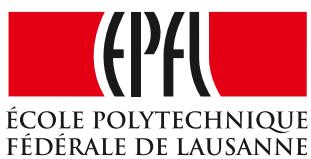


# **Measurement of time-dependent $CP$ violation in $B^0 \rightarrow D^\mp \pi^\mp$ decays and optimisation of Flavour Tagging algorithms at LHCb**

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, LHCb, LHC,  $CP$  violation, flavour tagging, mixing.



# Résumé

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



# Sommario

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



# Acknowledgements

*Lausanne, XX XXXX 2018*

V. B.



# Contents

<b>Abstract (English/Français/Italian)</b>	<b>v</b>
<b>Acknowledgements</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	2
1.2 The Standard Model of particle physics . . . . .	3
1.3 The Cabibbo-Kobayashi-Maskawa matrix . . . . .	5
1.4 Physics of the neutral $B$ mesons . . . . .	8
1.4.1 Oscillation of neutral mesons . . . . .	8
1.4.2 Decay of neutral mesons . . . . .	10
1.4.3 $CP$ violation in neutral meson systems . . . . .	12
1.5 $B^0 \rightarrow D^\pm \pi^\mp$ analysis strategy . . . . .	15
<b>2 The Large Hadron Collider and the LHCb experiment</b>	<b>17</b>
2.1 The Large Hadron Collider . . . . .	18
2.2 The LHCb experiment . . . . .	20
2.2.1 The tracking system . . . . .	20
2.3 Particle Identification (PID) . . . . .	24
2.3.1 The trigger system . . . . .	26
2.3.2 Event reconstruction, simulation and software . . . . .	29
2.3.3 Data collected by LHCb . . . . .	30
<b>3 Flavour tagging</b>	<b>33</b>
3.1 Flavour tagging algorithms . . . . .	34
3.2 Flavour tagging strategy for the $B^0 \rightarrow D^\pm \pi^\mp$ time-dependent analysis . .	40
3.2.1 Calibration of the opposite-side tagger combination . . . . .	42
3.2.2 Calibration of the same-side tagger combination . . . . .	46
3.3 Optimisation of the Opposite-Side electron tagger . . . . .	47
3.3.1 Introduction . . . . .	47
3.3.2 Sample selection . . . . .	49
3.3.3 Preselection optimisation . . . . .	49
3.3.4 BDT classifier implementation . . . . .	50

## Contents

---

3.3.5	Performance evaluation . . . . .	58
<b>4</b>	<b>Selection of <math>B^0 \rightarrow D^\mp\pi^\pm</math> decays</b>	<b>67</b>
4.1	Data sample and preselection . . . . .	68
4.1.1	Stripping and trigger requirements . . . . .	68
4.1.2	Preselection and sample definitions . . . . .	70
4.1.3	Vetoos for physics backgrounds . . . . .	70
4.1.4	Wrongly associated primary vertices . . . . .	72
4.1.5	Development of an MVA classifier . . . . .	73
4.1.6	BDT selection optimisation . . . . .	75
4.1.7	Multiple candidates . . . . .	77
4.1.8	Selection performance . . . . .	77
4.2	Simulation and expected sample composition . . . . .	80
4.2.1	PIDK correction . . . . .	80
4.2.2	Surviving physics backgrounds . . . . .	81
4.3	Fits to the $B^0$ invariant mass . . . . .	83
4.3.1	Probability density functions . . . . .	84
4.3.2	Fit to data . . . . .	85
4.3.3	$sWeight$ calculation . . . . .	87
4.3.4	Fits of subsamples . . . . .	93
<b>5</b>	<b>Measurement of <math>CP</math> violation in <math>B^0 \rightarrow D^\mp\pi^\pm</math> decays</b>	<b>97</b>
5.1	Time resolution . . . . .	98
5.1.1	Companion track momentum reweighting . . . . .	98
5.1.2	Resolution determination from decay time error parameterisation .	99
5.2	Time-dependent selection efficiency . . . . .	99
5.3	Decay-time fit to data . . . . .	102
5.3.1	Fit validation . . . . .	106
5.4	Systematics . . . . .	116
5.4.1	Systematic uncertainties from Gaussian constraints . . . . .	116
5.4.2	Systematic uncertainties estimated with pseudoexperiments . . . . .	117
5.4.3	Systematics related to the background subtraction . . . . .	121
5.5	Summary and interpretation of the result . . . . .	124
<b>Appendices</b>		<b>127</b>
A	Opposite-side tagging studies . . . . .	128
A.1	Mass fit of $B^\pm \rightarrow D^0\pi^\pm$ . . . . .	128
A.2	Reweighting of $B^\pm \rightarrow D^0\pi^\pm$ to $B^0 \rightarrow D^\mp\pi^\pm$ . . . . .	129
A.3	GOF tests for OS calibration on $B^\pm \rightarrow D^0\pi^\pm$ data . . . . .	131
B	Opposite-side electron optimisation . . . . .	134
C	Selection studies . . . . .	139
C.1	BDT input features . . . . .	139
C.2	Multiple candidates . . . . .	139

D	Particle identification plots . . . . .	143
E	Invariant mass fit studies . . . . .	146
F	PDF definitions . . . . .	147
G	Signal PDF for the decay-time fit . . . . .	149
H	Correlation matrix of decay-time fit of $B^0 \rightarrow D^\mp\pi^\pm$ . . . . .	151
I	Decay-time fits of $B^0 \rightarrow D^\mp\pi^\pm$ data subsamples . . . . .	152
J	Correlation between systematic uncertainties . . . . .	159
J.1	Correlation of $\Delta m$ systematics . . . . .	159
J.2	Correlation of systematics due to fit biases . . . . .	159
J.3	Correlation of systematics due to background subtraction . . . . .	159
K	Decay-time fit validation with bootstrapping . . . . .	161
L	Test of the decay-time fit via a toy tagger . . . . .	165
M	Inputs for $B^0 \rightarrow D^\pm\pi^\mp$ simulation . . . . .	167
<b>References</b>		<b>168</b>



# 1 Introduction

2

**3    1.1 Introduction**

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

- 7**     • opposite-side flavour tagging calibration (Sec. 3.2.1);  
**8**     • correction of particle identification (Sec. 4.2.1);  
**9**     • mass fit for *sWeights* calculation (Sec. 4.3);  
**10**    • time-dependent analysis, in particular acceptance parameterisation (Sec. 5.2), decay-  
**11**    time fit (Sec. 5.3), and estimation of systematic uncertainties (Sec. 5.4).

**12** Moreover, I was responsible for:

- 13**    • calibration and performance studies of the LHCb Silicon Tracker (Sec. 2.2.1);  
**14**    • opposite-side electron tagger optimisation (Sec. 3.3).

## 15 1.2 The Standard Model of particle physics

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

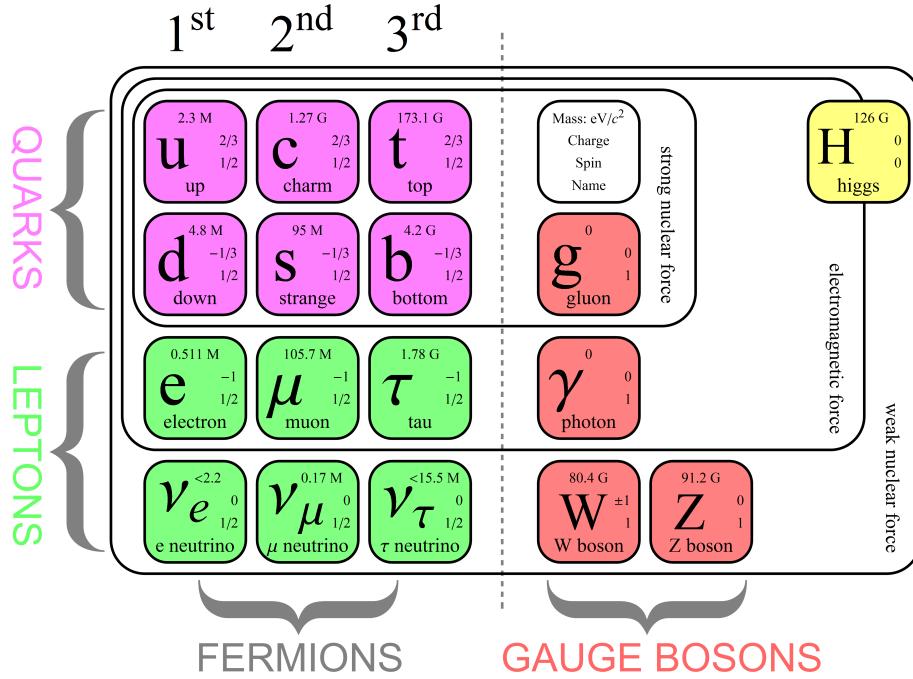


Figure 1.1 – Elementary particles described by the SM [4].

21 All particles are either *fermions* or *bosons*. The fermions (leptons, quarks) have half-  
 22 integer spin and follow the Fermi-Dirac statistics [5, 6], whereas Bosons (gauge bosons,  
 23 Higgs boson) have integer spin and follow the Bose-Einstein statistics [7].

24 Leptons (spin- $\frac{1}{2}$ ) include three charged <sup>1</sup>, massive particles (electron  $e^-$ , muon  $\mu^-$  and  
 25 tau  $\tau^-$ ), which interact via the electromagnetic and weak interactions, and three neutral,  
 26 (nearly) massless particles, called *neutrinos* ( $\nu_e$ ,  $\nu_\mu$  and  $\nu_\tau$ ), which only experience weak  
 27 interactions.

28 Six different types of quarks (spin- $\frac{1}{2}$ ) exist: the *up-type* quarks up ( $u$ ), charm ( $c$ ) and  
 29 top/truth ( $t$ ), having charge  $+\frac{2}{3}$ , and the *down-type* quarks down ( $d$ ), strange ( $s$ ), and  
 30 bottom/beauty ( $b$ ), which have charge  $-\frac{1}{3}$ . They can interact via electromagnetic, weak  
 31 and strong interactions, and they are all massive.

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

## Chapter 1. Introduction

---

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

<sup>61</sup> **1.3 The Cabibbo-Kobayashi-Maskawa matrix**

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

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

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  $b'$ :

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

<sup>62</sup> The CKM matrix is unitary ( $V_{\text{CKM}}^\dagger V_{\text{CKM}} = 1$ ), so it can be written in terms of four independent parameters, namely three angles and a *weak phase*  $\delta$ . The latter is the source of all *CP-violating* phenomena in the SM, i.e. asymmetries between particles and anti-particles; in fact, the *complexity* of  $V_{\text{CKM}}$  implies that the SM Lagrangian density is non *CP*-invariant, in agreement with the experimentally observed *CP* violation.

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

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

<sup>67</sup> 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:

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

## Chapter 1. Introduction

---

- 68 From Eq. 1.6, one can see that quark transitions within the same family (e.g.  $u \rightarrow d$ )  
 69 are more probable, whereas transitions between different families (e.g.  $b \rightarrow c$ ) are more  
 70 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. Having defined the following angles:

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

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

- 71 the triangles given by Eqs. 1.7 and 1.8 can be depicted as shown in Fig. 1.2. The first  
 72 triangle, defined by Eq. 1.7, is known as the *Unitarity Triangle* (UT) and its elements  
 73 can be measured from analyses of  $B^0$ ,  $B_s^0$  and  $B^\pm$  decays. The other triangle (Eq. 1.8)  
 74 can be studied from decays of  $B_s^0$  mesons. These measurements allow tests of the CKM  
 75 paradigm for  $CP$  violation in the SM. Any deviation from unitarity would be a signature  
 76 of new physics beyond the SM.

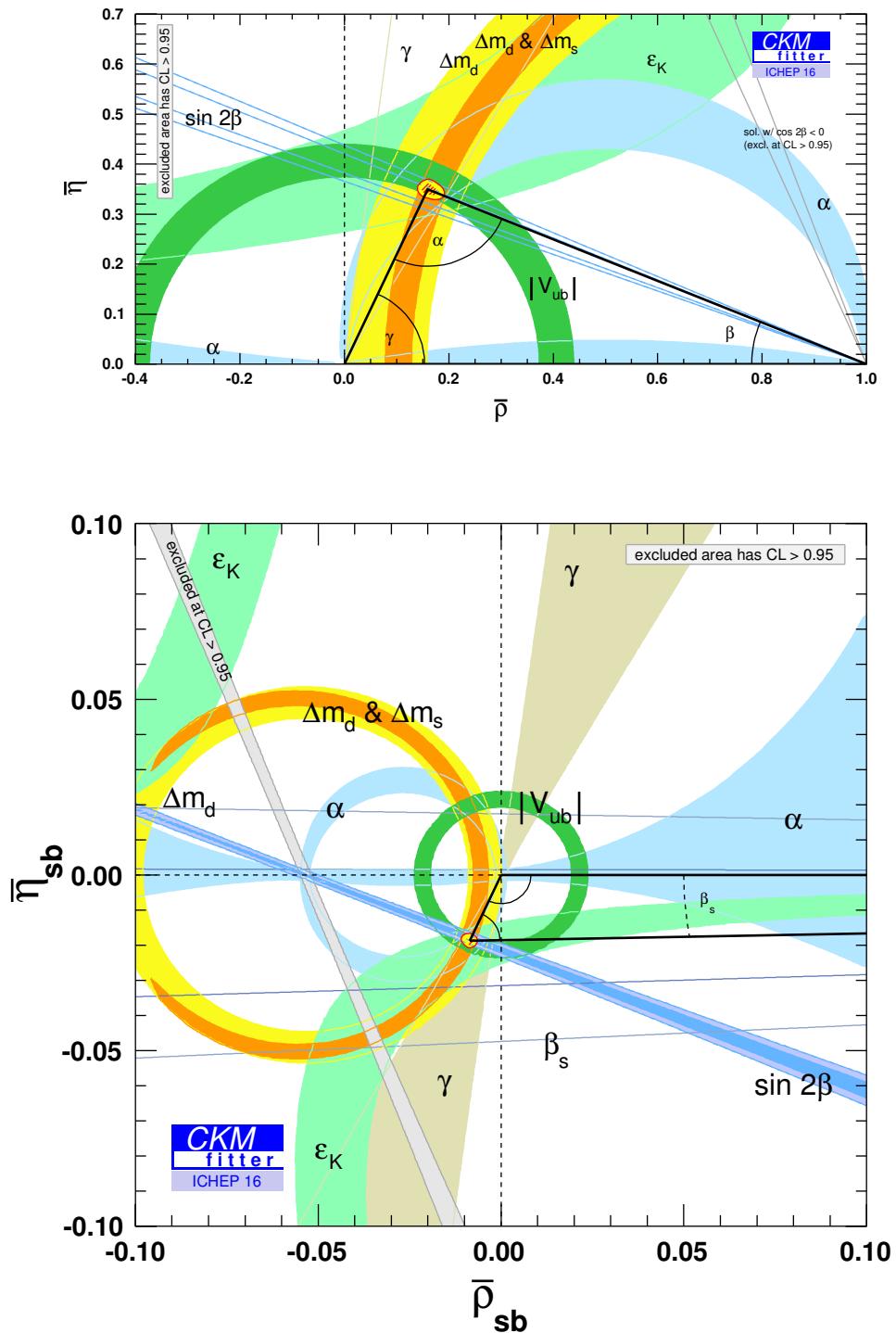


Figure 1.2 – Graphical representation of two of the six unitarity conditions of the CKM matrix, superimposed with the current experimental constraints [19].

<sup>77</sup> **1.4 Physics of the neutral  $B$  mesons**

<sup>78</sup> The theory of neutral  $B$  meson oscillation, decays and  $CP$  violation presented here is  
<sup>79</sup> derived from Refs. [20] and [21].

<sup>80</sup> **1.4.1 Oscillation of neutral mesons**

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

$$|B^0\rangle = |d\bar{b}\rangle, \quad |\bar{B}^0\rangle = |\bar{d}b\rangle, \quad (1.11)$$

$$|B_s^0\rangle = |s\bar{b}\rangle, \quad |\bar{B}_s^0\rangle = |\bar{s}b\rangle. \quad (1.12)$$

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 follows:

$$|\Psi(t)\rangle = a(t)|P^0\rangle + b(t)|\bar{P}^0\rangle + \sum_i c_i(t)|f_i\rangle, \quad (1.13)$$

<sup>81</sup> where  $|f_i\rangle$  are all the possible final state, and  $c(0) = 0$  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 Schroedinger equation for  $|\Psi(t)\rangle$  in terms of an effective, non-hermitian hamiltonian  $\mathcal{H}$ :

$$i\partial_t \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \mathcal{H} \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} a(t) \\ b(t) \end{pmatrix}. \quad (1.14)$$

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

$$\mathcal{H} = M - \frac{i}{2}\Gamma = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}. \quad (1.15)$$

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

eigenvalues of  $\mathcal{H}$  are:

$$\lambda_L = m_L - \frac{i}{2}\Gamma_L = H_0 + \sqrt{H_{12}H_{21}} = H_0 + \sqrt{\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\left(M_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}, \quad (1.16)$$

$$\lambda_H = m_H - \frac{i}{2}\Gamma_H = H_0 - \sqrt{H_{12}H_{21}} = H_0 - \sqrt{\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\left(M_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}. \quad (1.17)$$

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

$$|P_L\rangle = p|P^0\rangle + q|\bar{P}^0\rangle, \quad |P_H\rangle = p|P^0\rangle - q|\bar{P}^0\rangle \quad (1.18)$$

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

$$\frac{q}{p} = \sqrt{\frac{H_{21}}{H_{12}}} = \sqrt{\frac{M_{21} - \frac{i}{2}\Gamma_{21}}{M_{12} - \frac{i}{2}\Gamma_{12}}} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{21}^* - \frac{i}{2}\Gamma_{21}^*}}. \quad (1.19)$$

The relative phase  $\phi_M$  between  $M_{12}$  and  $\Gamma_{12}$  is an observable quantity describing indirect  $CP$  violation (Sec. 1.4.3):

$$M_{12} = M_{12}^* e^{i\phi_{CP}}, \quad \Gamma_{12} = \Gamma_{12}^* e^{i\phi_{CP}} e^{i\phi_M}. \quad (1.20)$$

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

$$\frac{q}{p} = -e^{-i\phi_M} \left[ 1 - \frac{1}{2} \left| \frac{\Gamma_{12}}{M_{12}} \right| \sin \phi_M + \mathcal{O} \left( \left| \frac{\Gamma_{12}}{M_{12}} \right|^2 \right) \right], \quad (1.21)$$

<sup>82</sup> which implies  $|q/p| \sim 1$ .

The difference and the average of the mass and width of the two mass eigenstates can be defined as:

$$\Delta m = m_H - m_L = \Re(\lambda_H - \lambda_L), \quad m = \frac{m_L + m_H}{2} = \frac{\Re(\lambda_H + \lambda_L)}{2}, \quad (1.22)$$

$$\Delta\Gamma = \Gamma_L - \Gamma_H = -2\Im(\lambda_L - \lambda_H), \quad \Gamma = \frac{\Gamma_L + \Gamma_H}{2} = -\frac{\Im(\lambda_H + \lambda_L)}{4}. \quad (1.23)$$

<sup>83</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:

$$|P^0(t)\rangle = g_+(t)|P^0(t)\rangle + \frac{q}{p}g_-(t)|\bar{P}^0(t)\rangle, \quad (1.24)$$

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

$$(1.26)$$

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  $|P^0(0)\rangle$  or  $|\bar{P}^0(0)\rangle$  becomes a  $|P^0(t)\rangle$  or  $|\bar{P}^0(t)\rangle$  are the following:

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

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

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

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

<sup>86</sup> The equations above describe the *oscillation* or *mixing* of the  $B^0$  and  $B_s^0$  mesons.

<sup>87</sup> **1.4.2 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  $\mathcal{H}$ :

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

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

After defining the following 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  $\bar{f}$ :

$$\begin{aligned} \frac{d\Gamma(P^0 \rightarrow f)}{dt}(t) = N_f |A_f|^2 \frac{1 + |\lambda_f|^2}{2} e^{-\Gamma t} \\ \left[ \cosh\left(\frac{\Delta\Gamma t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma t}{2}\right) + C_f \cos(\Delta m t) - S_f \sin(\Delta m t) \right], \end{aligned} \quad (1.35)$$

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

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

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

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

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

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

are known as *CP coefficients*.

<sup>89</sup> 1.4.3  $CP$  violation in neutral meson systems

<sup>90</sup> Three types of  $CP$  violation can occur. They are briefly sketched in Fig. 1.3 and described  
<sup>91</sup> in the following paragraphs.

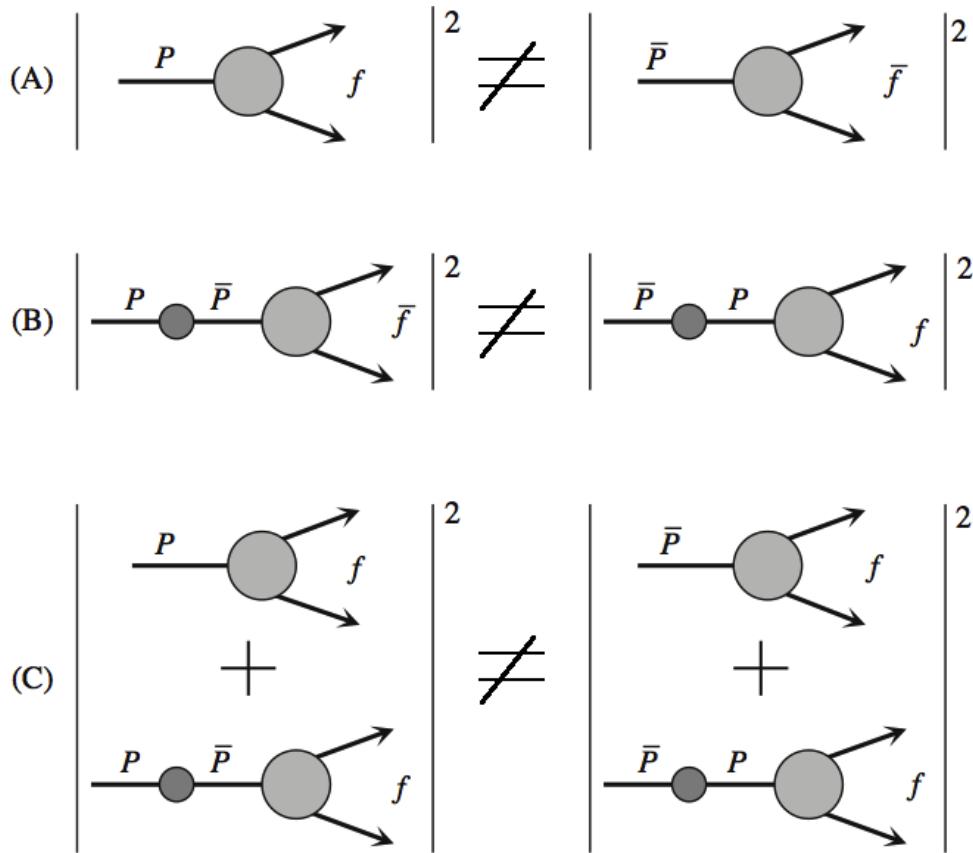


Figure 1.3 – Graphical representation of  $CP$  violation in decay (A), mixing (B) and interference between mixing and decay (C) [21].

<sup>92</sup>  $CP$  violation in decays

$CP$  violation in decays, also known as *direct CP* violation, happens when the decay amplitude for  $P \rightarrow f$  is different from 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 get the opposite sign. The following asymmetry between final states can be measured to determine direct  $CP$  violation experimentally for *charged mesons*, where mixing effects are absent:

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

### <sup>93</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 the  $CP$ -conjugated process  $P^0 \rightarrow \bar{P}^0$ . These two oscillation probabilities are given by Eq. 1.29 and Eq. 1.30 respectively. It turns out that they are identical unless

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

From Eq. 1.21, 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:

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

### <sup>94</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 a 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

## Chapter 1. Introduction

---

and decay can be accessed by studying the following asymmetries:

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

$$\mathcal{A}_{\bar{f}}(t) = \frac{\frac{d\Gamma(P^0 \rightarrow \bar{f})}{dt} - \frac{d\Gamma(\bar{P}^0 \rightarrow \bar{f})}{dt}}{\frac{d\Gamma(P^0 \rightarrow \bar{f})}{dt} + \frac{d\Gamma(\bar{P}^0 \rightarrow \bar{f})}{dt}} \quad (1.47)$$

A relevant example is the case of neutral  $B$  mesons, where  $|q/p| = 1$ . Using Eq. 1.35, 1.36, 1.37 and 1.38, the asymmetries 1.46 and 1.47 take the following forms:

$$\mathcal{A}_f(t) = \frac{C_f \cos(\Delta mt) - S_f \sin(\Delta mt)}{\cosh\left(\frac{\Delta\Gamma t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma t}{2}\right)}, \quad (1.48)$$

$$\mathcal{A}_{\bar{f}}(t) = \frac{-C_{\bar{f}} \cos(\Delta mt) + S_{\bar{f}} \sin(\Delta mt)}{\cosh\left(\frac{\Delta\Gamma t}{2}\right) + D_{\bar{f}} \sinh\left(\frac{\Delta\Gamma t}{2}\right)}. \quad (1.49)$$

<sup>95</sup> So, the  $CP$  coefficients can be directly measured from a time-dependent analysis of certain  
<sup>96</sup>  $B$  decays.

## 97 1.5 $B^0 \rightarrow D^\pm \pi^\mp$ analysis strategy

98 In this thesis, a decay-time dependent analysis of the decay  $B^0 \rightarrow D^\pm \pi^\mp$  is presented,  
 99 where the  $D^\pm$  meson is reconstructed as  $D^\pm \rightarrow K^\mp \pi^\pm \pi^\pm$ . The pion produced together  
 100 with the  $D^\pm$  meson will be named *bachelor* or *companion* hereafter. The objective of  
 101 this study is to perform a measurement of CP asymmetries, in order to constrain the  
 102 CKM angle  $\gamma$  [15, 16]. Measurements of  $\gamma$  via tree-level decays such as  $B^0 \rightarrow D^\pm \pi^\mp$   
 103 are theoretically clean [22, 23, 24]. CP violation appears in the interference between  
 104 the Cabibbo favoured  $b \rightarrow c$  amplitude without mixing,  $A(B^0 \rightarrow D^- \pi^+)$ , and the  
 105 Cabibbo suppressed  $b \rightarrow u$  amplitude with mixing,  $A(B^0 \rightarrow \bar{B}^0 \rightarrow D^- \pi^+)$ . Two of the  
 106 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  
 $\Delta\Gamma = 0$ , the time-dependent decay rates for  $B$  mesons initially produced as  $B^0$  or  $\bar{B}^0$  can  
 be written as follows:

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

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

$$\frac{d\Gamma(B^0 \rightarrow \bar{f})}{dt}(t) = \frac{1}{4\tau} e^{-t/\tau} [1 + C_{\bar{f}} \cos(\Delta mt) - S_{\bar{f}} \sin(\Delta mt)], \quad (1.52)$$

$$\frac{d\Gamma(\bar{B}^0 \rightarrow \bar{f})}{dt}(t) = \frac{1}{4\tau} e^{-t/\tau} [1 - C_{\bar{f}} \cos(\Delta mt) + S_{\bar{f}} \sin(\Delta mt)], \quad (1.53)$$

where  $\Delta m$  and  $\tau = 1/\Gamma$  are given by Eqs. 1.22 and 1.23. The CP coefficients  $C_{f/\bar{f}}$  and

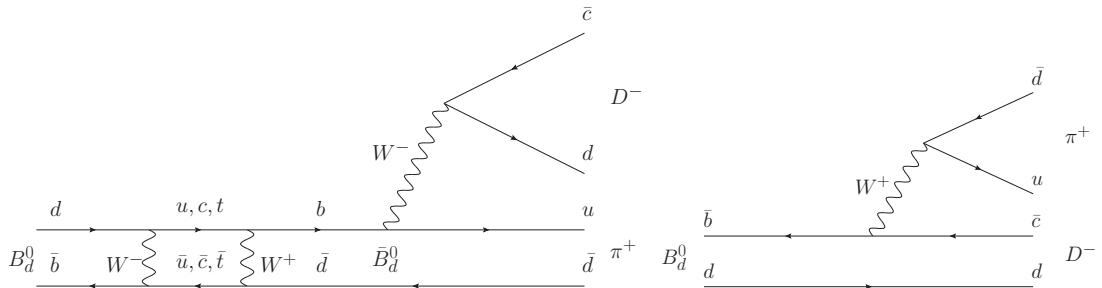


Figure 1.4 – Feynman diagrams contributing to  $B^0 \rightarrow D^- \pi^+$ , with (left) and without (right) mixing

## Chapter 1. Introduction

---

$S_{f/\bar{f}}$ , defined in Eqs. 1.39 and 1.40, can be expressed as

$$S_f = -\frac{2r_{D\pi} \sin[\delta - (\gamma + 2\beta)]}{1 + r_{D\pi}^2}, \quad S_{\bar{f}} = \frac{2r_{D\pi} \sin[\delta + (\gamma + 2\beta)]}{1 + r_{D\pi}^2}, \quad (1.54)$$

$$C_f = -C_{\bar{f}} = C = \frac{1 - r_{D\pi}^2}{1 + r_{D\pi}^2}, \quad (1.55)$$

where  $\beta$  (Eq. 1.9) is related to the  $B^0$  mixing phase,  $r_{D\pi} = |A(\bar{B}^0 \rightarrow D^- \pi^+)/A(B^0 \rightarrow D^- \pi^+)|$  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 [25]. An estimation of  $r_{D\pi}$  was performed by the BaBar and Belle collaborations [26, 27], 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. For this analysis,  $C_f$  ( $C_{\bar{f}}$ ) is simply fixed to +1 (-1) due to a lack of *sensitivity*.

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 expected signal yield (about  $5 \times 10^5$ ) is sufficiently large to compensate for this lack of sensitivity.

Measurements of  $\sin(2\beta + \gamma)$  in  $B^0 \rightarrow D^{(*)\pm} \pi^\mp$  were performed previously by the BaBar and Belle collaborations [28, 29, 30, 31]. A measurement of  $2\beta_s - \gamma$  was performed by LHCb in  $B_s^0 \rightarrow D_s^\mp K^\pm$  decays with  $3 \text{ fb}^{-1}$  of data [32].

