$  candidate instead of the tagging decision and mistag  
1312 probability: no biases are found on  $S_f$  and  $S_{\bar{f}}$ ;
- 1313 • using a *toy* (or *cheated*) tagger, as explained in Appendix K: no biases are found  
1314 on  $S_f$  and  $S_{\bar{f}}$ ;
- 1315 • using the calibration parameters obtained in the signal MC sample using the true  
1316 flavour information (see Sec. 3.2.1 and Sec. 3.2.2): no biases are found on  $S_f$  and  
1317  $S_{\bar{f}}$ ;
- 1318 • fixing the calibration parameters to the values obtained from the MC samples of  
1319 the control channels: biases of the order  $1\sigma$  on  $S_f$  and  $S_{\bar{f}}$  are found;
- 1320 • applying Gaussian-constraints on the calibration parameters using the values ob-  
1321 tained from the MC samples of the control channels: biases of the order of half the  
1322 statistical uncertainty of  $S_f$  and  $S_{\bar{f}}$  are found;

1323 This study confirms that the strategy of floating the calibration parameters in the fit is  
1324 the optimal choice. Other than the biases related to the flavour-tagging calibrations, the  
1325 origin of the small bias observed on the  $S_f$  parameter in the nominal configuration could  
1326 not be clarified. To confirm this bias, the study is repeated by fitting additional 1000  
1327 bootstrapped samples using an independent fitter. The mean of the distribution of the  
1328 residuals in this second study is confirmed to be of the same size, namely  $0.0064 \pm 0.0007$   
1329 for  $S_f$  and  $-0.0024 \pm 0.0007$  for  $S_{\bar{f}}$ . Hence, the weighted average of the small residuals  
1330 on  $S_f$  ( $0.0068 \pm 0.0005$ ) and  $S_{\bar{f}}$  ( $-0.0018 \pm 0.0005$ ) of both studies are considered as  
1331 systematic uncertainties. As described in Appendix L, the correlation between the  
1332 systematic uncertainties on  $S_f$  and  $S_{\bar{f}}$  associated to the fit biases reported here is 0.4.

## 5.4. Fit validation

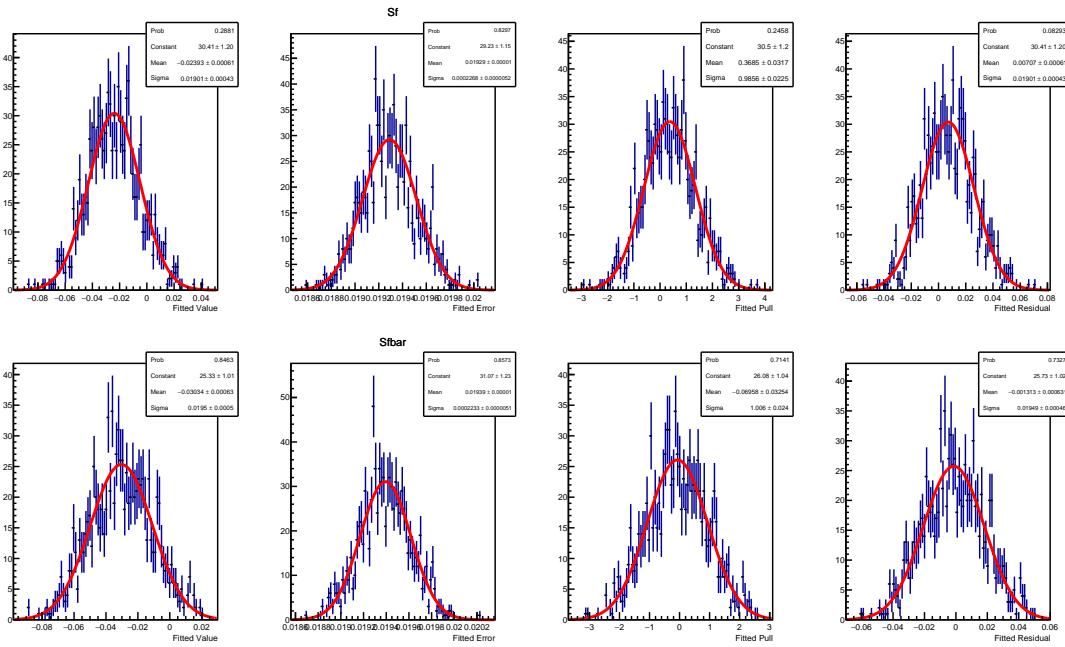


Figure 5.14 – Distributions of the fitted value, error, pull and residual for  $S_f$  (top) and  $S_{\bar{f}}$  (bottom) obtained from fits to bootstrapped Monte Carlo samples. Each distribution is fitted with a Gaussian function. Pulls and residuals are computed by taking the Monte Carlo generation values as reference (Appendix M).

---

## 1333 5.5 Systematics

1334 The identified systematic uncertainties are listed in Table 5.5 in decreasing order of their  
 1335 size. Their quadratic sum is 0.011 and 0.007 for  $S_f$  and  $S_{\bar{f}}$ , respectively. A description  
 1336 of each systematic effect is given in the following subsections. The “fit biases” are the  
 1337 residuals observed in the Monte Carlo bootstrap study discussed in Sec. 5.4.

Table 5.5 – Systematic uncertainties on the  $CP$  parameters  $S_f$  and  $S_{\bar{f}}$ .

Source	$S_f$	$S_{\bar{f}}$
uncertainty of $\Delta m$	0.0073	0.0061
fit biases	0.0068	0.0018
background subtraction	0.0042	0.0023
flavour tagging calibration models	0.0011	0.0015
flavour tagging efficiency asymmetries	0.0012	0.0015
PIDK efficiencies	0.0008	0.0008
acceptance model	0.0007	0.0007
assumption $\Delta \Gamma = 0$	0.0007	0.0007
assumption $C_f = -C_{\bar{f}} = 1$	0.0006	0.0006
decay time resolution	0.0012	0.0008
total systematic uncertainty	0.0111	0.0073
statistical uncertainties	0.0198	0.0199

### 1338 5.5.1 Systematic uncertainties from Gaussian constraints

1339 Systematic uncertainties due to external measurements used in the PDF are accounted  
 1340 for through Gaussian constraints in the likelihood. These parameters are the mixing  
 1341 frequency,  $\Delta m$ , and the  $B^0$  lifetime,  $\tau$ . The fit has been repeated by fixing the Gaussian-  
 1342 constrained parameters to their central values, in order to not propagate the uncertainty  
 1343 of these parameters to the statistical uncertainties of the fit. The statistical uncertainties  
 1344 of  $S_f$  and  $S_{\bar{f}}$  with  $\Delta m$  fixed are 0.0198 and 0.0199, respectively, whereas the statistical  
 1345 correlation is 0.6. Considering the difference in quadrature between the uncertainty  
 1346 from the nominal fit and that from this fit, the systematic uncertainty due to  $\Delta m$  are  
 1347 0.0073 and 0.0061 for  $S_f$  and  $S_{\bar{f}}$ , respectively. The correlation between the systematic  
 1348 uncertainties due to  $\Delta m$  on  $S_f$  and  $S_{\bar{f}}$  is  $-1$ , as described in Appendix L. The fit with  $\tau$   
 1349 fixed shows that the systematics uncertainty due to this parameter is negligible.

1350 Systematic uncertainty associated with the PIDK efficiencies (Table 4.10) are taken into  
 1351 account in the mass fit by means of Gaussian constraints on these parameters (Sec. 4.3.2).  
 1352 The mass fit is repeated by neglecting these uncertainties in the Gaussian constraints.  
 1353 Then, the time fit is performed with this new set of *sWeights*. The difference in quadrature  
 1354 between the uncertainty from this fit and that from the nominal fit gives the systematic

1355 due to the binning scheme in the PIDK resampling, which is 0.0008 for both  $S_f$  and  $S_{\bar{f}}$ .

### 1356 5.5.2 Systematic uncertainties estimated with pseudoexperiments

1357 When computing the systematic uncertainties with pseudoexperiments (or *toys*), a sample  
1358 with the same size as the data is generated by sampling the PDF with parameters fixed  
1359 to the value found in the data fit. The values of  $S_f$  and  $S_{\bar{f}}$  are fixed to those used in the  
1360 generation of the Monte Carlo sample (Appendix M). In the generation of the samples  
1361 the PDF is modified to consider alternative models according to the source of systematic  
1362 uncertainty under investigation. The generated sample is then fitted with the nominal  
1363 model. For each parameter, the mean of the distribution of the residuals from 1000 toys  
1364 is taken as a symmetric systematic uncertainty. If the mean is consistent with zero within  
1365  $\pm 1 \sigma$ , the error on the mean is taken instead. The systematic uncertainties estimated  
1366 with this toy-based method are the following:

- 1367 • the flavour tagging calibration model;
- 1368 • the flavour tagging efficiency asymmetries;
- 1369 • the acceptance model;
- 1370 • the decay time resolution;
- 1371 • the assumption  $C_f = -C_{\bar{f}} = 1$ ;
- 1372 • the assumption  $\Delta\Gamma = 0$ .

#### 1373 Flavour tagging calibration model

1374 Toys are generated using for the SS calibration the nominal model with a first order  
1375 polynomial, and for the OS the model is reduced by one degree as compared to the  
1376 nominal one. In the fit, the calibration models of both taggers are increased by one degree  
1377 compared to what was used in the generation step. The distribution of the residuals  
1378 of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.15. The residuals are not compatible with zero and  
1379 therefore they are assigned as systematic uncertainties.

#### 1380 Flavour tagging efficiency asymmetries

1381 Toys are generated with the flavour tagging asymmetries set to their estimate from  
1382 simulation minus their uncertainty, namely  $-0.14\%$  and  $-0.13\%$  for the OS and SS  
1383 tagger, respectively. The distributions of the residuals of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.16.  
1384 The residuals are not compatible with zero and therefore they are assigned as systematic  
1385 uncertainties.

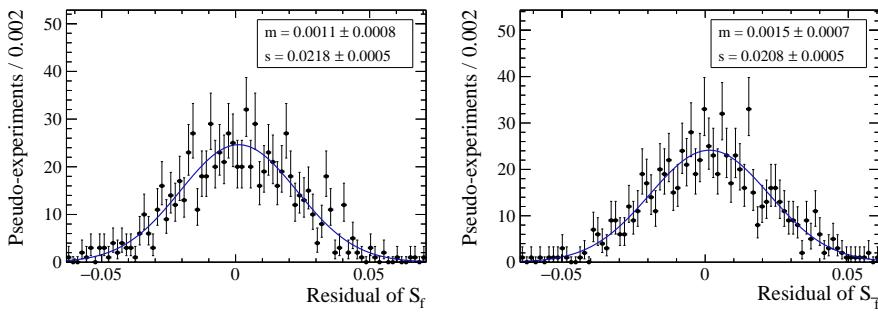


Figure 5.15 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the tagging calibration models.

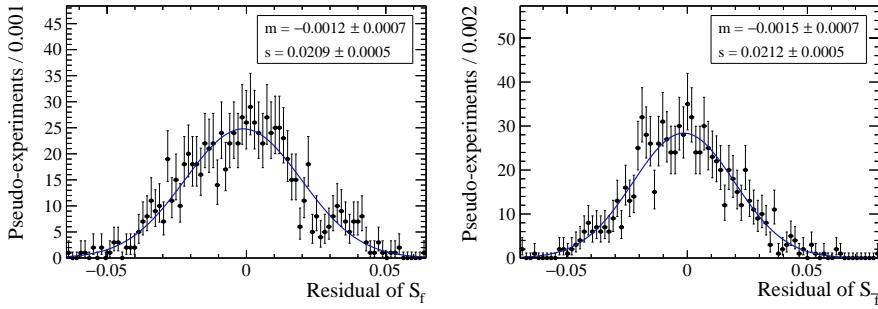


Figure 5.16 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the assumption on the flavour tagging efficiency asymmetry.

### 1386 Acceptance model

1387 The acceptance model is modified in the generation by replacing the nominal knots for  
 1388 the spline function with new knots, namely at 0.4, 0.45, 0.8, 1.3, 2.5, 6.0, and 12.0 ps. The  
 1389 distribution of the residuals of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.17. Residuals consistent with  
 1390 zero are found and therefore the uncertainty on the residuals is assigned as systematic  
 1391 uncertainty.

### 1392 Decay time resolution

1393 Toys are generated with time resolutions 20 fs larger and 20 fs smaller than the nominal  
 1394 value of 55 fs. The distributions of the fitted value of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.18.  
 1395 The largest residual is considered as overall systematic uncertainty.

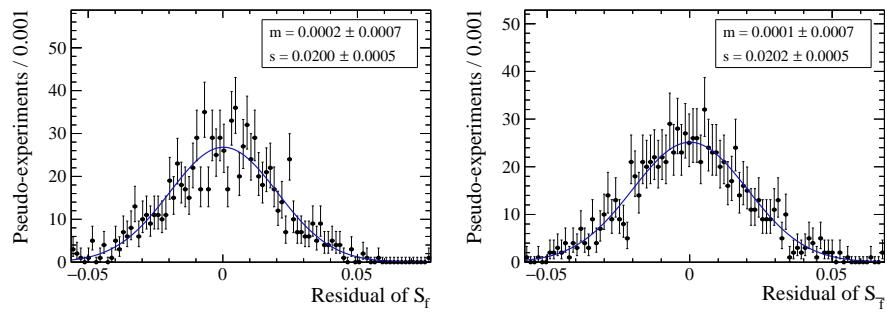


Figure 5.17 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the acceptance model.

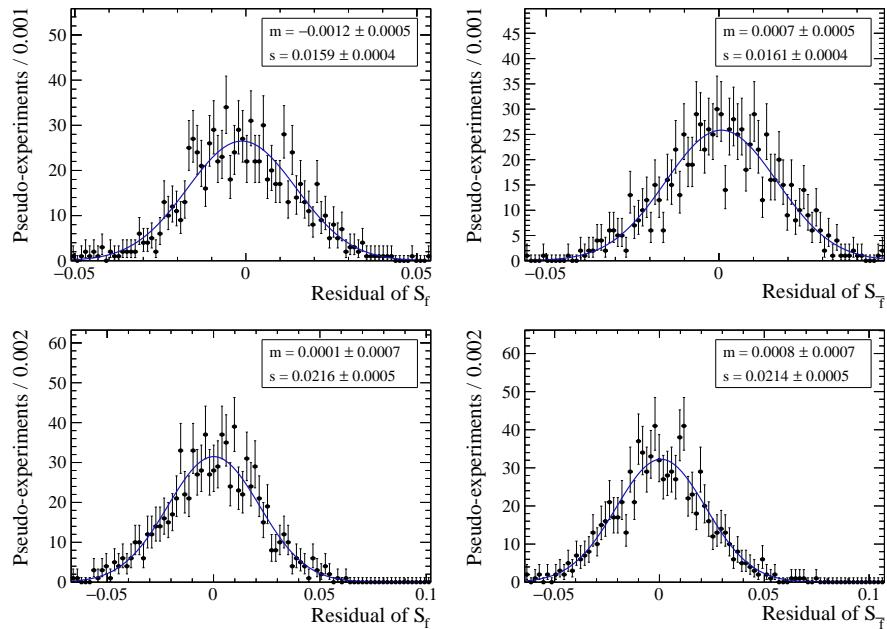


Figure 5.18 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the resolution model. Top: 75 fs resolution model. Bottom: 35 fs resolution model.

1396 Fixed  $C_f$ 

1397 Toys are generated with  $C_f = -C_{\bar{f}}$  set to the average of the measurements by Belle and  
 1398 BaBar minus the largest uncertainty among the two measurements, namely 0.993 [28, 29].  
 1399 The distributions of the residuals of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.19. Residuals consistent  
 1400 with zero are found, therefore the uncertainty on the residuals is assigned as systematic  
 uncertainty.

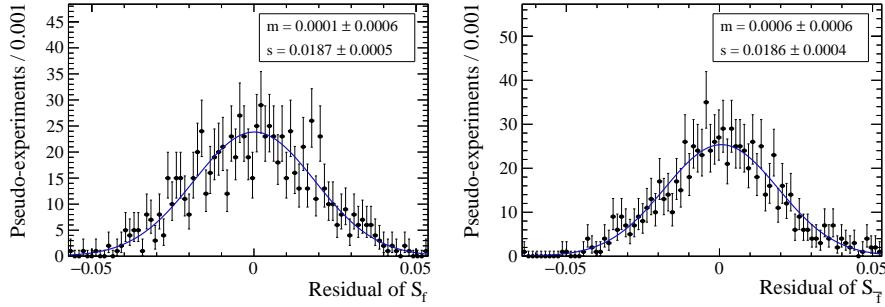


Figure 5.19 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the assumption  $C_f = -C_{\bar{f}} = 1$ .

1401

 1402 Fixed  $\Delta\Gamma$ 

1403 Toys are generated with  $\Delta\Gamma$  set to the world average value plus its uncertainty, namely  
 1404  $0.0079 \text{ ps}^{-1}$  [85]. Moreover, the  $D_f$  and  $D_{\bar{f}}$  coefficients (defined in Eqs. 1.35-1.38) have  
 1405 been fixed to their expected values of  $-0.0103$  and  $-0.0155$ , the same used in the Monte  
 1406 Carlo production of the  $B^0 \rightarrow D^- \pi^+$  sample (Appendix M). The distribution of the  
 1407 residuals of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.20. Residuals consistent with zero are found,  
 therefore the uncertainty on the residuals is assigned as systematic uncertainty.

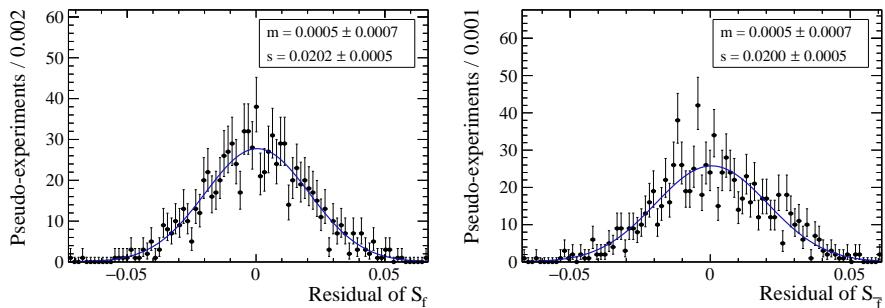


Figure 5.20 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the assumption  $\Delta\Gamma = 0$ .

1408

<sup>1409</sup> 5.5.3 Systematics related to the background subtraction

<sup>1410</sup> Systematic uncertainties can arise from the choice of the mass fit strategy adopted to calcu-  
<sup>1411</sup> late *sWeights* (Sec. 4.3). Fit B, used to compute the *sWeights*, is repeated in the full mass  
<sup>1412</sup> window ( $[5090, 6000]$  MeV/ $c^2$ ) instead of the narrow signal region ( $[5220, 5600]$  MeV/ $c^2$ ).  
<sup>1413</sup> In this way, the resulting sample is enriched in background events. The aim of this test  
<sup>1414</sup> is to estimate how much background events (with negative *sWeights*) affect the result  
<sup>1415</sup> for  $S_f$  and  $S_{\bar{f}}$  in the final decay time fit. The fitted total background yield in this new  
<sup>1416</sup> mass fit configuration is  $199\,767 \pm 481$ , compared to  $34\,102 \pm 299$  in the nominal fit  
<sup>1417</sup> configuration (Table 4.15). The projection of the PDF used for Fit B in the wide mass  
<sup>1418</sup> range is shown in Fig. 5.21. The new *sWeights* are then used in a decay time fit performed  
<sup>1419</sup> on the full sample following the same strategy as reported in Sec. 5.3. The correlated  
<sup>1420</sup> disagreement, defined as the difference between the fit results divided by the difference  
<sup>1421</sup> in quadrature between the fitted uncertainties, between the result of this fit and that of  
<sup>1422</sup> the nominal fit is  $2.3\sigma$  for  $S_f$  and  $1.8\sigma$  for  $S_{\bar{f}}$ . Because of this discrepancy, the  
<sup>1423</sup> difference between the newly obtained  $S_f$  and  $S_{\bar{f}}$  coefficients and the nominal values is  
<sup>1424</sup> taken as systematic uncertainty, yielding 0.0042 and 0.0023 for  $S_f$  and  $S_{\bar{f}}$  respectively.  
<sup>1425</sup> The correlation between the systematic uncertainties on  $S_f$  and  $S_{\bar{f}}$  is estimated to be 0.7,  
<sup>1426</sup> as shown in Appendix L.

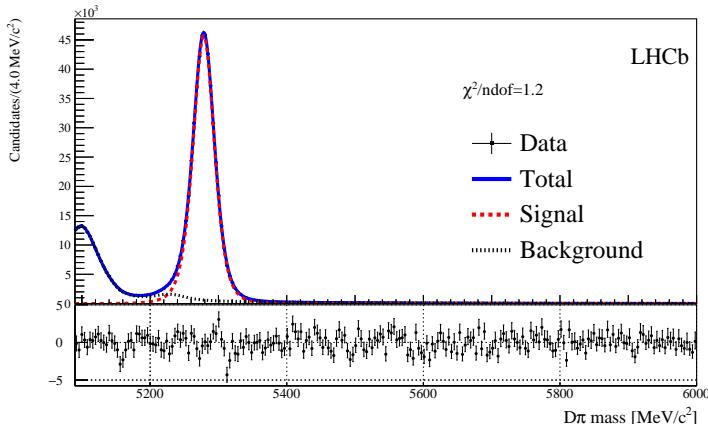


Figure 5.21 –  $D^\mp\pi^\pm$  mass distribution of the  $\pi$  sample with the results of Fit B in the large mass window superimposed.

<sup>1427</sup> Another test is made by repeating the mass fit with a different strategy:

- <sup>1428</sup> • a  $PIDK < 0$  cut (instead of  $PIDK < 5$ ) is applied on the pion PID in order to  
<sup>1429</sup> define the pion sample;
- <sup>1430</sup> • both Fit A and Fit B are performed in the narrow signal region ( $[5220, 5600]$  MeV/ $c^2$ );
- <sup>1431</sup> • during Fit A, only the pion sample is considered (no simultaneous fit in kaon and

1432 pion samples is performed);

- 1433 • only  $B^0 \rightarrow D^\mp K^\pm$  and combinatorial background are considered, whereas all the  
1434 other physical background are neglected;
- 1435 • the  $B^0 \rightarrow D^\mp K^\pm$  yield is Gaussian constrained to be  $0.0101 \pm 0.0012$  of the signal  
1436 yield, based on the selection efficiencies (including the  $\text{PID}K < 0$  cut) found on  
1437 Monte Carlo.

1438 The signal and total background yield obtained in this fit are  $406\,818 \pm 674$  and  $23\,938 \pm 266$   
1439 respectively. The projection of the PDF used for Fit A and Fit B in this configuration is  
1440 shown in Fig. 5.22. A decay time fit is performed on the resulting sample with *sWeights* by  
1441 following the same strategy as reported in Sec. 5.3. The correlated discrepancy between  
1442 the result of this fit and that of the nominal fit is  $0.4\,\sigma$  and  $1.6\,\sigma$  for  $S_f$  and  $S_{\bar{f}}$  respectively.  
1443 Given the good level of agreement, and the fact that systematic uncertainties on the PID  
1444 efficiencies are already considered, no further systematics are assigned.

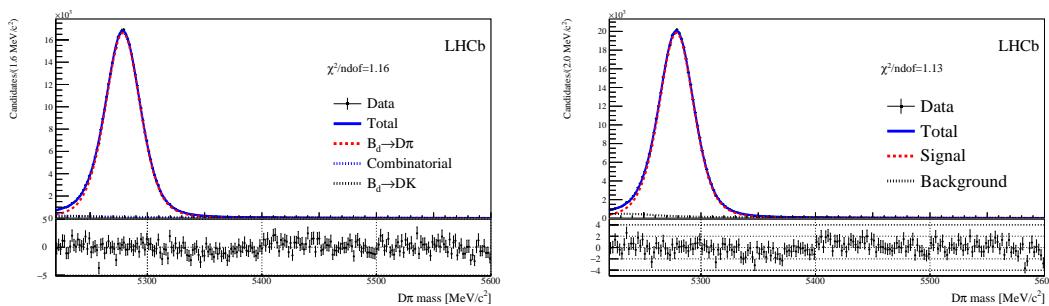


Figure 5.22 –  $D^\mp\pi^\pm$  mass distribution of the alternative  $\pi$  sample defined by the cut  $\text{PID}K < 0$  on the bachelor pion with the result of Fit A (left) and Fit B (right) superimposed.