The measurement presented in this thesis is performed in terms of a flavour-tagged, decay-time dependent analysis of the Run 1 dataset (Sec. 2.3.3). 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 data, 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 distribution of the resulting dataset is performed to extract sWeights for the signal component. The fit is described in detail in Sec. 4.3. The training and calibration of the flavour-tagging algorithms, which infer the initial flavour of the reconstructed  $B^0$  candidates, is summarised in Sec. 3.1. The final estimation of the  $CP$  coefficients is the result of an unbinned, sWeighted likelihood fit to the distributions of the decay time and the flavour tagging observables.

## 2 The Large Hadron Collider and the <sup>139</sup> LHCb experiment

**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 and made up with  $10^{11}$  protons, 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. 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

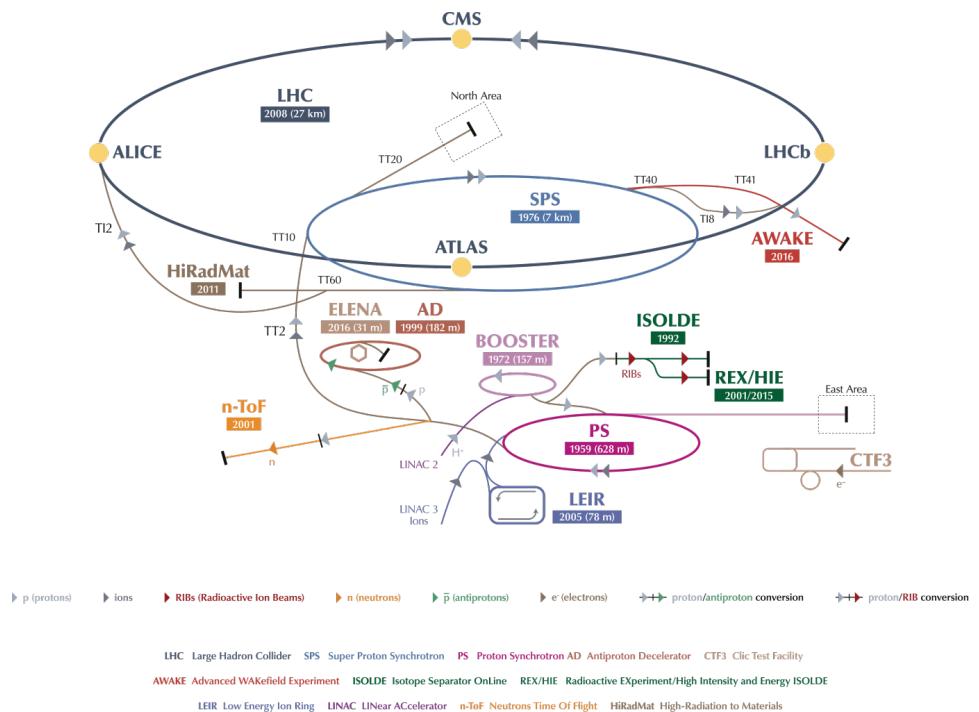


Figure 2.1 – Overview of the CERN accelerators complex.

<sup>160</sup> **2.2 The LHCb experiment**

<sup>161</sup> The *Large Hadron Collider beauty* (LHCb) experiment [?] is a single-arm forward spec-  
<sup>162</sup> trometer (see Fig. 2.2) that exploits the forward production of the  $b$ - and  $c$ - quarks pairs  
<sup>163</sup> from  $pp$  collisions (Fig. 2.3). The LHCb angular coverage is comprised between 15 mrad  
<sup>164</sup> and 250 (300) mrad in the vertical (horizontal) plane. The LHCb coordinate system  
<sup>165</sup> consists of an orthogonal set of coordinates,  $x$ ,  $y$ ,  $z$ , where  $z$  is aligned with the beam  
<sup>166</sup> pipe,  $y$  is perpendicular to the LHCb cavern ground and  $x$  is orthogonal to the other two.

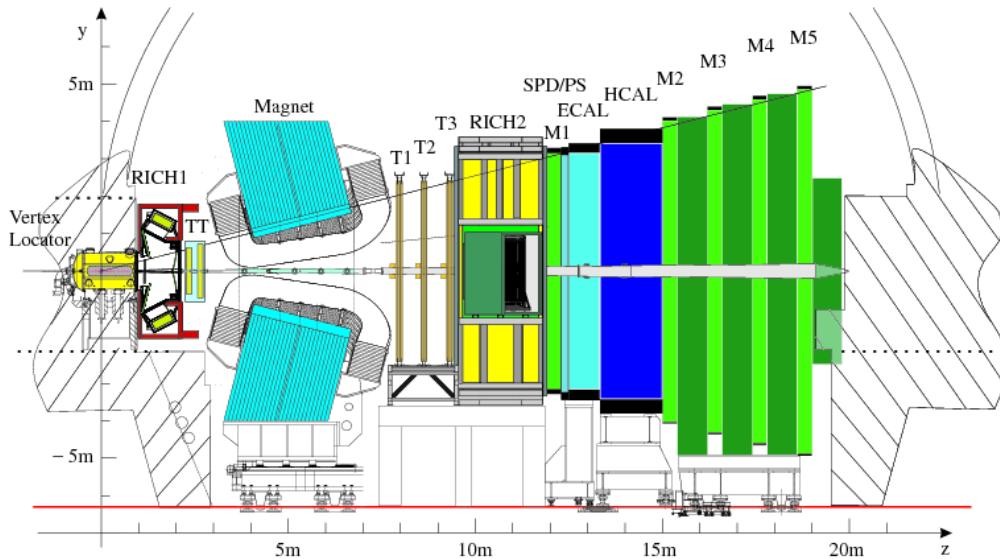


Figure 2.2 – Side view of the LHCb detector.

<sup>167</sup> The LHCb experiment is composed of different sub-detectors. The tracking system  
<sup>168</sup> includes a vertex and tracking detector called *VErtex LOcator* (VELO), the *Tracker*  
<sup>169</sup> *Turicensis* (TT), located upstream a magnetic dipole with an integrated field of 4 Tm, the  
<sup>170</sup> *Inner Tracker* (IT), situated downstream the magnet in three separated stations around  
<sup>171</sup> the beryllium beam-pipe, and the *Outer Tracker* (OT), installed in the same stations as  
<sup>172</sup> the IT. The *Particle IDentification* (PID) system comprises two *Ring Imaging CHerenkov*  
<sup>173</sup> detectors (RICH), an *Electromagnetic CALorimeter* (ECAL), which also includes a *Pre-*  
<sup>174</sup> *Shower* (PS) and *Scintillator Pad Detector* (SPD), a *Hadronic CALorimeter* (HCAL)  
<sup>175</sup> and five *muon stations* (M1-M5).

<sup>176</sup> **2.2.1 The tracking system**

<sup>177</sup> **The VErtox LOcator (VELO)**

<sup>178</sup> The VELO [33] (Fig. 2.4) is a silicon micro-strip detector for charged particles surround-  
<sup>179</sup> ing the interaction point, which identifies decay vertices and performs the first track

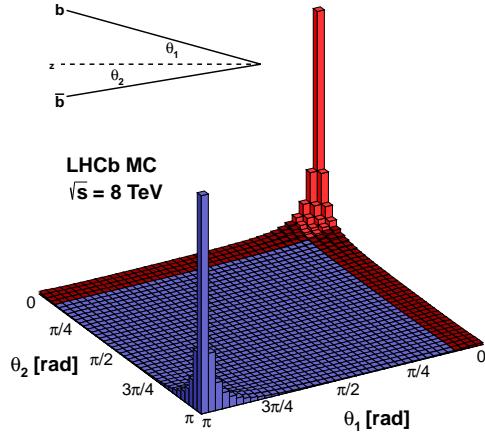


Figure 2.3 – Distribution of the  $b$  and  $\bar{b}$  pair simulated production angles with respect to the beam direction at  $\sqrt{s} = 8 \text{ TeV}$ .

reconstruction step. 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, which provide an excellent measurement of vertex positions and impact parameters (IP), 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 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 IP resolution of a track is measured to be  $\sigma_{\text{IP}} = 11.6 \pm 23.4/p_{\text{T}}$   $\mu\text{m}$  in  $x$  and  $\sigma_{\text{IP}} = 11.2 \pm 23.2/p_{\text{T}}$   $\mu\text{m}$  in  $y$ , where  $p_{\text{T}}$  is the *transverse momentum* (in  $\text{GeV}/c$ ) of the particle with respect to the beam axis.

### 193 The Tracker Turicensis (TT)

194 The TT [34] (Fig. 2.5) is a silicon micro-strip detector covering a total area of about  
 195  $7.9 \text{ m}^2$  upstream the magnet and divided into two separate stations (TTa, TTb). Each  
 196 station has two layers. The TT helps in improving the track momentum resolution and  
 197 detecting long-lived particles that decay outside the VELO acceptance. TTa is composed  
 198 by X and U layers, while TTb includes V and X layers. The X layers have read-out strips  
 199 aligned vertically, whereas the U and V *stereo* layers are rotated by  $+5^\circ$  and  $-5^\circ$  with  
 200 respect to the vertical in the  $xy$  plane.

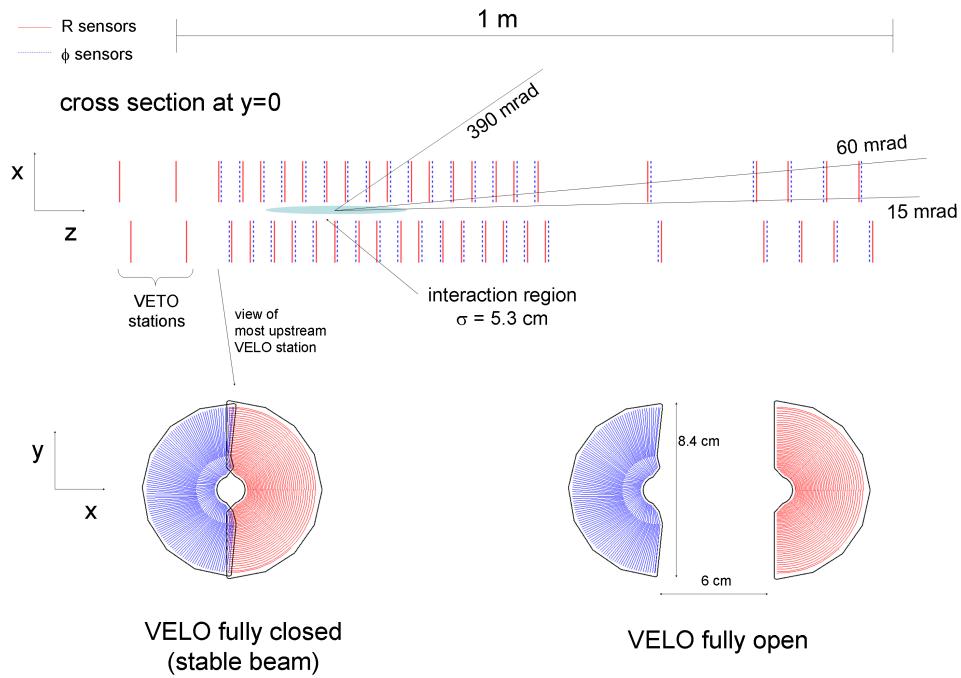


Figure 2.4 – Schematic overview of the VELO detector and its sensors.

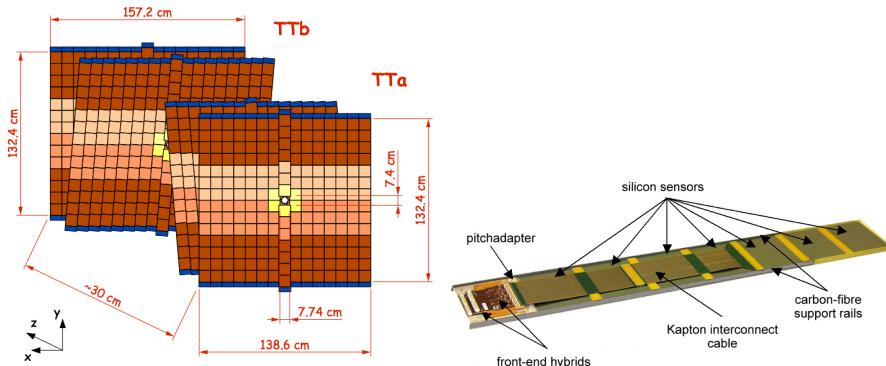


Figure 2.5 – Schematic overview of the TT stations/layers and one of the TT read out modules.

- 201 The TT active area is made of p-on-n silicon micro-strip sensors. Since the sensors are  
 202 exposed to a significant radiation due to a high track multiplicity, they are cooled to 0°  
 203 in order to minimise the damage.
- 204 The TT read-out module contain from one to four sensors in a row, resulting in read-out  
 205 strips up to 37 cm long. The strip pitch is 183  $\mu\text{m}$  and the sensor thickness is 500  $\mu\text{m}$ .  
 206 The hit resolution is about 50  $\mu\text{m}$ .

207 **The Inner Tracker (IT)**

208 The IT [35] (Fig. 2.6) is also a silicon micro-strip detector. Together with the TT, it forms  
 209 the *Silicon Tracker* (ST). It is dedicated to detect charged particles in the high track  
 210 density region around the beam-pipe downstream the magnet. It is separated into three  
 211 stations, where each station consists of four boxes. Each box has four layers made of seven  
 212 read-out modules arranged in a X-U-V-X layout similar to that of the TT. The total  
 213 coverage of the IT is about 4.2 m<sup>2</sup>. The boxes directly above and below the beam-pipe  
 214 are made of single-sensor modules, called *short modules*, whereas the side boxes are made  
 215 of two bonded silicon sensor modules, called *long modules*. The IT strip pitch is 198  $\mu\text{m}$ ,  
 216 while the p-on-n sensor thickness is 320 (410)  $\mu\text{m}$  for the short (long) modules. The hit  
 217 resolution is about 50  $\mu\text{m}$ .

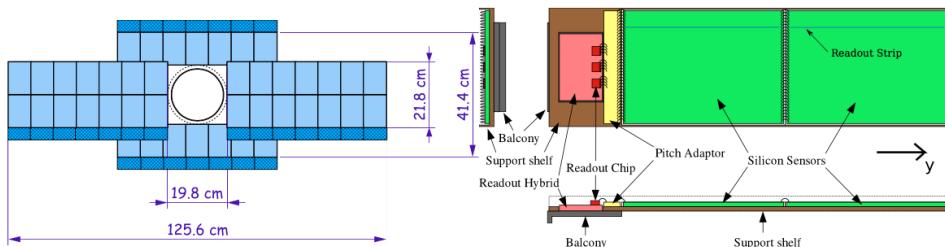


Figure 2.6 – Schematic overview of an IT station and one of the long IT read out modules.

218 **The Outer Tracker (OT)**

219 The OT [36] (Fig. 2.7) is a gaseous straw-tube detector filled with an Ar/CO<sub>2</sub>/O<sub>2</sub>  
 220 (70%/28.5%/1.5%) gas mixture. It is dedicated to the detection of charged particles in  
 221 the low track density region outside the IT acceptance and covers a large area of about  
 222 340 m<sup>2</sup>. The OT is composed of three stations, where each station has four layers in  
 223 a X-U-V-X configuration. Each station is separated physically for the left and right  
 224 sides with respect to the beam-pipe into two C-frames. Each layer is divided into two  
 225 mono-layers. The OT has different types of modules, the long F modules and the S1, S2,  
 226 S3 modules that are cut in two pieces to leave space for the IT. The straw tube and anode

227 wire diameters are 5 mm and 25  $\mu\text{m}$  respectively. The hit resolution is about 200  $\mu\text{m}$ .

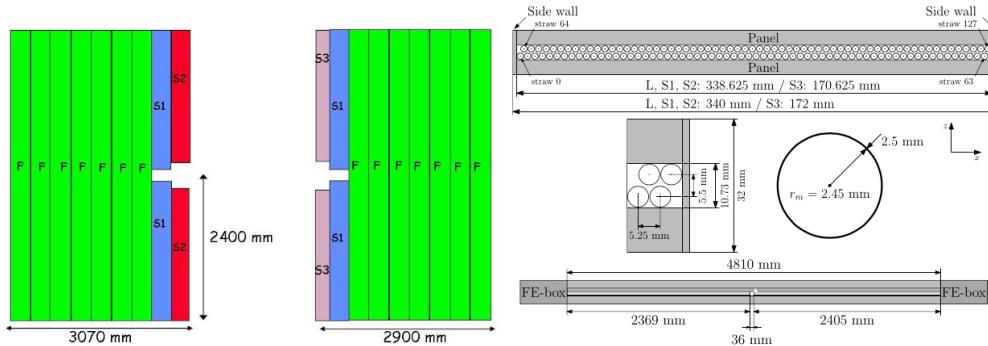


Figure 2.7 – Schematic overview of an OT layer and an OT module layout.

### 228 Spillover noise in the Silicon Tracker

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

238 In the first part of my PhD activity, I studied the effect of spillover clusters in the ST  
 239 using both simulated events and real collision data. This study showed that this time  
 240 spacing has little impact on the detector *occupancy*, and that the increase of ghost tracks  
 241 is negligible. Moreover, in the same study it was shown that the charge deposited by  
 242 particles in the detector can be exploited as a *feature* in multivariate analyses in order  
 243 to further reduce the ghost track contamination. These results were documented in an  
 244 internal note [37].

## 245 2.3 Particle Identification (PID)

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

247 When a charged particle is travelling faster than the speed of light in a medium, Cherenkov  
 248 light is produced at an angle that depends on the velocity of the particle and the refractive  
 249 index of the medium (*radiator*). By knowing the momentum from the tracker and the  
 250 velocity from the RICH detectors, the mass can be determined and therefore provide

## 2.3. Particle Identification (PID)

251 particle identification. Two RICH detectors [38] (Fig. 2.8) are used in order to provide  
 252 PID in different momentum ranges.

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

258 The average kaon identification efficiency in the momentum range [2, 100] GeV/c is  $\sim 95\%$ .  
 259 The average probability that pions are wrongly identified as kaons is  $\sim 5\%$  in the same  
 260 momentum range.

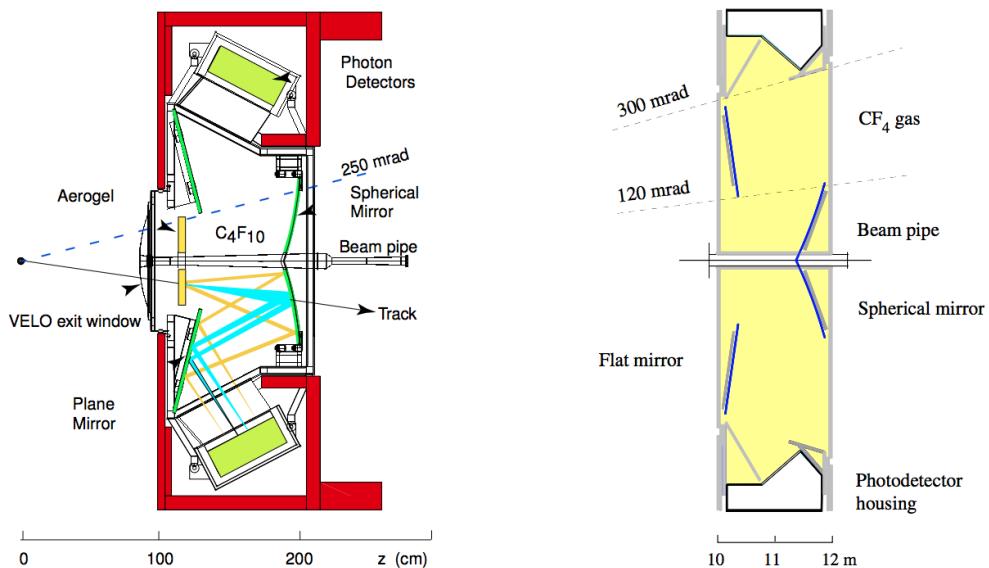


Figure 2.8 – Side view of RICH1 (left) and top view of RICH2 (right).

261 RICH2 is optimised for the momentum range from 15 to 100 GeV/c. The angular accep-  
 262 tance ranges from 15 mrad to 100 (120) mrad in the vertical (horizontal) plane where  
 263 most of the high-momentum tracks are produced. RICH2 uses tetrafluoromethane (CF<sub>4</sub>)  
 264 as radiator.

## 265 The Electromagnetic CALorimeter (ECAL), Pre-Shower (PS) and Scintillator 266 Pad Detector (SPD)

267 The ECAL [39] is used for the detection and measurement of the energy of electrons and  
 268 photons. The ECAL is built as a sandwich of alternating scintillators and lead layers in  
 269 the  $xy$  plane. Scintillation light produced by the shower of particles generated by the  
 270 lead plates is read out by Wave-Length Shifter (WLS) fibres coupled to PhotoMultiplier

271 Tubes (PMTs). The SPD is installed upstream the ECAL to separate electrons from  
272 photons. The PS is installed between the SPD and the ECAL. Both SPD and PS use  
273 scintillator pads read out by WLS fibres coupled to MultiAnode PhotoMultiplier Tubes  
274 (MAPMT). The acceptance range of the ECAL is from 25 mrad up to 300 (250) mrad in  
275 the horizontal (vertical) plane. The relative energy resolution of the ECAL is given by  
276  $\sigma_E/E = 10\%/\sqrt{E} \oplus 1\%$ , where  $E$  is given in GeV.

### 277 The Hadronic CALorimeter (HCAL)

278 The HCAL [39] is used for the detection and measurement of the energy of hadrons (such  
279 as pions and kaons) for the first level trigger. A HCAL cell is a sampling device made  
280 of alternating iron and scintillator tiles, where the latter are located along the beam  
281 direction. The HCAL has the same acceptance coverage of the ECAL. The relative energy  
282 resolution of the HCAL is given by  $\sigma_E/E = (69 \pm 5)\%/\sqrt{E} \oplus (9 \pm 2)\%$ , where  $E$  is given  
283 in GeV.

### 284 Muon detector

285 The muon system [40] (Fig. 2.9) is a gaseous detector composed of five stations (M1  
286 to M5) interleaved by 80 cm thick iron filters. The gaseous detectors are Multi-Wire  
287 Proportional Chambers (MWPC), except for the innermost part of M1, where triple GEM  
288 detectors are used to cope with the higher track density. The angular acceptance ranges  
289 from 20 (16) to 308 (256) mrad in the horizontal (vertical) plane. The muon detector has  
290 1380 chambers and covers a total area of 435 m<sup>2</sup>. Each muon chamber is composed of  
291 four layers of MPWC, except for M1, where two layers are used. The hit efficiency of the  
292 chambers is higher than 99% and allows to achieve a trigger efficiency greater than 95%  
293 for muons. The adopted gas mixture (Ar/CO<sub>2</sub>/CF<sub>4</sub> 40/55/5%) allows a fast triggering  
294 on muons (40 MHz).

#### 295 2.3.1 The trigger system

296 The bunch crossing rate of LHC is very high (40 MHz) because more than 99% of the  
297  $pp$  collisions do not produce interesting events. It's not possible to record events with  
298 such a high rate: therefore, a trigger system [41] is required to reduce the rate from  
299 40 MHz down to a few kHz. The rate reduction is achieved via selection criteria which  
300 ensure that events containing heavy flavour decays are stored. The signatures of these  
301 interesting decays include high  $p_T$  and transverse energy of the decay products, as well  
302 as displaced decay vertices (with large IP) due to the relative long lifetimes of  $b$ - and  
303  $c$ -hadrons. The trigger is divided into two sequential stages: a hardware stage called  
304 Level-0 (L0), and a software stage called the High Level Trigger (HLT). Different trigger  
305 decisions are separated into various *lines*, each of which provides information on different

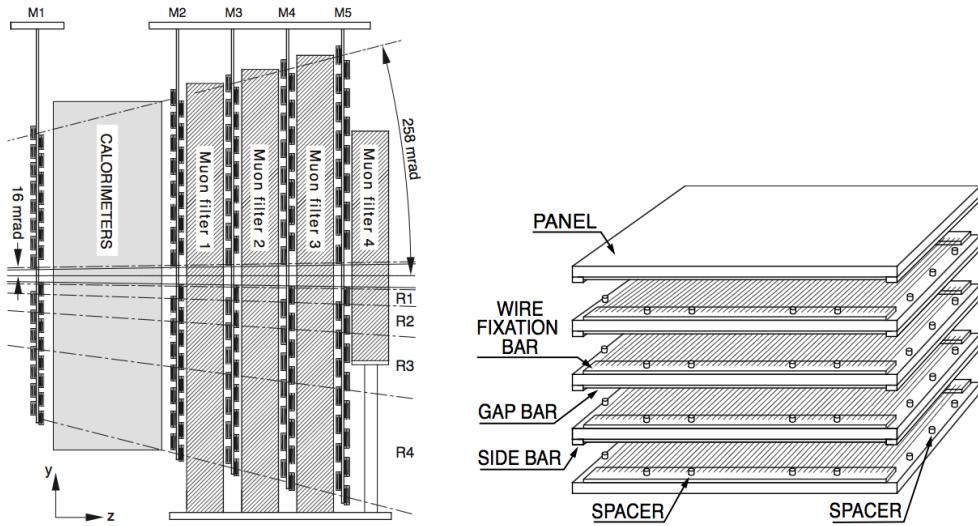


Figure 2.9 – Schematic overview of the muon system and a MWPC.

306 physics processes (e.g. decay topology, presence of muons etc. . . ). All the trigger steps  
 307 are summarised in Fig. 2.10.

308 Two types of trigger response are assigned offline, when some physics channel is analysed.  
 309 The TOS (*Trigger On Signal*) trigger occurs when the presence of the signal is sufficient  
 310 to have a positive trigger decision. The TIS (*Trigger Independent of Signal*) trigger occurs  
 311 when the "rest" of the event is sufficient to have a positive trigger decision, where the  
 312 "rest" of the event is defined after removing signal tracks and hits associated to them.

313 After the trigger stage, the data go through further offline selection steps, where exclusive  
 314 (e.g.  $B^0 \rightarrow D^- \pi^+$ ,  $B^+ \rightarrow J/\psi K^+$ ) and inclusive (e.g.  $J/\psi \rightarrow \mu^+ \mu^-$ ) decays are  
 315 reconstructed. This offline selection step is known as *stripping*, and each set of selection  
 316 requirement is called *stripping line*.

#### 317 Level-0 (L0)

318 The L0 trigger mainly exploits the calorimeters and muon chambers. The idea behind the  
 319 L0 trigger is to select events that contain high  $p_T$  muons and high  $E_T$  hadrons, electrons  
 320 and photons, which very likely come from  $b$ - and  $c$ -hadron decays. The L0 trigger reduces  
 321 the data rate from 40 MHz down to 1 MHz.

#### 322 High Level Trigger (HLT)

323 The HLT is separated into two stages, HLT1 and HLT2, and runs on about 29000  
 324 commercial CPU cores.

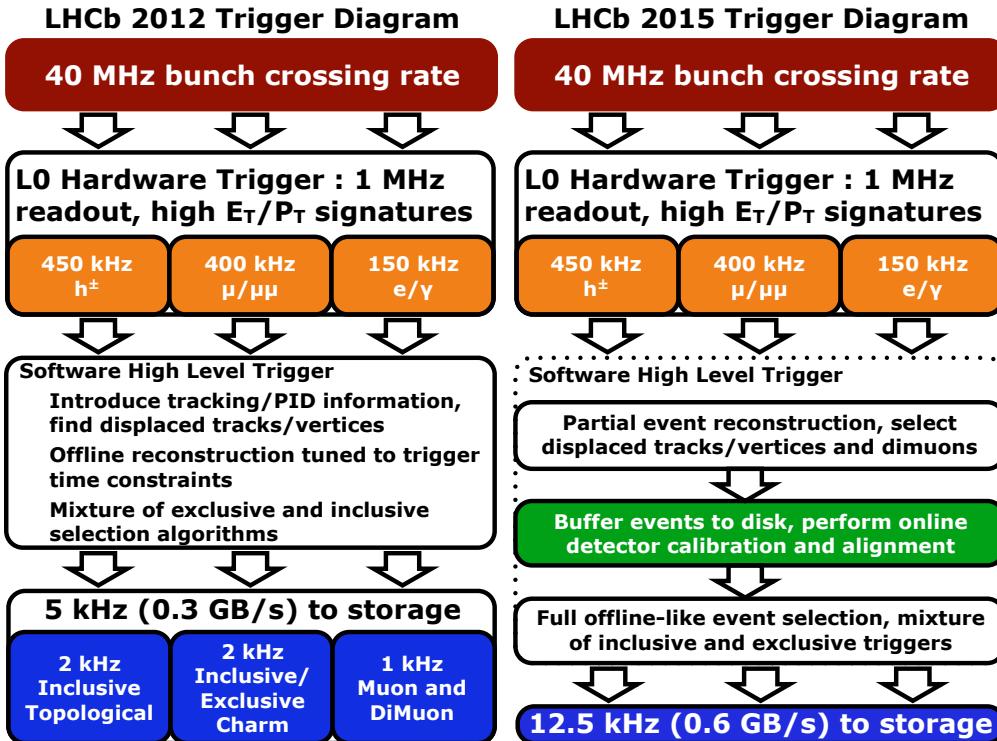


Figure 2.10 – Summary of the trigger strategy followed during the Run 1 (2011-2012, left) and Run 2 (2015-2017, right) data taking periods. During Run 2, an online detector and calibration alignment was introduced, plus full event selections (both inclusive and exclusive).

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

<sup>336</sup> **2.3.2 Event reconstruction, simulation and software**

<sup>337</sup> **Track and vertex reconstruction**

<sup>338</sup> Starting from the *hits* in the tracking detectors, tracks and vertices are reconstructed  
<sup>339</sup> via dedicated algorithms. Different track types are reconstructed, as shown in Fig. 2.11.  
<sup>340</sup> Each track is characterised by hits collected in different sub-detectors. For example,  
<sup>341</sup> downstream tracks, with no hits in the VELO, are typically associated to long-lived  
<sup>342</sup> particles such as  $\Lambda$  and  $K_S^0$ . Because of the presence of a magnetic field, tracks are bend  
<sup>343</sup> in the  $xz$  plane. By knowing the reconstructed particle trajectory and the magnetic field  
<sup>344</sup> map, it is possible to measure the momentum of the particle.