1445 As additional cross-check, the decay time fit is repeated for  $B^0 \rightarrow D^\mp\pi^\pm$  candidates  
1446 restricted in the  $[5250, 5330]\,\text{MeV}/c^2$  invariant mass region, very close to the  $B^0 \rightarrow D^\mp\pi^\pm$   
1447 signal peak position. No *sWeights* are applied on this subsample. The correlated  
1448 disagreement between the result of this fit and that of the nominal fit is  $0.2\,\sigma$  and  $1.3\,\sigma$   
1449 for  $S_f$  and  $S_{\bar{f}}$  respectively. Given the good level of agreement, no further systematics are  
1450 assigned, and the following conclusions are drawn:

- 1451 • the amount of combinatorial and  $B^0 \rightarrow D^\mp K^\pm$  backgrounds in the signal region  
1452 is very small, and their presence doesn't affect significantly the fitted  $S_f$  and  $S_{\bar{f}}$   
1453 coefficients as these are compatible with the nominal fit result;
- 1454 • any systematics due to a wrong modelling of signal and/or background PDF in the  
1455  $B^0 \rightarrow D^\mp\pi^\pm$  signal peak region is negligible, since the fitted value obtained from

1456 the nominal fit (with *sWeights*) and this alternative fit (with no mass fit at all) are  
1457 compatible.



## 6 Summary and interpretation of the results

<sup>1459</sup>

Table 6.1 – Comparison of the measurements of  $S_f$  and  $S_{\bar{f}}$ . The first uncertainty is statistical, the second is systematic.

	$S_f$ [%]	$S_{\bar{f}}$ [%]
Belle [36]	$+6.8 \pm 2.9 \pm 1.2$	$+3.1 \pm 3.0 \pm 1.2$
Babar [33]	$-2.3 \pm 4.8 \pm 1.4$	$+4.3 \pm 4.8 \pm 1.4$
LHCb (this analysis) [86]	$+5.8 \pm 2.0 \pm 1.1$	$+3.8 \pm 2.0 \pm 0.7$

## 6.1 Interpretation of the results

In this thesis, a time-dependent analysis of the decay  $B^0 \rightarrow D^\pm \pi^\mp$  in order to extract the  $CP$  observables  $S_f$  and  $S_{\bar{f}}$  was presented. The values obtained are

$$S_f = 0.058 \pm 0.020(\text{stat}) \pm 0.011(\text{syst}), \quad (6.1)$$

$$S_{\bar{f}} = 0.038 \pm 0.020(\text{stat}) \pm 0.007(\text{syst}). \quad (6.2)$$

The statistical and systematic correlations between  $S_f$  and  $S_{\bar{f}}$  are  $+60\%$  and  $-41\%$ , respectively, and the total correlation is  $+44\%$ . These values are in agreement with and more precise than the measurements from Belle and Babar [36, 33]. A direct comparison is shown in Table 6.1.

The measurements of  $S_f$  and  $S_{\bar{f}}$  are interpreted in terms of the angle  $\gamma$  and the strong phase  $\delta$ , by following the definition of  $S_f$  and  $S_{\bar{f}}$  given in Sec. 1.4:

$$S_f = -\frac{2r_{D\pi} \sin[\delta - (\gamma + 2\beta)]}{1 + r_{D\pi}^2}, \quad S_{\bar{f}} = \frac{2r_{D\pi} \sin[\delta + (\gamma + 2\beta)]}{1 + r_{D\pi}^2}. \quad (6.3)$$

A frequentist method, called PLUGIN and described in Ref. [87], is adopted to derive confidence intervals for  $\gamma$  and  $\delta$  by using external inputs for  $\beta$  and  $r_{D\pi}$ . Given the observed values  $\vec{A}^{\text{obs}} = (S_f, S_{\bar{f}})$  and the parameters  $\vec{\alpha} = (\gamma, \delta)$ , a  $\chi^2(\vec{\alpha})$  function is built as

$$\chi^2(\vec{\alpha}) = -2 \ln \mathcal{L}(\vec{\alpha}) = -2 \ln f(\vec{A}^{\text{obs}} | \vec{\alpha}) \propto \left( \vec{A}(\vec{\alpha}) - \vec{A}^{\text{obs}} \right)^T V^{-1} \left( \vec{A}(\vec{\alpha}) - \vec{A}^{\text{obs}} \right), \quad (6.4)$$

where  $V$  is the covariance matrix of  $S_f$  and  $S_{\bar{f}}$ . The best fit point,  $\vec{\alpha}_{\min}$ , is the one that minimises the expression of Eq. 6.4. The  $p$ -value, or 1-CL, is computed for each possible value of each component of  $\vec{\alpha}$  ( $\gamma$  and  $\delta$ ) as follows:

- a value for a given component of  $\vec{\alpha}$  (e.g.  $\gamma = \gamma_0$ ) is chosen;
- the corresponding test statistics  $\Delta\chi^2 = \chi^2(\vec{\alpha}(\gamma_0)) - \chi^2(\vec{\alpha}_{\min})$  is built;
- pseudoexperiments are generated to sample values for  $\vec{A}$ , called  $\vec{A}_j$ , from the PDF  $f(\vec{A} | \vec{\alpha}(\gamma_0))$ ;

- for each  $\vec{A}_j$ , a new value for the test statistics,  $\Delta\chi_j^2$ , is computed by replacing  $\vec{A}^{\text{obs}} \rightarrow \vec{A}_j$  in Eq. 6.4, and by taking the difference between the minimised values of  $\chi^2$  with respect to  $\vec{\alpha}$ , once with  $\gamma$  as free parameter, and once with  $\gamma = \gamma_0$ ;
- the  $p$ -value is computed as the fraction of pseudoexperiments for which  $\Delta\chi^2 < \Delta\chi_j^2$ .

In all the minimisation steps described above, the values  $\beta = (22.2 \pm 0.7)^\circ$  [85] and  $r_{D\pi} = (1.82 \pm 0.12 \pm 0.36(\text{SU}(3))\%$  are used as external Gaussian constraints. The latter is calculated from the branching fraction of  $B^0 \rightarrow D_s^- \pi^+$  decays, assuming SU(3) symmetry, following the same relation of Refs. [29, 28]:

$$r_{D\pi} = \tan \theta_c \frac{f_{D^+}}{f_{D_s^-}} \sqrt{\frac{\mathcal{B}(B^0 \rightarrow D_s^- \pi^+)}{\mathcal{B}(B^0 \rightarrow D^- \pi^+)}} \quad (6.5)$$

where  $\tan \theta_c = 0.23101 \pm 0.00032$  [21] is the tangent of the Cabibbo angle,  $\frac{f_{D_s^-}}{f_{D^+}} = 1.173 \pm 0.003$  [88, 89, 90] is the ratio of decay constants,  $\mathcal{B}(B^0 \rightarrow D_s^- \pi^+) = (2.16 \pm 0.26) \times 10^{-5}$  [22] and  $\mathcal{B}(B^0 \rightarrow D^- \pi^+) = (2.52 \pm 0.13) \times 10^{-3}$  [22]. An additional 20% relative error is added on  $r_{D\pi}$  to account for uncertainties due to possible non-factorizable SU(3)-breaking effects, as reported in Ref. [91].

The angle  $\gamma$  is determined to be in the interval  $[5, 86]^\circ \cup [185, 266]^\circ$  and  $\delta$  to be in the interval  $[-41, 41]^\circ \cup [140, 220]^\circ$ , both at the 68% CL. These intervals are illustrated in Fig. 6.1. In Fig. 6.2, contours are shown in the two-dimensional plane  $(\gamma, \delta)$ .

In addition to  $\gamma$  and  $\delta$ , the interval of  $|\sin(2\beta + \gamma)|$  is determined as well. This quantity does not rely on any external input for  $\beta$ , and it is thus experimentally cleaner. This interval is found to be  $[0.77, 1.0]$  at the 68% CL. as shown in Fig. 6.3.

The intervals for  $\gamma$ ,  $\delta$  and  $|\sin(2\beta + \gamma)|$  are also determined by assuming a SU(3)-breaking uncertainty of 0%, 20% and 100% on the value of  $r_{D\pi}$ . These are presented in Figs. 6.4 and 6.5.

## 6.2 Summary and perspectives

The  $B^0 \rightarrow D^\mp \pi^\pm$  analysis presented in this thesis was performed on the full LHCb Run 1 (2011–2012) dataset, corresponding to  $3 \text{ fb}^{-1}$  of collected data.

As can be seen from Eqs. 6.1 and 6.2, the dominant contribution in the error budget of  $S_f$  and  $S_{\bar{f}}$  is the statistical uncertainty. For this reason, the precision of the measurement can be easily improved by using a larger data sample.

The expected statistics that will be collected during Run 2 (2015–2018) is  $\sim 6 \text{ fb}^{-1}$ , with a luminosity of  $4 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ . The centre-of-mass energy during Run 2 (13 TeV)

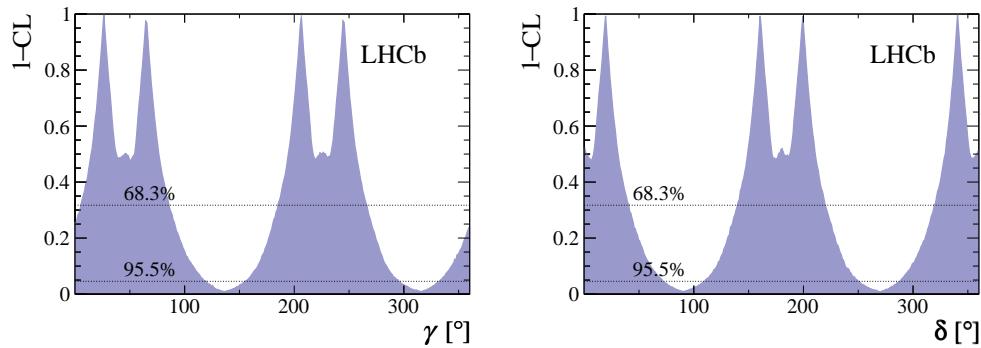


Figure 6.1 –  $p$ -value, or  $1-CL$ , as a function of  $\gamma$  (left) and  $\delta$  (right) obtained using the measured values of  $S_f$  and  $S_{\bar{f}}$ .

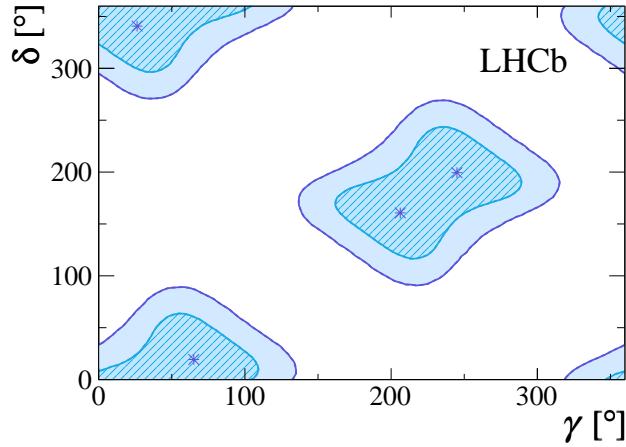


Figure 6.2 – Contours in the two-dimensional plane ( $\gamma$ ,  $\delta$ ) obtained using the measured values of  $S_f$  and  $S_{\bar{f}}$ . The crosses indicates the preferred values ( $1 - CL = 1$ ). The blue hatched (solid) areas hold 39% (87%) confidence level.

is about twice the Run 1 value; for this reason, the  $B^0$  production cross-section is also increased approximately by a factor two compared to Run 1. So, the increase in the number of reconstructed  $B^0 \rightarrow D^\mp \pi^\pm$  decays in Run 2 compared to Run 1 is about  $\sim 2 \times \frac{6}{3} = 4$ , which corresponds to a factor  $\sim \sqrt{4} = 2$  of decrease in statistical uncertainty.

After a long shutdown of three years (2018–2021), the LHC will restart with an increased luminosity of  $2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ , and  $pp$  collisions will be delivered during Run 3 (2021–2023) and Run 4 (2026–2029). In the meanwhile, an important upgrade of several LHCb detectors will take place. Particularly relevant for the  $B^0 \rightarrow D^\mp \pi^\pm$  analysis are the VELO upgrade [92], which will imply an improvement of hit efficiency and IP resolution thanks to the pixel geometry, the Upstream Tracker [93] (replacing the TT), which will improve the particle acceptance, and the scintillating fibre tracker [93] (replacing IT and OT), which is designed to cope with the expected higher occupancy due to the higher luminosity. At the

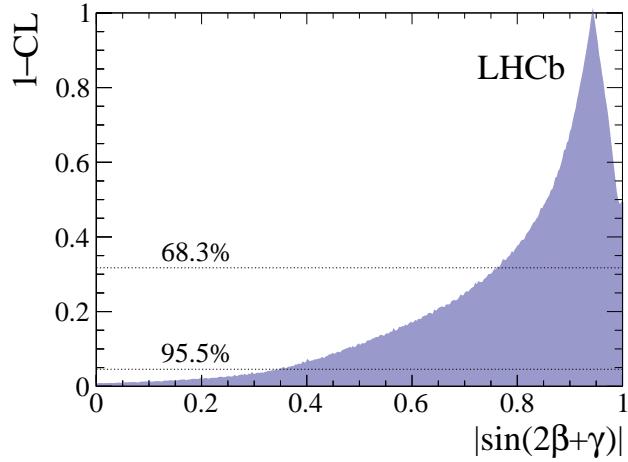


Figure 6.3 –  $p$ -value, or 1–CL, as a function of  $|\sin(2\beta + \gamma)|$  using the measured values of  $S_f$  and  $S_{\bar{f}}$ .

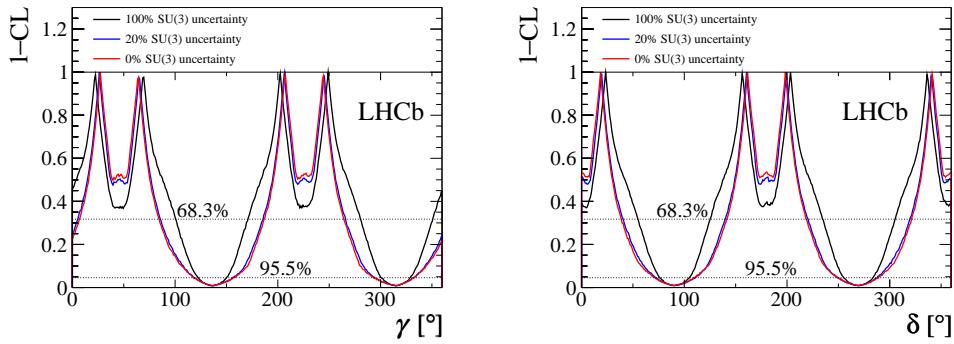


Figure 6.4 –  $p$ -value, or 1–CL, as a function of (left)  $\gamma$  and (right)  $\delta$  for assumptions of 0 %, 20 % and 100 % for the SU(3) uncertainty on the parameter  $r_{D\pi}$ .

end of Run 4,  $\sim 50 \text{ fb}^{-1}$  of total collected data are expected. The centre-of-mass energy of the  $pp$  collisions will be 13 TeV. Assuming to have the same signal-to-background ratio as of Run 1, the increase in the number of reconstructed  $B^0 \rightarrow D^\mp \pi^\pm$  decays at the end of Run 4 is about  $\sim 2 \times \frac{47}{3} \sim 30$ , which corresponds to a factor  $\sim \sqrt{30} \sim 5$  of decrease in statistical uncertainty compared to the Run 1 result.

Finally, an upgrade of the LHC will allow to increase the luminosity up to  $1 - 2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  starting from 2031. The expected statistics that will be collected in this high-luminosity scenario is  $\sim 300 \text{ fb}^{-1}$  of total collected data from  $pp$  collisions. Given the same centre-of-mass energy of 13 TeV, the increase in the number of  $B^0 \rightarrow D^\mp \pi^\pm$  decays will be about  $\sim 2 \times \frac{297}{3} \sim 200$ , corresponding to a reduction of the Run 1 statistical uncertainty by a factor  $\sim \sqrt{200} \sim 14$ .

These extrapolations are made by assuming the same tagging power as obtained on Run 1

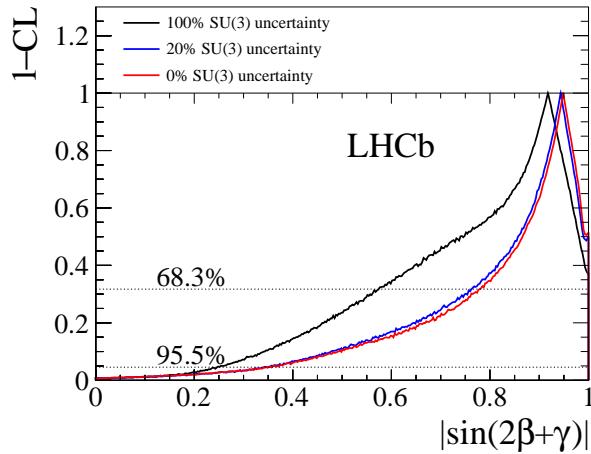


Figure 6.5 – 1-CL as a function of  $|\sin(2\beta + \gamma)|$  for assumptions of 0 %, 20 % and 100 % for the SU(3) uncertainty on the parameter  $r_{D\pi}$ .

1522 data; future developments of flavour tagging algorithms are thus crucial to further improve  
 1523 these projections. As it was described in Sec. 3.3.4, the performance of OS taggers on Run  
 1524 2 data is compatible with, and not worse than, the one obtained on Run 1 data thanks  
 1525 to the reoptimisation campaign on Run 2 data, while the tagging power of the SS $\pi$  and  
 1526 SS $\rho$  taggers is increased. Preliminary studies on simulated samples showed that both  
 1527 SS taggers and Run 2-optimised OS taggers have similar performances in Run 3 as the  
 1528 ones of Run 2; further improvements can be achieved by tuning these taggers specifically  
 1529 on Run 3 (and beyond) conditions. In parallel, a new approach, called *inclusive tagger*,  
 1530 is under development. This algorithm, consisting of a deep neural network, exploits all  
 1531 tracks and vertices reconstructed in the events in order to provide a tagging decision and  
 1532 a mistag estimate. Preliminary results on Run 1  $B^\pm \rightarrow J/\psi K^\pm$  data indicate a tagging  
 1533 power of the order of  $\sim 7 - 8\%$ , which would represent an absolute increase of  $\sim 2\%$   
 1534 compared to the combination of the standard OS and SS taggers.

1535 Concerning the systematic uncertainties on  $S_f$  and  $S_{\bar{f}}$ , the external constraint on  $\Delta m$ ,  
 1536 which is the dominant contribution to the systematic error budget, will also benefit  
 1537 from the high statistics collected by LHCb, since the world-leading measurement of this  
 1538 parameter is already obtained by the LHCb collaboration from semileptonic  $B^0$  decays [82],  
 1539 and this result will be updated with new data. In addition to  $\Delta m$ , other sources of  
 1540 systematic uncertainties that will be reduced thanks to the increased statistics foreseen  
 1541 for the next decades are  $\Delta\Gamma$  and the decay-time resolution.

The precision on the values of  $\gamma$  and  $\delta$  extracted from the measured values for  $S_f$  and  $S_{\bar{f}}$  will benefit from the increased knowledge on  $\beta$  and  $r_{D\pi}$ . The  $\beta$  angle will be measured with unprecedent precision also thanks to the Belle II experiment, which will start its operations between 2018 and 2019. The precision on the  $r_{D\pi}$  parameter will increase thanks to the improvement in the measurements of the  $B^0 \rightarrow D_s^- \pi^+$  and  $B^0 \rightarrow D^- \pi^+$

## 6.2. Summary and perspectives

---

branching ratios, according to Eq. 6.5. Moreover, if the  $B_s^0 \rightarrow D^+ K^-$  decay will be observed, an independent estimation of  $r_{D\pi}$  will be available by following the relation

$$r_{D\pi} = \frac{f_\pi}{f_K} \sqrt{\frac{\mathcal{B}(B_s^0 \rightarrow D^+ K^-)}{\mathcal{B}(B^0 \rightarrow D^- \pi^+)}} , \quad (6.6)$$

where SU(3) symmetry is assumed as for Eq. 6.5 and  $\frac{f_\pi}{f_K} = 1.1956(10)^{+26}_{-18}$  [89]. A search for  $B_s^0 \rightarrow D^+ K^-$  on Run 1 data gave a null result [94].

The value of  $\gamma$  that will be extracted by future time-dependent analyses of  $B^0 \rightarrow D^\mp \pi^\pm$  decays, and similarly from  $B_s^0 \rightarrow D_s^\mp K^\pm$  decays, will be one of the inputs for the global combination of direct measurements performed by LHCb. The current sensitivity, which is of the order of  $\sim 5 - 6^\circ$  [87], is expected to go down at the level of  $\sim 0.4^\circ$  with  $\sim 300 \text{ fb}^{-1}$  of total collected data. This sensitivity will allow to test the theoretical predictions for  $\gamma$ , which are characterised by negligible uncertainties, and to compare these direct measurements with the indirect determination from other CKM parameters: a discrepancy in any of these tests will be a clear signature of new physics beyond the SM.



## Appendices

1554 **A Opposite-side tagging studies**

1555 **A.1 Mass fit of  $B^\pm \rightarrow D^0\pi^\pm$**

1556 A fit to the mass distribution of  $B^\pm$  candidates is done to calculate *sWeights*, used in the  
1557 subsequent steps of the analysis to subtract the backgrounds surviving the selection. A  
1558 two-step procedure similar to that adopted for the  $B^0 \rightarrow D^\mp\pi^\pm$  analysis (“Fit A” in a  
1559 wide mass window to account for all backgrounds, and “Fit B” in a subset to calculate the  
1560 weights, as described in Sec. 4.3, where the same mass windows are adopted) is used for  
1561 the fit of the  $B^+ \rightarrow D^0\pi^+$  candidates. The projection of the total PDF on the  $\pi$  sample  
1562 and the  $K$  sample (“Fit A”) is shown in Fig. A.1, as well as the projection of the total  
1563 PDF in the reduced sample (“Fit B”). The  $\pi$  sample and  $K$  sample are defined by the  
1564 PID requirement on the companion track,  $\text{PIDK} < 5$  and  $\text{PIDK} > 5$ , respectively.

1565 The background components expected in the  $\pi$  sample for the  $B^+ \rightarrow D^0\pi^+$  mass fit are  
1566 listed below, together with the PDF used for each component:

- 1567 •  $B^\pm \rightarrow D^0\pi^\pm$ : double-sided Hypatia function.
- 1568 •  $B^+ \rightarrow D^0K^+$ : double-sided Hypatia function.
- 1569 •  $B^0 \rightarrow D^0\pi^\pm\pi^\mp$ : Crystal ball function plus Gaussian function.
- 1570 •  $B^\pm \rightarrow D^{*0}\pi^\pm$ : Johnson SU function plus Gaussian function.
- 1571 • Combinatorial background: single exponential function.

1572 The list of components expected in the  $K$  sample is the following:

- 1573 •  $B^\pm \rightarrow D^0\pi^\pm$ : double-sided Hypatia function.
- 1574 •  $B^\pm \rightarrow D^0K^\pm$ : single-sided Hypatia function.
- 1575 •  $B^\pm \rightarrow D^{*0}\pi^\pm$ : Crystal ball function plus exponential function.
- 1576 •  $B^\pm \rightarrow D^0K^{*\pm}$ : Gaussian function.
- 1577 • Combinatorial background: single exponential function.

1578 All the PDFs listed above are defined in Appendix F.

1579 The values for the fitted parameters floated in the fit are reported in Table A.1 for Fit A  
1580 and Table A.2 for Fit B. The naming convention for each parameter is similar to the one  
1581 used in Sec. 4.3.2.

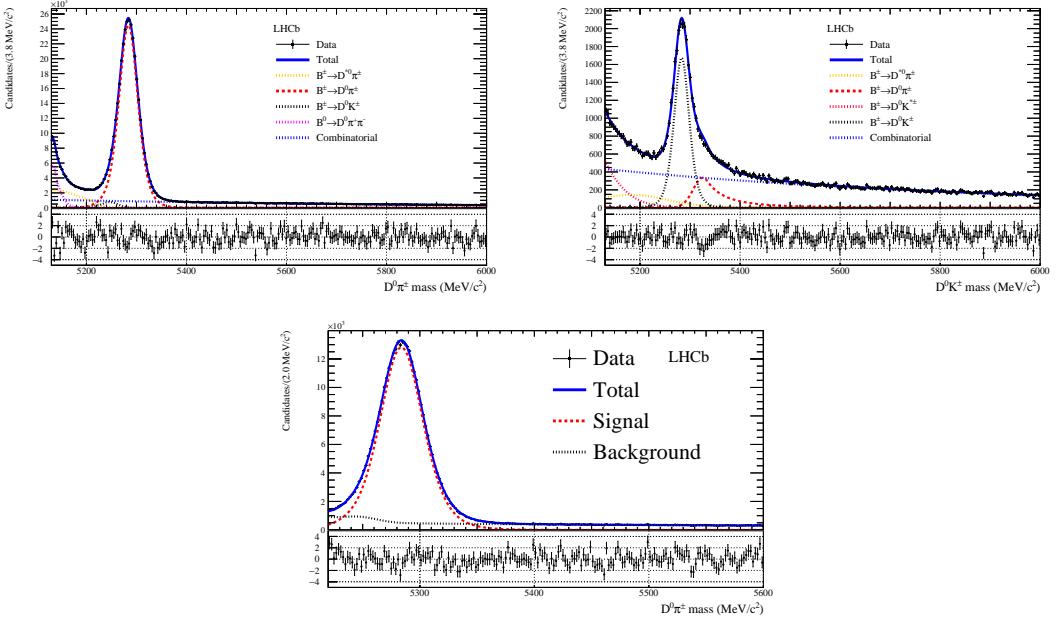


Figure A.1 – Top left:  $D^0\pi^\pm$  mass distribution of the  $\pi$  sample. Top right:  $D^0\pi^\pm$  mass distribution of the  $K$  sample. The result of the simultaneous fit (Fit A) to both samples is superimposed. Bottom:  $D^\mp\pi^\pm$  mass distribution of the  $\pi$  sample with the results of Fit B superimposed.

## 1582 A.2 Reweighting of $B^\pm \rightarrow D^0\pi^\pm$ to $B^0 \rightarrow D^\mp\pi^\pm$

1583 In order to improve the OS calibration portability, a multi-dimensional reweighting of the  
 1584 *sWeighted*  $B^\pm \rightarrow D^0\pi^\pm$  distributions is made to match the  $B^0 \rightarrow D^\mp\pi^\pm$  kinematics.

1585 The reweighting is made in two steps. In the first step, the variables considered in the  
 1586 reweighting are the transverse momentum, the pseudo-rapidity  $\eta$  and the decay time  $\tau_B$  of  
 1587 the  $B$  candidate, as well as the number of tracks and the number of primary vertices of the  
 1588 events. A BDT-based approach is followed in order to cope with the high dimensionality  
 1589 of the space as well as with the correlations among variables [95]. A comparison between  
 1590 weighted and unweighted distributions is provided in Figs. A.2 and A.3.

1591 In the second step, a new weight is computed by comparing the two-dimensional distribu-  
 1592 tions of the  $D$  meson decay time and HLT2 trigger composition between  $B^\pm \rightarrow D^0\pi^\pm$   
 1593 and  $B^0 \rightarrow D^\mp\pi^\pm$  after *sWeights* and the weights from the first step are applied. The  
 1594 HLT2 composition observable is a categorical variable which describes which HLT2 trigger  
 1595 line has been fired by the  $B$  candidate:

- 1596     • Hlt2Topo2BodyBBDTDecision only (value 0);
- 1597     • Hlt2Topo3BodyBBDTDecision or Hlt2Topo4BodyBBDTDecision only (value 1);

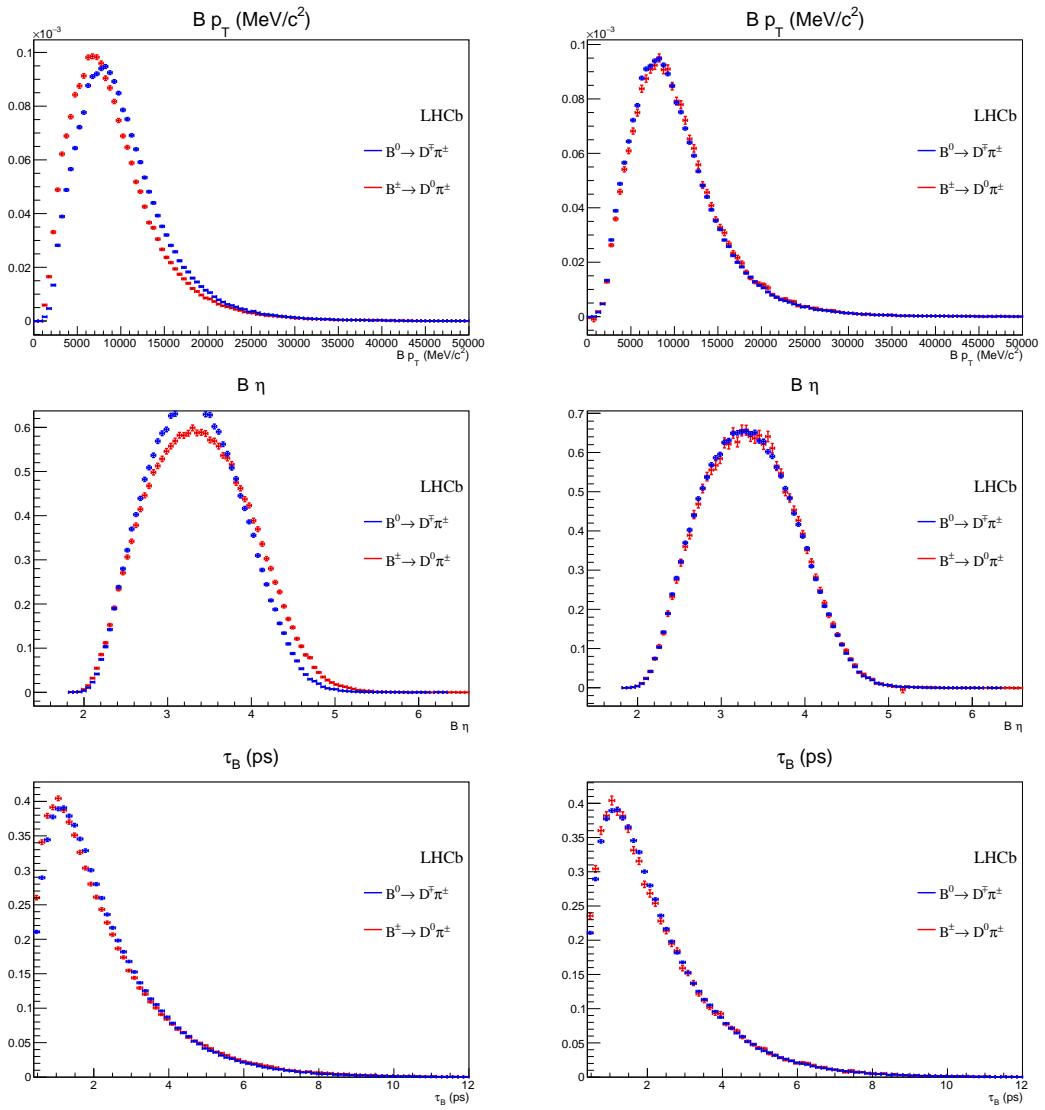


Figure A.2 – Normalised *s*Weighted distributions of the transverse momentum, the pseudo-rapidity  $\eta$  and the decay time  $\tau_B$  of the  $B^0$  and  $B^{\pm}$  mesons. Left: unweighted distributions. Right: distributions after reweighting the  $B^{\pm} \rightarrow D^0 \pi^{\pm}$  events.

Table A.1 – Results of the  $B^\pm \rightarrow D^0\pi^\pm$  mass fit (Fit A).

Parameter name	Fitted value
$\mu_{B^0 \rightarrow D\pi\pi}^\pi$	$5132.61 \pm 0.23$
$s\sigma_{B^0 \rightarrow D\pi\pi}^\pi$	$0.780 \pm 0.015$
$\sigma_{B^+ \rightarrow D^0 K}^K$	$19.47 \pm 0.30$
$\sigma_{B^+ \rightarrow D^0 K}^\pi$	$16.62 \pm 0.69$
$\mu_{B^+ \rightarrow D^0 K^*}^K$	$4960 \pm 150$
$\sigma_{B^+ \rightarrow D^0 K^*}^K$	$88 \pm 37$
$c_{\text{comb}}^K$	$-0.001305 \pm 0.000035$
$c1_{\text{comb}}^\pi$	$-0.001279 \pm 0.000022$
$\mu_{B^+ \rightarrow D^0 \pi}^K$	$5283.18 \pm 0.22$
$\mu_{B^+ \rightarrow D^0 \pi}^\pi$	$5283.880 \pm 0.046$
$sa_{B^+ \rightarrow D^0 \pi}^\pi$	$0.804 \pm 0.016$
$\mu_{B^+ \rightarrow D^0 \pi}^K$	$5325.4 \pm 1.2$
$sn_{B^+ \rightarrow D^0 \pi}^\pi$	$2.70 \pm 0.94$
$\sigma_{B^+ \rightarrow D^0 \pi}^\pi$	$22.850 \pm 0.054$
$N_{B^0 \rightarrow D^0 \pi\pi}^\pi$	$27245 \pm 430$
$N_{B^+ \rightarrow D^0 K}^K$	$18030 \pm 296$
$N_{LM}^K$	$5154 \pm 944$
$N_{B^+ \rightarrow D^{*0} \pi}^K$	$5704 \pm 1350$
$N_{B^+ \rightarrow D^{*0} \pi}^\pi$	$41871 \pm 578$
$N_{\text{comb}}^K$	$58761 \pm 555$
$N_{\text{comb}}^\pi$	$146824 \pm 793$
$N_{B^+ \rightarrow D\pi}^\pi$	$322597 \pm 812$

1598 • overlap of the first two categories (value 2).

1599 This reweighting is done separately from the first one in order to avoid a too fine partition  
 1600 of the samples, which would result in very low statistics in less populated bins. The  
 1601 reason why HLT2 trigger and  $D$  decay time are reweighted simultaneously is that these  
 1602 two observables are correlated. The result of this second reweighting is shown in Fig. A.4.

### 1603 A.3 GOF tests for OS calibration on $B^\pm \rightarrow D^0\pi^\pm$ data

1604 The number of free parameters (10) used in the GLM model for the OS calibration  
 1605 (Sec. 3.2.1) is the minimum number to obtain satisfactory GOF metrics. The GOF tests  
 1606 are performed automatically by the EPM; the metrics include the Pearson  $\chi^2$ , the deviance  
 1607  $G^2$ , the Cressie-Read ( $CR$ ) metric and the Le Cessie-van Houwelingen-Copas-Hosmer  
 1608 metric ( $S$ ), all described in Ref. [61].

1609 All these tests return a normally distributed score: this means that the score is equal

Table A.2 – Results of the  $B^\pm \rightarrow D^0\pi^\pm$  mass fit (Fit B).

Parameter	Fitted value
$N_{\text{bkg}}^\pi$	$85687 \pm 377$
$N_{B^\pm \rightarrow D\pi}^\pi$	$319974 \pm 612$

Table A.3 – GOF scores of two OS calibration fits of the reweighted  $B^\pm \rightarrow D^0\pi^\pm$  dataset.

GOF metric	Score (10 parameters)	Score (8 parameters)
$\chi^2$	-2.2	4.1
$G^2$	0.7	-3.9
$CR$	-1.7	2.9
$S$	1.8	-4.3

1610 to the distance (measured in standard deviations) from the perfect case, which is a null  
 1611 score. A comparison between the GOF scores obtained for the nominal calibration (10  
 1612 free parameters) and a simplified model (8 free parameters) is shown in Table A.3. In a  
 1613 simplified model, all scores are more than  $\sim 3$  standard deviations away from a perfect  
 1614 fit, whereas the scores for the nominal model are  $\sim 2$  standard deviations at most. For  
 1615 this reason, a calibration with 10 free parameters are chosen, and the model cannot be  
 1616 simplified further.

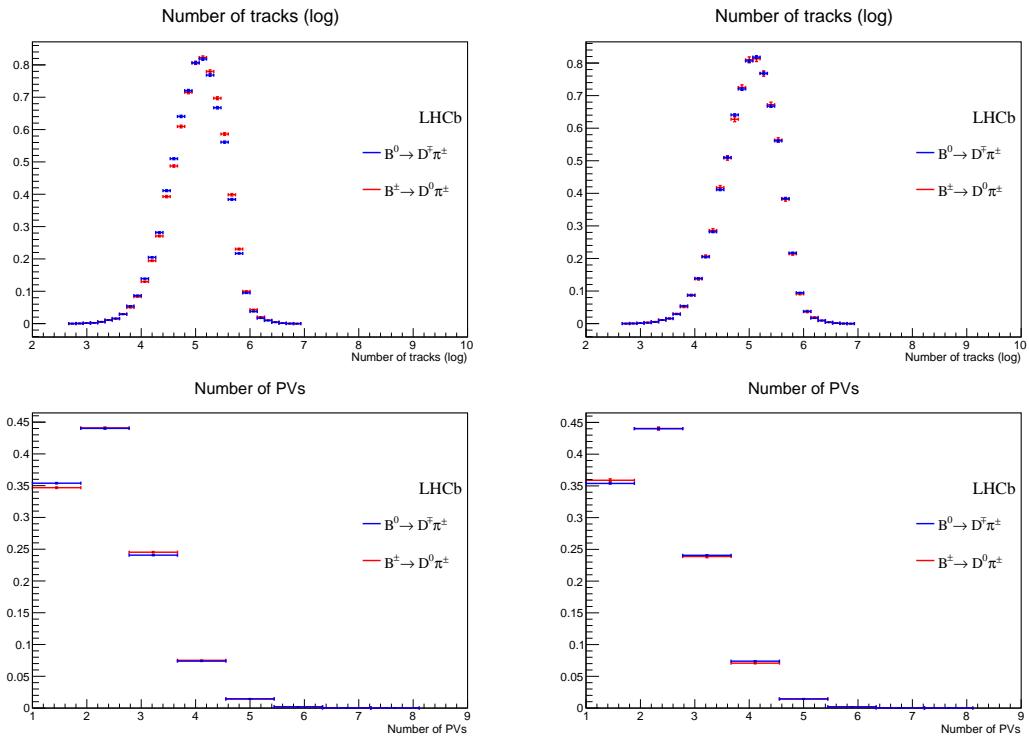


Figure A.3 – Normalised  $sWeighted$  distributions of the number of tracks and PVs in a  $B^0$  or  $B^\pm$  event. Left: unweighted distribution. Right: distributions after reweighting the  $B^\pm \rightarrow D^0\pi^\pm$  events.

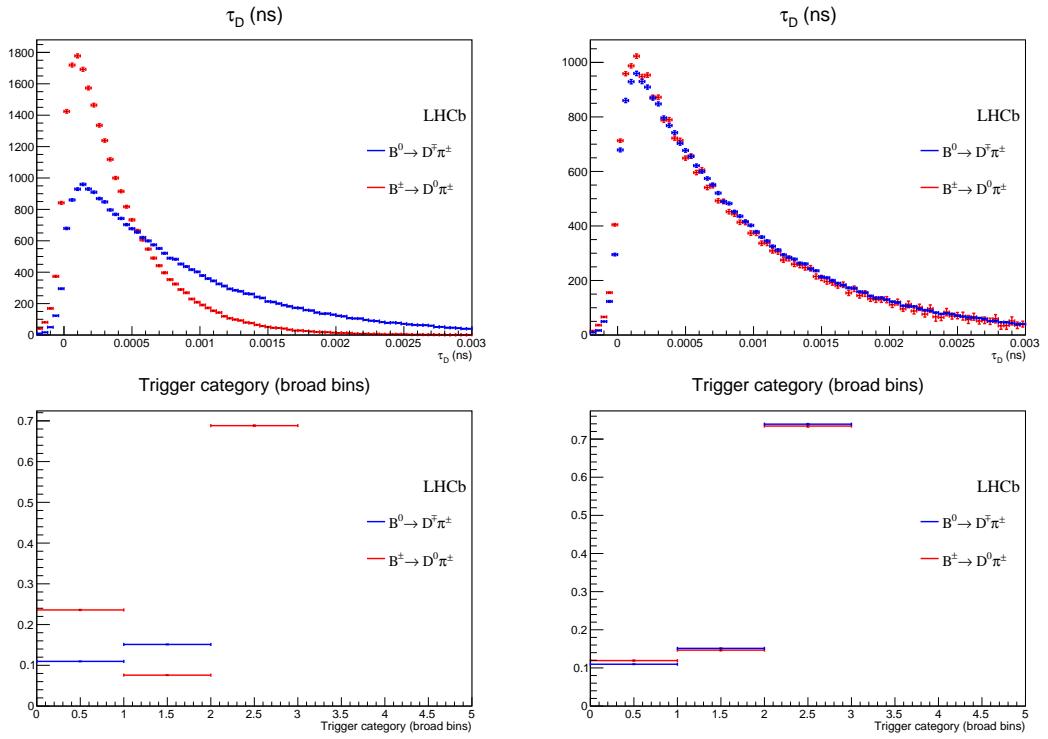


Figure A.4 – Normalised  $s\text{Weighted}$  distributions of the  $D^\mp$  and  $D^0$  mesons decay time and HLT2 trigger composition, where the weight obtained from the first reweighting step is also applied. Left: unweighted distributions. Right: distributions after reweighting the  $B^\pm \rightarrow D^0\pi^\pm$  events.

1617 **B Opposite-side electron optimisation**

1618 The correlation between the predicted mistag  $\eta$  and the input features is shown in  
1619 Figs. B.1, B.2 and B.3 for the Run 1 new, Run 2 B2CC and Run 2 B2OC implementations  
1620 of the OSe tagger, respectively. This correlation, known as *partial dependence*, allows to  
1621 check the impact of each feature on the classifier, in addition to the F-score (Sec. 3.3.3).

1622 An important step in the BDT development is the hyperparameter tuning. In particular,  
1623 the maximum depth (md) of each tree of the ensemble and the number of trees (nt) are  
1624 optimised. In order to do so, a cross-validation+bootstrapping procedure is followed:

1625 • For a given set of maximum depth and number of trees values, the training set is  
1626 bootstrapped 10 times. Each bootstrapped sample is then divided in three exclusive  
1627 subsamples.

1628 • The first subsample is used to train a BDT. The BDT is then transformed into  
1629 a mistag probability, and a calibration is performed on the second subsample (a  
1630 simple second order logistic function is used). Finally, the calibration is applied on  
1631 the third sample, where the per-event tagging power is computed. The ROC AUC  
1632 is also obtained as additional performance metric.

1633 • The above procedure is repeated by permutating the 3 samples. This means that,  
1634 in total, there are  $3 \times 10 = 30$  approximately independent estimations of the BDT  
1635 performance for each set of hyperparameters. The average tagging power and ROC  
1636 AUC values are finally computed over the 30 estimations, together with the standard  
1637 error on the mean.

1638 The result for the Run 1 new OSe algorithm is shown in Fig. B.4. The performance is  
1639 weakly dependent on the hyperparameters. For this reason, the maximum depth and  
1640 the number of trees are fixed to 3 and 300 respectively, in order to reduce complexity. A  
1641 similar result is observed for the Run 2 B2OC and Run 2 B2CC OSe algorithms.

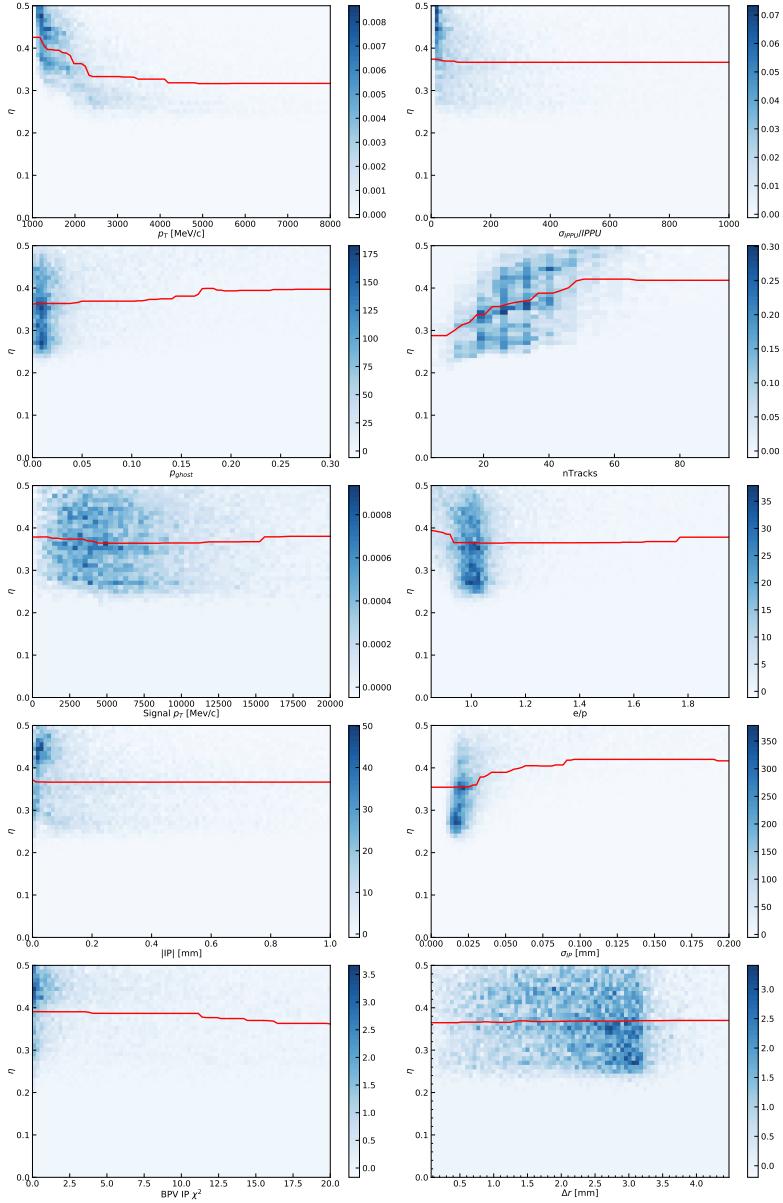


Figure B.1 – Partial dependence of the predicted mistag  $\eta$  (OSe Run 1 tagger) for each feature used as BDT input, marginalised over any other feature. The blue two-dimensional distributions represent the *sWeighted* data, whereas the red line shows the average  $\eta$  for each feature bin.

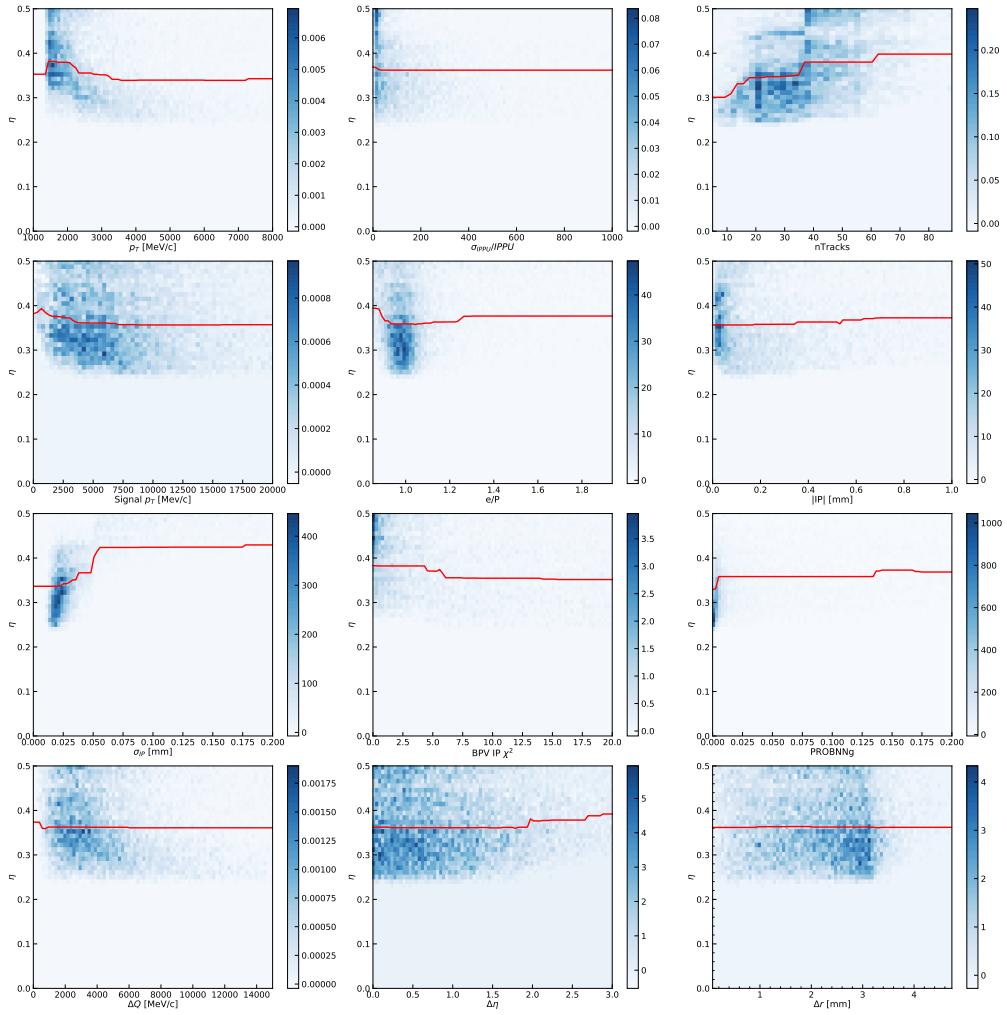


Figure B.2 – Partial dependence of the predicted mistag  $\eta$  (OSe Run 2 B2CC tagger) for each feature used as BDT input, marginalised over any other feature. The blue two-dimensional distributions represent the *sWeighted* data, whereas the red line shows the average  $\eta$  for each feature bin.