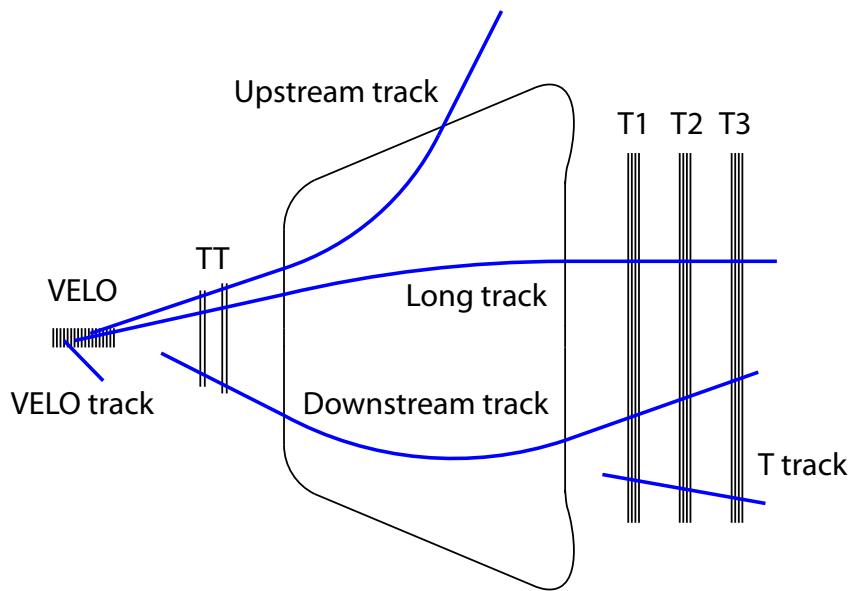


Figure 2.11 – Schematic description of the different track types reconstructed in LHCb.

<sup>345</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:

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

- 346 In typical LHCb analyses, requirements on the PID observables are applied to suppress  
 347 physical backgrounds due to wrongly-identified particles.

348 **Simulation**

- 349 The Monte Carlo (MC) simulation of  $pp$  collision, particle decays, and interaction with  
 350 the detector are crucial in the validation of physics analyses. The parton-parton collision  
 351 and hadronisation simulation is performed by PYTHIA [42], interfaced to EvtGen [43] for  
 352 the decay of the hadrons and leptons for standard productions. The QED corrections to  
 353 the decay (i.e. the emission of radiation photons) is generated by the PHOTOS package [44].  
 354 The interactions of particles with detector material and their tracking in the magnetic  
 355 field are simulated by GEANT4 [45, 46].

356 **2.3.3 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:

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

- 357 For a given data taking interval, the *integrated luminosity*  $L$  [ $\text{cm}^{-2}$ ] is a measure of  
 358 the amount of recorded data. The typical unit for luminosity is the inverse *barn*, which  
 359 corresponds to  $10^{24} \text{ cm}^{-2}$ . The LHCb integrated luminosity are of the order of the inverse  
 360 *femtobarn* ( $\text{fb}^{-1}$ ); one inverse femtobarn corresponds to the production of about  $10^{11} b\bar{b}$   
 361 quark pairs.

- 362 The LHCb detector collected data produced mainly from  $pp$  collisions in the 2010–2017  
 363 period, so far. During the 2011–2012 data taking period, called *Run 1*, about  $3 \text{ fb}^{-1}$  of  
 364 data were collected. The centre-of-mass energy  $\sqrt{s}$  of the  $pp$  interactions was 7 TeV and  
 365 8 TeV in 2011 and 2012 respectively, and the time spacing between *bunches* of protons  
 366 in the LHC was 50 ns. The 2013–2014 period, known as *Long Shutdown 1* (LS1), was  
 367 dedicated to some maintenance work for the LHC components. The *Run 2* data taking  
 368 period started in 2015, and it is planned to last until the end of 2018. The centre-of-mass  
 369 energy of the  $pp$  collisions during Run 2 is 13 TeV, whereas the time spacing between  
 370 proton bunches is 25 ns. In the 2015–2017 data taking period, about  $3.7 \text{ fb}^{-1}$  of data  
 371 were collected. The data collected for each year is summarised in Fig. 2.12.

### 2.3. Particle Identification (PID)

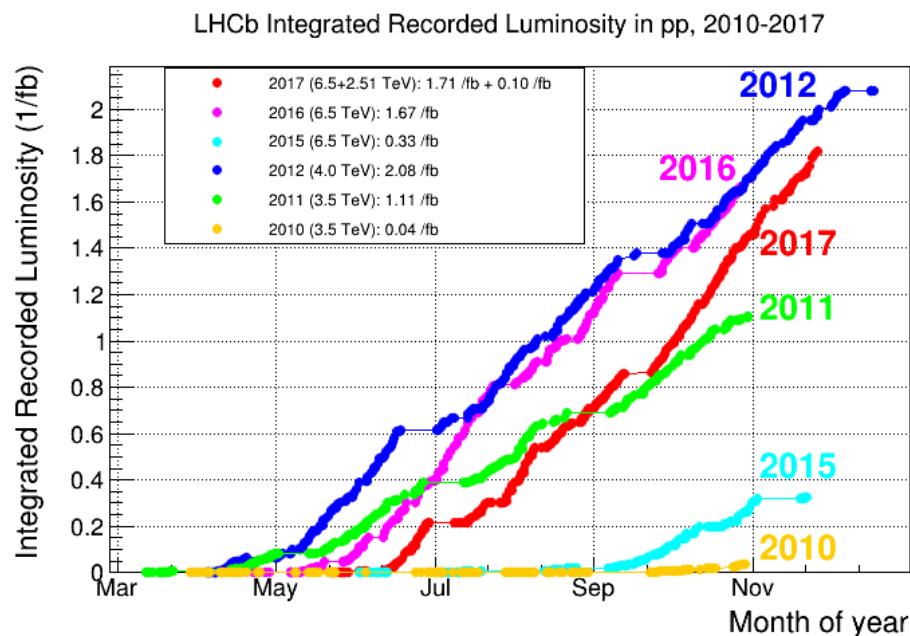


Figure 2.12 – Integrated luminosity of  $pp$  collisions collected each year by LHCb.



# 3 Flavour tagging

373

<sup>374</sup> **3.1 Flavour tagging algorithms**

<sup>375</sup> In this chapter, a description of the *flavour tagging* techniques at LHCb is reported.  
<sup>376</sup> After a brief introduction to the methods, the calibration of the *Opposite Side* (OS) and  
<sup>377</sup> *Same Side* (SS) algorithms for the  $B^0 \rightarrow D^\pm \pi^\mp$  time-dependent analysis are described.  
<sup>378</sup> Finally, a reoptimisation of the OS *electron* (OSe) tagger on both Run 1 and Run 2  
<sup>379</sup> data is reported. This work was made in collaboration with the University of Dortmund.  
<sup>380</sup> During my PhD work activity, I was mainly focused on the OS calibration and OSe  
<sup>381</sup> reoptimisation.

<sup>382</sup> The identification of the flavour at production time of a neutral  $B$  meson, *i.e.* whether it  
<sup>383</sup> contained a  $b$  or a  $\bar{b}$  quark at production, is a key element for any time-dependent analysis  
<sup>384</sup> aiming at the measurement of oscillations and  $CP$  asymmetries, and in particular for the  
<sup>385</sup>  $B^0 \rightarrow D^\pm \pi^\mp$  analysis reported in this thesis. In fact, this information is needed in order  
<sup>386</sup> to measure the decay rates or asymmetries introduced in Sec. 1.4. Complications arise  
<sup>387</sup> from two facts:

- <sup>388</sup> • neutral  $B$  meson oscillate, so the flavour at the production time might differ from  
<sup>389</sup> the flavour at the decay time;
- <sup>390</sup> • many final states, such as  $D^\pm \pi^\mp$ , are not  $CP$  eigenstates. This means that both  $B$   
<sup>391</sup> and  $\bar{B}$  mesons can decay to the same final state particles, and so the flavour cannot  
<sup>392</sup> be obtained from the charge of these particles, even if there were no oscillations.

<sup>393</sup> For these reasons, the flavour has to be reconstructed by exploiting particles not produced  
<sup>394</sup> in the decay of the neutral  $B$  meson being analysed.

<sup>395</sup> Techniques to infer the initial flavour of a reconstructed candidate are called flavour  
<sup>396</sup> tagging algorithms. Several flavour tagging algorithms exist in LHCb; they can be  
<sup>397</sup> categorised into same side taggers (SS taggers) and opposite side taggers (OS taggers). A  
<sup>398</sup> schematic representation of the taggers that can be used for tagging  $B^0$  mesons is shown  
<sup>399</sup> in Fig. 3.1.

<sup>400</sup> The SS taggers infer the production flavour of the signal  $B$  meson by selecting charged  
<sup>401</sup> particle candidates that have a high chance of being remnants of the hadronisation process  
<sup>402</sup> of the  $B$  candidate [47, 48]. For  $B^0$  mesons, the same-side pion tagger (SS $\pi$ ), which  
<sup>403</sup> exploits  $\pi^\mp$  mesons produced in the hadronisation of the  $B^0$  meson, and the same-side  
<sup>404</sup> proton tagger (SS $p$ ), which looks for co-produced protons, have been developed. For both  
<sup>405</sup> taggers, the charge of the pion or proton is correlated with the production flavour of  
<sup>406</sup> the signal  $B^0$  meson. The response of the two taggers is combined into a common SS  
<sup>407</sup> response.

<sup>408</sup> In contrast, the OS taggers exploit the predominant production process of  $B$  mesons  
<sup>409</sup> via  $b\bar{b}$  quark pair production [49]. They partially reconstruct the decay of the *other*

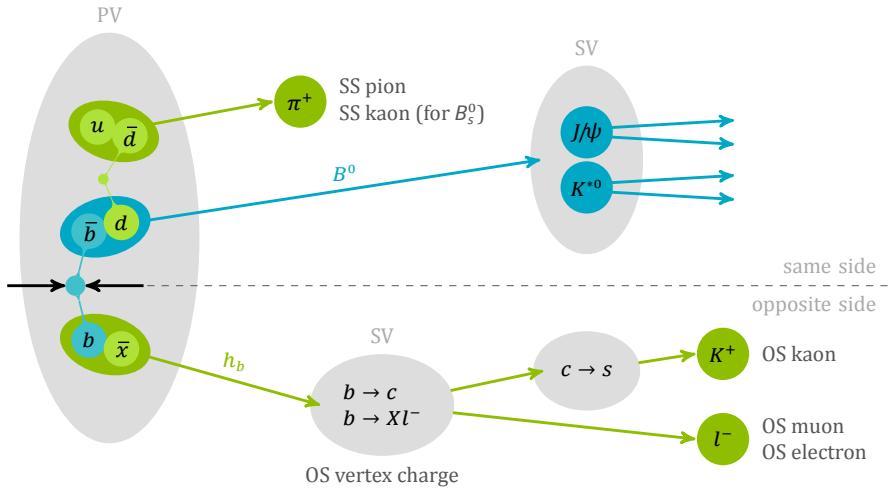


Figure 3.1 – Flavour tagging algorithms used in LHCb. In this cartoon, the signal channel is considered to be  $B^0 \rightarrow J/\psi K^{*0}$ .

410  $b$ -hadron produced along with each reconstructed signal  $B$  meson, and infer its initial  
 411 flavour. In fact, the flavour of the signal  $B$  meson and the other  $b$ -hadron produced  
 412 in the same collision are opposite. Several OS taggers have been developed in LHCb,  
 413 where the combination of the OS kaon (OS $K$ ), muon (OS $\mu$ ), electron (OS $e$ ), and vertex  
 414 charge (OSVtx) tagging algorithms represents the current standard OS combination. An  
 415 additional OS tagger, the OS Charm tagger (OSCharm) [50], can be exploited, and can  
 416 be combined with the OS standard combination.

417 Given a reconstructed candidate, each flavour tagging algorithm provides a flavour tag  $d$   
 418 and a prediction  $\eta$  for the probability of the tag to be wrong. This mistag probability  $\eta$  is  
 419 defined in the range  $[0, 0.5]$  and is based on the output of multivariate classifiers, which  
 420 are trained on datasets of flavour specific decays, and combine several kinematic and  
 421 geometric information on the tagging particle(s) and the event. The flavour tag takes the  
 422 values  $d = +1$  for an initial  $B^0$ ,  $d = -1$  for an initial  $\bar{B}^0$ , and  $d = 0$  when no tag could  
 423 be assigned; this happens, for example, if the tagging particle do not pass the selection  
 424 criteria of a given tagging algorithm, or if its trajectory lies outside the LHCb acceptance.

425 More details on flavour tagging at LHCb can be found in Refs. [51, 52, 53].

426 **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 (correctly) 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  $CP$ -conjugated 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:

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

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 occur. 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:

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

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

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

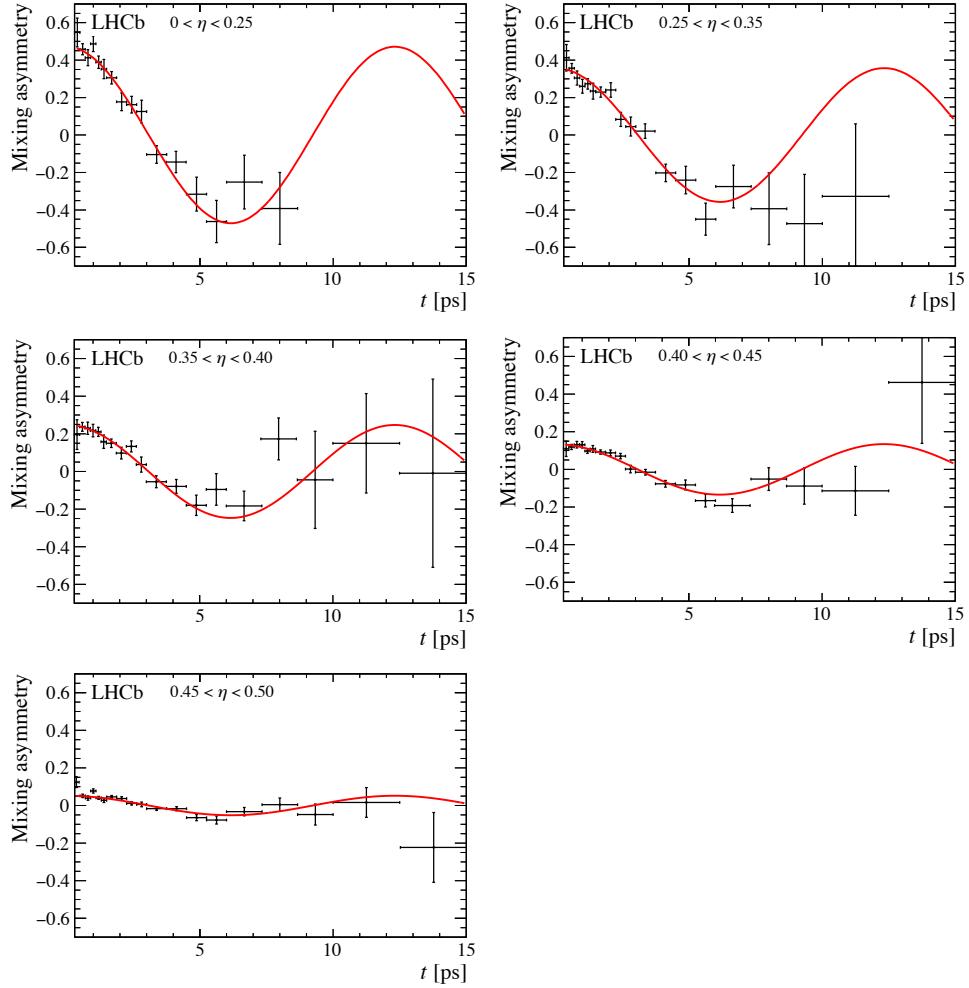


Figure 3.2 – Mixing asymmetry for SS pion tagged  $B_s^0 \rightarrow D_s^{\mp} \pi^{\pm}$  candidates in bins of increasing estimated mistag  $\eta$  [48].

<sup>434</sup> **Calibration of the tagging output**

<sup>435</sup> The output of the flavour tagging algorithms is the result of training multivariate classifiers  
<sup>436</sup> (MVA) using datasets of flavour specific  $B$  decays, and transforming the classifier output  
<sup>437</sup> into mistag estimates  $\eta$  through regression. However, as the training and validation  
<sup>438</sup> samples are different from the signal sample used in the  $CP$  measurement (*e.g.* in terms  
<sup>439</sup> of trigger and selection criteria that affect the distribution of the MVA input features),  
<sup>440</sup> the output needs to be calibrated. Again, using control samples of flavour specific decays,  
<sup>441</sup> calibration functions  $\omega(\eta)$  are obtained to transform the mistag estimate  $\eta$  of the algorithm  
<sup>442</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>443</sup> The use of the arithmetic mean  $\langle \eta \rangle$  of the  $\eta$  distribution aims at a decorrelation of  $p_0$   
<sup>444</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

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

<sup>445</sup> In this thesis, new models for the calibration functions are adopted instead of the standard  
<sup>446</sup> linear calibrations. These different parameterisations are called *Generalised Linear Models*  
<sup>447</sup> (GLM), and are implemented in the EPM (*Espresso Performance Monitor*) package [54].

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

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

- 448 The functions  $f_i(\eta)$  are called *basis functions*, and they can be chosen as polynomials
- 449 or spline functions. The set on basis functions is automatically orthogonalised by the
- 450 EPM by using the Gram-Schmidt method; this ensures that the corresponding calibration
- 451 parameters  $p_i$  and  $\Delta p_i$  are correlated as little possible.
- 452 The parameter  $d$  is the tagging decision, which is incorporated into the model in order to
- 453 parameterise  $\omega(\eta)$  for the two possible flavours.
- 454 The function  $g$  is known as *link function*. Usually, this is chosen as the inverse of a
- 455 cumulative distribution function in order to map input values into the interval  $[0, 1]$ , such
- 456 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:

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

- 457 where  $h$  is defined in Eq. 3.13. This link function is built such that the calibrated mistag
- 458 probability is defined in the interval  $(0, 0.5)$ . This choice solves a numerical issue that
- 459 often occurs when standard link functions (e.g. identity or logistic) are adopted. In fact,
- 460 if  $\omega > 0.5$ , then an arbitrary prescription has to be taken (e.g., label the candidate as
- 461 untagged, or flip the tagging decision and take  $1 - \omega$  as new calibrated mistag). If the
- 462 calibration parameters are free parameters in a time-dependent fit, this choice has to be
- 463 made during the minimisation process, according to the values  $\omega$  takes at each iteration.
- 464 This means that the relative number of  $B$  and  $\bar{B}$ , or the relative number of tagged and
- 465 untagged candidates, may change during the fit, which leads to numerical instabilities
- 466 due to discontinuous changes in the likelihood function.
- 467 The EPM estimates the calibration parameters  $p_i$  and  $\Delta p_i$  via an unbinned maximum
- 468 likelihood fit called *binomial regression*; this is an improvement over traditional, binned
- 469 least-squares fits, which are affected by a systematic uncertainty due to the binning
- 470 choice.

<sup>471</sup> **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 following 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

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

<sup>472</sup> If  $P(\bar{b}) > P(b)$  the combined tagging decision is  $d = +1$  and the final mistag probability  
<sup>473</sup> is  $\eta = 1 - P(\bar{b})$ . Otherwise if  $P(b) > P(\bar{b})$  the combined tagging decision and the mistag  
<sup>474</sup> probability are  $d = -1$  and  $\eta = 1 - P(b)$ .

<sup>475</sup> Equation 3.15 is valid under the assumption that all taggers in the combination are  
<sup>476</sup> independent. In the  $B^0 \rightarrow D^\pm \pi^\mp$  analysis presented in this thesis, the OS taggers are  
<sup>477</sup> combined in a single OS combination, and the same is done for the SS taggers. Effects  
<sup>478</sup> due to correlations among taggers within a combination are corrected by calibrating the  
<sup>479</sup> combined predicted mistag.

<sup>480</sup> **3.2 Flavour tagging strategy for the  $B^0 \rightarrow D^\pm \pi^\mp$  time-  
<sup>481</sup> dependent analysis**

<sup>482</sup> In the  $B^0 \rightarrow D^\pm \pi^\mp$  analysis presented in this thesis, the OS combination (including the OS  
<sup>483</sup> charm tagger) and the SS combination are used. The implementation of the OS algorithms  
<sup>484</sup> used in the combination are the same as described in Refs. [49, 50]; the OS algorithms  
<sup>485</sup> other than the OS charm tagger were built as neural networks trained on  $B^+ \rightarrow J/\psi K^+$   
<sup>486</sup> Run 1 data, whereas the OS charm tagger was implemented with a BDT trained on a  
<sup>487</sup> 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  
<sup>488</sup> taggers have been reimplemented for this specific analysis by exploiting  $B^0 \rightarrow J/\psi K^{*0}$   
<sup>489</sup> decays. The functional form of the tagging calibrations is studied in control samples of  
<sup>490</sup> flavour-specific decays properly corrected to resemble the signal decay. The calibration  
<sup>491</sup> parameters are determined directly in the decay-time-dependent fit of the signal described  
<sup>492</sup> in Sec. 5.3; they are nuisance parameters of the likelihood function. Determining the  
<sup>493</sup> calibration parameters from the data along with the  $CP$  observables is possible because  
<sup>494</sup> 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

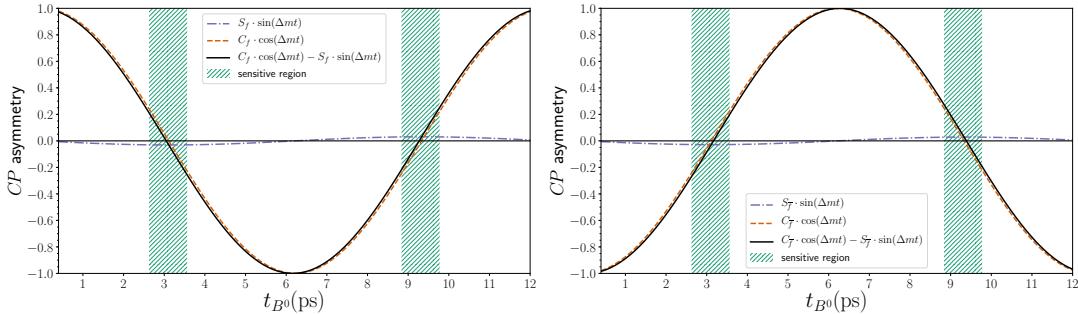


Figure 3.3 –  $\bar{B}^0$  versus  $B^0$  time-dependent asymmetries for the  $D^- \pi^+$  (left) and  $D^+ \pi^-$  (right) final states. The values of  $C_f$ ,  $C_{\bar{f}}$ ,  $S_f$  and  $S_{\bar{f}}$  are the ones used in simulation. The sensitivity to  $S_f$  and  $S_{\bar{f}}$  is maximised in 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.

495 –1 respectively). Hence, the cosine terms give sensitivity to the calibration parameters  
 496 independently of the sine terms, which are proportional to the  $S_f$  and  $S_{\bar{f}}$  coefficients. An  
 497 heuristic explanation is presented in Fig. 3.3. This strategy avoids any assumption on the  
 498 portability to the signal sample of the calibration parameters determined from the control  
 499 data. Such a strategy was studied extensively on simulation: the increase of the statistical  
 500 uncertainty of the  $S_f$  and  $S_{\bar{f}}$  coefficients given by the additional degrees of freedom of  
 501 the calibration parameters is smaller than the systematic uncertainties associated with  
 502 the calibration portability. Moreover, the use of the calibration parameters from the  
 503 control sample causes biases on  $S_f$  and  $S_{\bar{f}}$  of the order of their statistical uncertainty;  
 504 when letting the calibration parameters float in the fit, such biases are suppressed or  
 505 disappear, at the cost of a moderate increase of the statistical uncertainty. In addition,  
 506 while the precision of the OS tagger calibration from the control sample is similar to  
 507 the one from the signal sample, the calibration of the SS tagger derived from the signal  
 508 sample (Tab. 5.4) is much more precise than that from the control sample (Tab. 3.4).

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

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

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

#### 521 3.2.1 Calibration of the opposite-side tagger combination

##### 522 Data sample selection

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

528 Candidate  $B^+ \rightarrow D^0\pi^+$  decays are selected through the `B2DOPiD2HHBeauty2CharmLine`  
529 stripping line, versions `S21r1` (2011 data) and `S21` (2012 data), of the `BhadronCompleteEvent`  
530 stream. The  $B^+$  candidates are required to be TOS, *i.e.* to trigger on `Hlt1TrackAllL0Decision`  
531 at the HLT1 stage, and at least one among `Hlt2Topo2BodyBBDTDecision`, `Hlt2Topo3BodyBBDTDecision`,  
532 and `Hlt2Topo4BodyBBDTDecision` at HLT2. The additional requirements listed in Ta-  
533 ble. 3.1 are applied to further suppress backgrounds and enhance the signal purity.

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

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

##### 543 Calibration

544 The calibration of the estimated mistag  $\eta$  is performed on the fully reweighted  $B^+ \rightarrow D^0\pi^+$   
545 dataset. A GLM model with NSpline basis function [54] is adopted.

546 The projection of the fitted calibration function over the  $B^+ \rightarrow D^0\pi^+$  dataset is shown  
547 in Fig. 3.4, whereas the fitted calibration parameters are listed in Table. 3.2

548 The number of free parameters in the adopted GLM model (10) has been chosen in order  
549 to have satisfactory goodness-of-fit (GOF) metrics (more 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  $B^+ \rightarrow D^0 \pi^+$  candidates.

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

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

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

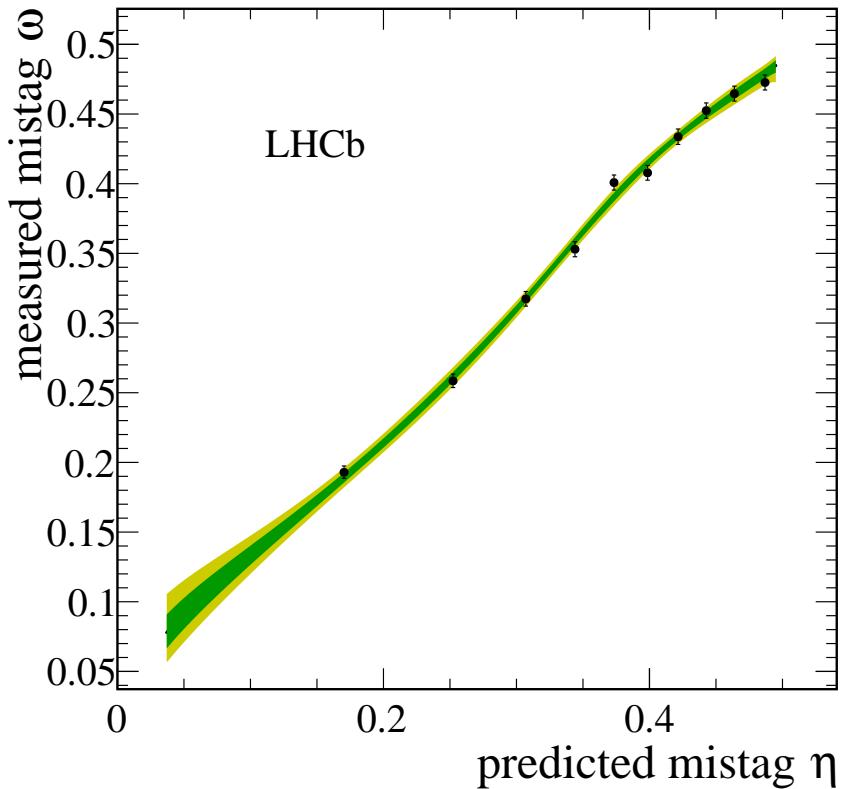


Figure 3.4 – Mistag  $\omega$  measured in bins of predicted mistag  $\eta$  for reweighted  $B^+ \rightarrow D^0\pi^+$  candidates (data points) and fitted calibration function. The green (yellow) band indicate the 68% (95%) confidence interval on the calibration function.

550 **Calibration portability**

551 The aim of the calibration is to return a mistag  $\omega$  as close as possible to the *true* mistag,  
 552 which would be given by a *true calibration*. The latter is not defined on data, but it is  
 553 possible to estimate it on MC by calibrating the predicted mistag  $\eta$  on the truth-matched  
 554 signal decay  $B^0 \rightarrow D^-\pi^+$ . In fact, since, in data, the true flavour of the  $B^0$  meson is  
 555 known in MC, this MC calibration can be done in the same way as  $B^+ \rightarrow D^0\pi^+$ , where  
 556 the true flavour is given by the  $B$  charge.

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

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

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

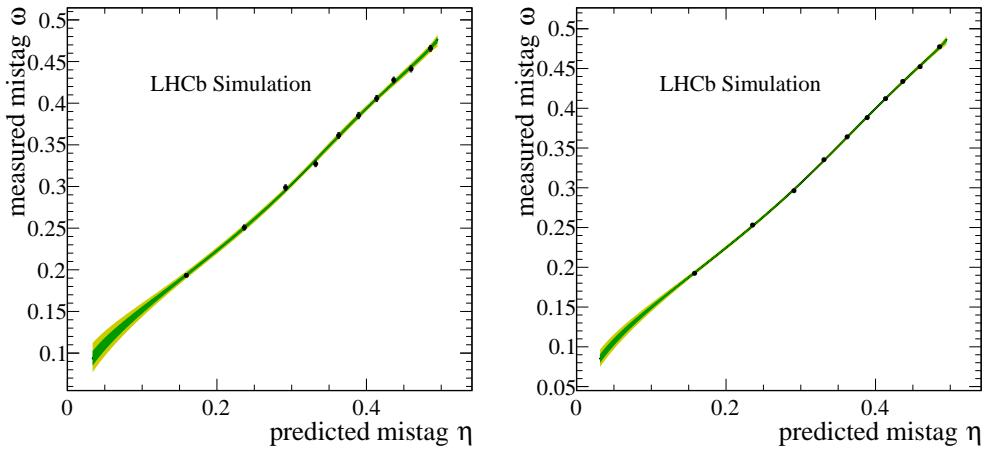


Figure 3.5 – Mistag  $\omega$  measured in bins of predicted mistag  $\eta$  for reweighted  $B^+ \rightarrow D^0\pi^+$  (left) and  $B^0 \rightarrow D^-\pi^+$  (right) candidates (data points) and fitted calibration functions. The green (yellow) band indicate the 68% (95%) confidence interval on the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

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

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

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

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

### 568 3.2.2 Calibration of the same-side tagger combination

569 As described in Ref. [48], the SS pion and proton taggers were both trained on the  
 570 2012 data sample of  $B^0 \rightarrow D^\mp \pi^\pm$  decays. As the effect of  $CP$  violation was neglected  
 571 during the training the algorithms and the underlying MVAs cannot be blindly used  
 572 when measuring  $CP$  violation in the same decay channel. Thus,  $B^0 \rightarrow J/\psi K^{*0}$  decays  
 573 are chosen instead, as they represent a flavour-specific  $B^0$  decay with a considerably large  
 574 signal yield.