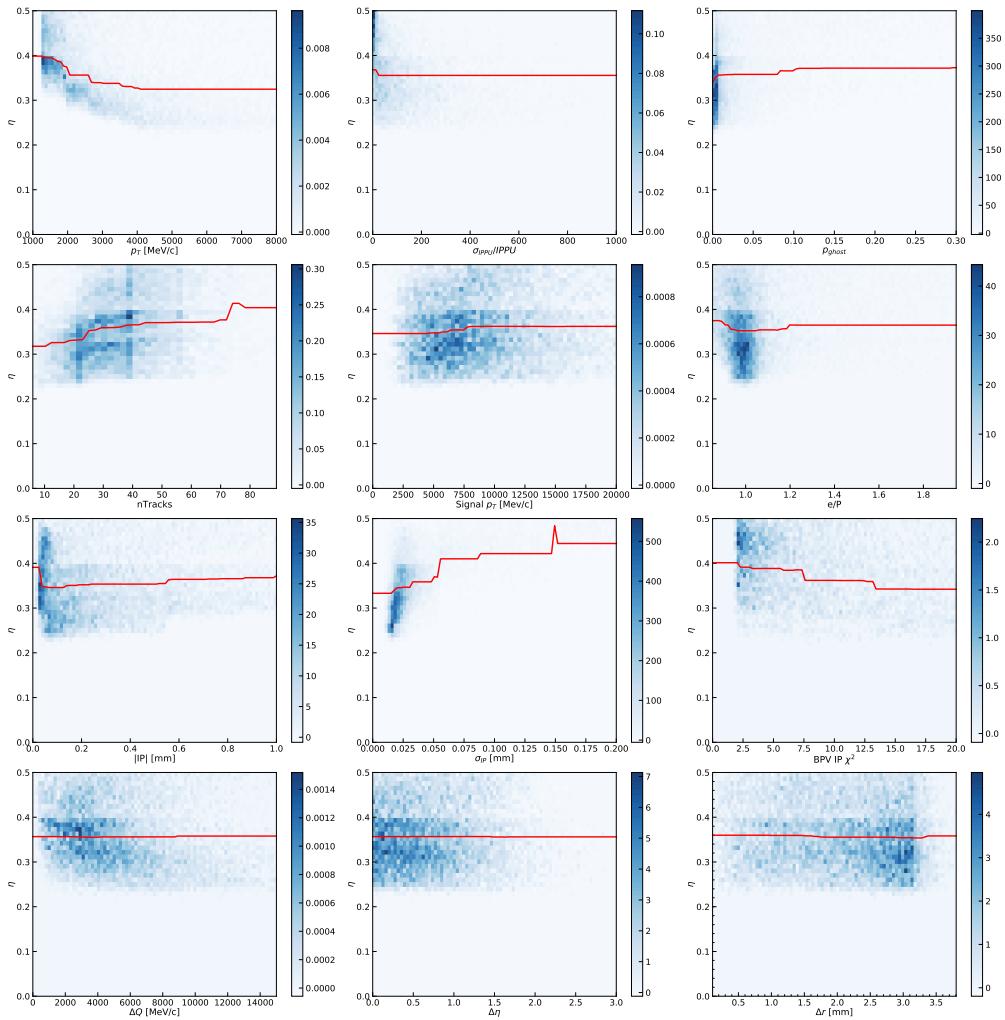


Figure B.3 – Partial dependence of the predicted mistag  $\eta$  (OSe Run 2 B2OC tagger) for each feature used as BDT input, marginalised over any other feature. The blue two-dimensional distributions represent the *sWeighted* data, whereas the red line shows the average  $\eta$  for each feature bin.

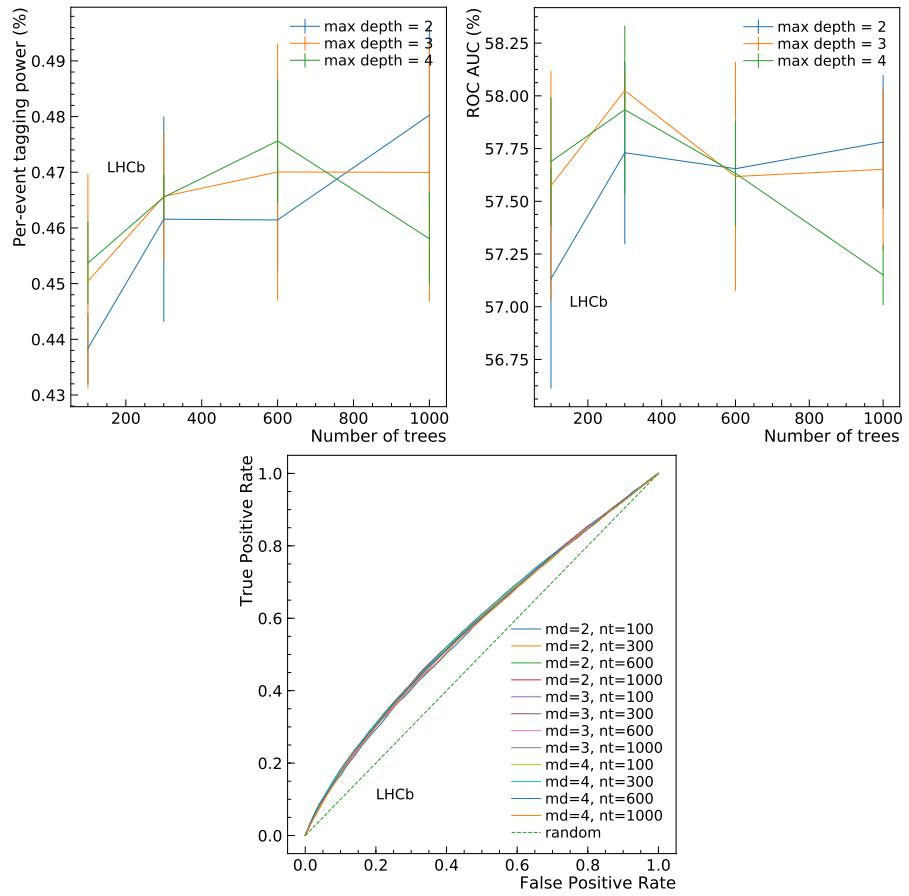


Figure B.4 – Tagging power (top left), ROC AUC (top right), and ROC curve (bottom) for each set of hyperparameters considered for the tuning of the BDT classifier used for the mistag estimation by the Run 1 new version of the OSe tagger (cross-validation).

## 1642 C $B^0 \rightarrow D^\mp \pi^\pm$ selection studies

### 1643 C.1 BDT input features

1644 The distributions of the input features for the BDT (listed and defined in Table 4.5) are  
 1645 shown in Figs. C.1, C.2 and C.3.

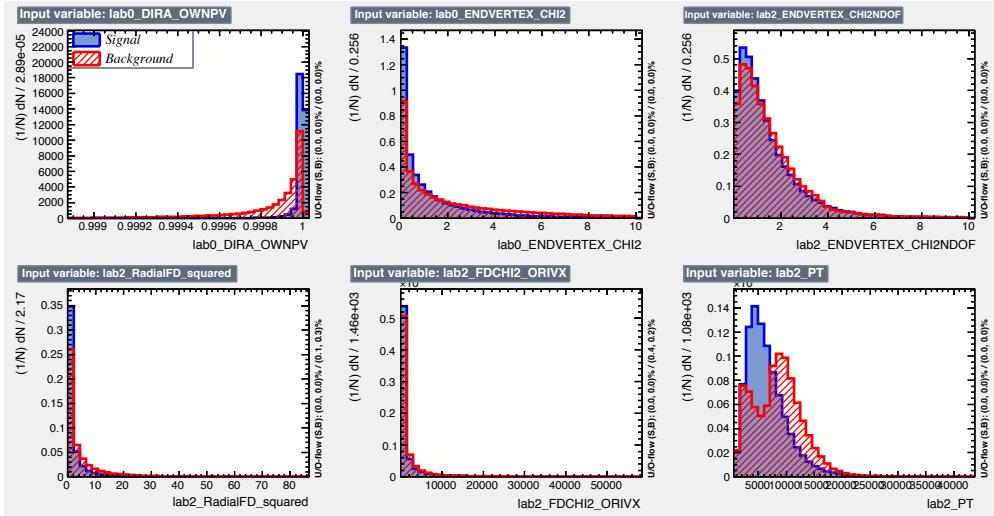


Figure C.1 – Input features used in the BDT training. From top left to bottom right: cosine of the direction angle of the  $B^0$ ,  $\chi^2$  of the  $B^0$  vertex,  $\chi^2/\text{ndof}$  of the  $D^\mp$  vertex,  $D^\mp$  radial flight distance,  $D^\mp$  flight distance  $\chi^2$  with respect to the  $B^0$  vertex and transverse momentum of the  $D^\mp$ .

### 1646 C.2 Multiple candidates

1647 Table C.1 gives a summary of the multiple candidates left after stripping and trigger  
 1648 selection, while Table C.2 reports the number of multiple candidates after stripping,  
 1649 trigger and offline selection.

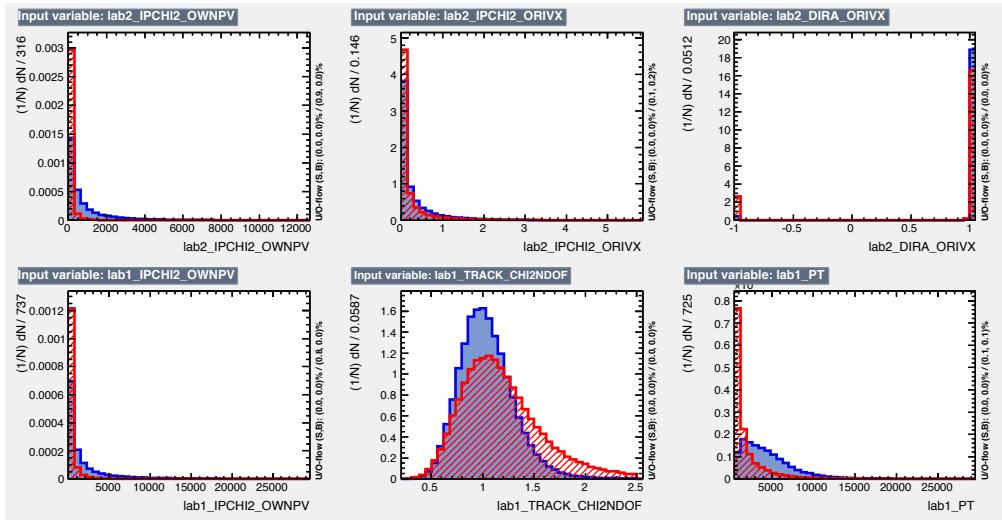


Figure C.2 – Input features used in the BDT training. From top left to bottom right:  $D^\mp$   $\text{IP}\chi^2$  with respect to the associated PV and the  $B^0$  vertex, cosine of the direction angle of the  $D^\mp$ , the  $\text{IP}\chi^2$  with respect to the associated PV of the bachelor pion, track  $\chi^2/\text{ndof}$  of the bachelor pion and the transverse momentum of the bachelor pion.

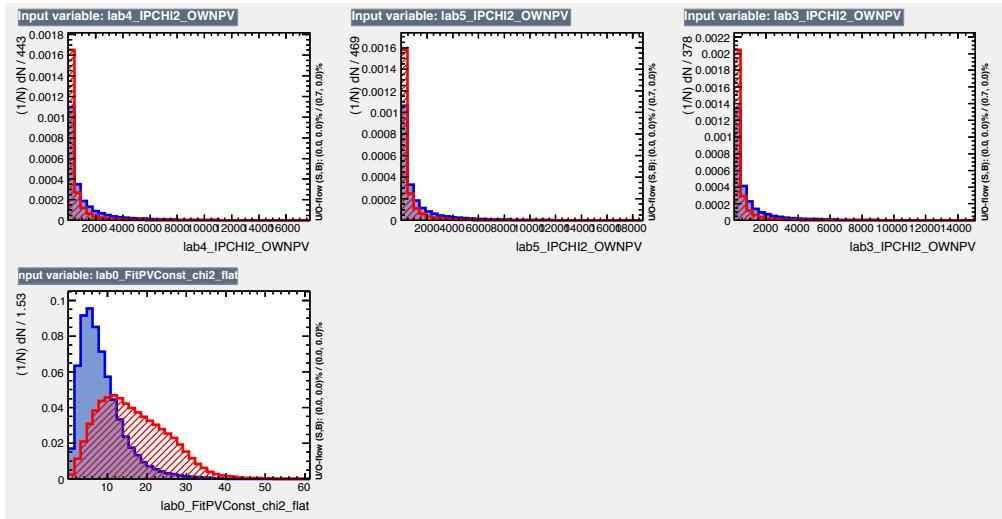


Figure C.3 – Input features used in the BDT training. From top left to bottom right:  $\text{IP}\chi^2$  of the associated primary vertex of the three  $D^\mp$  daughters and the  $\chi^2$  of the decay tree fit with PV constraint.

Table C.1 – Statistical information on multiple  $B^0 \rightarrow D^\mp\pi^\pm$  candidates left after stripping and trigger selection.

	2011	2012		
fraction of candidates that are not unique in a given event	18.3 %	19.5 %		
fraction of candidates to be discarded to maintain one candidate per event	10.1 %	11.0 %		
fraction of events with multiple candidates	9.0 %	9.6 %		
	#cands	#events	#cands	#events
1	5940804	1	16407228	
2	483991	2	1426286	
3	73902	3	226205	
4	20093	4	62640	
5	6132	5	19213	
6	2505	6	8044	
7	1087	7	3326	
8	528	8	1686	
9	251	9	839	
10	146	10	461	
11	78	11	279	
12	40	12	178	
13	28	13	109	
14	32	14	85	
15	10	15	53	
16	12	16	24	
17	7	17	16	
18	4	18	20	
19	5	19	9	
20	1	20	11	
21	3	21	5	
22	2	22	3	
23	0	23	2	
24	2	24	2	
25	1	25	1	
26	0	26	4	
30	0	30	1	
33	0	33	1	
40	1	40	0	
41	0	41	1	

Table C.2 – Statistical information on multiple  $B^0 \rightarrow D^\mp\pi^\pm$  candidates left after stripping, trigger and offline selection.

	2011	2012	
fraction of candidates pairs that are not unique in an event	0.8 %	0.8 %	
fraction of candidates to be discarded to maintain one candidate per event	0.4 %	0.4 %	
fraction of events with multiple candidates	0.4 %	0.4 %	
#cands	#events	#cands	#events
1	483074	1	1200956
2	1886	2	4962
3	38	3	98
4	4	4	9
5	1	5	3

## 1650 D Particle identification plots

1651 The  $p$  and  $\eta$  distributions, as well as the PID $K$  efficiency and misidentification rates  
 1652 before the resampling in bins of  $p$  and  $\eta$  are shown in Figs. D.1 and D.2 for the bachelor  
 particle of the signal, and in Fig. D.3 for the  $D^\mp$  daughters.

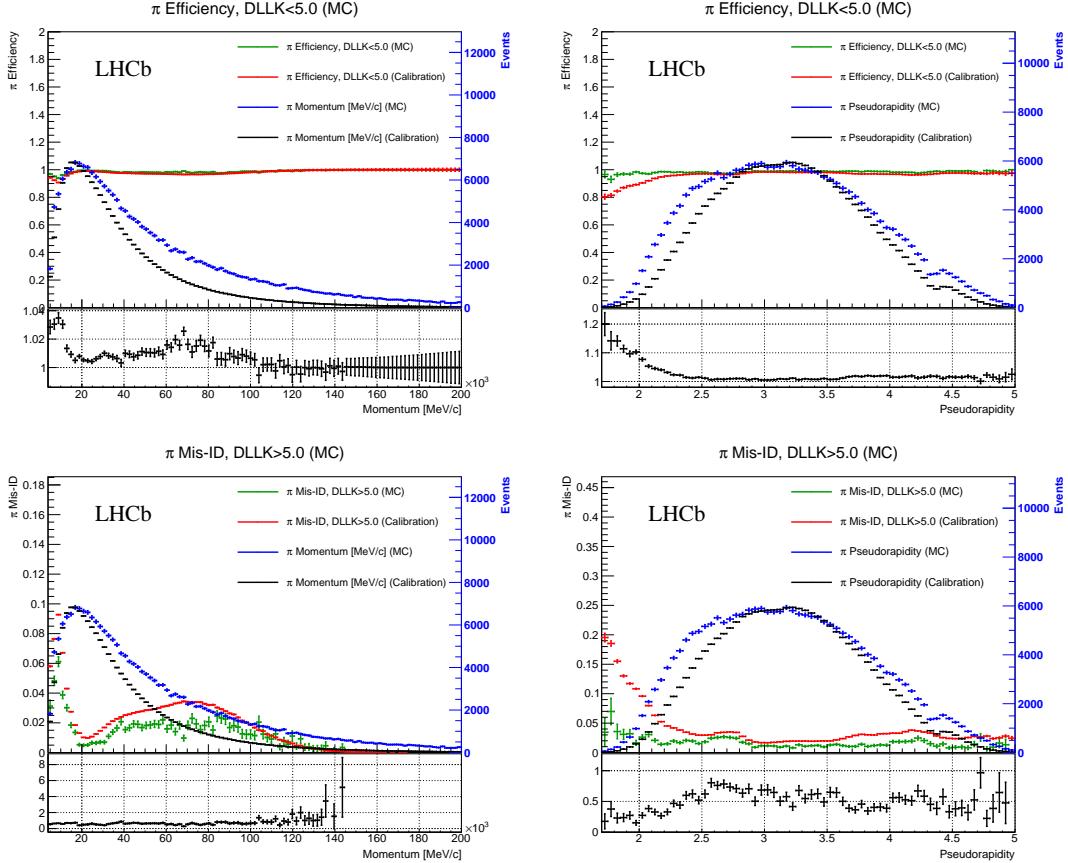


Figure D.1 – Efficiencies of the requirements  $PIDK < 5$  (top) and  $PIDK > 5$  (bottom) for bachelor pions as a function of momentum  $p$  (left) and pseudorapidity  $\eta$  (right), both for  $B^0 \rightarrow D^\mp \pi^\pm$  signal MC (green) and calibration mode (red). The superimposed histograms show the  $p$  and  $\eta$  distributions of the MC signal (blue) and calibration (black) samples. The ratio of the efficiency or misidentification rate between the MC signal and data calibration samples is shown in the lower pad (black).

1653

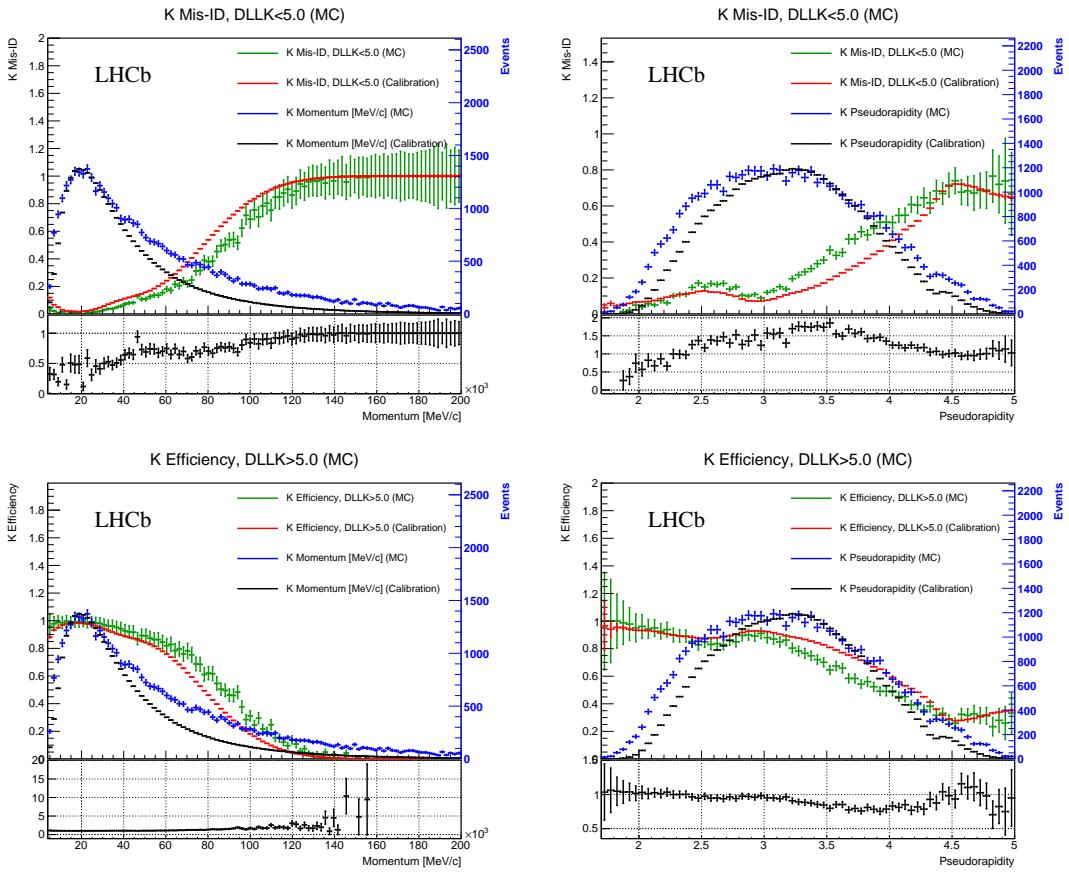


Figure D.2 – Efficiencies of the requirements  $\text{PID}K < 5$  (top) and  $\text{PID}K > 5$  (bottom) for bachelor kaons as a function of momentum  $p$  (left) and pseudorapidity  $\eta$  (right), both for  $B^0 \rightarrow D^\mp K^\pm$  MC (green) and calibration mode (red). The superimposed histograms show the  $p$  and  $\eta$  distributions of the  $B^0 \rightarrow D^\mp K^\pm$  MC (blue) and calibration (black) samples. The ratio of the efficiency or misidentification rate between the MC signal and data calibration samples is shown in the lower pad (black).

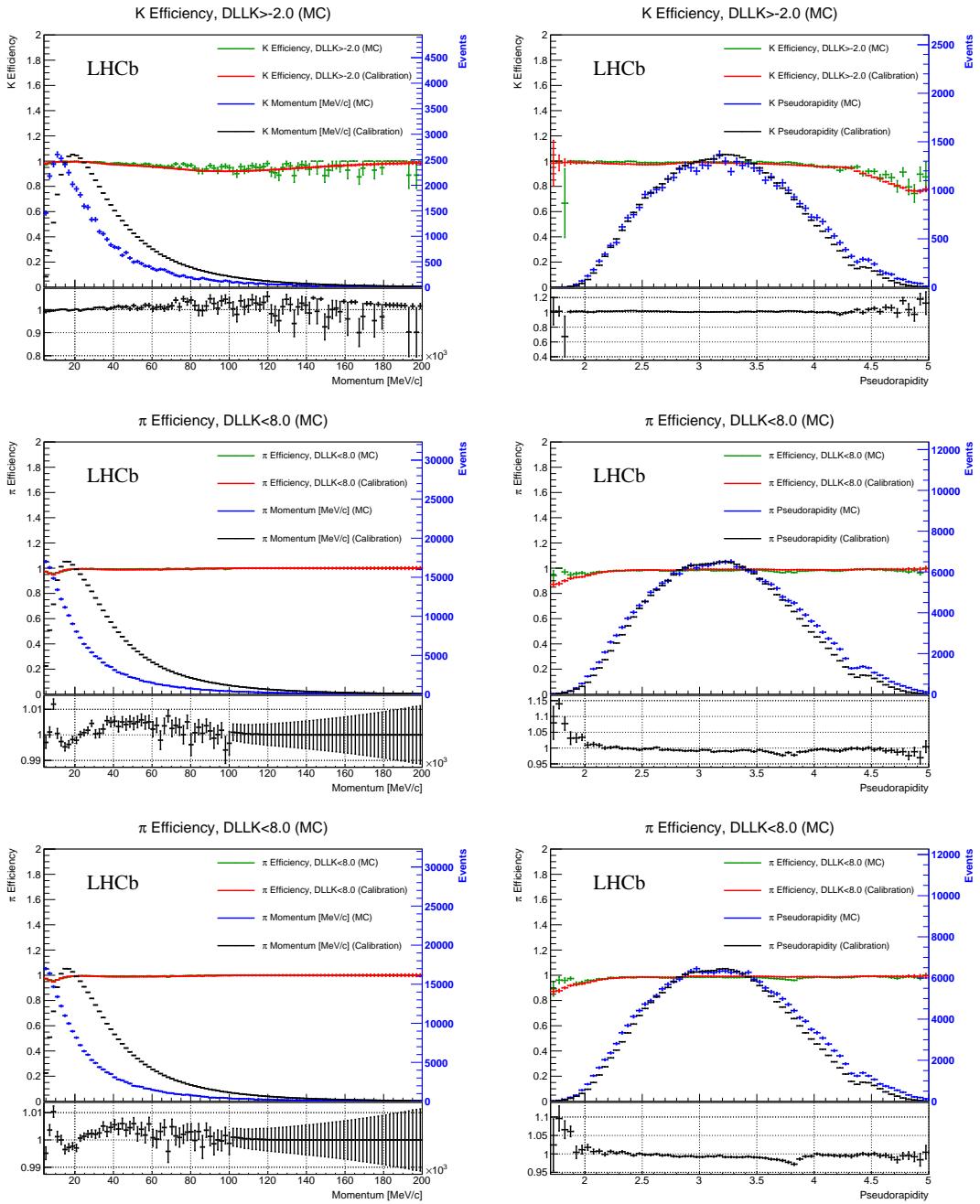


Figure D.3 – PID $K$  efficiencies for the kaon (top) and the two pions (middle, bottom) produced in the  $D^\mp$  decay as a function of momentum  $p$  (left) and pseudorapidity  $\eta$  (right), both for  $B^0 \rightarrow D^\mp \pi^\pm$  signal MC (green) and calibration mode (red). The superimposed histograms show the  $p$  and  $\eta$  distributions of the MC signal (blue) and calibration (black) samples. The ratio of the efficiency between the MC signal and data calibration samples is shown in the lower pad (black).

1654 **E Correlation between  $B^0$  mass and decay time**

1655 The small correlation between the  $B^0$  invariant mass and decay time is shown by comparing  
1656 the distribution of the decay time in bins of the invariant mass after applying the full  
1657 selection. This is done separately for signal and background. For the signal distribution  
1658 simulated data is used and the decay time is shown in six bins of the invariant mass.  
1659 (Fig. E.1). In order to account for the combinatorial background, the upper mass sideband  
1660 is chosen as a proxy. Figure E.2 shows the decay time in four bins of the invariant mass.  
1661 The physics background contribution in the signal region is considered to be small enough,  
1662 so that even a large correlation does not matter. Given the small differences for all  
1663 distributions, the correlations between decay time and invariant mass is assumed to be  
1664 small enough to justify the use of the invariant mass in the *sPlot* [74] technique for  
1665 disentangling signal from background.

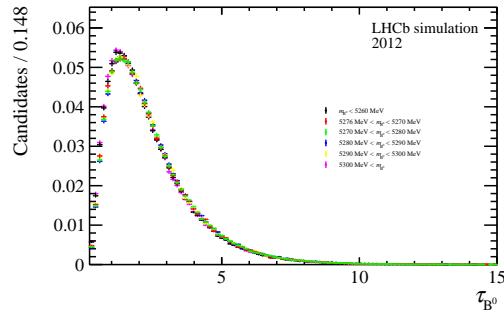


Figure E.1 – Normalised MC signal decay time distributions in six bins of the reconstructed invariant mass.

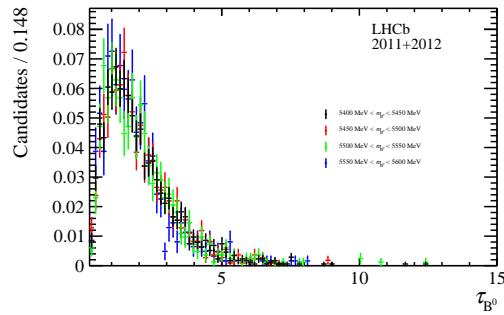


Figure E.2 – Upper mass sideband decay time distribution in four bins of the invariant mass. The shapes are shown normalised.

## 1666 F PDF definitions

1667 Throughout this section, each parameter  $p$  defined inside a PDF used for the mass fit is  
 1668 labelled as  $p_c^s$ , where  $s = \pi, K$  indicates the sample and  $c = B^0 \rightarrow D^\mp \pi^\pm, B^0 \rightarrow D^\mp \rho^\pm \dots$   
 1669 indicates the component. For sake of clarity, the  $s$  and  $c$  labels are dropped in the equations  
 1670 that follow. The mass *observable* is always indicated as  $m$ . The  $\propto$  symbol indicates that  
 1671 all PDFs are defined up to a normalisation constant, which depends on the interval chosen  
 1672 for  $m$ .

- Exponential function

$$E(m, c) \propto e^{-cm}.$$

- Gaussian function

$$G(m, \mu, \sigma) \propto e^{-\frac{(m-\mu)^2}{2\sigma^2}}.$$

- Double Gaussian function

$$DG(m, \mu, \sigma_1, \sigma_2, f) \propto \frac{f}{\sigma_1} e^{-\frac{(m-\mu)^2}{2\sigma_1^2}} + \frac{(1-f)}{\sigma_2} e^{-\frac{(m-\mu)^2}{2\sigma_2^2}}.$$

- Single-sided Crystal ball function

Having defined

$$A = \left( \frac{n}{|\alpha|} \right)^n e^{-\frac{|\alpha|^2}{2}}, \quad B = \frac{n}{|\alpha|} - |\alpha|,$$

the single-sided Crystal Ball function [77] is expressed as follows:

$$CB(m, \mu, \sigma, \alpha, n) \propto \begin{cases} e^{-\frac{(m-\mu)^2}{2\sigma^2}}, & \text{if } \frac{\alpha}{|\alpha|} \frac{m-\mu}{\sigma} \geq -|\alpha|, \\ A(B - \frac{m-\mu}{\sigma})^{-n}, & \text{if } \frac{\alpha}{|\alpha|} \frac{m-\mu}{\sigma} < -|\alpha|. \end{cases}$$

- Double-sided Hypatia function

Having defined

$$h(m, \mu, \sigma, \lambda, \zeta, \beta) \propto \left( (m - \mu)^2 + A_\lambda(\zeta) \sigma^2 \right)^{\frac{1}{2}\lambda - \frac{1}{4}} e^{\beta(m-\mu)} K_{\lambda - \frac{1}{2}} \left( \zeta \sqrt{1 + \left( \frac{m - \mu}{A_\lambda(\zeta) \sigma} \right)^2} \right),$$

and its first derivative with respect to  $m$ ,  $h'$ , then the double-sided Hypatia function  $H$  [75] is expressed as

$$H(m, \mu, \sigma, \lambda, \zeta, \beta, a_1, n_1, a_2, n_2) \propto$$

$$\begin{cases} h(m, \mu, \sigma, \lambda, \zeta, \beta) & \text{if } \frac{m-\mu}{\sigma} > -a_1 \text{ and } \frac{m-\mu}{\sigma} < a_2, \\ \frac{h(\mu-a_1\sigma, \mu, \sigma, \lambda, \zeta, \beta)}{\left(1-m/\left(n^{\frac{h(\mu-a_1\sigma, \mu, \sigma, \lambda, \zeta, \beta)}{h'(\mu-a_1\sigma, \mu, \sigma, \lambda, \zeta, \beta)}-a_1\sigma}\right)\right)^{n_1}} & \text{if } \frac{m-\mu}{\sigma} \leq -a_1, \\ \frac{h(\mu-a_2\sigma, \mu, \sigma, \lambda, \zeta, \beta)}{\left(1-m/\left(n^{\frac{h(\mu-a_2\sigma, \mu, \sigma, \lambda, \zeta, \beta)}{h'(\mu-a_2\sigma, \mu, \sigma, \lambda, \zeta, \beta)}-a_2\sigma}\right)\right)^{n_2}} & \text{if } \frac{m-\mu}{\sigma} \geq a_2. \end{cases}$$

The  $K_\lambda$  functions are special Bessel functions of the third kind, whereas  $A_\lambda$  is defined as

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

<sup>1673</sup> • **Single-sided Hypatia function**

<sup>1674</sup> A single-sided Hypatia function is obtained from a double-sided Hypatia function in  
<sup>1675</sup> the limit  $a_2 \rightarrow +\infty$ ,  $n_2 = 0$  (and by labelling  $a_1$  and  $n_1$  as  $a$  and  $n$ , respectively).

• **Johnson SU function**

Having defined the parameters

$$w = e^{\tau^2},$$

$$\omega = -\nu\tau,$$

$$c = \frac{1}{\sqrt{\frac{1}{2}(w-1)(w \cosh 2\omega + 1)}},$$

$$z = \frac{m - (\mu + c + \sigma\sqrt{w} \sinh \omega)}{c\sigma},$$

$$r = -\nu + \frac{\sinh^{-1} z}{\tau},$$

the Johnson SU function [76] is expressed as

$$J(m, \mu, \sigma, \nu, \tau) \propto \frac{1}{2\pi c(\nu, \tau)\sigma} e^{-\frac{1}{2}r(m, \mu, \sigma, \nu, \tau)^2} \frac{1}{\tau\sqrt{z(m, \mu, \sigma, \nu, \tau)^2 + 1}}.$$

## 1676 G Signal PDF for the decay-time fit

1677 The PDF describing the  $B^0$  decay time distribution can be written, in the most general  
 1678 case, as

$$P(t|d_{\text{OS}}, d_{\text{SS}}, \eta_{\text{OS}}, \eta_{\text{SS}}, q) = \frac{1}{4\tau} e^{-\frac{t}{\tau}} \left[ C_{\cosh}^{\text{eff}} \cosh \left( \frac{\Delta\Gamma t}{2} \right) + C_{\sinh}^{\text{eff}} \sinh \left( \frac{\Delta\Gamma t}{2} \right) \right. \\ \left. + C_{\cos}^{\text{eff}} \cos(\Delta m t) + C_{\sin}^{\text{eff}} \sin(\Delta m t) \right]. \quad (\text{G.1})$$

1679 The four “effective”  $CP$  coefficients inside Eq. G.1 depend on the final state ( $q = f = D^- \pi^+$   
 1680 or  $q = \bar{f} = D^+ \pi^-$ ), the tagging decision ( $d_i = -1, 0, +1$ , for  $i = \text{OS}, \text{SS}$ ), the mistag and  
 1681 the tagging efficiency of the OS and SS taggers, and the asymmetries (production, detection  
 1682 and tagging efficiency). In the ideal case (no asymmetries, zero mistag, 100% tagging  
 1683 efficiency), these effective coefficients become the  $CP$  coefficients already introduced in  
 1684 Sec. 1.3.2.

1685 The OS and SS taggers are combined on the fly during the time fit. All the steps to build  
 1686 the final PDF are described in details below.

The tagging efficiency  $\varepsilon_{\text{tag}}^i$  of the  $i$ th tagger ( $i = \text{OS}, \text{SS}$ ) is corrected for the tagging  
 efficiency difference  $\Delta\varepsilon^i$  as follows:

$$\varepsilon_B^i = \varepsilon_{\text{tag}}^i - \frac{1}{2}\Delta\varepsilon^i, \quad (\text{G.2})$$

$$\varepsilon_{\bar{B}}^i = \varepsilon_{\text{tag}}^i + \frac{1}{2}\Delta\varepsilon^i, \quad (\text{G.3})$$

1687 where  $\varepsilon_B^i$  and  $\varepsilon_{\bar{B}}^i$  are the two different tagging efficiencies for  $B^0$  and  $\bar{B}^0$ , respectively.  
 1688 Moreover, the OS and SS mistag are calibrated taking into account asymmetries between  
 1689  $B^0$  and  $\bar{B}^0$  according to Eq. 3.10. We will refer to them here as  $\omega_B^i$  and  $\omega_{\bar{B}}^i$  (for  $i = \text{OS}, \text{SS}$ ).  
 1690 It’s now convenient to define the quantity  $\Delta^\pm(d_{\text{OS}}, d_{\text{SS}})$  in the following way according  
 1691 to the OS and SS tagging decisions.

- If  $d_{\text{OS}} = d_{\text{SS}} = 0$  (un-tagged  $B^0$  candidate):

$$\Delta^\pm = (1 - \varepsilon_B^{\text{OS}} - \varepsilon_B^{\text{SS}} + \varepsilon_B^{\text{OS}}\varepsilon_B^{\text{SS}}) \pm (1 - \varepsilon_{\bar{B}}^{\text{OS}} - \varepsilon_{\bar{B}}^{\text{SS}} + \varepsilon_{\bar{B}}^{\text{OS}}\varepsilon_{\bar{B}}^{\text{SS}}). \quad (\text{G.4})$$

- If  $d_i = 0$ ,  $d_j \neq 0$  and  $i \neq j$  ( $B^0$  candidate tagged by one tagger):

$$\Delta^\pm = \frac{1}{2}\varepsilon_B^j \left[ 1 - \varepsilon_B^i + d_j \left( 1 - \varepsilon_B^i - 2\omega_B^j + 2\omega_B^j\varepsilon_B^i \right) \right] \\ \pm \frac{1}{2}\varepsilon_{\bar{B}}^j \left[ 1 - \varepsilon_{\bar{B}}^i + d_j \left( 1 - \varepsilon_{\bar{B}}^i - 2\omega_{\bar{B}}^j + 2\omega_{\bar{B}}^j\varepsilon_{\bar{B}}^i \right) \right]. \quad (\text{G.5})$$

- If  $d_{\text{OS}} = d_{\text{SS}} = 1$  ( $B^0$  candidates tagged by both taggers):

$$\begin{aligned}\Delta^\pm &= \frac{1}{4} \varepsilon_B^{\text{SS}} \varepsilon_B^{\text{OS}} \left[ 1 + d_{\text{SS}} \left( 1 - 2\omega_B^{\text{SS}} \right) + d_{\text{OS}} \left( 1 - 2\omega_B^{\text{OS}} \right) \right. \\ &\quad \left. + d_{\text{OS}} d_{\text{SS}} \left( 1 - 2\omega_B^{\text{SS}} - 2\omega_B^{\text{OS}} + 4\omega_B^{\text{SS}} \omega_B^{\text{OS}} \right) \right] \\ &\pm \frac{1}{4} \varepsilon_B^{\text{SS}} \varepsilon_B^{\text{OS}} \left[ 1 + d_{\text{SS}} \left( 1 - 2\omega_B^{\text{SS}} \right) + d_{\text{OS}} \left( 1 - 2\omega_B^{\text{OS}} \right) \right. \\ &\quad \left. + d_{\text{OS}} d_{\text{SS}} \left( 1 - 2\omega_B^{\text{SS}} - 2\omega_B^{\text{OS}} + 4\omega_B^{\text{SS}} \omega_B^{\text{OS}} \right) \right].\end{aligned}\tag{G.6}$$

Finally, the effective  $CP$  coefficients can be written as

$$C_{\sin, \cos}^{\text{eff}} = \begin{cases} (1 + A_D) C_{\sin, \cos}^{\text{phys}} (\Delta^- - A_P \Delta^+) & \text{for } q = f = D^- \pi^+, \\ -(1 - A_D) C_{\sin, \cos}^{\text{phys}} (\Delta^- - A_P \Delta^+) & \text{for } q = \bar{f} = D^+ \pi^-, \end{cases}\tag{G.7}$$

$$C_{\sinh, \cosh}^{\text{eff}} = \begin{cases} (1 + A_D) C_{\sinh}^{\text{phys}} (\Delta^+ - A_P \Delta^-) & \text{for } q = f = D^- \pi^+, \\ (1 - A_D) C_{\sinh}^{\text{phys}} (\Delta^+ - A_P \Delta^-) & \text{for } q = \bar{f} = D^+ \pi^-, \end{cases}\tag{G.8}$$

where  $A_D$  and  $A_P$  are the production and detection asymmetries defined in Eqs 5.9 and 5.10, respectively.

<sub>1694</sub> H Correlation matrix of the  $B^0 \rightarrow D^\mp\pi^\pm$  decay-time fit

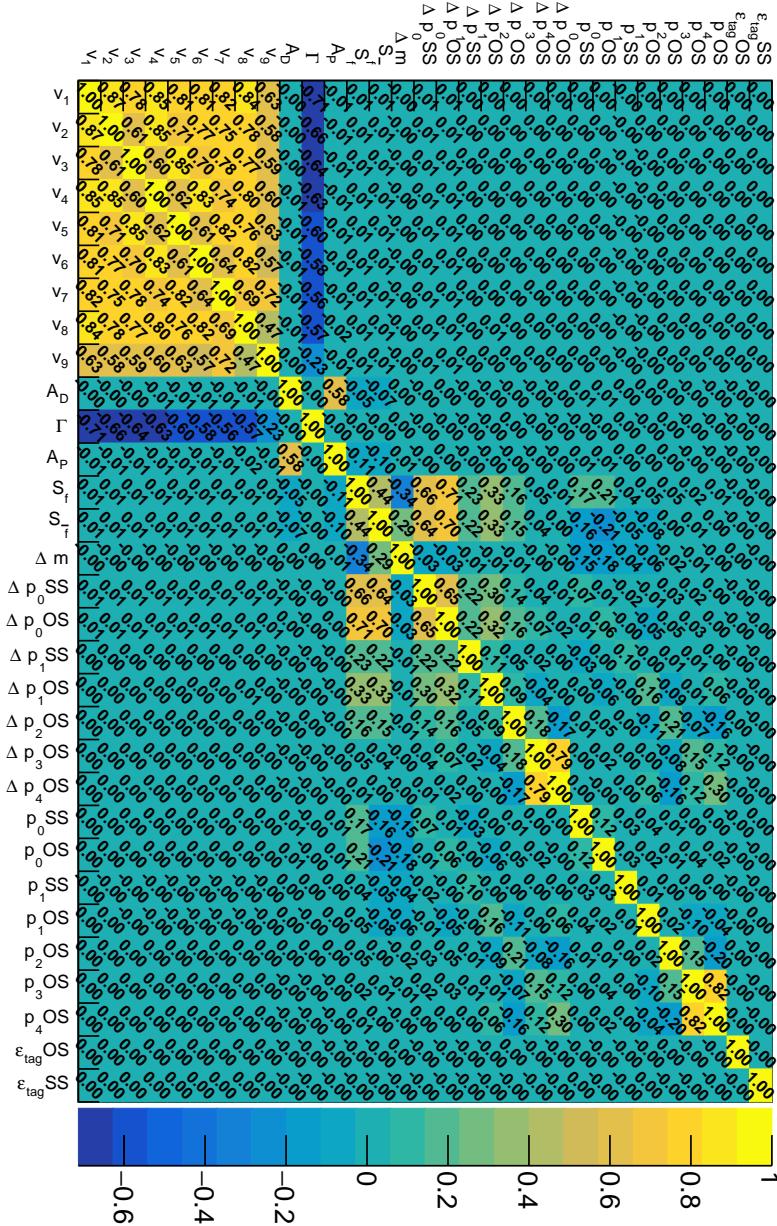


Figure H.1 – Correlation matrix of the  $B^0 \rightarrow D^\mp\pi^\pm$  decay-time fit.

<sup>1695</sup> **I Decay-time fits of  $B^0 \rightarrow D^\mp \pi^\pm$  data subsamples**

Table I.1 – Results of the decay-time fits to the 2011, 2012, magnet-up and down subsamples.

Parameter	2011	2012	Magnet Up	Magnet Down
$S_f$	$0.054 \pm 0.039$	$0.059 \pm 0.024$	$0.032 \pm 0.029$	$0.079 \pm 0.029$
$S_{\bar{f}}$	$0.031 \pm 0.039$	$0.041 \pm 0.024$	$-0.010 \pm 0.029$	$0.080 \pm 0.029$
$A_P$	$-0.0022 \pm 0.0054$	$-0.0079 \pm 0.0033$	$-0.0056 \pm 0.0041$	$-0.0072 \pm 0.0039$
$A_D$	$0.0137 \pm 0.0036$	$0.0065 \pm 0.0022$	$0.0075 \pm 0.0027$	$0.0096 \pm 0.0026$
$\Gamma$	$0.6588 \pm 0.0017$	$0.6587 \pm 0.0017$	$0.6587 \pm 0.0017$	$0.6588 \pm 0.0017$
$\Delta m$	$0.5047 \pm 0.0023$	$0.5058 \pm 0.0023$	$0.5054 \pm 0.0023$	$0.5050 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.126 \pm 0.038$	$-0.166 \pm 0.025$	$-0.143 \pm 0.030$	$-0.160 \pm 0.029$
$p_1^{\text{OS}}$	$-0.095 \pm 0.044$	$-0.011 \pm 0.028$	$-0.054 \pm 0.034$	$-0.014 \pm 0.033$
$p_2^{\text{OS}}$	$-0.025 \pm 0.018$	$-0.000 \pm 0.010$	$-0.013 \pm 0.013$	$-0.004 \pm 0.012$
$p_3^{\text{OS}}$	$-0.03 \pm 0.21$	$-0.43 \pm 0.14$	$-0.20 \pm 0.18$	$-0.40 \pm 0.15$
$p_4^{\text{OS}}$	$1.08 \pm 0.94$	$-1.41 \pm 0.70$	$-0.11 \pm 0.92$	$-0.63 \pm 0.64$
$\Delta p_0^{\text{OS}}$	$-0.117 \pm 0.092$	$-0.065 \pm 0.057$	$-0.103 \pm 0.070$	$-0.063 \pm 0.068$
$\Delta p_1^{\text{OS}}$	$0.064 \pm 0.067$	$0.172 \pm 0.042$	$0.187 \pm 0.051$	$0.101 \pm 0.050$
$\Delta p_2^{\text{OS}}$	$-0.036 \pm 0.025$	$-0.018 \pm 0.015$	$-0.020 \pm 0.019$	$-0.031 \pm 0.017$
$\Delta p_3^{\text{OS}}$	$-0.08 \pm 0.30$	$-0.29 \pm 0.20$	$-0.17 \pm 0.26$	$-0.34 \pm 0.21$
$\Delta p_4^{\text{OS}}$	$0.8 \pm 1.4$	$-1.5 \pm 1.2$	$-0.2 \pm 1.5$	$-0.72 \pm 0.89$
$p_0^{\text{SS}}$	$-0.046 \pm 0.039$	$-0.041 \pm 0.025$	$0.020 \pm 0.030$	$-0.097 \pm 0.030$
$p_1^{\text{SS}}$	$0.005 \pm 0.042$	$-0.019 \pm 0.026$	$-0.022 \pm 0.031$	$-0.004 \pm 0.032$
$\Delta p_0^{\text{SS}}$	$-0.142 \pm 0.083$	$-0.061 \pm 0.052$	$-0.102 \pm 0.063$	$-0.076 \pm 0.061$
$\Delta p_1^{\text{SS}}$	$0.068 \pm 0.062$	$0.030 \pm 0.039$	$0.048 \pm 0.046$	$0.036 \pm 0.046$
$\varepsilon_{\text{tag}}^{\text{OS}}$	$0.4270 \pm 0.0014$	$0.43458 \pm 0.00091$	$0.4321 \pm 0.0011$	$0.4326 \pm 0.0011$
$\varepsilon_{\text{tag}}^{\text{SS}}$	$0.92753 \pm 0.00076$	$0.93163 \pm 0.00046$	$0.92978 \pm 0.00058$	$0.93107 \pm 0.00054$
$v_1$	$0.350 \pm 0.011$	$0.3082 \pm 0.0066$	$0.3199 \pm 0.0078$	$0.3186 \pm 0.0075$
$v_2$	$0.535 \pm 0.018$	$0.479 \pm 0.011$	$0.506 \pm 0.014$	$0.483 \pm 0.013$
$v_3$	$0.886 \pm 0.029$	$0.760 \pm 0.018$	$0.786 \pm 0.021$	$0.800 \pm 0.020$
$v_4$	$1.055 \pm 0.034$	$0.973 \pm 0.021$	$1.013 \pm 0.026$	$0.978 \pm 0.024$
$v_5$	$1.186 \pm 0.038$	$1.060 \pm 0.023$	$1.096 \pm 0.028$	$1.089 \pm 0.026$
$v_6$	$1.176 \pm 0.037$	$1.095 \pm 0.023$	$1.140 \pm 0.028$	$1.096 \pm 0.026$
$v_7$	$1.231 \pm 0.039$	$1.108 \pm 0.023$	$1.156 \pm 0.028$	$1.126 \pm 0.027$
$v_8$	$1.267 \pm 0.035$	$1.143 \pm 0.021$	$1.189 \pm 0.026$	$1.163 \pm 0.024$
$v_9$	$1.203 \pm 0.050$	$1.135 \pm 0.029$	$1.174 \pm 0.037$	$1.136 \pm 0.035$

Table I.2 – Results of the decay-time fits when only OS tagged, SS tagged or OS and SS candidates are considered exclusively.