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

### 577 Calibration

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

### 586 Calibration portability

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

593 The two calibrations using the  $B^0 \rightarrow D^- \pi^+$  and  $B^0 \rightarrow J/\psi K^{*0}$  Monte Carlo samples are  
 594 shown in Fig. 3.7 and compared in Table 3.5. A full comparison that takes into account the  
 595 correlation between the parameters is obtained from a  $\chi^2$  test similar to the one described  
 596 in Sec. 3.2.1. The agreement is around  $0.1\sigma$ . Despite this test doesn't hint to issues of

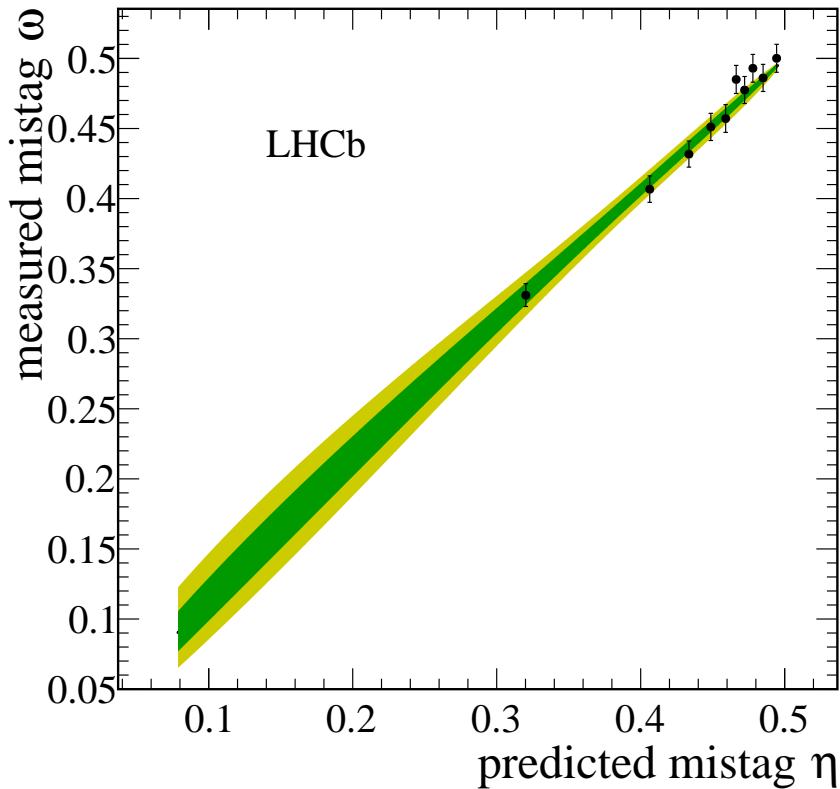


Figure 3.6 – Mistag  $\omega$  measured in bins of predicted mistag  $\eta$  for reweighted  $B^0 \rightarrow J/\psi K^{*0}$  candidates (data points) and fitted calibration function. The green (yellow) band indicate the 68% (95%) confidence interval on the calibration function.

portability between the decay modes, the same strategy used for the OS calibrations is followed, i.e fitting the parameter directly in data with the  $CP$  asymmetries. This is motivated by the fact that the  $B^0 \rightarrow D^- \pi^+$  signal sample has much more sensitivity to determine the parameters than the  $B^0 \rightarrow J/\psi K^{*0}$  sample. In addition, with this approach no systematic related to calibration portability is necessary, consistently with the case of the OS tagger.

### 3.3 Optimisation of the Opposite-Side electron tagger

#### 3.3.1 Introduction

The performance of the flavour tagging algorithms is correlated with the data taking conditions.

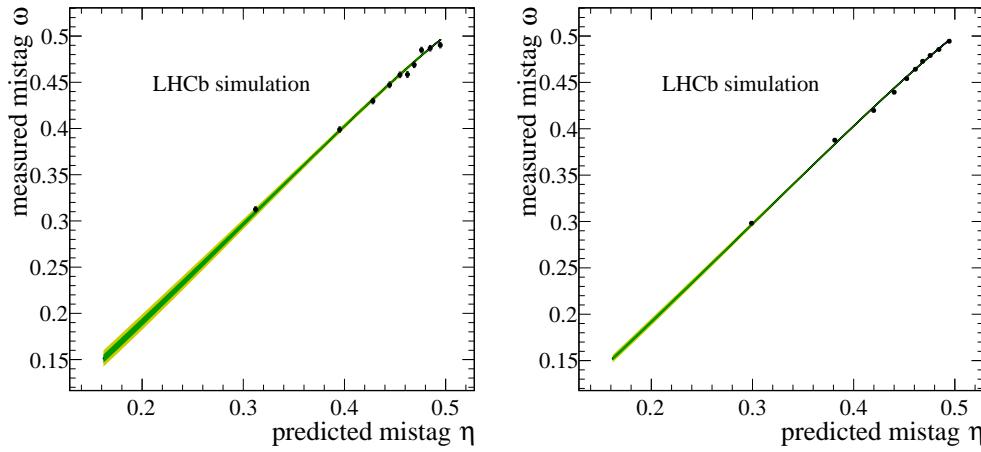


Figure 3.7 – Mistag  $\omega$  measured in bins of predicted mistag  $\eta$  for reweighted  $B^0 \rightarrow J/\psi K^{*0}$  (left) and  $B^0 \rightarrow D^- \pi^+$  (right) candidates (data points) and fitted calibration functions. The green (yellow) band indicate the 68% (95%) confidence interval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

607 On one hand, the tagging power of the SS taggers shows an increase on Run 2 data  
 608 compared to Run 1 thanks to either a higher tagging efficiency (SS $\pi$  and SS $p$ ) or a lower  
 609 mistag rate (SS $K$ ). This is due to the higher boost of the system at 13 TeV, which makes  
 610 the momentum spectrum of  $B$  meson and fragmentation tracks harder, and increases the  
 611 acceptance of the fragmentation tracks.

612 On the other hand, the tagging power of the existing OS $e$ , OS $\mu$ , and OS $K$  taggers  
 613 decreases on Run 2 data. The reason for this degradation is mainly due to the higher  
 614 track multiplicity, which increases the probability to have a wrong tag decision. Moreover,  
 615 because of the different Run 2 kinematics, the criteria to select the tagging particles  
 616 become obsolete, thus giving a lower tagging efficiency.

617 The performance of the OSCharm and OSVtx algorithms is, on average, compatible or  
 618 better on Run 2 than Run 1.

### **3.3. Optimisation of the Opposite-Side electron tagger**

---

619 In this section, the reoptimisation of the OSe tagger is presented. This reoptimisation is  
620 performed both on Run 2 data, in order to recover the observed loss in tagging power,  
621 and Run 1 data, to further improve the already existing algorithm. This reoptimisation  
622 consists of two main steps. First, sequential selection criteria are required to select  
623 electron-like particles yielding a sample of  $B$  signal candidates with a low average tagging  
624 power. Then, a BDT classifier is applied to discriminate between  $B$  candidates with right  
625 and wrong tag decisions for each selected track. Finally, for each  $B$  candidate, the BDT  
626 output for the track with highest transverse momentum is converted into a predicted  
627 mistag probability.

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

630 **3.3.2 Sample selection**

631 The OSe algorithm is developed in a data-driven fashion by using *sWeighted* samples of  
632  $B^\pm \rightarrow J/\psi K^\pm$  data. The full Run 1 dataset (2011+2012) is used to optimise the algorithm  
633 on Run 1 conditions, whereas the 2016 dataset is exploited to optimise the tagger on  
634 Run 2 conditions. An alternative optimisation on *sWeighted* 2016  $B^\pm \rightarrow D^0\pi^\pm$  data is  
635 performed in parallel. The motivation for this is to cross-check the Run 2 implementation  
636 on an independent decay mode, which is characterised by a different kinematics compared  
637 to  $B^\pm \rightarrow J/\psi K^\pm$ . Hereafter, the OSe tagger optimised on Run 1  $B^\pm \rightarrow J/\psi K^\pm$  data  
638 will be indicated as “Run 1 new” version, in order to distinguish it from the previous  
639 “Run 1 old” version introduced in Ref. [49], which was based on simple selection criteria  
640 and a neural network for the mistag estimation. The OSe tagger optimised on Run 2  
641  $B^\pm \rightarrow J/\psi K^\pm$  data, and Run 2  $B^\pm \rightarrow D^0\pi^\pm$  data will be denoted as “Run 2 B2CC” and  
642 “Run 2 B2OC” versions, respectively. Moreover, it is understood that the tunings of all  
643 the PROBNN features mentioned in this section are MC12TuneV2 for Run 1 data and  
644 MC15TuneV1 for Run 2 data.

645 **3.3.3 Preselection optimisation**

646 Electron-like particles are selected by means of a set of sequential requirements. A first  
647 preliminary selection is applied in order to exclude muon tracks (already exploited by  
648 the OS $\mu$  algorithm) and to choose tracks with good quality and reduced combinatorial  
649 background contamination. The details of this preliminary requirements are listed in  
650 Table 3.6.

An additional selection is applied in order to enhance the average tagging power of the

Table 3.6 – Preliminary requirements on tracks selected for the OSe algorithm.

Requirements	Comment
IsSignalDaughter = 0	Track not associated to the $B$ signal decay tree.
MuonPIDIsMuon = 0	Track with no hits in the muon detector.
Track $\chi^2/\text{ndof} < 3$	Track fit quality.
InAccCal = 1	Track in the acceptance of the electromagnetic calorimeter.
$0 < \text{VeloCharge} < 1.4$	Charge deposited in the Velo detector.
$\sigma_{\text{IP}} \neq 0$	Fit for the track IP w.r.t the primary vertex converged.
$0.85 < e/p < 2.0$	Rigidity.
TrackType = 3	Long track.

resulting sample, which is defined as:

$$\langle \varepsilon_{\text{eff}} \rangle = \frac{\sum_{i=1}^N w_i f_{\text{R}}^i}{\sum_{i=1}^N w_i (f_{\text{R}}^i + f_{\text{W}}^i)}, \quad (3.17)$$

where  $w_i$  are the *sWeights*, while  $f_{\text{R}}^i$  ( $f_{\text{W}}^i$ ) is the fraction of particles giving the right (wrong) flavour for the  $i$ th  $B$  candidate.

The expression of Eq. 3.17 is taken as the figure of merit to maximise during the selection optimisation. This maximisation is performed numerically by using gradient boosted regression trees to model  $\langle \varepsilon_{\text{eff}} \rangle$  as a function of the applied cuts [55]. The cuts are optimised separately for the Run 1 new, Run 2 B2CC and Run 2 B2OC algorithms; in all cases, about 25% of the available data for each sample is used.

The resulting, optimised requirements are reported in Table 3.7, while the convergence plots of the minimisation are shown in Fig. 3.8.

After the optimisation, the performance of the selection (including the average tagging power given by Eq. 3.17) is evaluated on the remaining 75% of data for each sample, yielding the results shown in Table 3.8.

### 3.3.4 BDT classifier implementation

The selection described in Sec. 3.3.3 is applied on the remaining part of the data ( $\sim 75\%$ ) used by each OSe implementation. Each of these remaining samples is then split in three further subsamples for the following purposes:

- The first 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. This sample is also used for the hyperparameters tuning.

### 3.3. Optimisation of the Opposite-Side electron tagger

Table 3.7 – Optimised requirements for the selection of tracks used by the OSe algorithm.

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

Table 3.8 – Performance of the preliminary selection (OSe algorithm) applied on the data not used for the selection optimisation ( $\sim 75\%$  of the total dataset for each sample). The average tagging power  $\langle \varepsilon_{\text{eff}} \rangle$  is given by 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	$3.664 \pm 0.024$	$34.32 \pm 0.32$	$0.360 \pm 0.006$

- 670 • The second and the third subsets (*evaluation set 1* and *2*), each including  $\sim 12.5\%$   
671 of the total data, are adopted together as test set to check for overtraining. The  
672 evaluation set 1 is also used to calibrate the obtained tagger, which is then applied  
673 to the second evaluation set in order to measure the performance; the procedure is  
674 then repeated by swapping the two samples (*two-fold validation*).
- 675 The BDT classifier is trained to identify  $B$  candidates as rightly or wrongly tagged. The  
676 list of the features adopted to build the BDTs are reported in Table 3.9. The distributions  
677 of the input features for the two possible values of the target are shown in Figs. 3.9, 3.10  
678 and 3.11 for the Run 1 new, Run 2 B2CC and Run 2 B2OC samples respectively. The  
679 Pearson correlation coefficients between the input features are reported in Fig. 3.12.
- 680 The BDT classifier consists of an ensemble of 300 gradient-boosted decision trees [56],  
681 where each tree can have a maximum depth of 3. The objective of the classifier is a binary  
682 logistic loss function plus a quadratic regularisation term to control model complexity  
683 (with regularisation parameter  $\lambda = 1$ ). Some hyperparameters were tested by means of a  
684 cross-validation+bootstrapping method on the training set, as described in Appendix B).
- 685 The importance (or F score) of each feature, defined as the total number of times a feature  
686 is chosen as split node by any tree in the BDT ensemble, is presented in Fig. 3.14, while

### Chapter 3. Flavour tagging

---

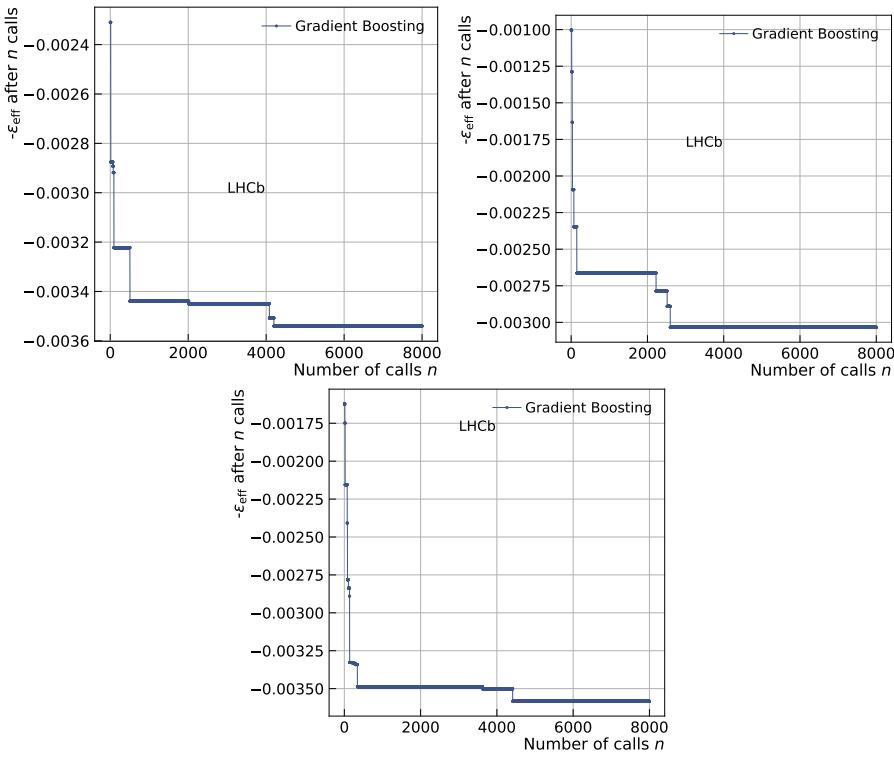


Figure 3.8 – Minimised value of the average tagging power as a function of the gradient boosted regression tree algorithm iteration for the Run 1 new (top left), Run 2 B2CC (top right) and Run 2 B2OC (bottom) implementations of the OSe tagger.

Table 3.9 – Features adopted 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 ✓(✗) indicates if a given feature is included (discarded).

Feature	Comment/description	Run 1 new	Run 2 B2CC	Run 2 B2OC
nTracks	Number of reconstructed tracks	✓	✓	✓
$p_T$	transverse momentum of tagging track	✓	✓	✓
$\sigma_{\text{IP}}$	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	✓	✓	✓
$p_{\text{ghost}}$	Ghost probability	✓	✗	✓
$e/p$	rigidity	✓	✓	✓
$\Delta R$	difference in $R$ -coordinate between $B$ and tagging track	✓	✓	✓
$ \text{IP} $	absolute value of tagging track IP	✓	✓	✓
$\sigma_{\text{IPPU}}/\text{IPPU}$	significance of the IP w.r.t pile-up vertex for tagging track	✓	✓	✓
PROBNN ghost	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. Optimisation of the Opposite-Side electron tagger

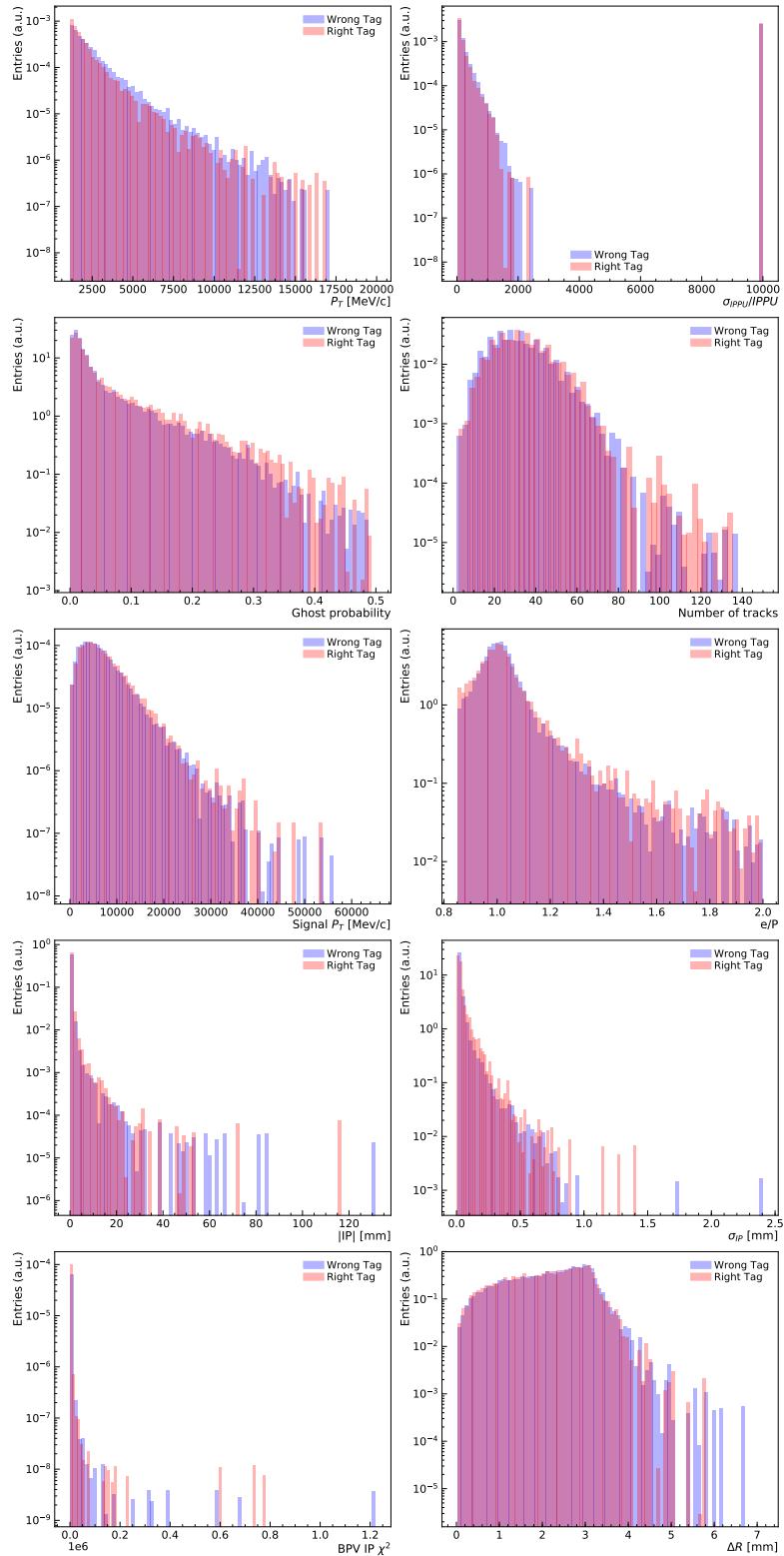


Figure 3.9 – Distributions (for the *sWeighted*, Run 1  $B^\pm \rightarrow J/\psi K^\pm$  sample) of the input features of the BDT classifier (OSe tagger) for candidates with a right (red) and wrong (blue) tag decision.

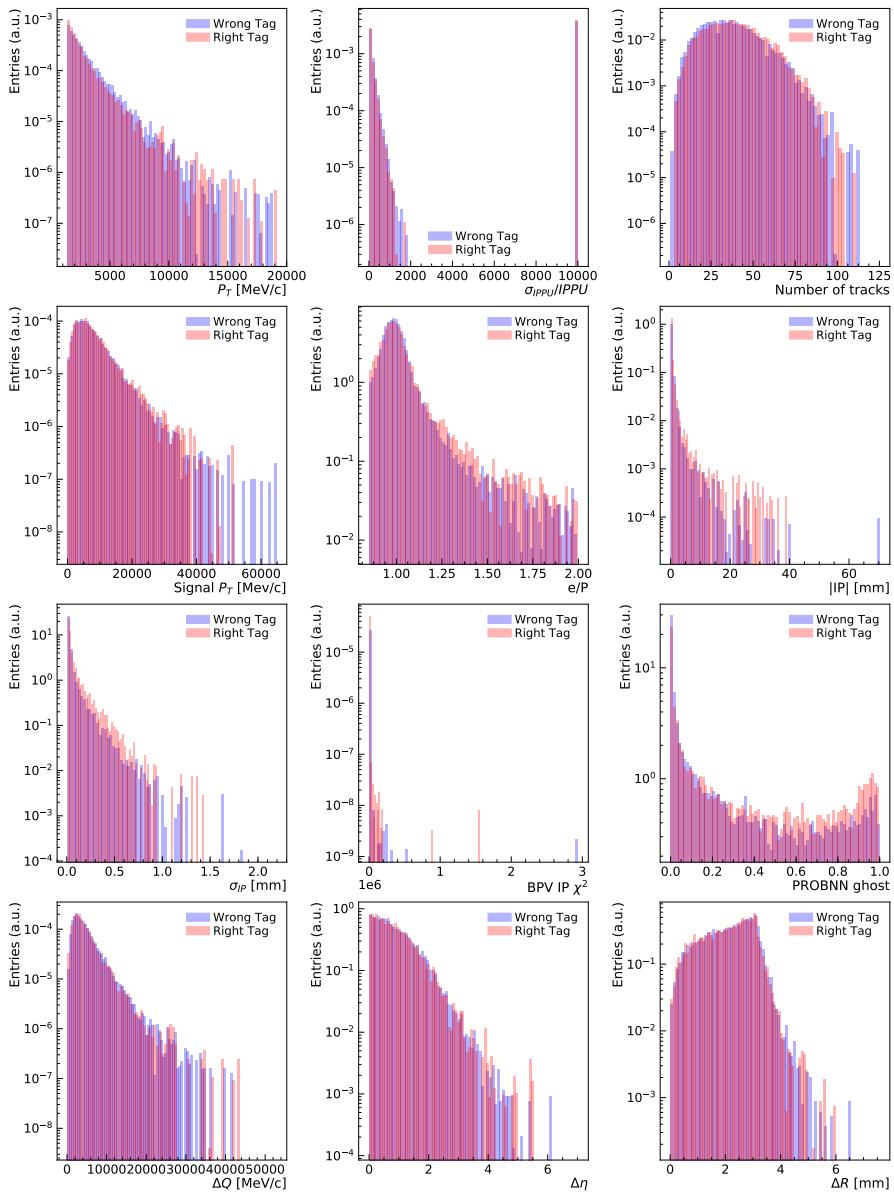


Figure 3.10 – Distributions (for the  $sWeighted$ , Run 2  $B^\pm \rightarrow J/\psi K^\pm$  sample) of the input features of the BDT classifier (OSe tagger) for candidates with a right (red) and wrong (blue) tag decision.

### 3.3. Optimisation of the Opposite-Side electron tagger

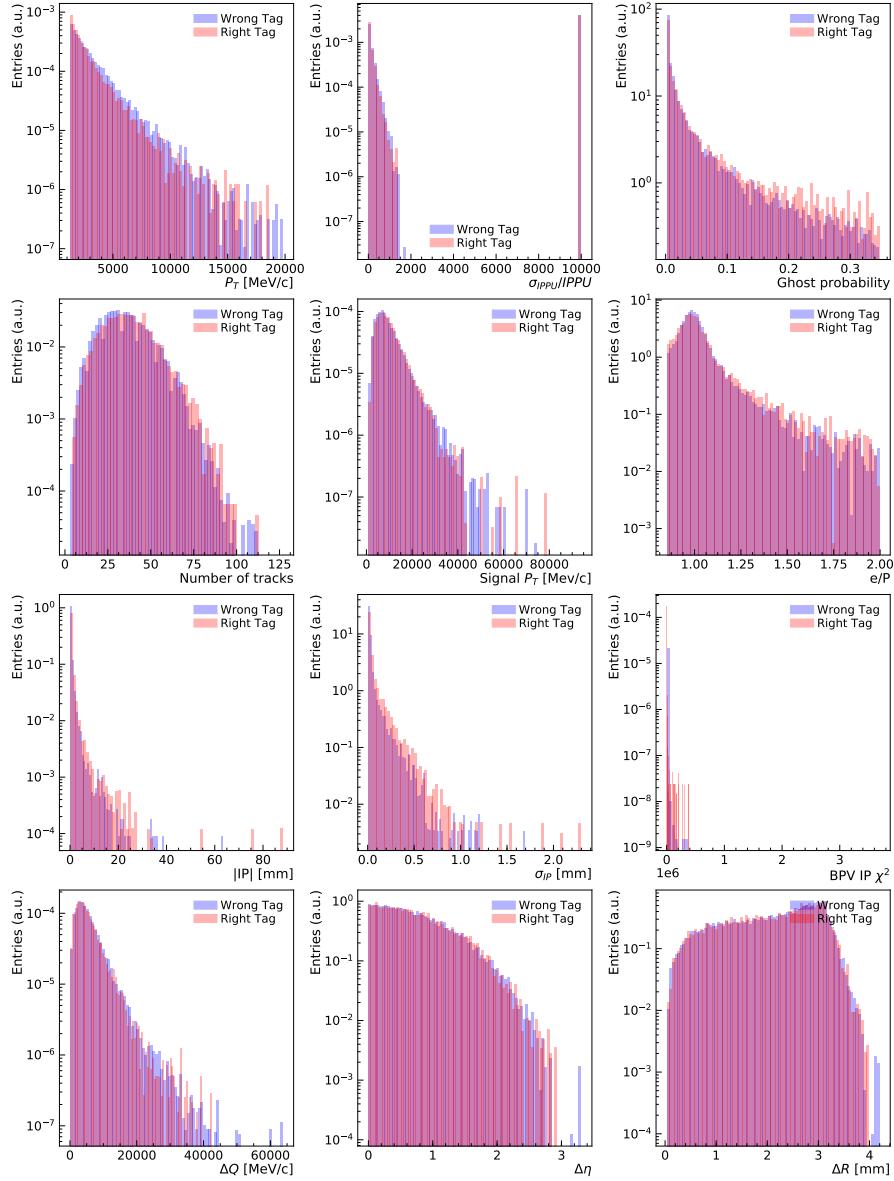


Figure 3.11 – Distributions (for the *sWeighted*, Run 2  $B^\pm \rightarrow D^0\pi^\pm$  sample) of the input features of the BDT classifier (OSe tagger) for candidates with a right (red) and wrong (blue) tag decision.