Parameter	OS only	SS only	both OS and SS
$S_f$	$0.046 \pm 0.056$	$0.037 \pm 0.040$	$0.064 \pm 0.024$
$S_{\bar{f}}$	$-0.015 \pm 0.056$	$0.019 \pm 0.040$	$0.048 \pm 0.024$
$A_P$	$0.009 \pm 0.010$	$-0.0110 \pm 0.0036$	$-0.0022 \pm 0.0044$
$A_D$	$0.0100 \pm 0.0066$	$0.0086 \pm 0.0024$	$0.0083 \pm 0.0030$
$\Gamma$	$0.6588 \pm 0.0017$	$0.6587 \pm 0.0017$	$0.6587 \pm 0.0017$
$\Delta m$	$0.5051 \pm 0.0023$	$0.5055 \pm 0.0023$	$0.5049 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.164 \pm 0.051$		$-0.148 \pm 0.022$
$p_1^{\text{OS}}$	$-0.017 \pm 0.058$		$-0.036 \pm 0.024$
$p_2^{\text{OS}}$	$-0.007 \pm 0.023$		$-0.0077 \pm 0.0091$
$p_3^{\text{OS}}$	$-0.16 \pm 0.26$		$-0.34 \pm 0.11$
$p_4^{\text{OS}}$	$0.2 \pm 1.4$		$-0.55 \pm 0.50$
$\Delta p_0^{\text{OS}}$	$-0.10 \pm 0.13$		$-0.059 \pm 0.055$
$\Delta p_1^{\text{OS}}$	$0.120 \pm 0.089$		$0.154 \pm 0.038$
$\Delta p_2^{\text{OS}}$	$-0.071 \pm 0.033$		$-0.015 \pm 0.013$
$\Delta p_3^{\text{OS}}$	$-0.34 \pm 0.37$		$-0.24 \pm 0.16$
$\Delta p_4^{\text{OS}}$	$0.9 \pm 2.1$		$-0.59 \pm 0.71$
$p_0^{\text{SS}}$		$-0.023 \pm 0.028$	$-0.072 \pm 0.031$
$p_1^{\text{SS}}$		$-0.027 \pm 0.027$	$0.008 \pm 0.034$
$\Delta p_0^{\text{SS}}$		$-0.102 \pm 0.075$	$-0.112 \pm 0.058$
$\Delta p_1^{\text{SS}}$		$0.014 \pm 0.041$	$0.075 \pm 0.050$
$v_1$	$0.276 \pm 0.014$	$0.3155 \pm 0.0070$	$0.3344 \pm 0.0087$
$v_2$	$0.403 \pm 0.024$	$0.500 \pm 0.012$	$0.504 \pm 0.015$
$v_3$	$0.668 \pm 0.040$	$0.793 \pm 0.019$	$0.821 \pm 0.023$
$v_4$	$0.903 \pm 0.050$	$0.992 \pm 0.023$	$1.017 \pm 0.028$
$v_5$	$1.070 \pm 0.059$	$1.083 \pm 0.024$	$1.112 \pm 0.030$
$v_6$	$1.045 \pm 0.058$	$1.100 \pm 0.024$	$1.159 \pm 0.031$
$v_7$	$1.154 \pm 0.062$	$1.137 \pm 0.025$	$1.142 \pm 0.030$
$v_8$	$1.138 \pm 0.054$	$1.165 \pm 0.022$	$1.199 \pm 0.028$
$v_9$	$1.117 \pm 0.086$	$1.148 \pm 0.032$	$1.171 \pm 0.040$

Table I.3 – Results of the decay-time fits in bins of the reconstructed  $B^0$  transverse momentum ( $\text{GeV}/c$ ).

Parameter	$p_T \in [0, 6.5)$	$p_T \in [6.5, 9.3)$	$p_T \in [9.3, 12.8)$	$p_T \in [12.8, +\infty)$
$S_f$	$0.032 \pm 0.039$	$0.074 \pm 0.040$	$0.088 \pm 0.038$	$0.046 \pm 0.036$
$S_{\bar{f}}$	$0.008 \pm 0.039$	$0.041 \pm 0.041$	$0.066 \pm 0.039$	$0.025 \pm 0.036$
$A_P$	$-0.0052 \pm 0.0053$	$-0.0041 \pm 0.0053$	$-0.0158 \pm 0.0054$	$-0.0013 \pm 0.0056$
$A_D$	$0.0092 \pm 0.0033$	$0.0140 \pm 0.0035$	$0.0008 \pm 0.0037$	$0.0098 \pm 0.0040$
$\Gamma$	$0.6587 \pm 0.0017$	$0.6587 \pm 0.0017$	$0.6588 \pm 0.0017$	$0.6588 \pm 0.0017$
$\Delta m$	$0.5048 \pm 0.0023$	$0.5054 \pm 0.0023$	$0.5056 \pm 0.0023$	$0.5045 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.088 \pm 0.041$	$-0.123 \pm 0.040$	$-0.149 \pm 0.039$	$-0.237 \pm 0.037$
$p_1^{\text{OS}}$	$-0.008 \pm 0.045$	$0.044 \pm 0.048$	$-0.052 \pm 0.046$	$-0.117 \pm 0.042$
$p_2^{\text{OS}}$	$0.006 \pm 0.016$	$-0.027 \pm 0.019$	$-0.005 \pm 0.021$	$-0.004 \pm 0.016$
$p_3^{\text{OS}}$	$-0.60 \pm 0.20$	$-0.44 \pm 0.23$	$0.06 \pm 0.31$	$-0.19 \pm 0.20$
$p_4^{\text{OS}}$	$-1.61 \pm 0.83$	$-0.9 \pm 1.3$	$-0.8 \pm 1.5$	$0.07 \pm 0.81$
$\Delta p_0^{\text{OS}}$	$-0.094 \pm 0.099$	$-0.104 \pm 0.099$	$-0.064 \pm 0.090$	$-0.062 \pm 0.082$
$\Delta p_1^{\text{OS}}$	$0.074 \pm 0.066$	$0.183 \pm 0.073$	$0.218 \pm 0.071$	$0.078 \pm 0.064$
$\Delta p_2^{\text{OS}}$	$-0.040 \pm 0.022$	$-0.012 \pm 0.027$	$-0.042 \pm 0.032$	$-0.001 \pm 0.023$
$\Delta p_3^{\text{OS}}$	$-0.20 \pm 0.27$	$-0.58 \pm 0.37$	$-0.08 \pm 0.58$	$-0.00 \pm 0.29$
$\Delta p_4^{\text{OS}}$	$-0.0 \pm 1.1$	$-3.2 \pm 1.9$	$-2.8 \pm 2.6$	$0.4 \pm 1.2$
$p_0^{\text{SS}}$	$-0.198 \pm 0.084$	$-0.223 \pm 0.044$	$-0.100 \pm 0.038$	$0.146 \pm 0.032$
$p_1^{\text{SS}}$	$-0.142 \pm 0.081$	$-0.045 \pm 0.051$	$-0.034 \pm 0.040$	$-0.014 \pm 0.033$
$\Delta p_0^{\text{SS}}$	$-0.24 \pm 0.13$	$-0.075 \pm 0.087$	$-0.047 \pm 0.080$	$-0.091 \pm 0.074$
$\Delta p_1^{\text{SS}}$	$-0.03 \pm 0.11$	$-0.041 \pm 0.071$	$0.047 \pm 0.059$	$0.073 \pm 0.053$
$\varepsilon_{\text{tag}}^{\text{OS}}$	$0.4461 \pm 0.0015$	$0.4210 \pm 0.0014$	$0.4176 \pm 0.0014$	$0.4452 \pm 0.0015$
$\varepsilon_{\text{tag}}^{\text{SS}}$	$0.92426 \pm 0.00078$	$0.94508 \pm 0.00066$	$0.93922 \pm 0.00070$	$0.91274 \pm 0.00084$
$v_1$	$0.0492 \pm 0.0018$	$0.2480 \pm 0.0073$	$0.463 \pm 0.014$	$0.779 \pm 0.025$
$v_2$	$0.1206 \pm 0.0048$	$0.433 \pm 0.014$	$0.702 \pm 0.023$	$1.030 \pm 0.036$
$v_3$	$0.493 \pm 0.015$	$0.757 \pm 0.024$	$0.937 \pm 0.032$	$1.307 \pm 0.047$
$v_4$	$0.749 \pm 0.021$	$0.955 \pm 0.029$	$1.127 \pm 0.037$	$1.344 \pm 0.048$
$v_5$	$0.902 \pm 0.025$	$1.079 \pm 0.033$	$1.162 \pm 0.038$	$1.432 \pm 0.051$
$v_6$	$0.981 \pm 0.027$	$1.033 \pm 0.032$	$1.211 \pm 0.039$	$1.389 \pm 0.049$
$v_7$	$0.993 \pm 0.027$	$1.114 \pm 0.033$	$1.203 \pm 0.039$	$1.406 \pm 0.050$
$v_8$	$1.062 \pm 0.025$	$1.142 \pm 0.030$	$1.215 \pm 0.035$	$1.403 \pm 0.045$
$v_9$	$1.033 \pm 0.038$	$1.112 \pm 0.045$	$1.245 \pm 0.053$	$1.365 \pm 0.063$

Table I.4 – Results of the decay-time fits in bins of the number of reconstructed primary vertices (PV) in the event.

Parameter	1 PV	2 PV	> 2 PV
$S_f$	$0.007 \pm 0.034$	$0.061 \pm 0.031$	$0.118 \pm 0.035$
$S_{\bar{f}}$	$0.027 \pm 0.033$	$0.016 \pm 0.031$	$0.080 \pm 0.036$
$A_P$	$-0.0055 \pm 0.0047$	$-0.0112 \pm 0.0043$	$-0.0009 \pm 0.0049$
$A_D$	$0.0064 \pm 0.0032$	$0.0053 \pm 0.0029$	$0.0152 \pm 0.0033$
$\Gamma$	$0.6587 \pm 0.0017$	$0.6587 \pm 0.0017$	$0.6588 \pm 0.0017$
$\Delta m$	$0.5052 \pm 0.0023$	$0.5051 \pm 0.0023$	$0.5052 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.193 \pm 0.034$	$-0.177 \pm 0.032$	$-0.067 \pm 0.037$
$p_1^{\text{OS}}$	$-0.011 \pm 0.043$	$-0.033 \pm 0.035$	$-0.042 \pm 0.042$
$p_2^{\text{OS}}$	$-0.034 \pm 0.018$	$0.011 \pm 0.013$	$-0.002 \pm 0.016$
$p_3^{\text{OS}}$	$-0.37 \pm 0.20$	$-0.35 \pm 0.17$	$-0.11 \pm 0.20$
$p_4^{\text{OS}}$	$-0.05 \pm 0.93$	$-1.25 \pm 0.97$	$-0.4 \pm 1.1$
$\Delta p_0^{\text{OS}}$	$-0.189 \pm 0.079$	$-0.122 \pm 0.074$	$0.120 \pm 0.087$
$\Delta p_1^{\text{OS}}$	$0.119 \pm 0.065$	$0.088 \pm 0.053$	$0.273 \pm 0.065$
$\Delta p_2^{\text{OS}}$	$-0.055 \pm 0.026$	$-0.020 \pm 0.019$	$0.000 \pm 0.023$
$\Delta p_3^{\text{OS}}$	$0.16 \pm 0.28$	$-0.53 \pm 0.26$	$-0.17 \pm 0.30$
$\Delta p_4^{\text{OS}}$	$1.9 \pm 1.3$	$-2.6 \pm 1.4$	$-1.2 \pm 1.6$
$p_0^{\text{SS}}$	$-0.031 \pm 0.035$	$-0.057 \pm 0.032$	$-0.031 \pm 0.037$
$p_1^{\text{SS}}$	$0.020 \pm 0.038$	$-0.034 \pm 0.034$	$-0.015 \pm 0.039$
$\Delta p_0^{\text{SS}}$	$-0.102 \pm 0.072$	$-0.175 \pm 0.066$	$0.065 \pm 0.078$
$\Delta p_1^{\text{SS}}$	$-0.009 \pm 0.056$	$0.046 \pm 0.049$	$0.094 \pm 0.057$
$\varepsilon_{\text{tag}}^{\text{OS}}$	$0.4394 \pm 0.0013$	$0.4330 \pm 0.0012$	$0.4240 \pm 0.0013$
$\varepsilon_{\text{tag}}^{\text{SS}}$	$0.93394 \pm 0.00065$	$0.93134 \pm 0.00059$	$0.92557 \pm 0.00071$
$v_1$	$0.3144 \pm 0.0084$	$0.3265 \pm 0.0082$	$0.3150 \pm 0.0090$
$v_2$	$0.484 \pm 0.014$	$0.491 \pm 0.014$	$0.510 \pm 0.016$
$v_3$	$0.754 \pm 0.022$	$0.802 \pm 0.022$	$0.829 \pm 0.026$
$v_4$	$0.935 \pm 0.026$	$1.012 \pm 0.026$	$1.040 \pm 0.031$
$v_5$	$1.027 \pm 0.029$	$1.093 \pm 0.029$	$1.170 \pm 0.035$
$v_6$	$1.038 \pm 0.029$	$1.121 \pm 0.029$	$1.204 \pm 0.035$
$v_7$	$1.087 \pm 0.030$	$1.144 \pm 0.029$	$1.197 \pm 0.035$
$v_8$	$1.103 \pm 0.027$	$1.173 \pm 0.026$	$1.262 \pm 0.032$
$v_9$	$1.118 \pm 0.041$	$1.150 \pm 0.038$	$1.203 \pm 0.046$

Table I.5 – Results of the decay-time fits in bins of the logarithm of the total number of reconstructed tracks ( $\ln N_{\text{tr}}$ ) in the event.

Parameter	$\ln N_{\text{tr}} \in [0, 4.8)$	$\ln N_{\text{tr}} \in [4.8, 5.2)$	$\ln N_{\text{tr}} \in [5.2, +\infty)$
$S_f$	$0.036 \pm 0.029$	$0.038 \pm 0.034$	$0.116 \pm 0.039$
$S_{\bar{f}}$	$0.046 \pm 0.029$	$0.000 \pm 0.034$	$0.066 \pm 0.039$
$A_P$	$-0.0031 \pm 0.0046$	$-0.0071 \pm 0.0046$	$-0.0093 \pm 0.0047$
$A_D$	$0.0092 \pm 0.0031$	$0.0073 \pm 0.0031$	$0.0091 \pm 0.0031$
$\Gamma$	$0.6587 \pm 0.0017$	$0.6588 \pm 0.0017$	$0.6588 \pm 0.0017$
$\Delta m$	$0.5056 \pm 0.0023$	$0.5049 \pm 0.0023$	$0.5050 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.077 \pm 0.031$	$-0.174 \pm 0.035$	$-0.221 \pm 0.038$
$p_1^{\text{OS}}$	$-0.050 \pm 0.041$	$-0.038 \pm 0.038$	$-0.003 \pm 0.042$
$p_2^{\text{OS}}$	$-0.047 \pm 0.020$	$0.013 \pm 0.015$	$0.018 \pm 0.016$
$p_3^{\text{OS}}$	$-0.15 \pm 0.28$	$-0.27 \pm 0.19$	$-0.22 \pm 0.20$
$p_4^{\text{OS}}$	$-0.3 \pm 1.4$	$-1.33 \pm 0.94$	$-1.13 \pm 0.97$
$\Delta p_0^{\text{OS}}$	$-0.077 \pm 0.071$	$-0.190 \pm 0.081$	$0.056 \pm 0.091$
$\Delta p_1^{\text{OS}}$	$0.105 \pm 0.062$	$0.099 \pm 0.059$	$0.239 \pm 0.065$
$\Delta p_2^{\text{OS}}$	$-0.030 \pm 0.031$	$-0.042 \pm 0.021$	$-0.010 \pm 0.023$
$\Delta p_3^{\text{OS}}$	$0.26 \pm 0.59$	$-0.11 \pm 0.26$	$-0.71 \pm 0.29$
$\Delta p_4^{\text{OS}}$	$-1.6 \pm 2.3$	$1.8 \pm 1.4$	$-3.8 \pm 1.5$
$p_0^{\text{SS}}$	$0.015 \pm 0.031$	$-0.028 \pm 0.035$	$-0.126 \pm 0.039$
$p_1^{\text{SS}}$	$-0.038 \pm 0.032$	$-0.037 \pm 0.037$	$0.028 \pm 0.044$
$\Delta p_0^{\text{SS}}$	$-0.100 \pm 0.063$	$-0.057 \pm 0.073$	$-0.089 \pm 0.083$
$\Delta p_1^{\text{SS}}$	$0.003 \pm 0.047$	$0.074 \pm 0.054$	$0.069 \pm 0.063$
$\varepsilon_{\text{tag}}^{\text{OS}}$	$0.4082 \pm 0.0012$	$0.4370 \pm 0.0013$	$0.4523 \pm 0.0013$
$\varepsilon_{\text{tag}}^{\text{SS}}$	$0.90933 \pm 0.00073$	$0.93877 \pm 0.00061$	$0.94361 \pm 0.00059$
$v_1$	$0.2904 \pm 0.0075$	$0.3406 \pm 0.0091$	$0.3303 \pm 0.0089$
$v_2$	$0.450 \pm 0.013$	$0.521 \pm 0.015$	$0.518 \pm 0.016$
$v_3$	$0.728 \pm 0.021$	$0.814 \pm 0.024$	$0.847 \pm 0.025$
$v_4$	$0.932 \pm 0.025$	$1.010 \pm 0.028$	$1.051 \pm 0.030$
$v_5$	$1.008 \pm 0.027$	$1.116 \pm 0.031$	$1.167 \pm 0.033$
$v_6$	$1.033 \pm 0.028$	$1.144 \pm 0.032$	$1.186 \pm 0.033$
$v_7$	$1.066 \pm 0.028$	$1.140 \pm 0.032$	$1.226 \pm 0.034$
$v_8$	$1.083 \pm 0.025$	$1.219 \pm 0.029$	$1.235 \pm 0.030$
$v_9$	$1.135 \pm 0.039$	$1.123 \pm 0.042$	$1.211 \pm 0.044$

Table I.6 – Results of the decay-time fits in bins of the difference in pseudorapidity  $\Delta\eta$  between the  $D^\mp$  meson and the bachelor pion.

Parameter	$\Delta\eta \in (-\infty, -0.36)$	$\Delta\eta \in [-0.36, 0.10)$	$\Delta\eta \in [0.10, 0.53)$	$\Delta\eta \in [0.53, +\infty)$
$S_f$	$0.047 \pm 0.038$	$0.050 \pm 0.039$	$0.103 \pm 0.038$	$0.034 \pm 0.039$
$S_{\bar{f}}$	$0.028 \pm 0.038$	$0.060 \pm 0.039$	$0.006 \pm 0.038$	$0.054 \pm 0.039$
$A_P$	$-0.0092 \pm 0.0053$	$-0.0121 \pm 0.0054$	$0.0004 \pm 0.0053$	$-0.0051 \pm 0.0053$
$A_D$	$0.0072 \pm 0.0036$	$0.0043 \pm 0.0037$	$0.0118 \pm 0.0036$	$0.0106 \pm 0.0035$
$\Gamma$	$0.6587 \pm 0.0017$	$0.6587 \pm 0.0017$	$0.6588 \pm 0.0017$	$0.6588 \pm 0.0017$
$\Delta m$	$0.5052 \pm 0.0023$	$0.5047 \pm 0.0023$	$0.5046 \pm 0.0023$	$0.5059 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.127 \pm 0.039$	$-0.145 \pm 0.039$	$-0.157 \pm 0.039$	$-0.188 \pm 0.041$
$p_1^{\text{OS}}$	$-0.032 \pm 0.049$	$-0.033 \pm 0.044$	$-0.044 \pm 0.045$	$-0.010 \pm 0.045$
$p_2^{\text{OS}}$	$-0.037 \pm 0.023$	$0.011 \pm 0.017$	$0.003 \pm 0.018$	$-0.016 \pm 0.015$
$p_3^{\text{OS}}$	$-0.01 \pm 0.32$	$-0.12 \pm 0.22$	$0.01 \pm 0.23$	$-0.95 \pm 0.20$
$p_4^{\text{OS}}$	$-0.7 \pm 2.0$	$0.16 \pm 0.92$	$0.66 \pm 0.96$	$-2.93 \pm 0.91$
$\Delta p_0^{\text{OS}}$	$-0.105 \pm 0.092$	$-0.082 \pm 0.093$	$0.001 \pm 0.089$	$-0.140 \pm 0.093$
$\Delta p_1^{\text{OS}}$	$0.185 \pm 0.074$	$0.276 \pm 0.069$	$0.029 \pm 0.067$	$0.082 \pm 0.067$
$\Delta p_2^{\text{OS}}$	$-0.021 \pm 0.037$	$-0.054 \pm 0.026$	$0.007 \pm 0.025$	$-0.032 \pm 0.022$
$\Delta p_3^{\text{OS}}$	$0.16 \pm 0.79$	$-0.31 \pm 0.30$	$-0.10 \pm 0.31$	$-0.56 \pm 0.28$
$\Delta p_4^{\text{OS}}$	$-2.5 \pm 2.9$	$-0.1 \pm 1.3$	$-0.4 \pm 1.3$	$-1.2 \pm 1.3$
$p_0^{\text{SS}}$	$-0.061 \pm 0.040$	$-0.009 \pm 0.038$	$-0.029 \pm 0.040$	$-0.071 \pm 0.041$
$p_1^{\text{SS}}$	$0.018 \pm 0.044$	$-0.060 \pm 0.039$	$0.055 \pm 0.043$	$-0.054 \pm 0.044$
$\Delta p_0^{\text{SS}}$	$-0.143 \pm 0.083$	$-0.121 \pm 0.082$	$0.024 \pm 0.081$	$-0.098 \pm 0.084$
$\Delta p_1^{\text{SS}}$	$0.157 \pm 0.066$	$0.033 \pm 0.058$	$-0.038 \pm 0.063$	$0.017 \pm 0.063$
$\varepsilon_{\text{tag}}^{\text{OS}}$	$0.4370 \pm 0.0015$	$0.4344 \pm 0.0015$	$0.4346 \pm 0.0015$	$0.4236 \pm 0.0014$
$\varepsilon_{\text{tag}}^{\text{SS}}$	$0.93563 \pm 0.00072$	$0.92611 \pm 0.00077$	$0.92349 \pm 0.00078$	$0.93669 \pm 0.00071$
$v_1$	$0.2796 \pm 0.0082$	$0.377 \pm 0.011$	$0.396 \pm 0.012$	$0.2429 \pm 0.0072$
$v_2$	$0.439 \pm 0.015$	$0.565 \pm 0.019$	$0.591 \pm 0.020$	$0.404 \pm 0.013$
$v_3$	$0.768 \pm 0.025$	$0.872 \pm 0.029$	$0.879 \pm 0.030$	$0.679 \pm 0.022$
$v_4$	$0.882 \pm 0.028$	$1.061 \pm 0.034$	$1.105 \pm 0.035$	$0.952 \pm 0.029$
$v_5$	$1.078 \pm 0.033$	$1.152 \pm 0.037$	$1.180 \pm 0.038$	$0.982 \pm 0.031$
$v_6$	$1.009 \pm 0.031$	$1.181 \pm 0.037$	$1.202 \pm 0.038$	$1.095 \pm 0.033$
$v_7$	$1.112 \pm 0.034$	$1.171 \pm 0.037$	$1.208 \pm 0.038$	$1.083 \pm 0.033$
$v_8$	$1.095 \pm 0.029$	$1.222 \pm 0.034$	$1.200 \pm 0.034$	$1.191 \pm 0.031$
$v_9$	$1.130 \pm 0.046$	$1.128 \pm 0.048$	$1.312 \pm 0.052$	$1.065 \pm 0.045$

Table I.7 – Results of the decay-time fits performed for TOS candidates on L0Hadron and all the other candidates.

Parameter	L0HadronTOS	not L0HadronTOS
$S_f$	$0.029 \pm 0.026$	$0.111 \pm 0.034$
$S_{\bar{f}}$	$0.029 \pm 0.026$	$0.048 \pm 0.034$
$A_P$	$-0.0092 \pm 0.0035$	$-0.0002 \pm 0.0049$
$A_D$	$0.0079 \pm 0.0024$	$0.0093 \pm 0.0032$
$\Gamma$	$0.6588 \pm 0.0017$	$0.6587 \pm 0.0017$
$\Delta m$	$0.5053 \pm 0.0023$	$0.5049 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.196 \pm 0.027$	$-0.083 \pm 0.034$
$p_1^{\text{OS}}$	$-0.066 \pm 0.030$	$-0.005 \pm 0.039$
$p_2^{\text{OS}}$	$-0.006 \pm 0.012$	$-0.010 \pm 0.015$
$p_3^{\text{OS}}$	$-0.20 \pm 0.16$	$-0.46 \pm 0.18$
$p_4^{\text{OS}}$	$-0.66 \pm 0.79$	$-0.70 \pm 0.80$
$\Delta p_0^{\text{OS}}$	$-0.145 \pm 0.060$	$0.044 \pm 0.084$
$\Delta p_1^{\text{OS}}$	$0.122 \pm 0.046$	$0.151 \pm 0.059$
$\Delta p_2^{\text{OS}}$	$-0.032 \pm 0.017$	$-0.012 \pm 0.020$
$\Delta p_3^{\text{OS}}$	$-0.34 \pm 0.25$	$0.04 \pm 0.25$
$\Delta p_4^{\text{OS}}$	$-1.8 \pm 1.3$	$0.7 \pm 1.1$
$p_0^{\text{SS}}$	$0.013 \pm 0.024$	$-0.229 \pm 0.046$
$p_1^{\text{SS}}$	$-0.024 \pm 0.025$	$-0.029 \pm 0.050$
$\Delta p_0^{\text{SS}}$	$-0.108 \pm 0.053$	$-0.061 \pm 0.083$
$\Delta p_1^{\text{SS}}$	$0.012 \pm 0.038$	$0.100 \pm 0.070$
$\varepsilon_{\text{tag}}^{\text{OS}}$	$0.40498 \pm 0.00094$	$0.4834 \pm 0.0014$
$\varepsilon_{\text{tag}}^{\text{SS}}$	$0.93331 \pm 0.00048$	$0.92502 \pm 0.00071$
$v_1$	$0.4111 \pm 0.0092$	$0.1889 \pm 0.0052$
$v_2$	$0.608 \pm 0.015$	$0.331 \pm 0.010$
$v_3$	$0.909 \pm 0.022$	$0.633 \pm 0.019$
$v_4$	$1.087 \pm 0.025$	$0.863 \pm 0.024$
$v_5$	$1.174 \pm 0.027$	$0.981 \pm 0.028$
$v_6$	$1.188 \pm 0.027$	$1.018 \pm 0.028$
$v_7$	$1.200 \pm 0.027$	$1.051 \pm 0.029$
$v_8$	$1.234 \pm 0.024$	$1.095 \pm 0.026$
$v_9$	$1.218 \pm 0.034$	$1.063 \pm 0.039$

## 1696 J Decay-time fit validation with bootstrapping

1697 In this appendix, further results of the Monte Carlo time fit validation are shown in  
 1698 addition to Fig. 5.14. The results for the production asymmetry, detection asymmetry,  
 1699 and acceptance parameters are not shown because no independent reference value for  
 1700 these parameters exists in Monte Carlo.

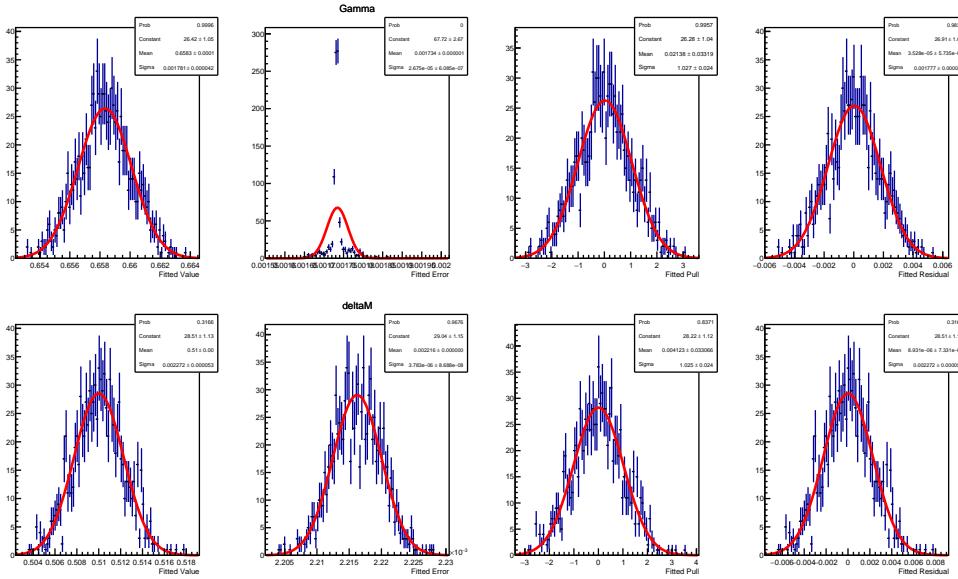


Figure J.1 – Distributions of the fitted value, error, pull and residual for  $\Gamma$  (top) and  $\Delta m$  (bottom). Each distribution is fitted with a Gaussian function. The result of the Gaussian fit is shown for the fitted error as well, even though uncertainties are not always Gaussian. Pulls and residuals are computed by taking the Monte Carlo generation value as reference (see Appendix M).

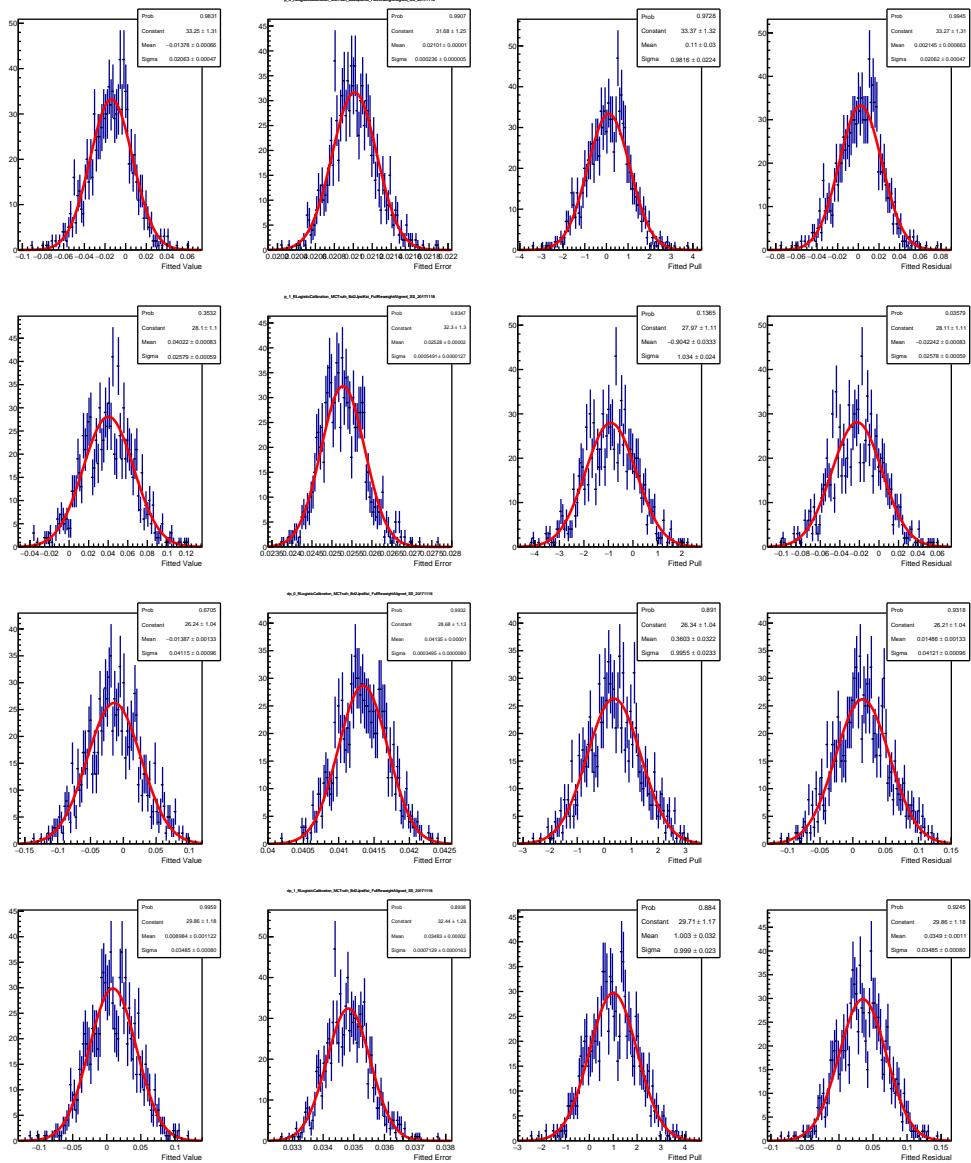


Figure J.2 – Distributions of the fitted value, error, pull and residual for the SS tagger calibration parameters  $p_0^{\text{SS}}$ ,  $p_1^{\text{SS}}$ ,  $\Delta p_0^{\text{SS}}$ , and  $\Delta p_1^{\text{SS}}$  (from top to bottom). Each distribution is fitted with a Gaussian function. The result of the Gaussian fit is shown for the fitted error as well, even though uncertainties are not always Gaussian. Pulls and residuals are computed by taking the values found on the  $B^0 \rightarrow J/\psi K^{*0}$  Monte Carlo calibration as reference (see Table 3.5).

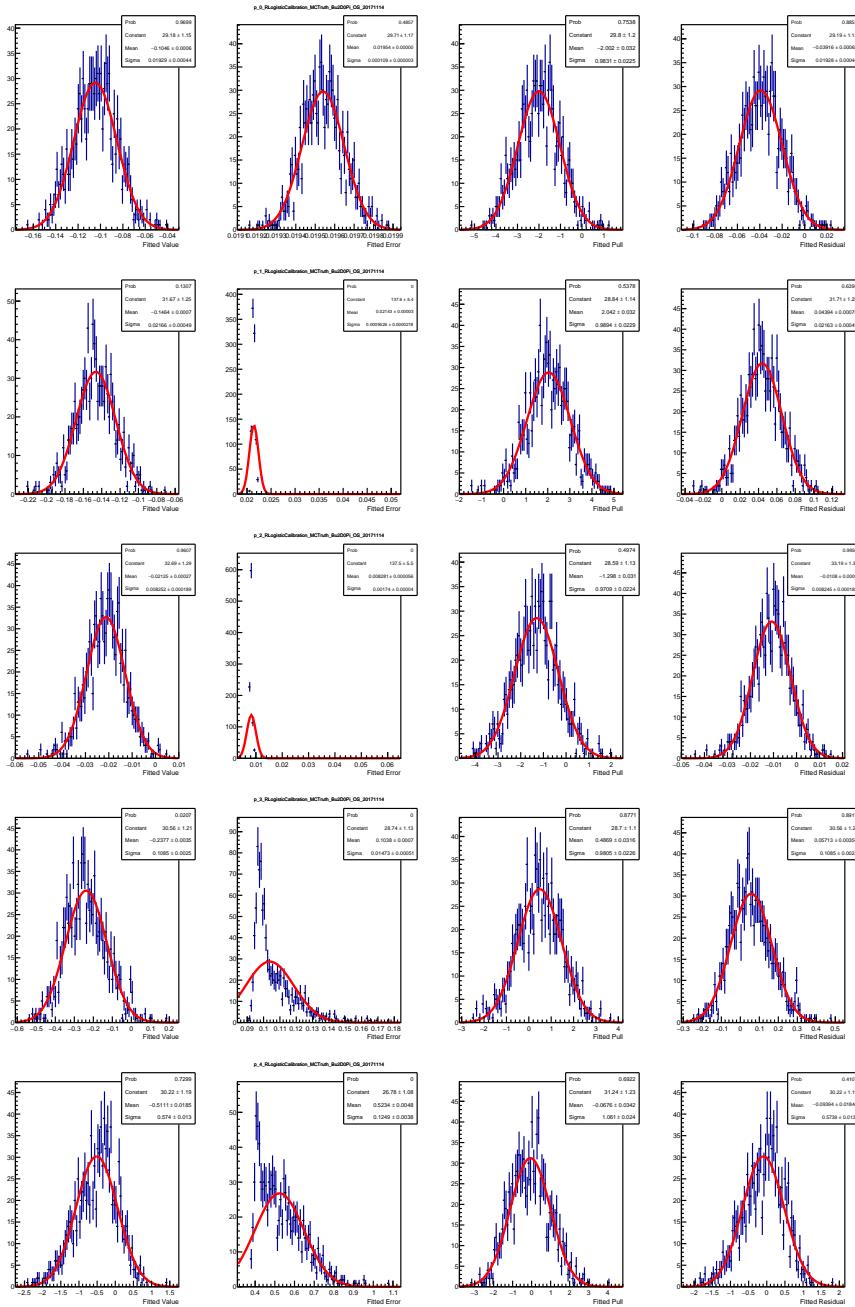


Figure J.3 – Distributions of the fitted value, error, pull and residual for the OS tagger calibration parameters  $p_0^{\text{OS}}$ ,  $p_1^{\text{OS}}$ ,  $p_2^{\text{OS}}$ ,  $p_3^{\text{OS}}$ , and  $p_4^{\text{OS}}$  (from top to bottom). Each distribution is fitted with a Gaussian function. The result of the Gaussian fit is shown for the fitted error as well, even though uncertainties are not always Gaussian. Pulls and residuals are computed by taking the values found on the  $B^+ \rightarrow D^0\pi^+$  Monte Carlo calibration as reference (see Table 3.3).

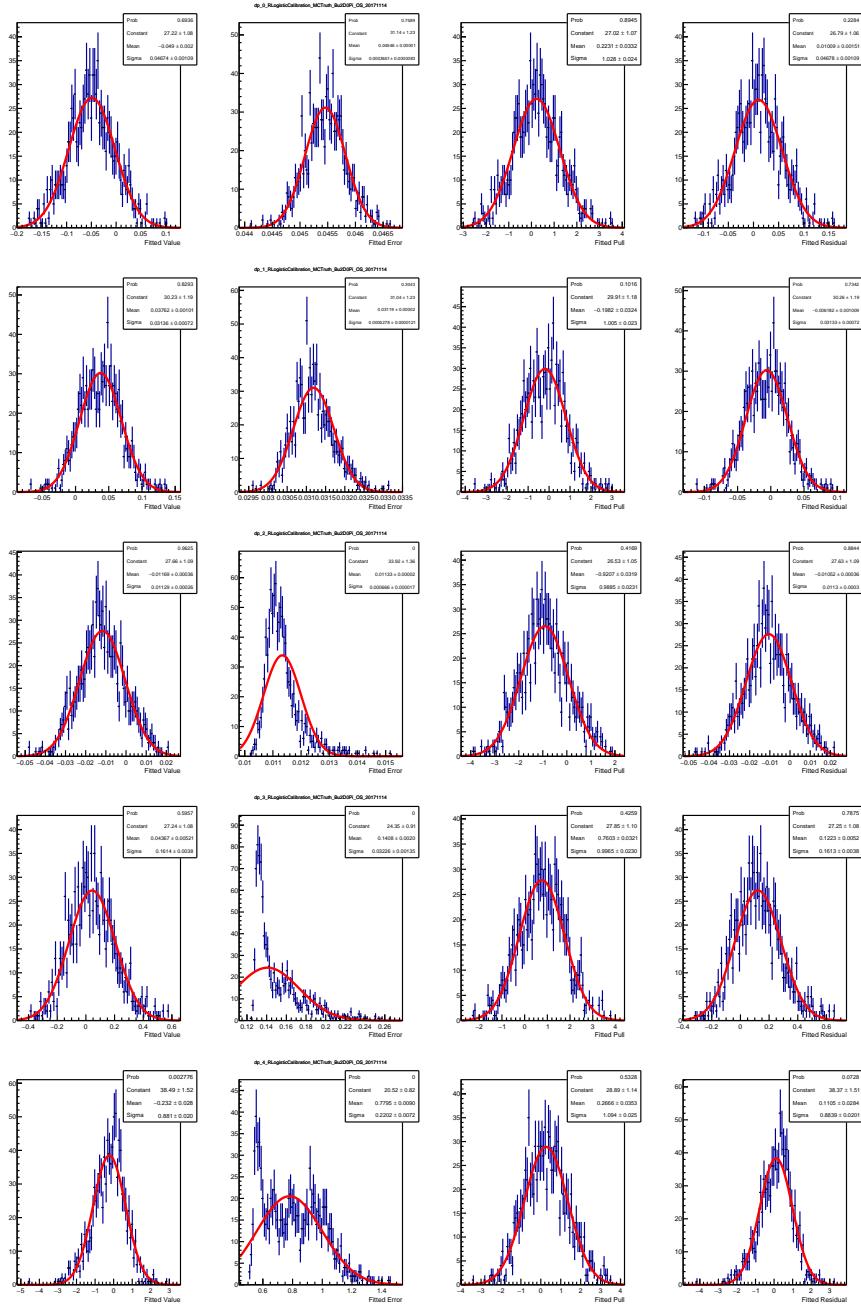


Figure J.4 – Distributions of the fitted value, error, pull and residual for the OS tagger calibration parameters  $\Delta p_0^{\text{OS}}$ ,  $\Delta p_1^{\text{OS}}$ ,  $\Delta p_2^{\text{OS}}$ ,  $\Delta p_3^{\text{OS}}$ , and  $\Delta p_4^{\text{OS}}$  (from top to bottom). Each distribution is fitted with a Gaussian function. The result of the Gaussian fit is shown for the fitted error as well, even though uncertainties are not always Gaussian. Pulls and residuals are computed by taking the values found on the  $B^+ \rightarrow D^0\pi^+$  Monte Carlo calibration as reference (see Table 3.3).

1701 **K Test of the decay-time fit via a toy tagger**

The *toy tagger* used to perform the test mentioned in Sec. 5.4 is created as follows. First, a mistag ( $\eta$ ) PDF is created from the *sWeighted*  $\eta$  distribution of the OS tagger on data. This template is created as a `RooHistPdf`. Then, for each candidate in each bootstrapped Monte Carlo sample, a value of  $\eta$  is drawn from this PDF. The decision of the toy tagger is initially taken from the true ID of the  $B^0$  meson, which is always correct by definition. In order to emulate wrong tagging decisions, a random number  $r_i$  is generated for the  $i^{th}$   $B^0$  candidate between 0 and 1. If  $\eta_i$  is the mistag assigned to this candidate, the tagging decision  $d_i$  may need to be flipped (and thus made wrong) according to the following criterion:

$$d_i \rightarrow \begin{cases} -d_i & \text{if } r_i \leq \eta_i \\ d_i & \text{otherwise} \end{cases}. \quad (\text{K.1})$$

1702 During the time fit, the mistag calibration is simply taken as a linear function (Eq. 3.9)  
1703 with  $p_0 = \langle \eta \rangle = 0.370029$  (taken from the adopted template) and  $p_1 = 1$ , which means  
1704  $\omega = \eta$  for all candidates. In fact, the per-event mistag  $\eta$  is the true mistag  $\omega$  probability by  
1705 construction. In this way, it is possible to test the time fit with a per-event mistag without  
1706 relying on any approximation or uncertainty coming from the calibration procedure.  
1707 Moreover, the tagging efficiency is 100% by construction.

1708 The distributions of the fitted value, error, pull and residual for the relevant parameters  
1709 are shown in Fig. K.1. All pull distributions have means compatible with 0 and widths  
1710 compatible with 1, meaning that the maximum likelihood estimation of the parameters is  
1711 unbiased and returns correct uncertainties.

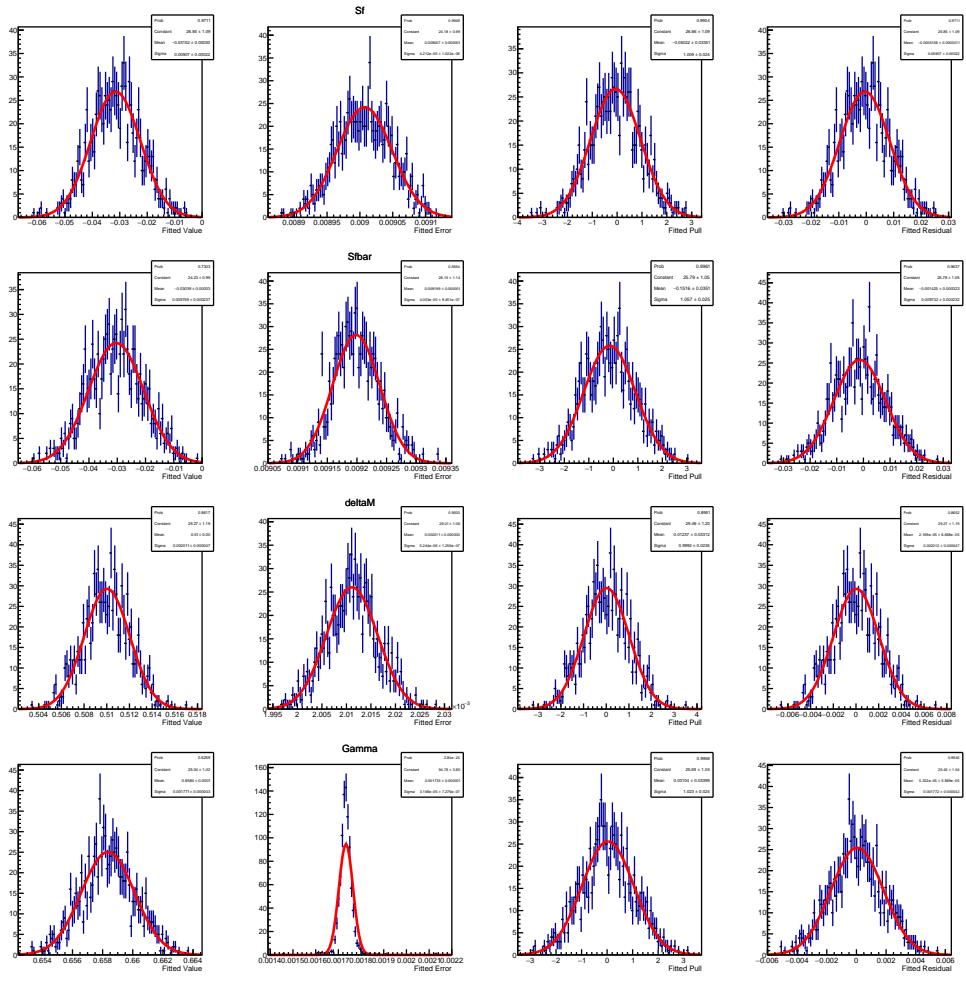


Figure K.1 – Distributions of the fitted value, error, pull and residual for the main parameters ( $S_f$ ,  $S_{\bar{f}}$ ,  $\Delta m$ , and  $\Gamma$ , from top to bottom) fitted on bootstrapped Monte Carlo samples with a toy tagger. Each distribution is fitted with a Gaussian function. Pulls and residuals are computed by taking the Monte Carlo generation value as reference.

## **L Correlation between systematic uncertainties**

As reported in Table 5.5, the main systematic uncertainties on  $S_f$  and  $S_{\bar{f}}$  are the ones due to the constraint on  $\Delta m$ , the fit biases, and the background subtraction. The correlations between the systematics uncertainties on  $S_f$  and  $S_{\bar{f}}$  due to these sources are described in details in this appendix. The correlation of the total systematic resulting from these three contributions is  $-0.41$ . The correlation between other sources of systematics is neglected.

### **L.1 Correlation of $\Delta m$ systematics**

The correlation of systematics uncertainties due to  $\Delta m$  between  $S_f$  and  $S_{\bar{f}}$  is estimated by comparing the nominal fit result, and the result obtained with  $\Delta m$  fixed as described in Sec. 5.5.1. This correlation  $\rho_{S_f, S_{\bar{f}}}^{\Delta m}$  is simply computed from the difference in statistical covariance between the two fit results:

$$\rho_{S_f, S_{\bar{f}}}^{\Delta m} = \frac{\sigma_{S_f}^{\text{nominal}} \sigma_{S_{\bar{f}}}^{\text{nominal}} \rho_{S_f, S_{\bar{f}}}^{\text{nominal}} - \sigma_{S_f}^{\Delta m \text{ fixed}} \sigma_{S_{\bar{f}}}^{\Delta m \text{ fixed}} \rho_{S_f, S_{\bar{f}}}^{\Delta m \text{ fixed}}}{\sqrt{(\sigma_{S_f}^{\text{nominal}})^2 - (\sigma_{S_f}^{\Delta m \text{ fixed}})^2} \sqrt{(\sigma_{S_{\bar{f}}}^{\text{nominal}})^2 - (\sigma_{S_{\bar{f}}}^{\Delta m \text{ fixed}})^2}} \quad (\text{L.1})$$

The value of  $\rho_{S_f, S_{\bar{f}}}^{\Delta m}$  so obtained is  $-1$ , meaning that  $S_f$  and  $S_{\bar{f}}$  are fully anticorrelated driven by the uncertainty on  $\Delta m$ .