### Chapter 3. Flavour tagging

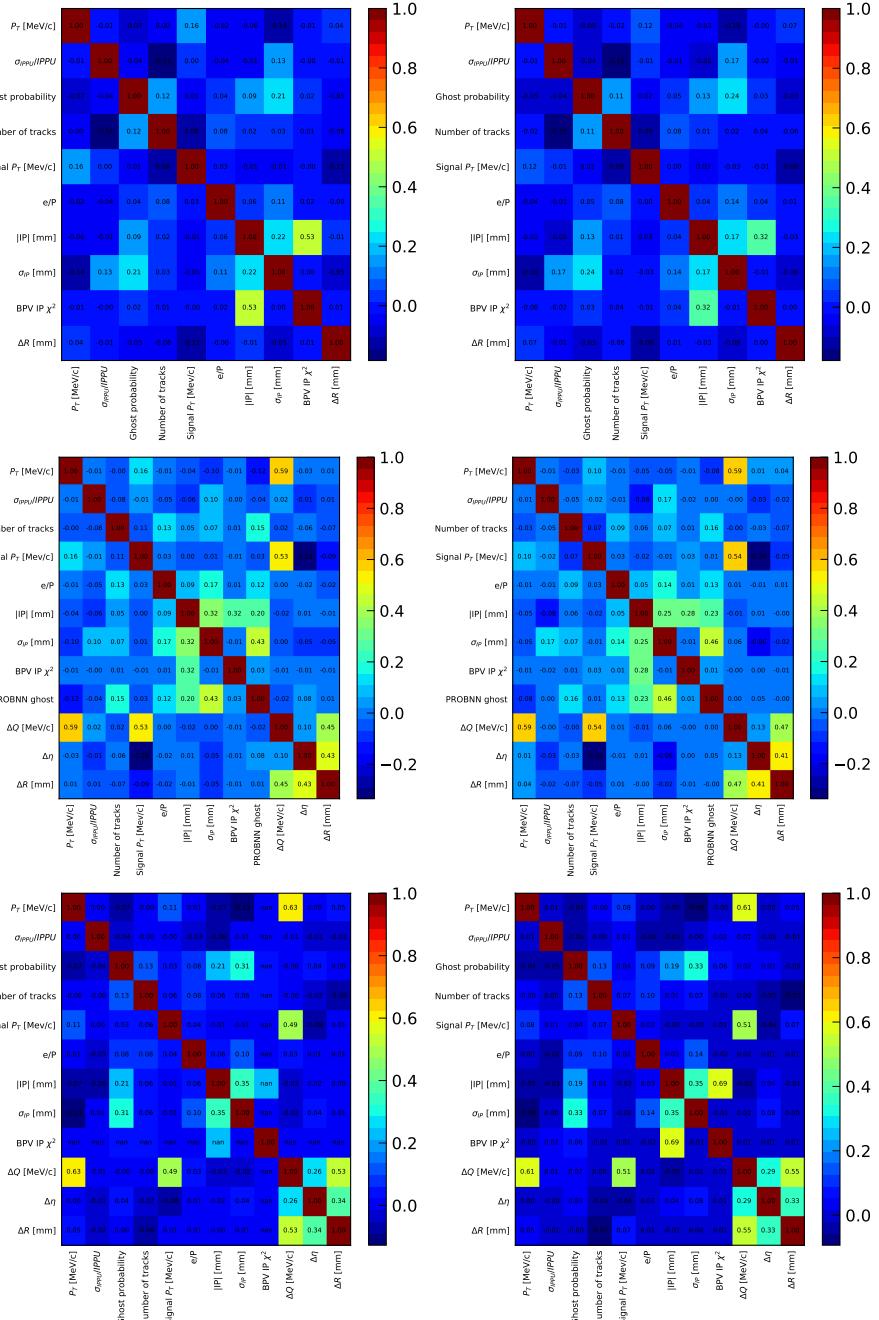


Figure 3.12 – Pearson correlation coefficients between the input features of the Run 1 new (top), Run 2 B2CC (middle) and Run 2 B2OC (bottom) BDT classifiers (OSe tagger) for candidates with a right (left) and wrong (right) tag decision.

### 3.3. Optimisation of the Opposite-Side electron tagger

687 the *partial dependence* of the predicted mistag  $\eta$  (on the training set) as a function of  
 688 each input feature is shown in Appendix B.

689 The receiver operating characteristic (ROC) curves, which report the rate of true right-  
 690 tagged candidates as a function of the rate of false right-tagged candidates, are reported  
 691 in Fig. 3.13.

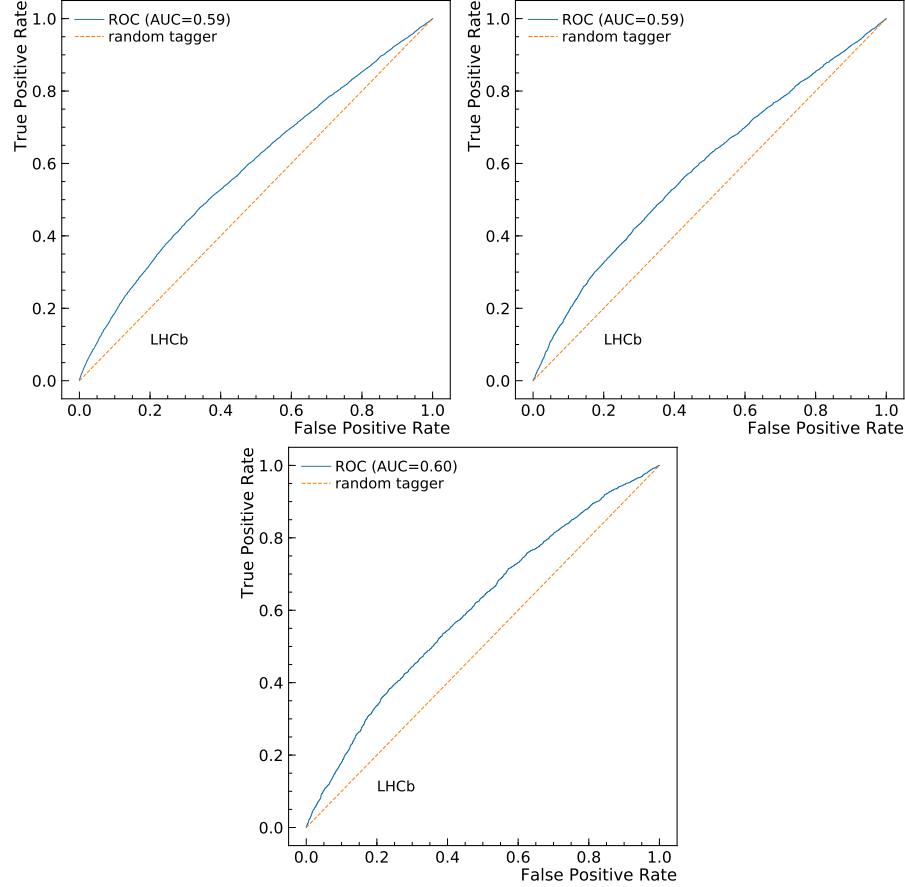


Figure 3.13 – Rate of true right-tagged candidates as a function of the rate of false right-tagged candidates (ROC curves) for the BDT classifiers of the Run 1 new (top left), Run 2 B2CC (top right) and Run 2 B2OC (bottom) implementations of OSe. The obtained ROC curves are represented in blue, while 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.

692 The feature selection, BDT training and feature importance evaluation chain has been  
 693 repeated iteratively in order to exclude highly-correlated and poorly-important features,  
 694 until the BDT performance didn't start 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

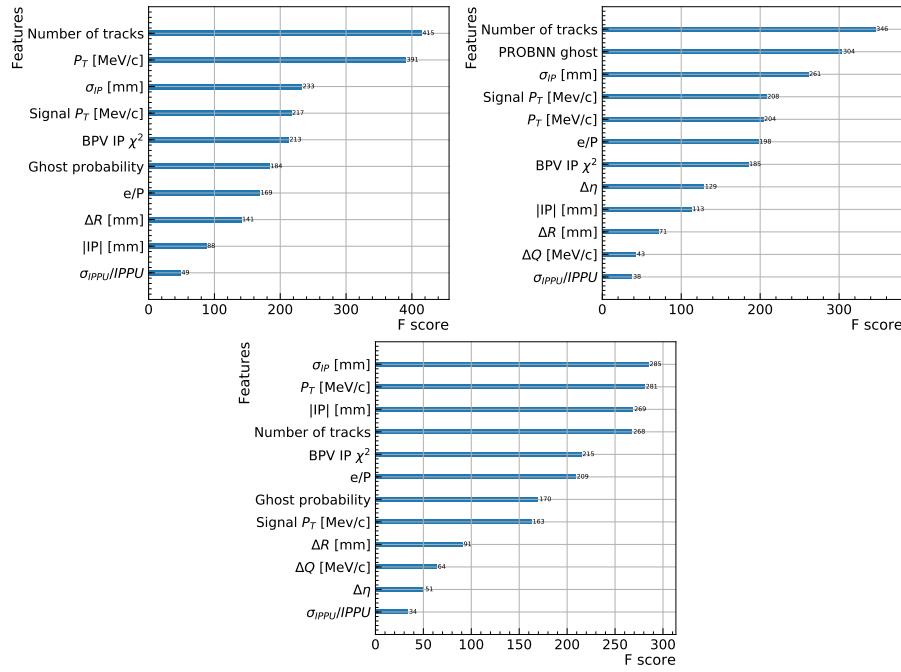


Figure 3.14 – Feature importance for the BDT classifiers of the Run 1 new (top left), Run 2 B2CC (top right) and Run 2 B2OC (bottom) implementations of OSe.

on both  $P$  and tagging decision  $d$ :

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

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

### 697 3.3.5 Performance evaluation

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

699 Once the BDT is trained on the training set, the mistag  $\eta$  is predicted for each candidate  
700 in the remaining part of the data for each sample. Then, a two-fold evaluation is applied:

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

### 3.3. Optimisation of the Opposite-Side electron tagger

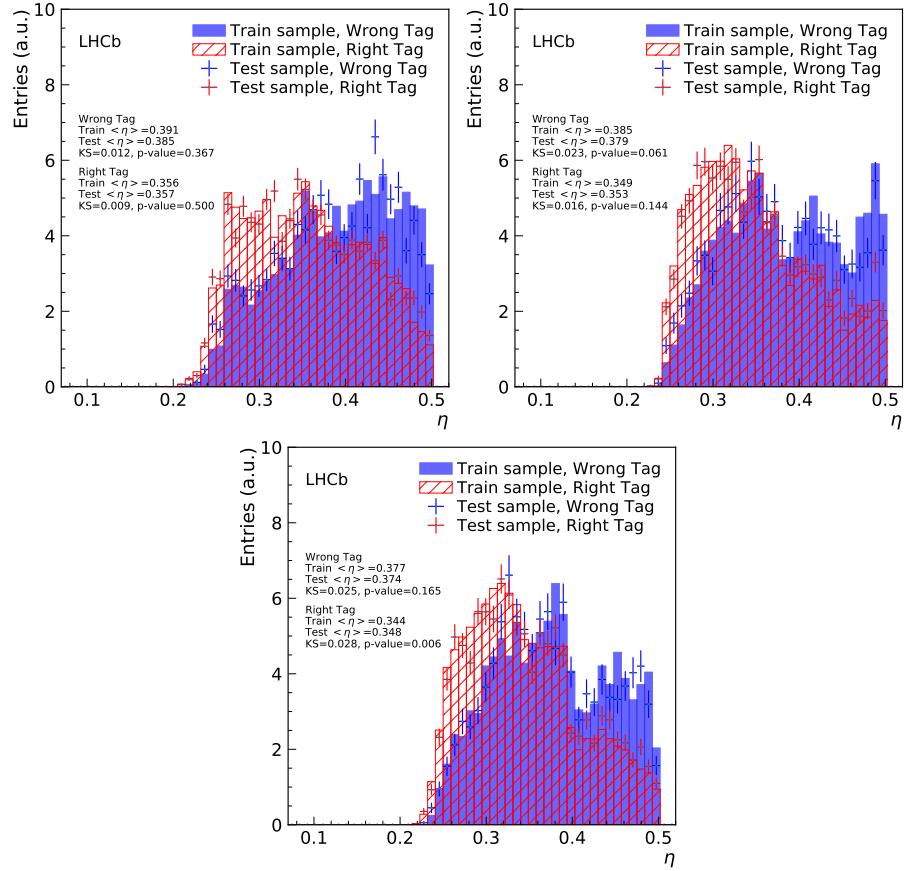


Figure 3.15 – *sWeighted* distributions of the  $\eta$  predicted by the BDT classifiers of the Run 1 new (top left), Run 2 B2CC (top right) and Run 2 B2OC (bottom) versions of OSe. The blue-solid (red-hatched) histogram represents the training data for candidates having the wrong (right) tag decision. The blue (red) points indicate the test data for candidates with wrong (right) tag decision. The overtraining is checked by means of a Kologorov-Smirnov (KS) test to measure the compatibility between training data and remaining data. This test is performed separately for candidates with wrong and right tag decision.

### Chapter 3. Flavour tagging

---

Table 3.10 – Calibrated, per-event tagging power 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	$\varepsilon_{\text{eff}} (\%)$ , set 1	$\varepsilon_{\text{eff}} (\%)$ , 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	$0.455 \pm 0.043$	$0.434 \pm 0.041$	$0.444 \pm 0.030$

705 The calibrated per-event tagging power is computed by considering, for each tagged  
706  $B$  candidate, only the tagging particle with the highest transverse momentum. The  
707 calibration model consists of a first order natural spline with a logistic link function. The  
708 result of these calibrations are shown in Fig. 3.16. The calibrated per-event tagging power  
709 is reported in Table 3.10.

710 **Performance on  $B^0 \rightarrow D^\mp \pi^\pm$  data**

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

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

- 720     • each sample (Run 1 and Run 2) is splitted randomly in two subsamples;  
721     • the calibrations are found on one subsample for all OS taggers;  
722     • the calibrations so obtained are applied to the other subsample, and the calibrated  
723        performance is evaluated.  
724     • the calibrated OS taggers are combined, the combination is calibrated in order to  
725        correct for effects due to correlations among taggers, and the performance of the  
726        calibrated combination is evaluated.

727 The calibrations are obtained via a time-dependent analysis of the  $B^0 \rightarrow D^\mp \pi^\pm$  decays,  
728 where acceptance and resolution effects are neglected as described in Ref. [54]. The chosen  
729 model  $\omega(\eta)$  for each tagger is a GLM model with a logistic link function, and a first order  
730 spline as basis function. The results of the calibration and the mistag distribution of each

### 3.3. Optimisation of the Opposite-Side electron tagger

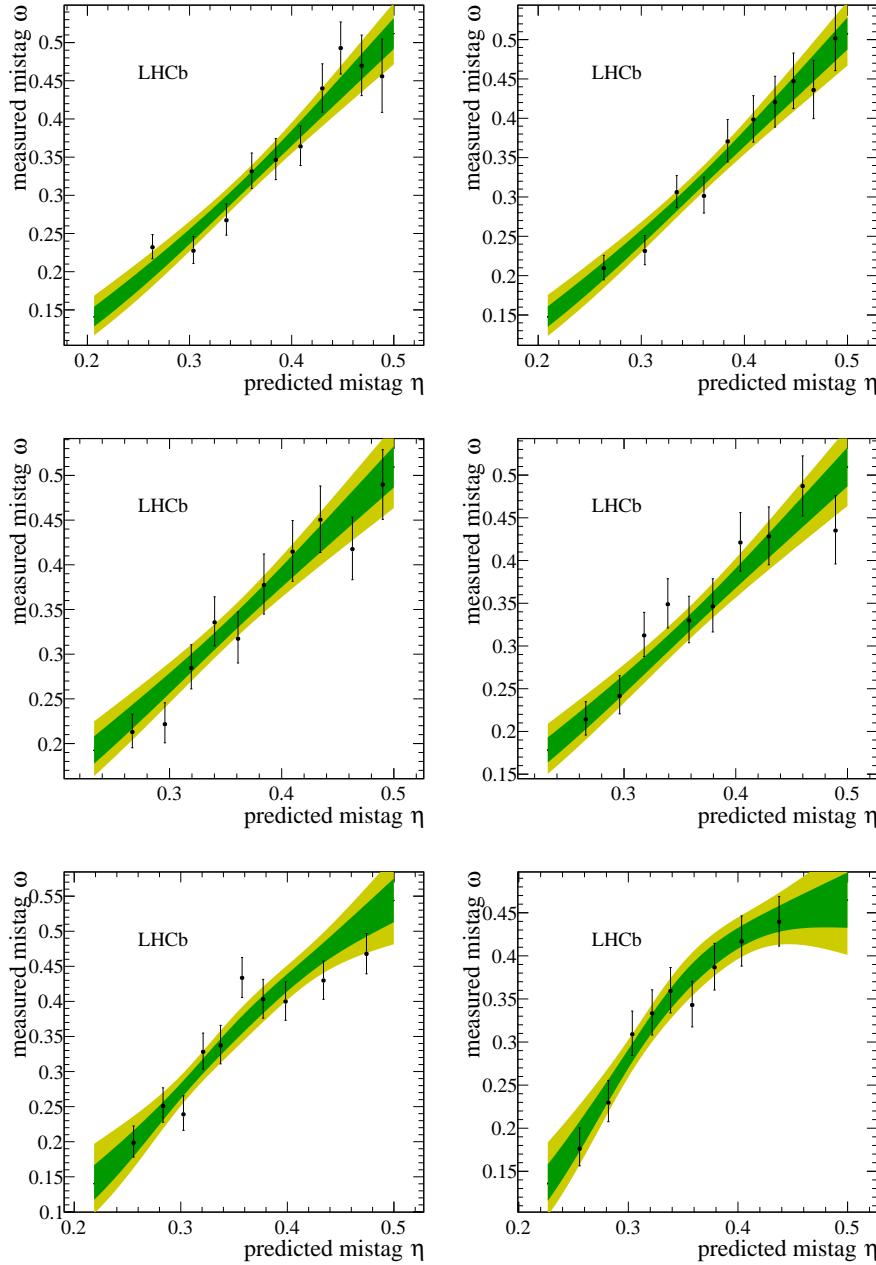


Figure 3.16 – OSe mistag calibration results for the (top) Run 1 new, (middle) Run 2 B2CC and (bottom) Run 2 B2OC optimisations. Left: calibration obtained on the second evaluation sample plotted together with the first evaluation sample. Right: calibration obtained on the first evaluation sample plotted 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.

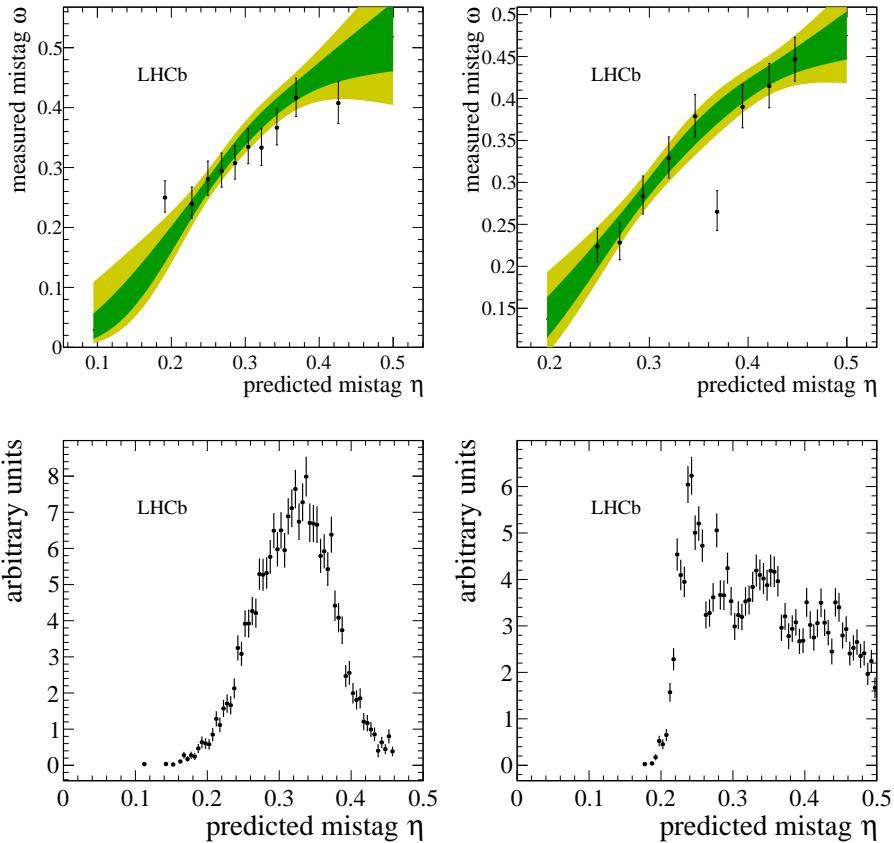


Figure 3.17 – Top: mistag calibration results on *sWeighted* Run 1  $B^0 \rightarrow D^\mp \pi^\pm$  data for the Run 1 old (left), and Run 1 new (right) versions of the OSe tagger. The *sWeighted* data sample is shown as black points. The green (yellow) band indicates the 68% (95%) C.L. interval for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

731 OSe implementation are shown in Figs. 3.17 and 3.19; the calibration and the mistag of  
 732 the corresponding OS combinations are also reported in Figs. 3.18 and 3.20.

733 The performance is reported in Tables 3.11 and 3.12. The Run 1 new tuning allows to gain a  
 734 relative 9% in tagging power for the OSe tagger on Run 1 data; the corresponding, relative  
 735 gain of the OS Combination is 3%. The tagging power of OSVtx and OSCharm increases  
 736 on Run 2 data compared to Run 1; for this reason, no optimisations are required. The  
 737 tagging power of OSe, OS $\mu$ , and OSK with the Run 1 tunings is lower on Run 2 data  
 738 compared to Run 1. However, compared to the Run 1 tunings, the Run 2 tunings show a  
 739 relative improvement in tagging power of about  $\sim 160\%$  for OSe, and  $\sim 6\%$  for OS $\mu$  and  
 740 OSK on Run 2 data. This allows to recover similar performances as the ones obtained on  
 741 Run 1 data with the Run 1 tunings, both for the individual taggers and their combination.  
 742 Moreover, the Run 2 B2CC and B2OC tunings show consistent tagging powers on Run  
 743 2 data, meaning that the optimisation is robust against the different kinematics of the

### 3.3. Optimisation of the Opposite-Side electron tagger

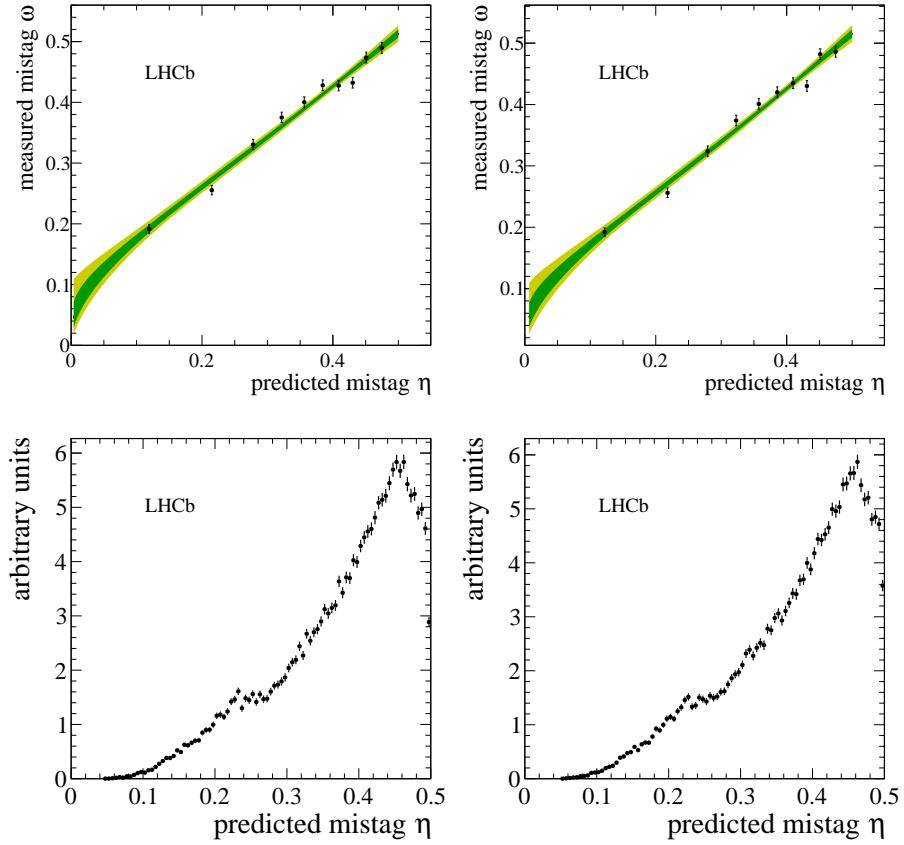


Figure 3.18 – Top: mistag calibration results on *sWeighted* Run 1  $B^0 \rightarrow D^\mp \pi^\pm$  data for the combination of the OS taggers. The results obtained with the Run 1 old (left), and Run 1 new (right) tunings of OSe are shown. The *sWeighted* data sample is shown as black points. The green (yellow) band indicates the 68% (95%) C.L. interval for the fitted calibration functions. Bottom: distributions of the uncalibrated mistag  $\eta$ .

744 adopted decays.

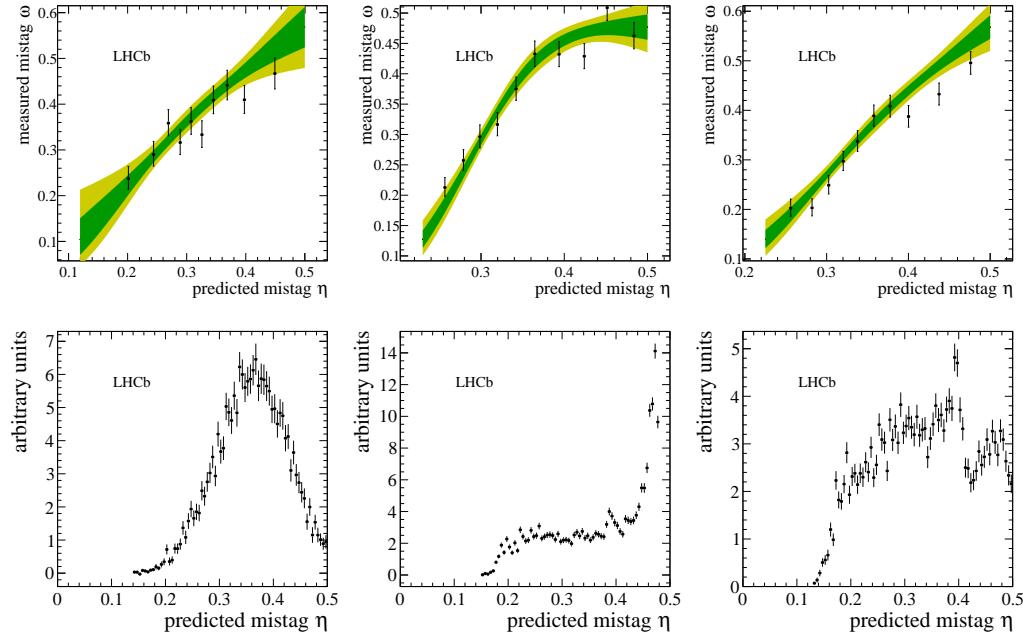


Figure 3.19 – Mistag calibration results on  $s\text{Weighted}$  Run 2  $B^0 \rightarrow D^\mp \pi^\pm$  data for the OSe taggers. The results obtained with the Run 1 (left), Run 2 B2CC (center), and Run 2 B2OC (right) tunings are shown. The  $s\text{Weighted}$  data sample is shown as black points. The green (yellow) band indicates the 68% (95%) C.L. interval for the fitted calibration functions. Bottom: distributions of the uncalibrated mistag  $\eta$ .

Table 3.11 – Performance (tagging efficiency, average mistag, and tagging power) of the OSe taggers on  $s\text{Weighted}$  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 2 new tunings.

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

### 3.3. Optimisation of the Opposite-Side electron tagger

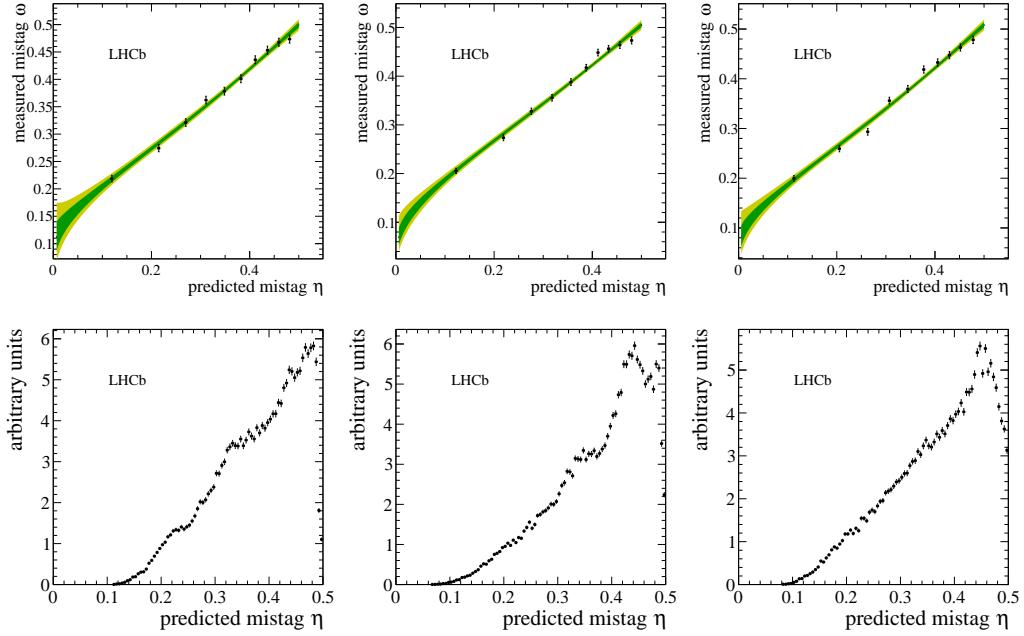


Figure 3.20 – Mistag calibration results on *sWeighted* Run 2  $B^0 \rightarrow D^\mp \pi^\pm$  data for the combination of the OS taggers. The results obtained with the Run 1 old (left), Run 2 B2CC (center), and Run 2 B2OC (right) tunings of OSe, OS $\mu$ , and OSK are shown. The *sWeighted* data sample 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  $\eta$ .

Table 3.12 – Performance (tagging efficiency, average mistag, and tagging power) 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.

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



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

746

747 **4.1 Data sample and preselection**

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

- 749 1. stripping and trigger requirements;
- 750 2. a *cut-based* preselection;
- 751 3. vetoes for misidentified backgrounds and wrongly associated *primary vertices* (PVs);
- 752 4. a multivariate classification (MVA);
- 753 5. a final randomised multiple candidate selection.