### **L.2 Correlation of systematics due to fit biases**

As described in Sec. 5.4, a bias on  $S_f$  and  $S_{\bar{f}}$  is observed from bootstrapped MC samples, and the size of this bias is assigned as systematic uncertainty. The associated correlation is estimated from the two-dimensional distribution of  $(S_f^{\text{fit}} - S_f^{\text{gen}})$  versus  $(S_{\bar{f}}^{\text{fit}} - S_{\bar{f}}^{\text{gen}})$  obtained on the same set of bootstrapped MC samples, as shown in Fig. L.1. The resulting correlation is  $0.4$ .

### **L.3 Correlation of systematics due to background subtraction**

A systematic due to *sWeighting* and background subtraction is assigned by repeating Fit B in a wider mass range, as described in Sec. 5.5.3. In order to estimate the correlation of this systematic uncertainty between  $S_f$  and  $S_{\bar{f}}$ , the data sample is bootstrapped in a similar way as done for the Monte Carlo (Sec. 5.4). Then, *sWeights* are obtained twice on each sample, once with the nominal strategy, and once by selecting a wide mass range in Fit B. Finally, the time fit is performed on each sample using both the sets of *sWeights* computed in the previous step. The correlation is estimated from the two-dimensional distribution of  $(S_{\bar{f}}^{\text{nominal}} - S_{\bar{f}}^{\text{wide mass}})$  versus  $(S_f^{\text{nominal}} - S_f^{\text{wide mass}})$ , which is shown in Fig. L.2. The resulting correlation is  $0.7$ .

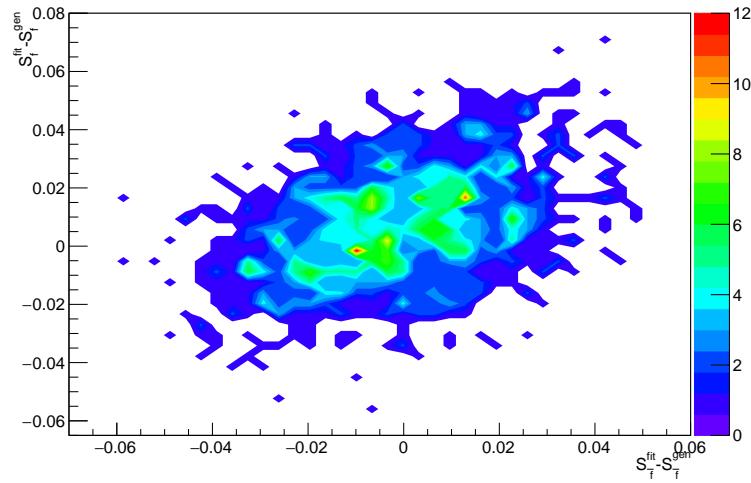


Figure L.1 – Two-dimensional distribution of  $(S_f^{\text{fit}} - S_f^{\text{gen}})$  versus  $(S_{\bar{f}\bar{f}}^{\text{fit}} - S_{\bar{f}\bar{f}}^{\text{gen}})$  obtained from the fits to bootstrapped MC samples (Sec. 5.4).

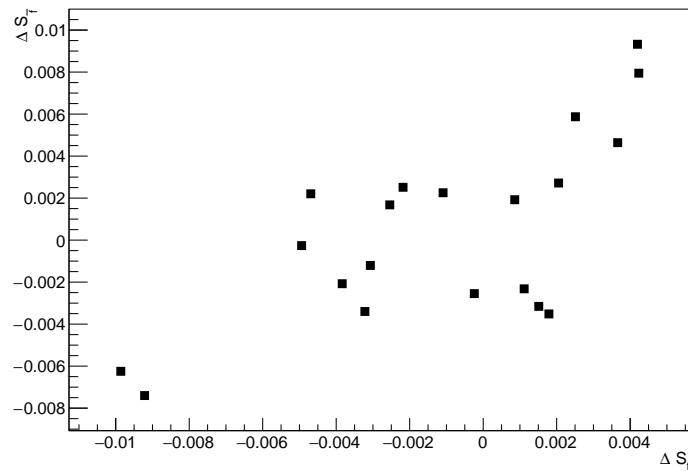


Figure L.2 – Two-dimensional distribution of  $S_{\bar{f}\bar{f}}^{\text{nominal}} - S_{\bar{f}\bar{f}}^{\text{wide mass}}$  ( $\Delta S_{\bar{f}\bar{f}}$ ) versus  $S_f^{\text{nominal}} - S_f^{\text{wide mass}}$  ( $\Delta S_f$ ) obtained from fits to bootstrapped data samples, where Fit B is performed with the nominal strategy or with a wider mass range for each sample.

<sup>1737</sup> M Inputs for the  $B^0 \rightarrow D^\pm \pi^\mp$  simulation

Table M.1 – Parameter values used in the generation of the  $B^0 \rightarrow D^\pm \pi^\mp$  MC sample, and resulting values for the  $CP$  coefficients.

Parameter	Generation value
$\Delta m$	$0.51 \text{ ps}^{-1}$
$\tau$	$1.519\,068 \text{ ps}$
$\Delta\Gamma/\Gamma$	0
$ q/p $	1
$\arg(q/p)$	-0.764
$ A_f $	0.0849
$\arg(A_f)$	0.002 278
$ \bar{A}_f $	0.001 37
$\arg(\bar{A}_f)$	-1.128 958
$ A_{\bar{f}} $	0.001 37
$\arg(A_{\bar{f}})$	1.3145
$ \bar{A}_{\bar{f}} $	0.0849
$\arg(\bar{A}_{\bar{f}})$	0.002 278
$S_f$	-0.0305
$S_{\bar{f}}$	-0.0282
$C_f$	0.9995
$C_{\bar{f}}$	-0.9995
$D_f$	-0.0103
$D_{\bar{f}}$	-0.0155

# Bibliography

- 1739 [1] A. Salam, *Weak and electromagnetic interactions*, Conf. Proc. **C680519** (1968) 367.
- 1740 [2] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264.
- 1741 [3] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.
- 1742 [4] M. Lubej, *personal website*, <http://www-f9.ijs.si/%7Elubej/SM.pdf>. Accessed: 2018-05-14.
- 1744 [5] E. Fermi, *Sulla quantizzazione del gas perfetto monoatomico*, Rendiconti Lincei **3** (1926) 145.
- 1746 [6] P. A. M. Dirac, *On the theory of quantum mechanics*, Proceedings of the Royal Society **A112** (**762**) (1926) 661.
- 1748 [7] S. Bose, *Plancks Gesetz und Lichtquantenhypothese*, Z. Phys. **26** (1924) 178.
- 1749 [8] LHCb collaboration, R. Aaij *et al.*, *Observation of the resonant character of the*  $Z(4430)^-$  *state*, Phys. Rev. Lett. **112** (2014) 222002, [arXiv:1404.1903](https://arxiv.org/abs/1404.1903).
- 1751 [9] LHCb collaboration, R. Aaij *et al.*, *Observation of  $J/\psi p$  resonances consistent with pentaquark states in  $\Lambda_b^0 \rightarrow J/\psi p K^-$  decays*, Phys. Rev. Lett. **115** (2015) 072001, [arXiv:1507.03414](https://arxiv.org/abs/1507.03414).
- 1754 [10] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.
- 1756 [11] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.
- 1758 [12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964) 585.
- 1760 [13] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, [arXiv:1207.7214](https://arxiv.org/abs/1207.7214).

- 1763 [14] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass*  
 1764 *of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30,  
 1765 arXiv:1207.7235.
- 1766 [15] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **178** (1963)  
 1767 531.
- 1768 [16] M. Kobayashi and T. Maskawa, *CP violation in the renormalizable theory of weak*  
 1769 *interaction*, Progr. Theor. Phys. **49** (1973) 652.
- 1770 [17] L. L. Chau and W. Y. Keung, *Comments on the parametrization of the Kobayashi-*  
 1771 *Maskawa matrix*, Phys. Rev. Lett. **53** (1984) 1802.
- 1772 [18] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa matrix*, Progr. Theor.  
 1773 Phys. **49** (1973) 652.
- 1774 [19] C. Jarlskog, *Commutator of the quark mass matrices in the standard electroweak*  
 1775 *model and a measure of maximal CP Violation*, Phys. Rev. Lett. **55** (1985) 1039.
- 1776 [20] A. D. Sakharov, *Violation of CP invariance, C asymmetry, and baryon asym-*  
 1777 *metry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32, [Usp. Fiz.  
 1778 Nauk161,no.5,61(1991)].
- 1779 [21] CKMfitter group, J. Charles *et al.*, *Current status of the Standard*  
 1780 *Model CKM fit and constraints on  $\Delta F = 2$  new physics*, Phys. Rev. **D91**  
 1781 (2015) 073007, arXiv:1501.05013, updated results and plots available at  
 1782 <http://ckmfitter.in2p3.fr/>.
- 1783 [22] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys.  
 1784 **C40** (2016) 100001, and 2017 update.
- 1785 [23] I. Bigi and A. I. Sanda, *CP violation*, Cambridge Monographs on Particle Physics,  
 1786 Nuclear Physics and Cosmology, 2009.
- 1787 [24] R. Fleischer, *New strategies to obtain insights into CP violation through*  
 1788  $B_{(s)} \rightarrow D_{(s)}^\pm K^\mp$ ,  $D_{(s)}^{*\pm} K^\mp$ , ... and  $B_{(d)} \rightarrow D^\pm \pi^\mp$ ,  $D^{*\pm} \pi^\mp$ , ... *decays*, Nucl. Phys.  
 1789 **B671** (2003) 459, arXiv:hep-ph/0304027.
- 1790 [25] R. Aleksan, I. Dunietz, and B. Kayser, *Determining the CP violating phase  $\gamma$* , Z.  
 1791 Phys. **C54** (1992) 653.
- 1792 [26] I. Dunietz and R. G. Sachs, *Asymmetry between inclusive charmed and anticharmed*  
 1793 *modes in  $B^0$ ,  $\bar{B}^0$  decay as a measure of CP violation*, Phys. Rev. **D37** (1988) 3186,  
 1794 erratum *ibid.* **D39** (1989) 3515.
- 1795 [27] LHCb collaboration, R. Aaij *et al.*, *Measurement of the time-dependent CP asymme-*  
 1796 *tries in  $B_s^0 \rightarrow J/\psi K_S^0$* , JHEP **06** (2015) 131, arXiv:1503.07055.

- 1797 [28] BaBar collaboration, B. Aubert *et al.*, *Measurement of the branching fractions of*  
 1798 *the rare decays  $B^0 \rightarrow D_s^{(*)+} \pi^-$ ,  $B^0 \rightarrow D_s^{(*)+} \rho^-$ , and  $B^0 \rightarrow D_s^{(*)-} K^{(*)+}$* , Phys. Rev.  
 1799 **D78** (2008) 032005, [arXiv:0803.4296](#).
- 1800 [29] Belle collaboration, A. Das *et al.*, *Measurements of branching fractions for*  
 1801  *$B^0 \rightarrow D_s^+ \pi^-$  and  $\bar{B}^0 \rightarrow D_s^+ K^-$* , Phys. Rev. **D82** (2010) 051103, [arXiv:1007.4619](#).
- 1802 [30] LHCb, B. Storaci, *Updated average  $f_s/f_d$  b-hadron production fraction ratio for 7*  
 1803 *TeV pp collisions*, .
- 1804 [31] Belle, R. Louvot *et al.*, *Measurement of the decay  $B_s0 \rightarrow D_s - \pi^+$  and evidence for*  
 1805  *$B_s0 \rightarrow D_s \pm K^\pm$  in  $e^+ e_-$  annihilation at  $\sqrt{s} = 10.87\text{-GeV}$* , Phys. Rev. Lett. **102**  
 1806 (2009) 021801, [arXiv:0809.2526](#).
- 1807 [32] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys.  
 1808 **C40** (2016) 100001, and 2017 update.
- 1809 [33] BaBar collaboration, B. Aubert *et al.*, *Measurement of time-dependent CP asymme-*  
 1810 *tries in  $B^0 \rightarrow D^{(*)\pm} \pi^\mp$  and  $B^0 \rightarrow D^\pm \rho^\mp$  decays*, Phys. Rev. D **73** (2006) 111101,  
 1811 [arXiv:hep-ex/0602049](#).
- 1812 [34] BaBar collaboration, B. Aubert *et al.*, *Measurement of time-dependent CP-violating*  
 1813 *asymmetries and constraints on  $\sin(2\beta + \gamma)$  with partial reconstruction of  $B \rightarrow$*   
 1814  *$D^{*\mp} \pi^\pm$  decays*, Phys. Rev. **D71** (2005) 112003, [arXiv:hep-ex/0504035](#).
- 1815 [35] Belle collaboration, S. Bahinipati *et al.*, *Measurements of time-dependent CP asym-*  
 1816 *metries in  $B \rightarrow D^{*\mp} \pi^\pm$  decays using a partial reconstruction technique*, Phys. Rev.  
 1817 **D84** (2011) 021101, [arXiv:1102.0888](#).
- 1818 [36] Belle collaboration, F. J. Ronga *et al.*, *Measurement of CP violation in  $B^0 \rightarrow D^{*-} \pi^+$*   
 1819 *and  $B^0 \rightarrow D^- \pi^+$  decays*, Phys. Rev. **D73** (2006) 092003, [arXiv:hep-ex/0604013](#).
- 1820 [37] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP asymmetry in  $B_s^0 \rightarrow D_s^\mp K^\pm$*   
 1821 *decays*, [arXiv:1712.07428](#), submitted to JHEP.
- 1822 [38] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*,  
 1823 Nucl. Instrum. Meth. **A555** (2005) 356, [arXiv:physics/0402083](#).
- 1824 [39] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys.  
 1825 **A30** (2015) 1530022, [arXiv:1412.6352](#).
- 1826 [40] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007,  
 1827 [arXiv:1405.7808](#).
- 1828 [41] LHCb collaboration, R. Antunes Nobrega *et al.*, *LHCb reoptimized detector design*  
 1829 *and performance: Technical Design Report*, CERN-LHCC-2003-030.

- 1830 [42] LHCb collaboration, A. Franca Barbosa *et al.*, *LHCb inner tracker: Technical Design*  
 1831 *Report*, CERN-LHCC-2002-029.
- 1832 [43] LHCb collaboration, P. R. Barbosa Marinho *et al.*, *LHCb outer tracker: Technical*  
 1833 *Design Report*, CERN-LHCC-2001-024.
- 1834 [44] V. Battista, F. Blanc, M. Martinelli, and M. Tobin, *A study of spillover clusters and*  
 1835 *ghost tracks in the Silicon Tracker with 25 ns bunch spacing*, CERN-LHCB-INT-  
 1836 2016-010.
- 1837 [45] A. Powell *et al.*, *Particle identification at LHCb*, Proceedings of ICHEP2010 **020**  
 1838 (2010), LHCb-PROC-2011-008.
- 1839 [46] LHCb collaboration, S. Amato *et al.*, *LHCb calorimeters: Technical Design Report*,  
 1840 CERN-LHCC-2000-036.
- 1841 [47] A. A. Alves Jr *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013)  
 1842 P02022.
- 1843 [48] LHCb collaboration, R. Antunes Nobrega *et al.*, *LHCb trigger system: Technical*  
 1844 *Design Report*, CERN-LHCC-2003-031.
- 1845 [49] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*,  
 1846 Comput. Phys. Commun. **178** (2008) 852, [arXiv:0710.3820](https://arxiv.org/abs/0710.3820).
- 1847 [50] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth.  
 1848 **A462** (2001) 152.
- 1849 [51] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections*  
 1850 *in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, [arXiv:hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026).
- 1851 [52] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl.  
 1852 Instrum. Meth. **A506** (2003) 250.
- 1853 [53] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE  
 1854 Trans. Nucl. Sci. **53** (2006) 270.
- 1855 [54] *Development of "same side" flavour tagging algorithms for measurements of flavour*  
 1856 *oscillations and CP violation in the  $B^0$  mesons system*, PhD thesis, Milano Bicocca  
 1857 U., 2015.
- 1858 [55] LHCb collaboration, R. Aaij *et al.*, *New algorithms to tag the flavour of  $B^0$  mesons*  
 1859 *using pions and protons*, Eur. Phys. J. **C77** (2016) 238, [arXiv:1610.06019](https://arxiv.org/abs/1610.06019).
- 1860 [56] LHCb collaboration, R. Aaij *et al.*, *Opposite-side flavour tagging of B mesons at the*  
 1861 *LHCb experiment*, Eur. Phys. J. **C72** (2012) 2022, [arXiv:1202.4979](https://arxiv.org/abs/1202.4979).
- 1862 [57] LHCb collaboration, R. Aaij *et al.*, *B flavour tagging using charm decays at the*  
 1863 *LHCb experiment*, JINST **10** (2015) P10005, [arXiv:1507.07892](https://arxiv.org/abs/1507.07892).

- 1864 [58] M. Grabalosa, *Flavour Tagging developments within the LHCb experiment*, PhD  
 1865 thesis, Barcelona U., 2012.
- 1866 [59] LHCb collaboration, *Optimization and calibration of the LHCb flavour tagging*  
 1867 *performance using 2010 data*, LHCb-CONF-2011-003.
- 1868 [60] V. Battista, *b-flavour tagging in pp collisions at LHCb*, Il Nuovo Cimento **39C** (2016)  
 1869 335.
- 1870 [61] J. Wimberley, *Calibration flavor tagging algorithms with binomial regression*, LHCb-  
 1871 INT-2017-002.
- 1872 [62] W. Cheney and D. R. Kincaid, *Linear algebra: theory and applications*, Jones &  
 1873 Bartlett Learning International Series in Mathematic, 2009.
- 1874 [63] T. Head *et al.*, *scikit-optimize python package*, <https://github.com/scikit-optimize>.  
 1875 Accessed: 2018-05-18.
- 1876 [64] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*,  
 1877 [arXiv:1603.02754](https://arxiv.org/abs/1603.02754).
- 1878 [65] L. Breiman, *Bagging predictors*, Machine Learning **24** (1996) 123.
- 1879 [66] LHCb collaboration, *A single track HLT1 trigger*, LHCb-PUB-2011-003.
- 1880 [67] LHCb collaboration, *The LHCb inclusive B triggers*, LHCb-INT-2011-030.
- 1881 [68] W. D. Hulsbergen *et al.*, *Decay chain fitting with a Kalman filter*, Nucl. Instrum.  
 1882 Meth. **A552** (2005) 566.
- 1883 [69] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and*  
 1884 *regression trees*, Wadsworth international group, Belmont, California, USA, 1984.
- 1885 [70] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neu-*  
 1886 *ral networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577,  
 1887 [arXiv:physics/0408124](https://arxiv.org/abs/physics/0408124).
- 1888 [71] A. Hoecker *et al.*, *TMVA: Toolkit for multivariate data analysis*, PoS **ACAT** (2007)  
 1889 040, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- 1890 [72] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning*  
 1891 *and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.
- 1892 [73] P. Koppenburg, *Statistical biases in measurements with multiple candidates*,  
 1893 [arXiv:1703.01128](https://arxiv.org/abs/1703.01128).
- 1894 [74] M. Pivk and F. R. L. Diberder, *sPlot: a statistical tool to unfold data distributions*,  
 1895 Nucl. Instrum. Meth. **A555** (2005) 356.

- 1896 [75] D. M. Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*,  
 1897 Nucl. Instrum. Meth. **A764** (2014) 150.
- 1898 [76] N. L. Johnson, *Systems of frequency curves generated by methods of translation*,  
 1899 Biometrika **36** (1949) 149.
- 1900 [77] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime*  
 1901 and *Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986,  
 1902 DESY-F31-86-02.
- 1903 [78] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP-violating phase  $\phi_s$  in*  
 1904  $\bar{B}_s^0 \rightarrow D_s^+ D_s^-$  *decays*, Phys. Rev. Lett. **113** (2014) 211801, [arXiv:1409.4619](#).
- 1905 [79] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of CP violation in*  
 1906  $B_s^0 \rightarrow J/\psi K^+ K^-$  *decays*, Phys. Rev. Lett. **114** (2015) 041801, [arXiv:1411.3104](#).
- 1907 [80] M. Karbach, G. Raven, and M. Schiller, *Decay time integrals in neutral meson mixing*  
 1908 and *their efficient evaluation*, [arXiv:1407.0748](#).
- 1909 [81] Y. Xie, *sFit: a method for background subtraction in maximum likelihood fit*,  
 1910 [arXiv:0905.0724](#).
- 1911 [82] LHCb collaboration, R. Aaij *et al.*, *A precise measurement of the  $B^0$  meson oscillation*  
 1912 *frequency*, Eur. Phys. J. **C76** (2016) 412, [arXiv:1604.03475](#).
- 1913 [83] LHCb collaboration, R. Aaij *et al.*, *Measurement of  $B^0$ ,  $B_s^0$ ,  $B^+$  and  $\Lambda_b^0$  produc-*  
 1914 *tion asymmetries in 7 and 8 TeV pp collisions*, Phys. Lett. **B774** (2017) 139,  
 1915 [arXiv:1703.08464](#).
- 1916 [84] P. Good, *Permutation, parametric and bootstrap tests of hypotheses*, Springer Series  
 1917 in Statistics, 2004.
- 1918 [85] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-*  
 1919 *hadron, and  $\tau$ -lepton properties as of summer 2016*, Eur. Phys. J. **C77**  
 1920 (2017) 895, [arXiv:1612.07233](#), updated results and plots available at  
 1921 <http://www.slac.stanford.edu/xorg/hflav/>.
- 1922 [86] LHCb, R. Aaij *et al.*, *Measurement of CP violation in  $B^0 \rightarrow D^\mp \pi^\pm$  decays*,  
 1923 [arXiv:1805.03448](#).
- 1924 [87] LHCb, R. Aaij *et al.*, *Measurement of the CKM angle  $\gamma$  from a combination of LHCb*  
 1925 *results*, JHEP **12** (2016) 087, [arXiv:1611.03076](#).
- 1926 [88] S. Aoki *et al.*, *Review of lattice results concerning low-energy particle physics*, Eur.  
 1927 Phys. J. **C77** (2017), no. 2 112, [arXiv:1607.00299](#).
- 1928 [89] Fermilab Lattice, MILC, A. Bazavov *et al.*, *Charmed and light pseudoscalar meson*  
 1929 *decay constants from four-flavor lattice QCD with physical light quarks*, Phys. Rev.  
 1930 **D90** (2014) 074509, [arXiv:1407.3772](#).

- 1931 [90] N. Carrasco *et al.*, *Leptonic decay constants  $f_K$ ,  $f_D$ , and  $f_{D_s}$  with  $N_f = 2 + 1 + 1$*
- 1932 *twisted-mass lattice QCD*, Phys. Rev. **D91** (2015) 054507, [arXiv:1411.7908](https://arxiv.org/abs/1411.7908).
- 1933 [91] K. De Bruyn *et al.*, *Exploring  $B_s \rightarrow D_s^{(*)\pm} K^\mp$  decays in the presence of a sizable*
- 1934 *width difference  $\Delta\Gamma_s$* , Nucl. Phys. **B868** (2013) 351, [arXiv:1208.6463](https://arxiv.org/abs/1208.6463).
- 1935 [92] LHCb collaboration, *LHCb VELO Upgrade Technical Design Report*, CERN-LHCC-
- 1936 2013-021. LHCb-TDR-013.
- 1937 [93] L. Collaboration, *LHCb Tracker Upgrade Technical Design Report*, Tech. Rep.
- 1938 CERN-LHCC-2014-001. LHCB-TDR-015, Feb, 2014.
- 1939 [94] M. Marinangeli, *Search for the  $B_s^0 \rightarrow D^+ K^-$  decay at LHCb*, Master Project, École
- 1940 Polytechnique Fédérale de Lausanne, 2016.
- 1941 [95] A. Rogozhnikov, *hep\_ml python package*, [https://arogozhnikov.github.io/hep\\_ml/](https://arogozhnikov.github.io/hep_ml/).
- 1942 Accessed: 2018-05-18.