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

755 **4.1.1 Stripping and trigger requirements**

756 Signal  $B^0 \rightarrow D^\mp\pi^\pm$  candidates are reconstructed using a dedicated stripping line (called  
757 `B02DPiD2HHHBeauty2CharmLine`). Each event is required to have less than 500 long  
758 tracks. The criteria that the charged tracks have to fulfill are listed in Table 4.1. Three of  
759 these hadrons have to form a common vertex to build a  $D^\mp$  meson. Further requirements  
760 on the  $D^\mp$  combination are given in Table 4.2. The  $B^0$  candidates are built by combining  
761 a  $D^\mp$  candidate and a bachelor particle if the requirements listed in Tab. 4.3 are fulfilled.  
762 Finally, a bagged boosted decision tree (BDT) classifier [57], which is trained on simulated  
763 candidates, is applied. A minimum value of 0.05 is required for the output value of the  
764 BDT. Stripped candidates are then filtered according to how they were selected at the  
765 trigger level: no specific requirements are made at L0; at HLT1,  $B$  candidates are required  
766 to be TOS from the `Hlt1TrackAllL0Decision` trigger line; at HLT2,  $B$  candidates  
767 are required to be TOS from one of the following lines: `Hlt2Topo2BodyBBDTDecision`,  
768 `Hlt2Topo3BodyBBDTDecision` or `Hlt2Topo4BodyBBDTDecision`. These trigger lines are  
769 described in details Refs. [58, 59].

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  $\text{IP}\chi^2$  w.r.t any PV is defined as the compatibility between the particle trajectory and all the primary vertices reconstructed in the same event (a large value indicates that the track does not originate from that vertex).

$\chi^2/\text{ndof}$	$< 3.0 (< 2.5)$
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  $\chi^2_{\text{IP}}$ , which is the vertex-fit  $\chi^2$  for the PV reconstructed with and without the  $B^0$  candidate.

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

Table 4.3 – Stripping requirements on the  $D^\mp\pi^\pm$  combination.

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

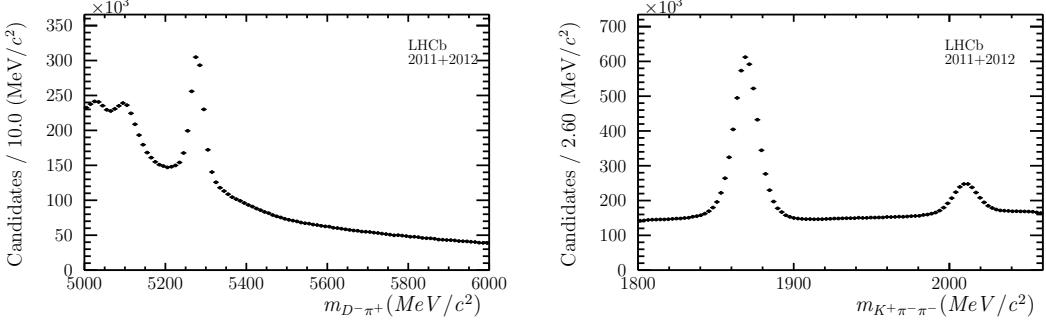


Figure 4.1 –  $D^\mp\pi^\pm$  and  $K^\pm\pi^\mp\pi^\mp$  mass distributions of the reconstructed  $B^0 \rightarrow D^\mp\pi^\pm$ ,  $D^\mp \rightarrow K^\pm\pi^\mp\pi^\mp$  candidates after the stripping and trigger selection.

770 Figure 4.1 shows the  $D^\mp\pi^\pm$  and  $K^\pm\pi^\mp\pi^\mp$  mass distributions of the reconstructed can-  
 771 didates after the stripping and trigger selection. In the  $D^\mp\pi^\pm$  mass distribution the  
 772  $B^0 \rightarrow D^\mp\pi^\pm$  signal peak is already visible at around 5280 MeV/ $c^2$ . The structure  
 773 at masses lower than the  $B$  peak originates from partially reconstructed  $B^0 \rightarrow D\rho$   
 774 and  $B^0 \rightarrow D^*\pi$  decays. The  $K^\pm\pi^\mp\pi^\mp$  mass distributions features a clearly visible  
 775  $D^\mp \rightarrow K^\pm\pi^\mp\pi^\mp$  peak at 1870 MeV/ $c^2$  and a  $D^{*\mp} \rightarrow D^0(K^\pm\pi^\mp)\pi^\mp$  peak around  
 776 2010 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/ $c^2$
PIDK for pions	$< +8$ from $D^\mp$
PIDK 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 and uncertainties on vertex positions, particle momenta, flight distances, decay times, and invariant masses, a Kalman filter, known as `DecayTreeFitter` (DTF) [60], 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 momenta and the invariant mass of the  $B^0$  meson, the invariant mass of the  $D^\mp$  meson is constrained to the PDG value ( $m_{D^\mp}^{\text{PDG}} = 1869.61$  MeV/ $c^2$  [20]) in a separate DTF fit. The PIDK 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 PIDK of the bachelor pion define two categories of candidates: those of the so-called *pion sample* ( $\text{PIDK} \leq 5$ ) and those of the so-called *kaon sample* ( $\text{PIDK} > 5$ ). This distinction will be useful in the fit to the  $B^0$  mass distribution for determining the sample composition described in Sec. 4.3. All the following selection steps are applied to both the pion and kaon samples.

#### 4.1.3 Vetoos for physics backgrounds

Misidentification of kaons and protons as pions leads to exclusive backgrounds in the data sample. These are suppressed by means of explicit *vetoos*.

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.

For the  $D^\mp$  mesons which are reconstructed in the  $D^\mp \rightarrow K^\pm\pi^\mp\pi^\mp$  final state, a  $p \rightarrow \pi$  mis-ID can lead to background contributions from  $\Lambda_b^0 \rightarrow \Lambda_c^\mp (\rightarrow K^\pm\pi^\mp p^\mp)\pi^\pm$ . To reduce these contributions, the proton mass hypothesis is applied separately to both pions 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/ $c^2$  window around the  $\Lambda_c^\mp$  mass,  $m_{\Lambda_c^\mp} = 2286.46$  MeV/ $c^2$  [20], it is required to have  $\text{PIDp} < -8.0$  and if it is inside a 50 MeV/ $c^2$  window around the  $\Lambda_c^\mp$  mass it is required to have  $\text{PIDp} < -5.0$ . A plot showing the distributions before and after applying the veto is given in Fig. 4.2. This requirement shows a signal efficiency of  $(93.48 \pm 0.06)$  %. For the rejection of  $\Lambda_b^0 \rightarrow \Lambda_c^\mp\pi^\pm$ , the efficiency on a signal MC for this decay is checked. After stripping and preselection alone,  $(99.720 \pm 0.004)$  % of the

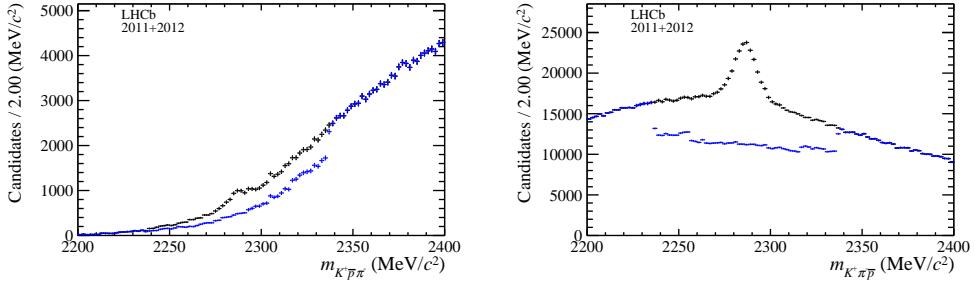


Figure 4.2 – Distributions of the invariant mass of the  $K^\pm\pi^\mp p^\mp$  combinations where each of the two daughter pion of the  $D^\mp$  meson candidate is assigned in turn the proton mass. The distribution is given without (black) and with (blue) the  $\Lambda_c^\mp$  veto described in the text.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

807     $\Lambda_b^0 \rightarrow \Lambda_c^\pm \pi^\mp$  candidates are rejected; this veto rejects  $(76.6 \pm 0.6)\%$  of these remaining  
 808     $\Lambda_b^0 \rightarrow \Lambda_c^\mp \pi^\pm$  candidates.

809    In the same way as protons may be misidentified as pions, it is possible for kaons to  
 810    be misidentified as pions. Such a mis-ID would lead to background contributions from  
 811     $B_s^0 \rightarrow D_s^\mp (\rightarrow K^\pm K^\mp \pi^\mp) \pi^\pm$ . To check for these contributions, the kaon mass hypothesis  
 812    is applied in turn to each of the two pions from the  $D^\pm$  final state. The invariant mass  
 813    of the resulting  $K^\pm K^\mp \pi^\mp$  system is recalculated and plotted for two different ranges of  
 814    the  $B^0$  mass: the first range, from 5330 to 5400 MeV/c<sup>2</sup>, covers the invariant mass signal  
 815    region of the  $B_s^0$  meson as possible background contribution. The second range, from 5500  
 816    to 5700 MeV/c<sup>2</sup> is the upper mass *sideband* for this possible background contamination.  
 817    As can be seen in Fig. 4.3 the distribution which covers the invariant mass of the  $B_s^0$   
 818    meson does not show any significant peaking structure compared to the upper mass  
 819    sideband region. The visible differences are expected as the distributions arise from  
 820    different kinematic ranges. To double check for possible resonant contributions from  
 821    a kaon mis-ID, the decay of the  $D^\mp$  meson after applying the kaon mass hypothesis  
 822    is investigated. Possible resonant decays of the  $D$  meson can take place via a  $K^*$  or  
 823     $\phi$  resonance. These resonances would be visible in the  $K\pi$  and  $KK$  invariant mass  
 824    distributions, which are plotted in the same two ranges in Fig. 4.4. As the distributions  
 825    in the signal and background ranges look compatible, the  $D_s^\mp$  contamination is negligible  
 826    and no veto is applied.

827    In the same way as for the  $D^\mp$  meson daughters, it is also possible that the bachelor  
 828    pion candidate is actually a misidentified kaon. This mis-ID could lead to background  
 829    contributions of  $B^0 \rightarrow D^0 (\rightarrow K^\mp \pi^\pm) K^+ \pi^-$ . The same background contribution could  
 830    arise if one  $D^\mp$  meson daughter pion is misidentified as a kaon and combined with the  
 831    bachelor pion. To check for this contribution, the kaon mass hypothesis is applied to the  
 832    bachelor pion and the  $D$  meson daughter pions, and the invariant mass distributions for

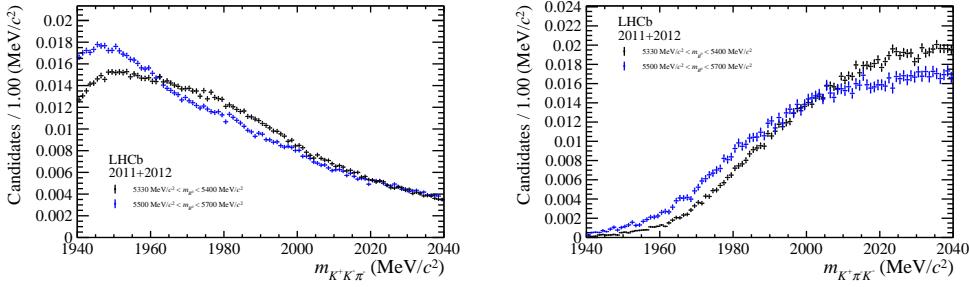


Figure 4.3 – Distributions of the invariant mass of the  $K^\pm K^\mp \pi^\mp$  combinations where each of the two daughter pion of the  $D^\mp$  meson candidate is assigned in turn the kaon mass. The distribution is given in the  $B^0$  meson mass range from  $5300$  to  $5400\text{ MeV}/c^2$  (black) and in the  $B^0$  meson mass range from  $5400$  to  $5500\text{ MeV}/c^2$  (blue) after applying the  $\Lambda_c^\mp$  veto. On the left (right) the kaon mass hypothesis is applied to the pion with lower (higher)  $p_T$ .

833 the four possible  $K\pi$  systems are plotted after applying the BDT classifier described in  
 834 Sec. 4.1.5 (Fig. 4.5). As the distributions show no significant peaking structures, this  
 835 contribution is neglected at this stage and no specific cuts are applied.

#### 836 4.1.4 Wrongly associated primary vertices

837 Given an average number of  $pp$  interactions per bunch crossing of  $\nu = 2.5$ , a large fraction  
 838 of events have more than one reconstructed PV. The PV to which the  $B^0$  candidate has  
 839 the smallest  $\chi_{\text{IP}}^2$  (*best* PV) is chosen. Furthermore, it is possible that more than one  $B^0$   
 840 candidate is reconstructed per event: in this case, the  $B^0$  candidate is chosen randomly  
 841 following the prescription of Ref. [61], which prevents any unexpected bias due to a more  
 842 specific choice.

843 In events where the association of the  $B^0$  candidate to the best PV  $B^0$  is wrong, the  
 844 reconstructed decay time of this candidate will be incorrect. These wrongly associated  
 845 candidates are causing a large tail in the decay time distribution, which can be clearly  
 846 observed in signal MC where the true decay time is known. Giving each candidate a  
 847 weight equal to  $e^{t/\tau}$ , where  $\tau$  is the true lifetime, leads to an excess of candidates at  
 848 high decay times. To remove these incorrect associations in MC, one can compare the  $z$   
 849 position of the associated PV with the  $z$  position of the true PV and reject the candidate  
 850 if the distance between those positions exceeds 5 times its uncertainty (Fig. 4.6). In real  
 851 data, the true PV is unknown, so a selection cut on the smallest  $B^0$  impact parameter  
 852  $\chi_{\text{DTF,PV}}^2$  (called  $\text{MinIP}\chi^2$ ) with respect to any other PV in the event is applied (Fig. 4.7).  
 853 Events with only a single PV are not affected. The main advantage of this  $\text{MinIP}\chi^2$   
 854 variable is that all PVs are treated equally, without any biasing choice. The cut on  
 855  $\text{MinIP}\chi^2$  is optimised to retain 98 % of the truth-matched signal candidates in MC. The

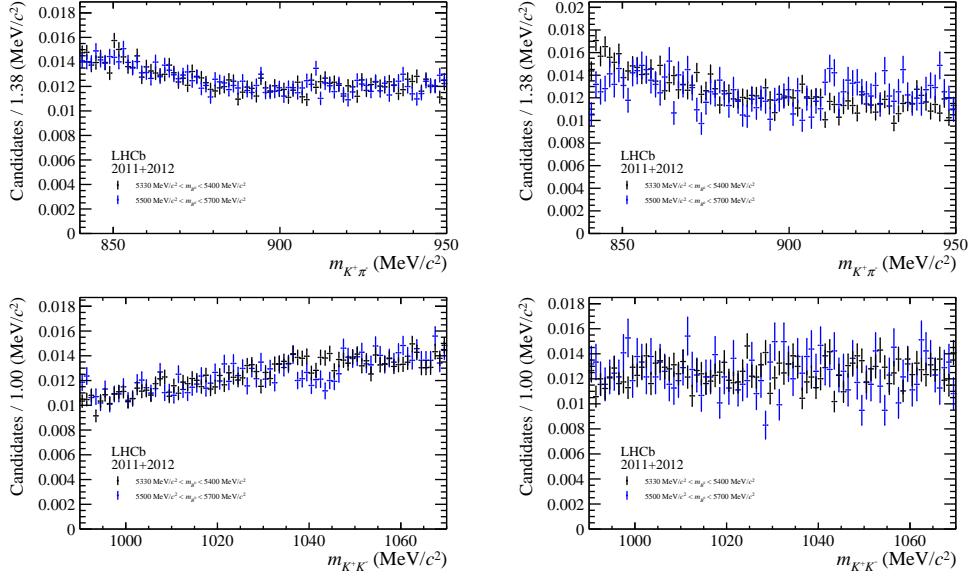


Figure 4.4 – Distributions of the invariant mass of the  $K^\mp\pi^\pm$  combination (top) and for the  $K^\mp K^\pm$  combination (bottom), where for the latter each of the two daughter pion of the  $D^\mp$  meson candidate is assigned in turn the kaon mass. The distributions are shown in the  $B^0$  meson mass range of [5300, 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 lower (higher)  $p_T$ .

optimal requirement is then found at  $\text{MinIP}\chi^2 > 16.5$ . A plot showing the signal MC weighted decay time distribution after applying this cut is given in Fig. 4.7.

#### 4.1.5 Development of an MVA classifier

The combinatorial background, consisting of candidates created from random combinations of tracks, is rejected by using a Boosted Decision Tree (BDT) classifier [62, 63]. The signal input to the training stage consists of signal MC candidates simulated under 2012 data-taking conditions, while the upper mass sideband above 5500 MeV/ $c^2$  from the 2012 data sample is used as template for the combinatorial background. The BDT is trained on one half of these samples, the other half being used to test its performance. Before the BDT training, all previous selection steps (the cut-based preselection, the mass vetoes and the wrongly associated PV veto) are applied. To reduce the number of input features, the ones with a correlation larger than 97% with any other feature are removed. The 16 final input features are listed in Table 4.5. The correlation matrices between the input features in the signal and the background samples are shown in Fig. 4.8, while the distributions of the input features can be found in Appendix C.1.

The BDT implementation of TMVA [64] is used. The BDT is built out of 1700 trees,

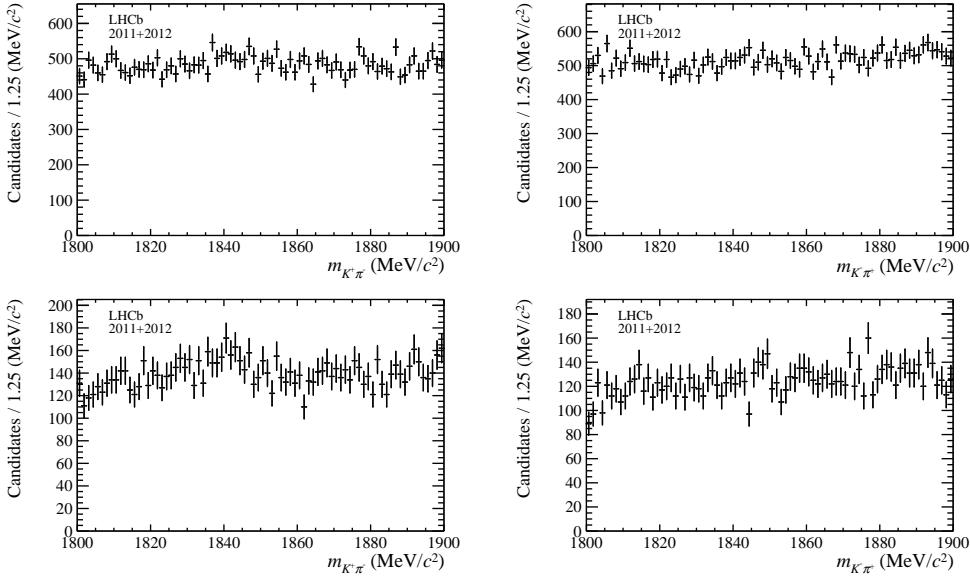


Figure 4.5 – Distributions of the invariant mass of the four possible  $K^\mp \pi^\pm$  combinations. In the top (bottom) plots the bachelor pion is combined with the pions from the  $D^\mp$  meson with lower (higher)  $p_T$ . In the left (right) plots the kaon mass hypothesis is applied to the bachelor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

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

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

with a depth limited to four. For each node, at least 2.5 % of the training events have to be present. The features are scanned at 20 points to find the optimal cut value. The chosen boosting method is the AdaBoost [65] algorithm with a boost factor  $\beta = 0.5$ . The

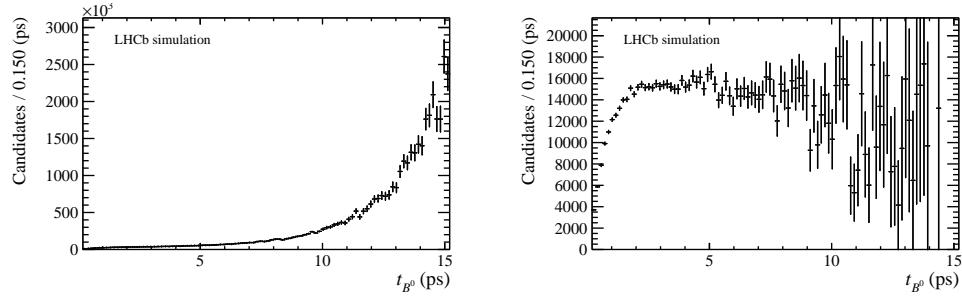


Figure 4.6 – Left: decay time distribution of signal MC events weighted with  $e^{t/\tau}$ , where  $\tau$  is the true lifetime. At high decay times an excess of candidates can be observed. Right: same distribution after requiring that the absolute difference between the best PV  $z$  position and the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

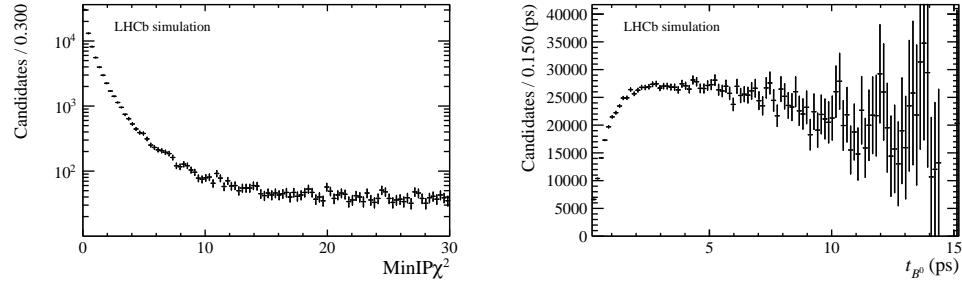


Figure 4.7 – Left: distribution of the smallest  $B^0$  impact parameter  $\text{MinIP}\chi^2$  with respect to any other primary vertex in the event for signal MC events. Right: decay time distribution of signal MC events weighted with  $e^{t/\tau}$ , where  $\tau$  is the true lifetime, after requiring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

number of trees and the maximal depth of trees have been increased iteratively until no significant increase of the performance without overtraining was observed. The BDT is tested on the events that are not used in the training. The plot of this overtraining check is given in Fig. 4.9.

#### 4.1.6 BDT selection optimisation

To estimate the best requirement on the output of the BDT classifier, the statistical uncertainty of the  $CP$  coefficients derived from the analysis of simulated samples is used as the figure of merit (FoM). To determine the sensitivity, the preselection, the mass vetoes and the wrongly associated PV veto are applied and the BDT classifier is calculated for every candidate. The BDT cut point is then scanned with a step size of 0.01 from -0.15 to 0.10 and a step size of 0.05 in the outer regions. For each cut point of the BDT

## 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 (left) and background (right).

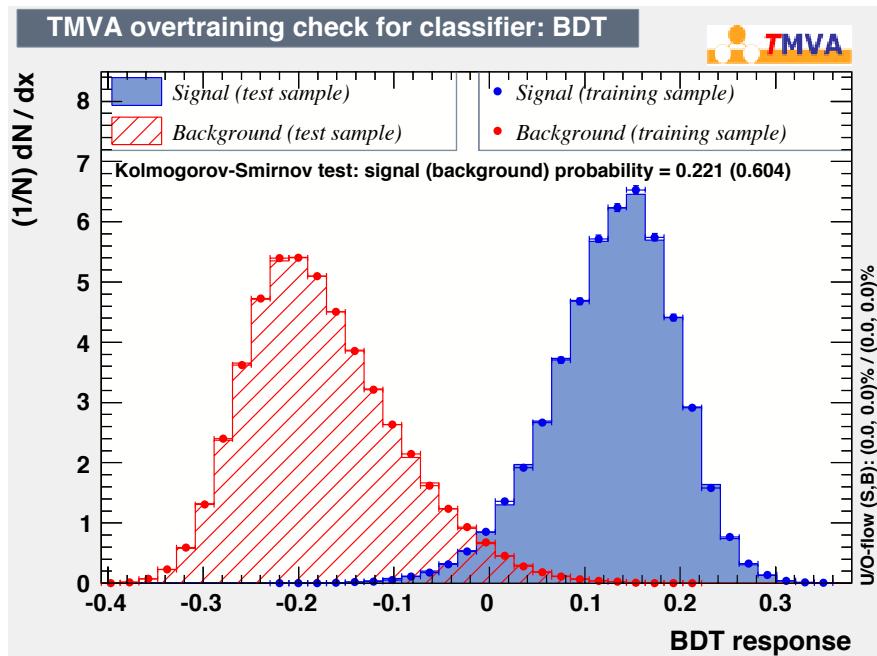


Figure 4.9 – Distributions of the BDT response on training and test samples.

886 classifier, a simulated (*toy*) sample is generated. This simulated sample contains the same  
 887 signal and combinatorial background yields as determined from the real dataset via a  
 888 maximum likelihood fit of the  $B^0$  mass distribution. Finally, a time-dependent analysis of  
 889 each toy dataset is performed in order to estimate the statistical uncertainty on  $S_f$  and  
 890  $S_{\bar{f}}$ . These statistical uncertainties as a function of the BDT cut are shown in Fig. 4.10.  
 Based on these distributions, the BDT cut point is chosen to be at 0.05.

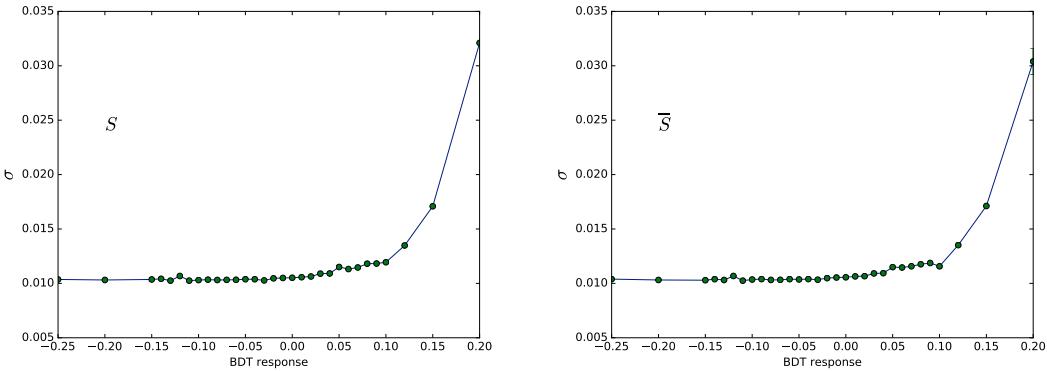


Figure 4.10 – Statistical uncertainty on of  $S_f$  (left) and  $S_{\bar{f}}$  (right) as a function of the BDT output classifier, as obtained from simulated samples

891

#### 892 4.1.7 Multiple candidates

893 After the stripping selection and trigger requirements, approximately 9% of events contain  
 894 at least two  $B^0$  candidates, and 18 – 20% of all  $B^0$  candidates share an event. If the  
 895 offline selection is also applied, around 0.4% of events contain multiple  $B^0$  candidates,  
 896 and 0.8% of all  $B^0$  candidates share an event. More details are given in Appendix C.2. In  
 897 order to be consistent with the prescription used in the stripping and trigger requirements,  
 898 only the best PV is chosen; all events in which the best PV is no longer present after the  
 899 offline selection are removed. Finally, since the remaining  $B^0$  candidates are considered to  
 900 be equally likely signal candidates, as single  $B^0$  candidate per event is chosen randomly.

#### 901 4.1.8 Selection performance

902 The offline selection performances are listed in Table 4.6. They are determined by using  
 903 data candidates of the 2012 sample with an invariant  $B^0$  mass above  $5500 \text{ MeV}/c^2$  to  
 904 represent combinatorial background, and signal MC candidates (see Sec. 4.2) to represent  
 905 the signal. In Fig. 4.11 a comparison of the mass distribution of the selected 2011 and  
 906 2012 data is shown. Additionally, the BDT performances are quoted in Table 4.7 split by  
 907 both magnet polarity and year of data taking. Finally, in order to check the contribution

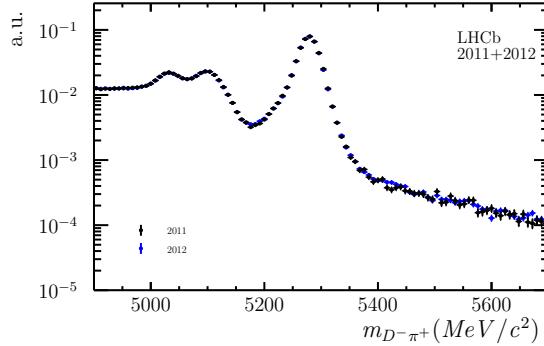


Figure 4.11 – Mass distributions of the  $B^0 \rightarrow D^\mp \pi^\pm$  candidates passing the full offline selection for the 2011 (blue) and 2012 (black) data samples.

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	$(98.96 \pm 0.03) \%$	$(7.66 \pm 0.03) \%$
mass vetoes combined	$(92.51 \pm 0.07) \%$	$(16.77 \pm 0.04) \%$
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	$(70.7 \pm 0.1) \%$	$(99.911 \pm 0.002) \%$

of non-resonant  $B^0 \rightarrow K^+ \pi^- \pi^- \pi^+$  decays, the  $B^0$  and  $D^\mp$  invariant mass distributions are analysed after applying the full offline selection in two ways. First, the  $D^\mp$  mass distribution is plotted for candidates falling in a  $B^0$  mass signal window (Fig. 4.12). From this plot, the maximal contamination from non-resonant decays can be estimated to be roughly 1 %. Then, the  $B^0$  distribution after excluding the  $D^-$  signal window is plotted. To quantify the non-resonant  $B^0 \rightarrow K^+ \pi^- \pi^- \pi^+$  decays, the sum of an exponential and a Gaussian with a fixed shape is used to fit the resulting  $B^0$  mass distribution, as shown in Fig. 4.13. As the fitted  $B^0$  yield is  $645 \pm 242$ , the non-resonant contribution is assumed to be negligible.

## 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
$\varepsilon_{\text{sig+bkg}}$	$(50.67 \pm 0.08)\%$	$(50.04 \pm 0.07)\%$	$(46.19 \pm 0.04)\%$	$(46.32 \pm 0.04)\%$

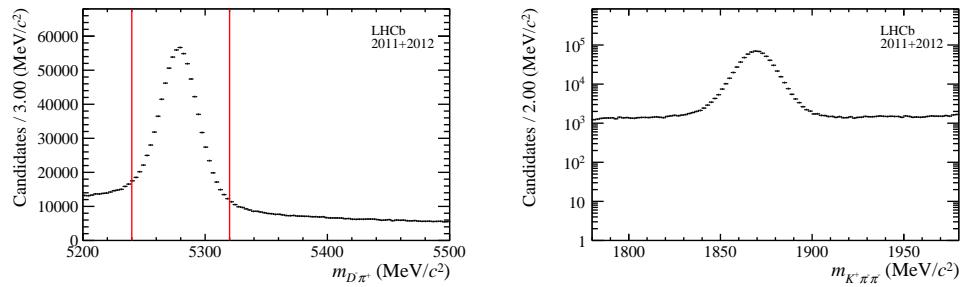


Figure 4.12 – Left:  $B^0$  mass distribution with red vertical lines indicating the selected signal window. Right: resulting  $D^\mp$  mass distribution in the  $B^0$  signal window.

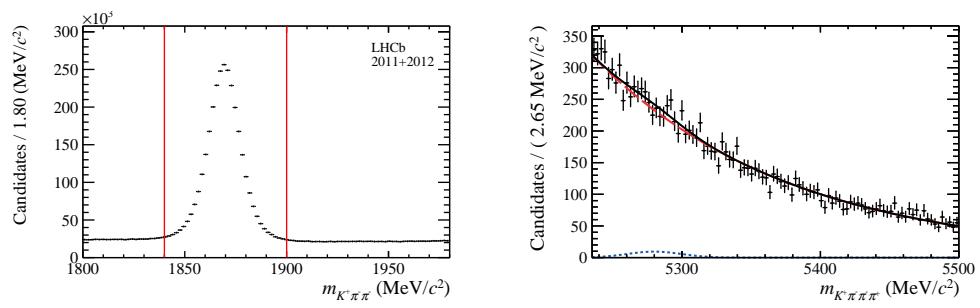


Figure 4.13 – Left:  $D^\mp$  mass distribution with red vertical lines indicating the excluded range. Right:  $B^0$  mass distribution outside the  $D^\mp$  mass window with the fitting function overlaid.

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	$0.721 \pm 0.006$
$B_s^0 \rightarrow D_s^\mp\pi^\pm$	13264021	7942	$0.1531 \pm 0.0017$
$\Lambda_b^0 \rightarrow \Lambda_c^\mp\pi^\pm$	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 selection, and the selection efficiencies. Efficiencies include generator level, stripping, offline selection and tagging efficiencies.

## 917 4.2 Simulation and expected sample composition

918 Simulated samples are used to (i) gain a detailed overview of all sources of  $b$ -hadron  
 919 decays that contribute to the sample and (ii) model the relevant distributions studied  
 920 in the analysis. Simulated data undergoes the same processing of reconstruction and  
 921 selection as real data. Each sample is split into "up" and "down" magnet polarities and  
 922 into 2011 and 2012 data taking conditions subsamples, in proportions similar to those  
 923 present in real data.

924 The used simulated samples are listed in Table 4.8, together with the number of true  
 925 signal events passing the final selection and the corresponding total efficiencies. The PID  
 926 requirements on the bachelor pion are not applied in order to compute these efficiencies.  
 927

### 928 4.2.1 PIDK correction

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

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

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

## 4.2. Simulation and expected sample composition

---

941 pseudorapidity,  $\eta$ , is defined. For each bin, the corresponding PIDK distribution in the  
942 calibration sample is built and for each event in the MC sample, a random PIDK value  
943 is sampled from the PIDK distribution associated with the corresponding bin in the  
944 calibration sample. More details are given in Appendix D.

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

### 948 4.2.2 Surviving physics backgrounds

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

- 951     •  $B^0 \rightarrow D^\mp K^\pm$ : Peaking background due to the bachelor kaon being wrongly  
952         identified as a pion.
- 953     •  $B^0 \rightarrow D^\mp \rho^\pm (\rightarrow \pi^\pm \pi^0)$ : Low mass background due to a missing neutral pion in  
954         the reconstruction.
- 955     •  $B^0 \rightarrow D^\mp (\rightarrow D^\mp \gamma/\pi^0) \pi^\pm$ : Low mass background due to a missing neutral  
956         particle in the reconstruction.

957 In the kaon sample, the following backgrounds are expected:

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

964 The background fractions expected in the pion sample with respect to the  $B^0 \rightarrow D^\mp \pi^\pm$   
965 signal are reported in Table 4.9. These fractions are computed using the branching  
966 fractions of the expected decay as inputs and from the ratio of efficiencies estimated from  
967 MC and corrected as described in Sec. 4.2.1. When relevant we consider also the ratio of  
968 the fragmentation probabilities of  $b$  quarks to different  $b$  hadrons. These expectations  
969 will be compared with the results from the fit to data described in the next section.

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

Decay	$\mathcal{B}$ [%]	$\epsilon_{\text{bkg}}$ [%]	$f_{B^0/B_s^0/\Lambda_b^0}$ [%]	$f_{\text{bkg}}$ [%]
$B^0 \rightarrow D^\mp K^\pm$	$0.00186 \pm 0.00020$	$0.684 \pm 0.008$	$33.9 \pm 3.9$	$2.61 \pm 0.31$
$B^0 \rightarrow D^\mp \rho^\pm (\rightarrow \pi^\pm \pi^0)$	$0.071 \pm 0.011$	$0.1149 \pm 0.0024$	$33.9 \pm 3.9$	$16.7 \pm 2.8$
$B^0 \rightarrow D^{\mp*} (\rightarrow D^\mp \gamma/\pi^0) \pi^\pm$	$0.0080 \pm 0.0004$	$0.705 \pm 0.006$	$33.9 \pm 3.9$	$11.6 \pm 0.8$
$\Lambda_b^0 \rightarrow \Lambda_c^\pm (K^\pm \pi^\mp p^\mp) \pi^\pm$	$0.032 \pm 0.004$	$0.0150 \pm 0.0008$	$21.2 \pm 6.9$	$0.62 \pm 0.24$
$B_s^0 \rightarrow D_s^\mp \pi^\pm$	$0.0164 \pm 0.0014$	$0.1493 \pm 0.0017$	$11.1 \pm 1.4$	$1.64 \pm 0.32$

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  $(0.254 \pm 0.014)\%$  and  $(1.924 \pm 0.006)\%$  respectively.

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

---

### 976 4.3 Fits to the $B^0$ invariant mass

977 The *sPlot* technique [66] is applied in order to statistically disentangle signal from  
 978 background for the subsequent decay time fit. The  $D^\mp\pi^\pm$  invariant mass, where the  $D^\mp$   
 979 mass is constrained to its known value in order to improve the mass resolution, is adopted  
 980 as discriminating observable thanks to its small correlation with the  $B^0$  decay time (see  
 981 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 for the pion sample and the kaon  
 sample (see Sec. 4.1.2). This approach is adopted in order to control the contamination of  
 $B^0 \rightarrow D^\mp K^\pm$  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:

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

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

982 The quantities  $\epsilon_{\text{PID}}(B^0 \rightarrow DX)_Y$  are the fractions of true  $B^0 \rightarrow D^\mp X^\pm$  candidates that  
 983 are selected in the  $Y$  sample by applying the relative PIDK cut. These fractions (or  
 984 efficiencies) are estimated on  $B^0 \rightarrow D^\mp\pi^\pm$  and  $B^0 \rightarrow D^\mp K^\pm$  MC samples where the  
 985 PIDK distributions are resampled from calibration data, as described in Sec. 4.2.1; the  
 986 results of these estimations are reported in Table 4.10.

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	PIDK requirement	fraction
$B^0 \rightarrow D^\mp\pi^\pm$	< 5 ( $\pi$ sample)	$0.979 \pm 0.004$
$B^0 \rightarrow D^\mp\pi^\pm$	> 5 ( $K$ sample)	$0.021 \pm 0.001$
$B^0 \rightarrow D^\mp K^\pm$	< 5 ( $\pi$ sample)	$0.373 \pm 0.005$
$B^0 \rightarrow D^\mp K^\pm$	> 5 ( $K$ sample)	$0.637 \pm 0.007$

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

### 4.3.1 Probability density functions

Before Fit A, the PDFs used to describe both the pion and kaon sample components are first estimated on MC samples. The parameters of the combinatorial background PDFs are instead determined directly on data. The PDFs used for the pion sample are:

- $B^0 \rightarrow D^\mp\pi^\pm$ : sum of a double-sided Hypatia [67] and a Johnson SU [68] functions ( $\text{PDF}_{B^0 \rightarrow D\pi}^\pi$ ).
- $B^0 \rightarrow D^\mp K^\pm$ : double-sided Hypatia function ( $\text{PDF}_{B^0 \rightarrow DK}^\pi$ ).
- $B^0 \rightarrow D^\mp\rho^\pm$ : Johnson SU function ( $\text{PDF}_{B^0 \rightarrow D\rho}^\pi$ ).
- $B^0 \rightarrow D^\mp\pi^\pm$ : sum of a single-sided Crystal Ball function [69] and a Gaussian function ( $\text{PDF}_{B^0 \rightarrow D^*\pi}^\pi$ ).
- Background: sum of an exponential function and a constant function.

For the kaon sample they are:

- $B^0 \rightarrow D^\mp\pi^\pm$ : double-sided Hypatia function ( $\text{PDF}_{B^0 \rightarrow D\pi}^K$ ).
- $B^0 \rightarrow D^\mp K^\pm$ : single-sided Hypatia function ( $\text{PDF}_{B^0 \rightarrow DK}^K$ ).
- $B^0 \rightarrow D^\mp\rho^\pm$ : double Gaussian function ( $\text{PDF}_{B^0 \rightarrow D\rho}^K$ ).
- $B^0 \rightarrow D^\mp K^{\pm*}$ : Gaussian function ( $\text{PDF}_{B^0 \rightarrow DK^*}^K$ ).
- Background: sum of an exponential function and a constant function.

The definitions of all the PDFs listed above are reported in Appendix F. The fits to the MC samples are shown in Figs. 4.14 and 4.15. The parameters obtained from these fits that are fixed in the data fits are listed in Tables 4.11 and 4.12.

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

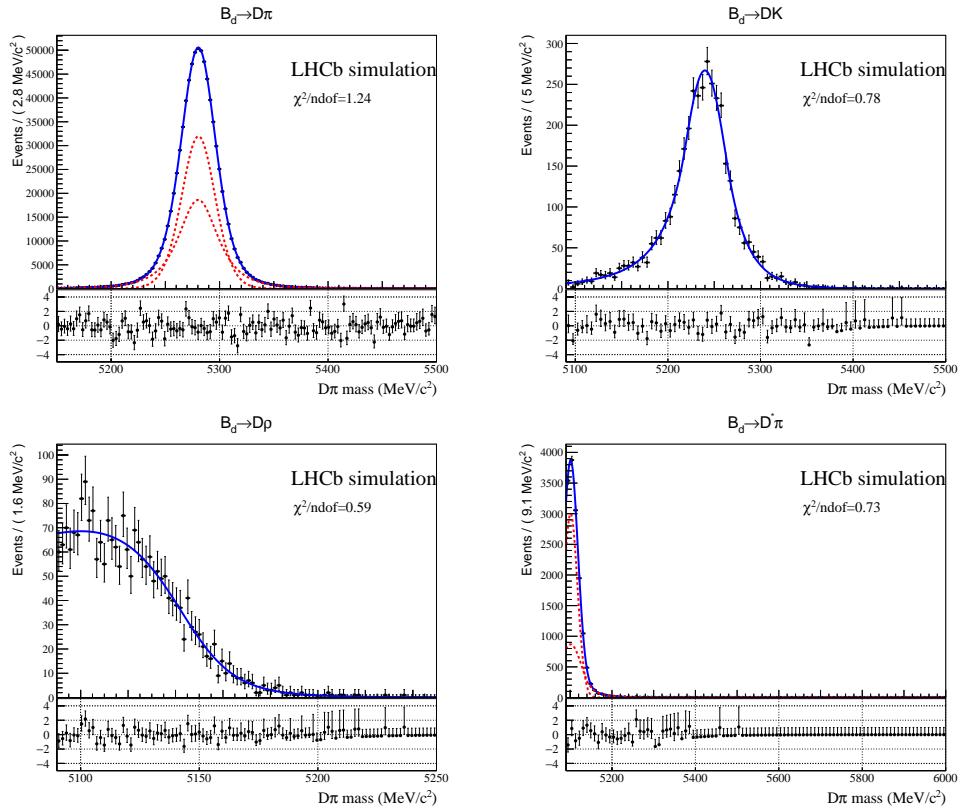


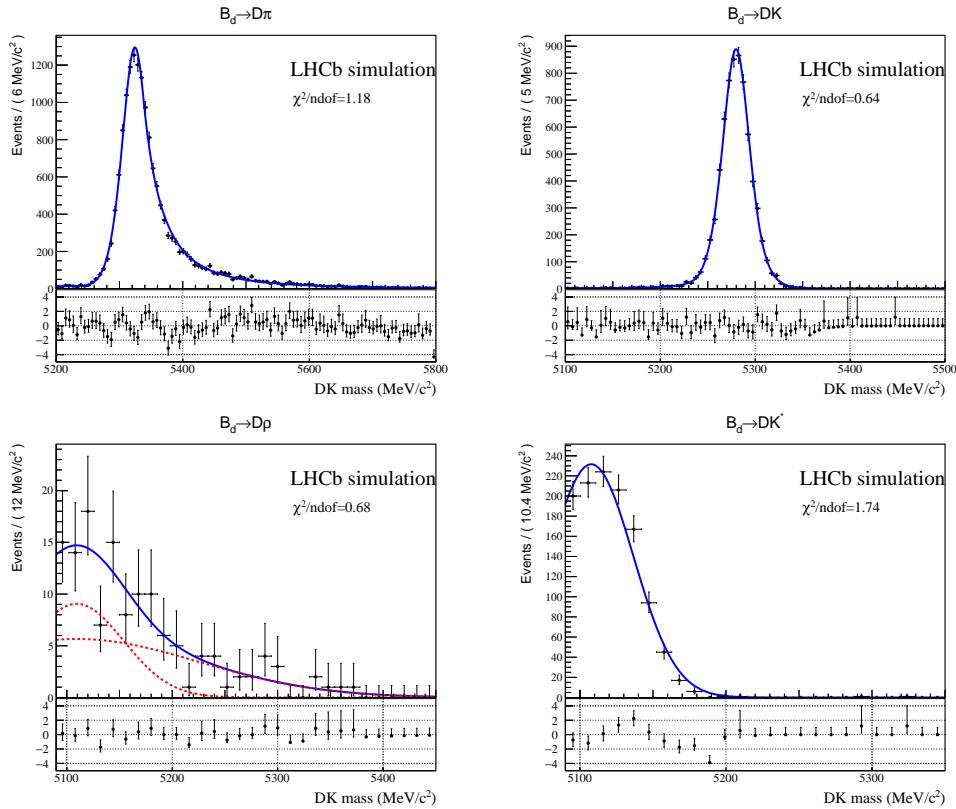
Figure 4.14 – Fits to MC samples of PDFs describing the  $\pi$  sample.

#### <sup>1013</sup> 4.3.2 Fit to data

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

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

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


 Figure 4.15 – Fits to MC samples of PDFs describing the  $K$  sample.

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

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

$$\mathcal{L}_K = \frac{e^{-N_{K,\text{exp}}} (N_{K,\text{exp}})^{N_{K,\text{obs}}}}{N_{K,\text{obs}}!} \prod_{i=1}^{N_{K,\text{obs}}} f_K(m_i), \quad (4.6)$$

where  $N_{\pi,\text{exp}}$  ( $N_{\pi,\text{obs}}$ ) is the total number of expected (observed) candidates in the pion sample, while  $N_{K,\text{exp}}$  ( $N_{K,\text{obs}}$ ) is the number of expected (observed) candidates in the kaon sample. The product  $\mathcal{L}_\pi \mathcal{L}_K$  is maximised during the fit.

The following strategy is adopted to perform Fit A:

- The mean and width parameters ( $\mu_{B^0 \rightarrow D\pi}^\pi, \sigma H_{B^0 \rightarrow D\pi}^\pi, \sigma J_{B^0 \rightarrow D\pi}^\pi, \mu_{B^0 \rightarrow DK}^K, \sigma_{B^0 \rightarrow DK}^K$ ) of  $\text{PDF}_{B^0 \rightarrow D\pi}^\pi$  and  $\text{PDF}_{B^0 \rightarrow DK}^K$  are floated in the fit.
- 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 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 MC and both multiplied by a floating scale factor  $sa_{B^0 \rightarrow D\pi}^\pi$ ; the same constraint

is applied to  $n1_{B^0 \rightarrow D\pi}^\pi$ ,  $n2_{B^0 \rightarrow D\pi}^\pi$ , 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  $(\mu_{B^0 \rightarrow D^*\pi}^\pi, \mu_{B^0 \rightarrow D\rho}^{\pi/K})$  of  $\text{PDF}_{B^0 \rightarrow D^*\pi}^\pi$ ,  $\text{PDF}_{B^0 \rightarrow D\rho}^{\pi/K}$ , 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  $B^0 \rightarrow D\pi$  peak in the  $\pi$  sample/ $B^0 \rightarrow DK$  peak in the  $K$  sample is denoted as  $\Delta\mu_{\text{comp}}^{K/\pi}$ . The mean parameters  $(\mu_{B^0 \rightarrow DK^*}^K, \mu_{B^0 \rightarrow DK}^K, \mu_{B^0 \rightarrow D\pi}^K)$  of  $\text{PDF}_{B^0 \rightarrow DK^*}^K$ ,  $\text{PDF}_{B^0 \rightarrow DK}^K$ ,  $\text{PDF}_{B^0 \rightarrow D\pi}^K$  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  $\text{PDF}_\pi$  and  $\text{PDF}_K$  in the  $D^\mp \pi^\pm$  and  $D^\mp K^\pm$  invariant mass observables (Fit A) are shown in Fig. 4.16, 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  $B^0 \rightarrow D^{\mp*} \pi^\pm$  in the pion sample are compared with the expected yields from Table 4.9, 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. 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  $B^0 \rightarrow D^{\mp*} \pi^\pm$  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 follows:

$$f_\pi(m) = N_{B^0 \rightarrow D\pi}^\pi \text{PDF}_{B^0 \rightarrow D\pi}^\pi + N_{\text{bkg}}^\pi \text{PDF}_{\text{bkg}}^\pi. \quad (4.7)$$

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

Parameter	Value	Comment
$a1_{B^0 \rightarrow D\pi}^\pi$	$0.722 \pm 0.091$	fitted in MC and constrained in data
$a2_{B^0 \rightarrow D\pi}^\pi$	$0.96 \pm 0.12$	fitted in MC and constrained in data
$n1_{B^0 \rightarrow D\pi}^\pi$	$5.92 \pm 0.92$	fitted in MC and constrained in data
$n2_{B^0 \rightarrow D\pi}^\pi$	$5.83 \pm 0.38$	fitted in MC and constrained in data
$\beta_{B^0 \rightarrow D\pi}^\pi$	0.0	fixed
$\lambda_{B^0 \rightarrow D\pi}^\pi$	$-1.240 \pm 0.060$	fitted in MC
$\zeta_{B^0 \rightarrow D\pi}^\pi$	0.0	fixed
$f_{B^0 \rightarrow D\pi}^\pi$	$0.436 \pm 0.060$	fixed
$\sigma_{B^0 \rightarrow DK}^\pi$	$23.43 \pm 0.42 \text{ MeV}/c^2$	fitted in MC
$a1_{B^0 \rightarrow DK}^\pi$	$0.898 \pm 0.025$	fitted in MC
$a2_{B^0 \rightarrow DK}^\pi$	$1.092 \pm 0.033$	fitted in MC
$n1_{B^0 \rightarrow DK}^\pi$	$3.83 \pm 0.40$	fitted in MC
$n2_{B^0 \rightarrow DK}^\pi$	$22.0 \pm 7.6$	fitted in MC
$\beta_{B^0 \rightarrow DK}^\pi$	0.0	fixed
$\lambda_{B^0 \rightarrow DK}^\pi$	$-24 \pm 10$	fitted in MC
$\zeta_{B^0 \rightarrow DK}^\pi$	0.0	fixed
$\nu_{B^0 \rightarrow D\rho}^\pi$	$-2.01 \pm 0.15$	fitted in MC
$\mu_{B^0 \rightarrow D\rho}^\pi$	$4828 \pm 80 \text{ MeV}/c^2$	fitted in MC and constrained into $\Delta\mu_{B^0 \rightarrow D\rho}^\pi$
$\sigma_{B^0 \rightarrow D\rho}^\pi$	$550 \pm 190 \text{ MeV}/c^2$	fitted in MC
$\tau_{B^0 \rightarrow D\rho}^\pi$	$1.163 \pm 0.090$	fitted in MC
$\alpha_{B^0 \rightarrow D^*\pi}^\pi$	$-1.443 \pm 0.031$	fitted in MC
$n_{B^0 \rightarrow D^*\pi}^\pi$	$4.65 \pm 0.30$	fitted in MC
$\mu_{B^0 \rightarrow D^*\pi}^\pi$	$5100.93 \pm 0.23 \text{ MeV}/c^2$	fitted in MC and constrained into $\Delta\mu_{B^0 \rightarrow D^*\pi}^\pi$
$\sigma G_{B^0 \rightarrow D^*\pi}^\pi$	$16.52 \pm 0.20 \text{ MeV}/c^2$	fitted in MC
$\sigma C B_{B^0 \rightarrow D^*\pi}^\pi$	$25.84 \pm 0.48 \text{ MeV}/c^2$	fitted in MC
$f_{B^0 \rightarrow D^*\pi}^\pi$	$0.302 \pm 0.011$	fitted in MC

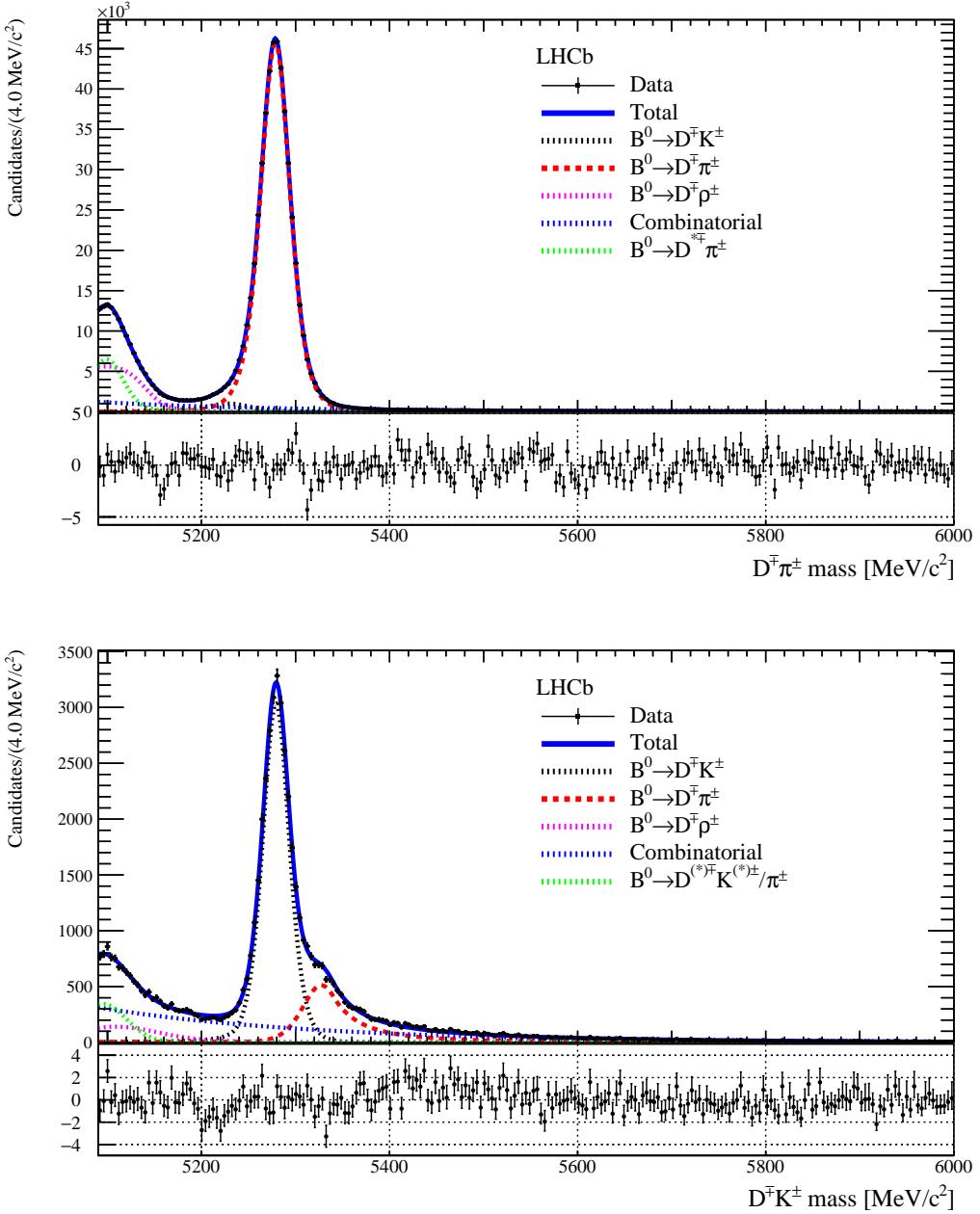


Figure 4.16 – Top:  $D^\mp\pi^\pm$  mass distribution of the  $\pi$  sample. Bottom:  $D^\mp K^\pm$  mass distribution of the  $K$  sample. The result of the simultaneous fit (Fit A) to both samples is superimposed. The plot below each histogram shows the normalised fit residuals (data minus fit divided by fit error).

The  $N_{\text{bkg}}^\pi$  coefficient describes the total number of background events in the new range. The PDF<sub>bkg</sub> <sup>$\pi$</sup>  term is defined as:

$$\begin{aligned} \text{PDF}_{\text{bkg}}^\pi = & f_{\text{comb}}^\pi \text{PDF}_{\text{comb}}^\pi \\ & + f_{B^0 \rightarrow DK}^\pi \text{PDF}_{B^0 \rightarrow DK}^\pi + f_{B^0 \rightarrow D\rho}^\pi \text{PDF}_{B^0 \rightarrow D\rho}^\pi \\ & + (1 - f_{\text{comb}}^\pi - f_{B^0 \rightarrow DK}^\pi - f_{B^0 \rightarrow D\rho}^\pi) \text{PDF}_{B^0 \rightarrow D^*\pi}^\pi. \end{aligned} \quad (4.89)$$

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

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

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

$$f_j^\pi = \frac{N_j^\pi \int_{5220 \text{ MeV}/c^2}^{5600 \text{ MeV}/c^2} \text{PDF}_j^\pi dm}{\sum_i N_i^\pi}, \quad (4.9)$$

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

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.

Fit B is used as starting point to apply the *sPlot* technique and extract *sWeights* used to subtract the total background component from the  $\pi$  sample. The projection of the fitted  $f_\pi(m)$  in Fit B, and a comparison between the weighted and unweighted datasets projected over the  $B^0$  decay time and  $D^\mp$  invariant mass observables, are reported in Fig. 4.17. The distribution of *sWeights* is shown in Fig. 4.18.

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

Table 4.13 – Results of Fit A.

Parameter	Fitted value	Reference
$\mu_{B^0 \rightarrow DK}^\pi$	$5228.62 \pm 0.92$	$5239.96 \pm 0.52$
$\sigma_{B^0 \rightarrow DK}^\pi$	$17.17 \pm 0.15$	$17.32 \pm 0.26$
$\mu_{B^0 \rightarrow D\pi}^\pi$	$5094.8 \pm 3.9$	$5107.7 \pm 2.2$
$\sigma_{B^0 \rightarrow D\pi}^\pi$	$25.5 \pm 2.6$	$28.9 \pm 1.2$
$c1_{\text{comb}}^\pi$	$-0.00576 \pm 0.00017$	/
$c2_{\text{comb}}^\pi$	$-0.0010 \pm 0.0010$	/
$f_{\text{comb}}^\pi$	$0.899 \pm 0.025$	/
$c_{\text{comb}}^K$	$-0.004397 \pm 0.000066$	/
$\mu_{B^0 \rightarrow DK}^K$	$5279.19 \pm 0.14$	$5279.84 \pm 0.20$
$\mu_{B^0 \rightarrow D\pi}^K$	$5278.360 \pm 0.032$	$5280.291 \pm 0.025$
$sa_{B^0 \rightarrow D\pi}^\pi$	$0.684 \pm 0.022$	/
$sn_{B^0 \rightarrow D\pi}^\pi$	$2.71 \pm 0.80$	/
$\sigma H_{B^0 \rightarrow D\pi}^\pi$	$37.69 \pm 0.69$	$43.9 \pm 4.9$
$\sigma J_{B^0 \rightarrow D\pi}^\pi$	$17.01 \pm 0.17$	$16.82 \pm 0.15$
$\mu_{B^0 \rightarrow D\pi}^K$	$5327.32 \pm 0.78$	$5324.37 \pm 0.44$
$\epsilon_{\text{PID}}(B^0 \rightarrow DK)_K$	$0.6197 \pm 0.0079$	$0.637 \pm 0.013$
$\epsilon_{\text{PID}}(B^0 \rightarrow D\pi)_\pi$	$0.98048 \pm 0.00041$	$0.979 \pm 0.004$
$N_{B^0 \rightarrow DK}^K$	$28820 \pm 242$	/
$N_{B^0 \rightarrow D\pi}^K$	$3164 \pm 110$	/
$N_{B^0 \rightarrow D\rho}^\pi$	$73766 \pm 1239$	/
$N_{B^0 \rightarrow D^* \pi}^\pi$	$52494 \pm 819$	/
$N_{\text{comb}}^K$	$17469 \pm 341$	/
$N_{\text{comb}}^\pi$	$56230 \pm 1336$	/
$N_{B^0 \rightarrow D\pi}^\pi$	$483398 \pm 1040$	/

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. The second columns report the yields obtained from Fit B, whereas 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)
$N_{B^0 \rightarrow D\pi}^\pi$	$479045 \pm 732$	$483398 \pm 1040$
$N_{\text{bkg}}^\pi$	$34381 \pm 300$	$34615 \pm 664$

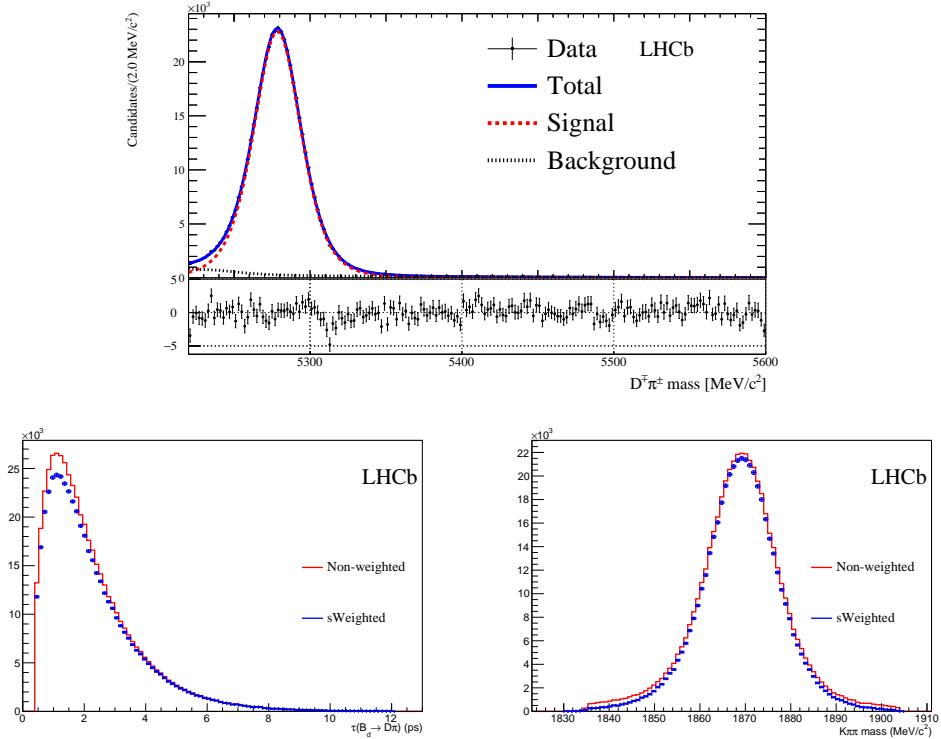


Figure 4.17 – Top:  $D^\mp\pi^\pm$  mass distribution of the  $\pi$  sample with the results of Fit B superimposed. The plot below the histogram shows the normalised fit residuals (data minus fit divided by fit error). Bottom:  $B^0$  decay time (left) and  $K^\mp\pi^\pm\pi^\pm$  mass (right) distributions of the  $\pi$  sample in the  $B^0$  mass region ( $[5220, 5600]$  MeV/c $^2$ ), with (blue) and without (red) *sWeights* from Fit B.

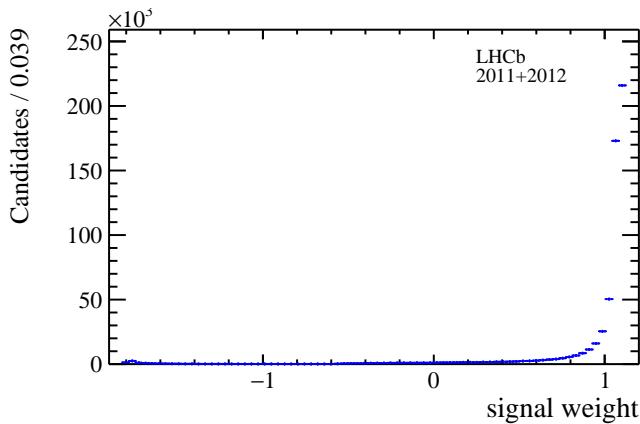


Figure 4.18 – Distribution of *sWeights* obtained from Fit B.

## 1060 4.3.4 Fits of subsamples

1061 In order to validate the data sample, selection and fit procedure, the Fit A is repeated in  
 1062 smaller subsamples. These subsamples are divided per year of data taking (2011, 2012),  
 1063 magnet polarity (up, down) and final state ( $D^+\pi^-$ ,  $D^-\pi^+$ ).

1064 In order to cope with the reduced statistics in the 2011 subsample, the combinatorial  
 1065 background PDF of the  $K$  sample ( $\text{PDF}_{\text{comb}}^K$ ) is taken as simple exponential (instead of  
 1066 an exponential plus constant function).

1067 The projections of the fitted PDFs describing the pion and kaon samples for each data  
 subsample are shown in Figs. 4.19 and 4.20, respectively.

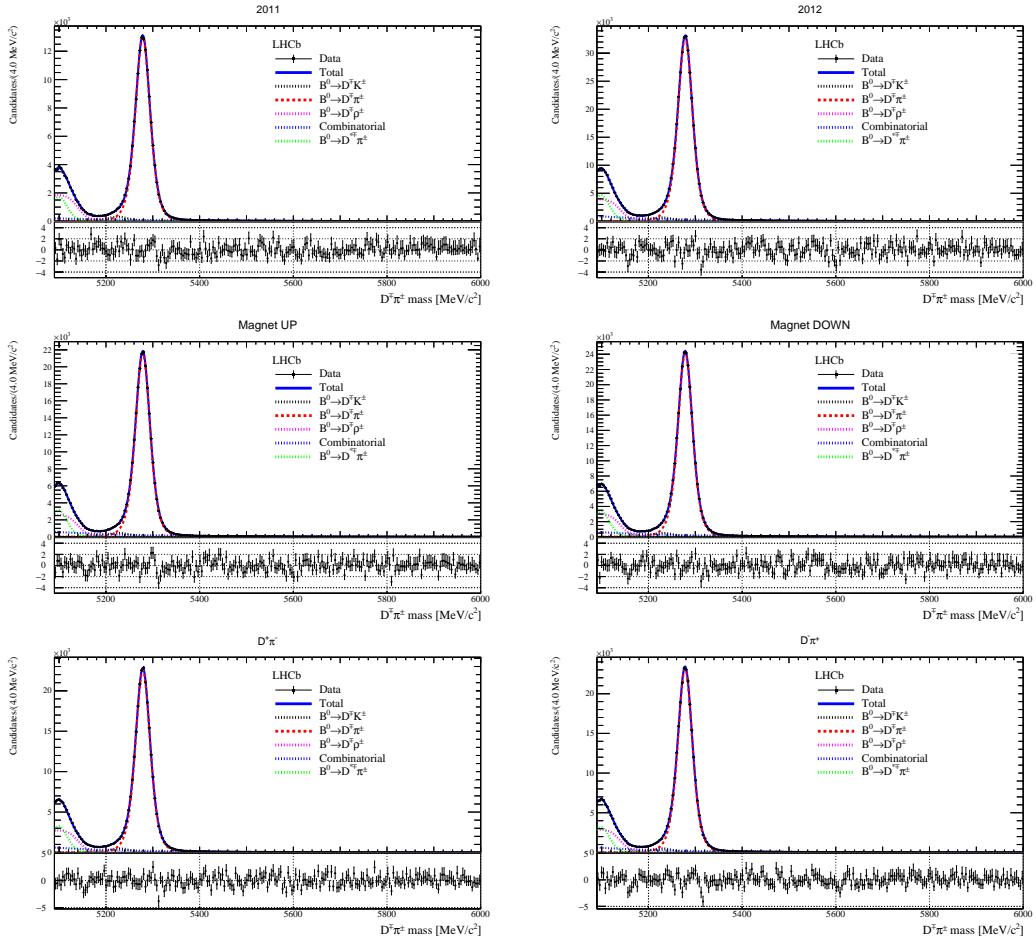


Figure 4.19 – Projections of the fitted PDFs describing the pion sample for each data subsample.

1068

1069 Fit B strategy is also repeated exactly as before for each subsample. The corresponding  
 1070 signal and background fitted yields are listed in Table 4.16. The sum of the yields for

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

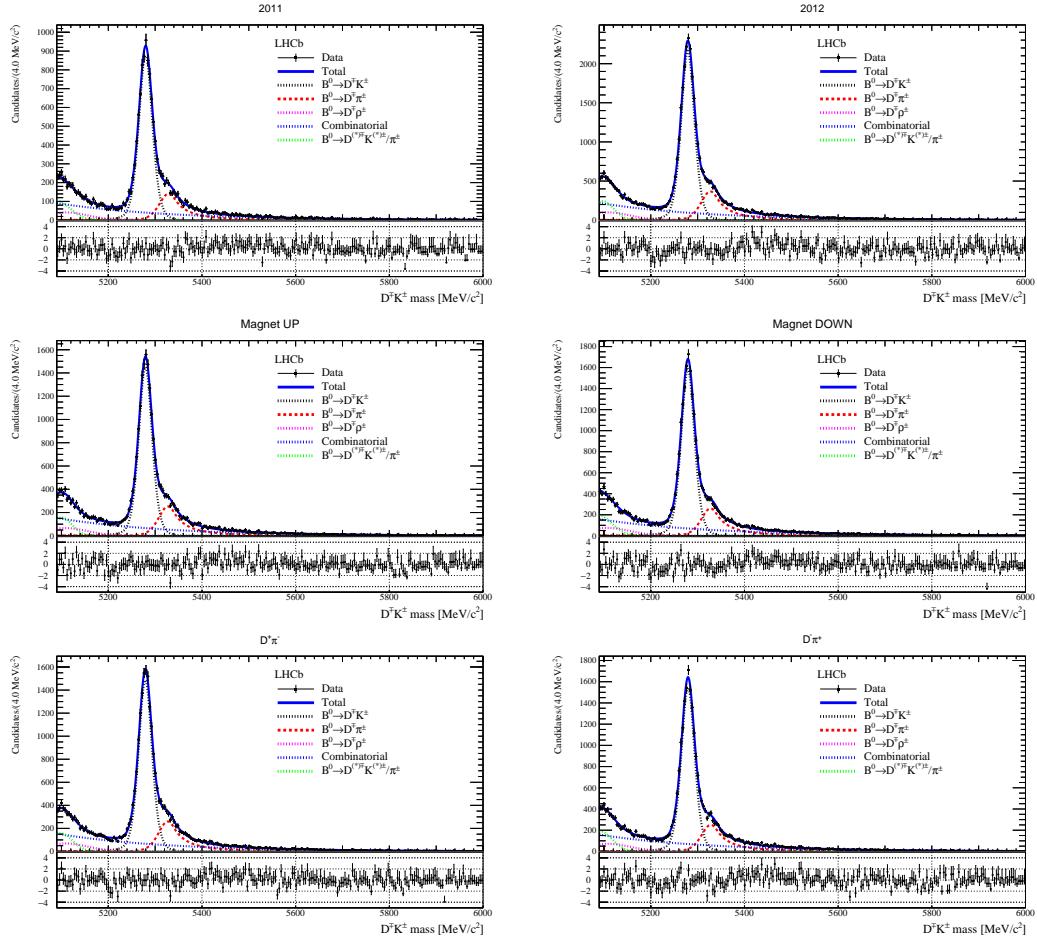


Figure 4.20 – Projections of the fitted PDFs describing the kaon sample for each data subsample.

each subsample is compatible with the total signal yield in the total sample (reported in Table 4.15), which is  $(4.790 \pm 0.007) \times 10^5$ . The asymmetry between the yields of the  $D^- \pi^+$  and  $D^+ \pi^-$  samples is  $0.0100 \pm 0.0015$ , which is in agreement with the detection asymmetry between  $\pi^+$  and  $\pi^-$  obtained in this analysis and by previous measurements (more details given in Sec. 5.3). The ratio between the fitted yields on the 2011 and 2012 samples is compatible with the different collected luminosities and data taking conditions between the two years (twice more luminosity is collected in 2012 compared to 2011, and the  $b$ -production cross section is increased by a factor  $8/7$  in 2012 because of the increase of the centre-of-mass energy). Moreover, the ratio between the yields obtained with the magnet up and down samples is in agreement with the ratio of luminosities collected with the two magnet polarities (same luminosity in 2012, +30 % more magnet down data in 2011).

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

---

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
$2.421 \pm 0.005$	$2.373 \pm 0.005$	$4.794 \pm 0.007$



## 5 Measurement of $CP$ violation in $^{1084} B^0 \rightarrow D^\mp \pi^\pm$ decays

## 5.1 Time resolution

The decay time resolution is determined from a sample of fake  $B^0$  candidates formed from a prompt  $D^\pm$  candidate combined with a track originating from the PV. This is referred to as “ $D^\pm$ +track”. The candidates are selected from the B02DKLTUBD2HHH stripping line. These candidates are subjected to the same offline selection as that of the signal sample without a BDT cut, with two additional requirements: the number of PVs per event are required to be 1 in order to reduce wrong PV associations, and the  $D^\pm$  IP  $\chi^2$  with respect to the PV is less than 9 to reduce the  $D^\mp$  from  $B^0$ . The combined stripping and offline selection yields 51053 candidates. True  $D^\pm$ +track candidates are unfolded from combinatorial background and nonresonant decays by means of *sWeights* computed via a fit to the  $K^\pm \pi^\mp \pi^\mp$  invariant mass distribution.

### 5.1.1 Companion track momentum reweighting

The decay time resolution is found to be dependent upon the companion track  $p_T$  which is considerably lower on average for the  $D^\pm$ +track candidates than it is for genuine  $B^0 \rightarrow D^\pm \pi^\mp$  signal. This is corrected for by reweighting the prompt sample to have the same  $\log(p_T)$  distribution as the  $B^0 \rightarrow D^\pm \pi^\mp$  signal. The logarithm is taken to compress the high- $p_T$  tails, and a binning scheme is chosen to have an equal number of signal events per bin. The  $\log(p_T)$  spectra for signal, prompt, and reweighted prompt candidates is shown in Fig. 5.1. Prior to reweighting the average proper time resolution is determined to be  $\sim 71$  fs, post-reweighting the resolution is found to be consistent with the value of 54 fs that was obtained in other  $B$  meson time-dependent analyses [70].

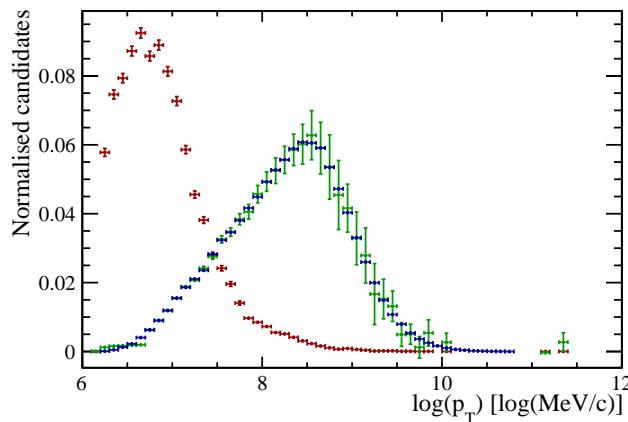


Figure 5.1 – Normalised  $\log(p_T)$  distributions of signal  $B^0 \rightarrow D^\pm \pi^\mp$ (blue), prompt  $D^\pm$ +track before reweighting (red) and after reweighting (green).

1106 **5.1.2 Resolution determination from decay time error parameterisa-**  
 1107 **tion**

In order to study potential second order corrections to the decay time error distribution, fits to the prompt decay time distribution 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 signal *sWeights* in each bin is equal. The fit is similar to that used to determine the resolution in Ref. [71], consisting of three components: a delta function convolved with a Gaussian resolution function accounts for the genuine prompt  $D^\pm$ + 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:

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

1108 where  $\langle\delta\rangle$  is the average per-event decay time error of the whole (unbinned) sample, while  
 1109  $\langle\delta\rangle_i$  is the average per-event decay time error in each bin. In the prompt  $D^\pm$ + track  
 1110 sample this is determined to be  $0.0307 \pm 0.0097$  ps, in good agreement with the signal  
 1111 sample value of  $0.034 \pm 0.011$  ps. The fit determines the average resolution,  $\langle\sigma\rangle$ , in addition  
 1112 to the trend. This fit is shown in Fig. 5.2, the result of which is presented in Table 5.2.  
 1113 The global average resolution is determined from this fit to be  $\langle\sigma\rangle = 0.05491 \pm 0.00038$   
 1114 ps. The procedure is found to be stable and yields compatible results with fits to 10 bins  
 1115 ( $0.05523 \pm 0.00041$  ps) and 30 bins ( $0.05464 \pm 0.00037$  ps).

1116 This method, which accounts for second-order corrections to the decay time error, is used to  
 1117 define the width of a single Gaussian in the decay time fit to data,  $\mathcal{R}(t-t') = G(t-t', \langle\sigma\rangle)$ ,  
 1118 with  $\langle\sigma\rangle = 0.05491 \pm 0.00038$  ps. The uncertainty stated here is statistical. Systematic  
 1119 uncertainties will be considered in Sec. 5.4.2.

1120 **5.2 Time-dependent selection efficiency**

1121 Because of some of the selection criteria described in Sec. 4.1, the  $B^0$  decay time  
 1122 distribution is biased, *i.e.* different from the shape it would have with a selection efficiency

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 for comparison.

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 – Resolution 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

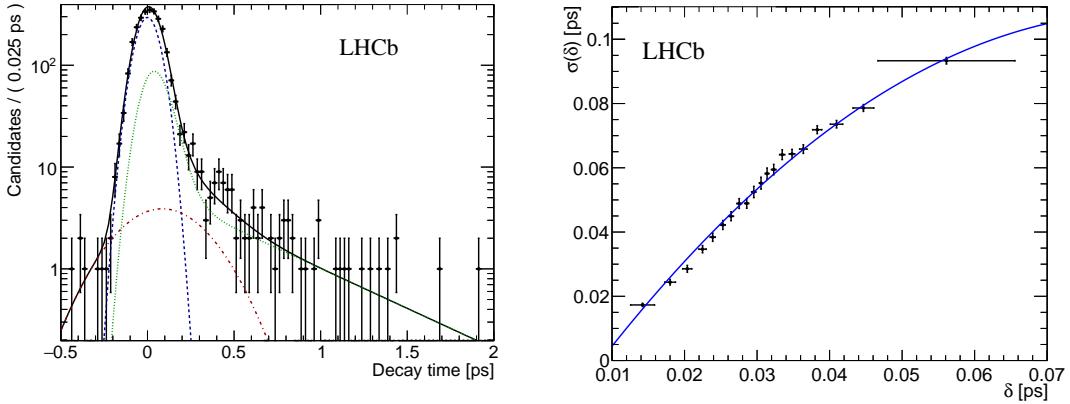


Figure 5.2 – Left:  $p_T$ -corrected and background-subtracted decay-time distribution for the 15th bin ( $[0.034, 0.035]$  ps) in per-event decay time error. The fit result is overlaid as the black solid curve: the wrong-PV, from- $b$ , and prompt components are shown as the red dot-dashed, green dotted, and blue dashed curves, respectively. The numerical results are presented in Table 5.1. Right: measured resolution vs. average per-event decay time error determined from fits to the decay time in bins of decay time error. The horizontal axis uncertainties are the standard deviation of the average per-event decay time error in each bin. The overlaid fit is described in the text.

equal to 100%. This efficiency, called here and after “acceptance”, is a function of the reconstructed proper-time. In particular, it goes very rapidly to zero at low decay times due to the impact parameter requirements which exclude short-lived  $B^0$  candidates; then, it reaches a “plateau” at intermediate decay times; finally, it drops at high decay times because of reconstruction effects from VELO.

The acceptance function  $a(t)$  is parameterised using splines defined analytically as described in Ref. [72]. 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 latter is constrained to be the linear extrapolation from the previous two coefficients:

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

The knot positions and the number of knots are optimized in order to fit the  $B^0 \rightarrow D^\mp \pi^\pm$

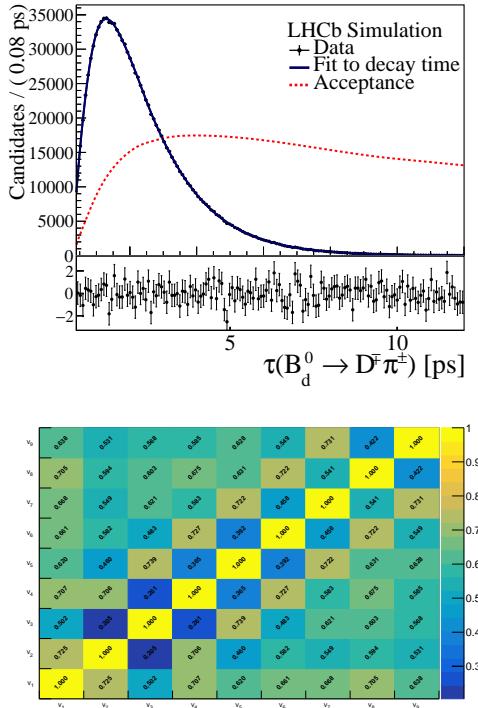


Figure 5.3 – Top: distribution of the reconstructed decay time of simulated and selected  $B^0 \rightarrow D^\mp \pi^\pm$  decays (data points), with fit model superimposed (blue curve), and fitted acceptance function (red dotted curve). Bottom: correlation matrix of the nine fitted acceptance parameters.

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 describing the signal

### 5.3. Decay-time fit to data

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

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

distribution. This is

$$f(t, d_{OS}, d_{SS}, \eta_{OS}, \eta_{SS}) = a(t) \int dt' \mathcal{R}(t - t') P(t'|d_{OS}, d_{SS}, \eta_{OS}, \eta_{SS}) P(\eta_{OS}) P(\eta_{SS}), \quad (5.4)$$

where  $\mathcal{R}(t - t')$  is the Gaussian resolution function, and  $P(\eta_i)$  ( $i = OS, SS$ ) is the PDFs of the predicted mistag probability of the taggers. The term  $P(t|d_{OS}, d_{SS}, \eta_{OS}, \eta_{SS})$  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  $d_i$  and the mistag probability  $\eta_i$ , and it contains the decay rates of Eqs. 1.50–1.53. A detailed description of the time PDF including the tagging parameters, and the detection and production asymmetries, is given in Appendix G. The function maximised during the fit is the logarithm of the likelihood (with *sWeights*  $s_W^i$ ) obtained from the PDF given in Eq. 5.4,

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

where  $N$  is the number of candidates in the fitted sample, and  $s$  is a correction factor given by

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

<sup>1134</sup> This parameter  $s$  allows to take into account the dilution due to the background subtraction  
<sup>1135</sup> with the *sWeights*, so that correctly-estimated uncertainties from the fit are obtained [73].

<sup>1136</sup> In the PDF,  $\Delta\Gamma$  is fixed to zero. Moreover, the  $C_f$  ( $C_{\bar{f}}$ ) coefficient is fixed to 1 ( $-1$ )

because the value of  $r_{D\pi}^2$  is such that the sensitivity to  $C_f$  ( $C_{\bar{f}}$ ) is negligible. Systematic uncertainties will be considered in Sec. 5.4 for all these assumptions.

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

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

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

where  $\tau$  is taken as the world average value [20], and  $\Delta m$  is the LHCb measurement from semileptonic  $B^0$  decays [74].

The free parameters of the fit are:

- the  $S_f$  and  $S_{\bar{f}}$  coefficients;
- the detection asymmetry<sup>1</sup>  $A_D$  and the production asymmetry<sup>2</sup>  $A_P$ ;
- the calibration parameters for both OS and SS taggers;
- the time acceptance coefficients.

The tagging efficiency asymmetries  $A_{\text{eff}}^i$  parameters are found to be consistent with zero in the  $B^0 \rightarrow D^- \pi^+$  Monte Carlo sample: for this reason, these coefficients are fixed to zero in the fit.

The value of the parameters obtained from the fit to data are listed in Table 5.4. The correlation matrix of the parameters is reported in Appendix H. The projection of the PDF on the decay-time distribution is shown in Fig 5.4, while Fig 5.5 shows the  $B^0$ - $\bar{B}^0$  asymmetries of Eqs. 1.46 and 1.47 for the two final states.

It is possible to define the  $CP$  asymmetries between Cabibbo-favoured (CF) and Cabibbo-suppressed (CS) rates as follows:

$$A_{\text{CF}} = \frac{\Gamma_{B^0 \rightarrow f}(t) - \Gamma_{\bar{B}^0 \rightarrow \bar{f}}(t)}{\Gamma_{B^0 \rightarrow f}(t) + \Gamma_{\bar{B}^0 \rightarrow \bar{f}}(t)}, \quad (5.9)$$

$$A_{\text{CS}} = \frac{\Gamma_{\bar{B}^0 \rightarrow f}(t) - \Gamma_{B^0 \rightarrow \bar{f}}(t)}{\Gamma_{\bar{B}^0 \rightarrow f}(t) + \Gamma_{B^0 \rightarrow \bar{f}}(t)}.. \quad (5.10)$$

The asymmetries defined above are plotted together with data in Fig. 5.6.

The projection of the PDF considering the decay-time distribution of the four independent

<sup>1</sup>The detection asymmetry is defined as  $\frac{N(f) - N(\bar{f})}{N(f) + N(\bar{f})}$ . This definition has the opposite sign compared to the one in Ref. [75].

<sup>2</sup>The production asymmetry is defined as  $\frac{N(\bar{B}^0) - N(B^0)}{N(\bar{B}^0) + N(B^0)}$

### 5.3. Decay-time fit to data

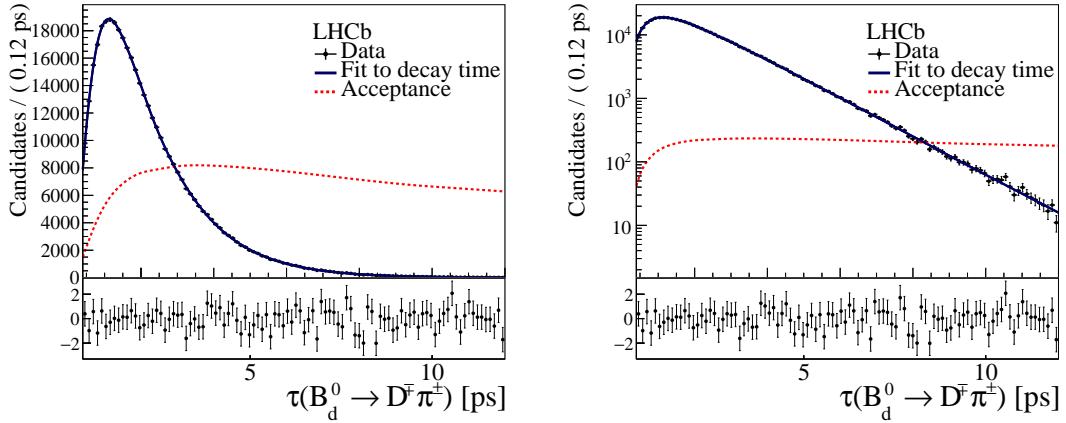


Figure 5.4 – Distribution of the reconstructed decay time of *sWeighted*  $B^0 \rightarrow D^\mp\pi^\pm$  decays (data points), with fit model superimposed (blue curve), and fitted acceptance function (red dotted curve), in linear (left) and logarithmic (right) scale

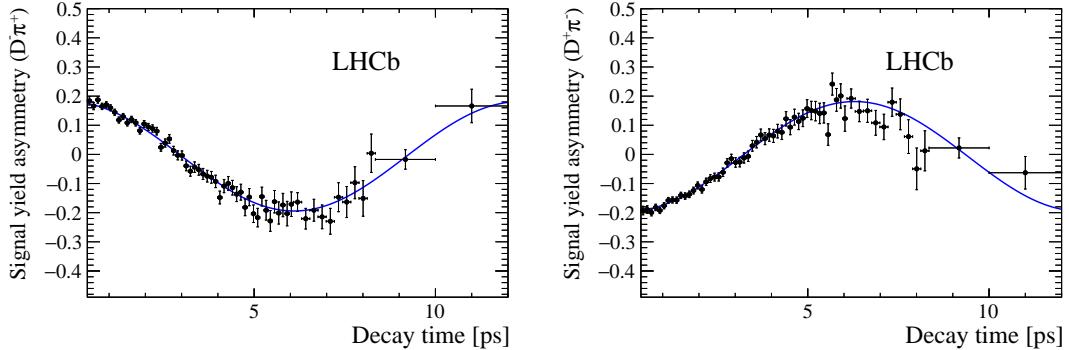


Figure 5.5 – Time-dependent asymmetry between  $B^0$  and  $\bar{B}^0$  decays (data points) for the  $D^-\pi^+$  (left) and  $D^+\pi^-$  (right) final states, with fit model superimposed (blue curve).

1155 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  
1156 tagged candidates, are reported in Figs. 5.7 and 5.8, respectively.

1157 The 2D contour plots for the  $CP$  coefficients  $S_f$  and  $S_{\bar{f}}$  and for the detection and  
1158 production asymmetry are shown in Fig. 5.9.

1159 Using the flavour tagging calibrations obtained from the fit, the tagging performance of the  
1160 signal sample is computed. The average of the total squared dilution is  $(6.554 \pm 0.017)\%$ .  
1161 Taking into account also untagged candidates, *i.e.* considering the tagging efficiency of  
1162  $(85.23 \pm 0.05)\%$ , the tagging power is  $(5.59 \pm 0.01)\%$ .

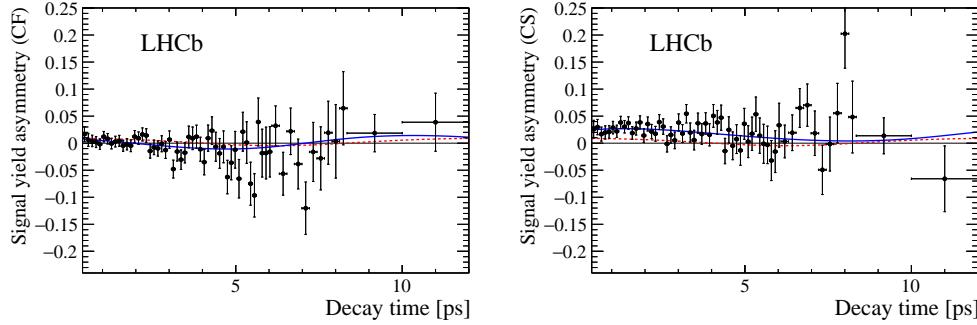


Figure 5.6 – Decay-time-dependent signal-yield asymmetries for (left) Cabibbo-favoured and (right) Cabibbo-suppressed decay topologies, defined in Eqs. 5.9 and 5.10. The point with error bars are the data, the blue solid curve is the projection of the signal PDF, and the red dotted curve indicates the projection of the fit when  $S_{\bar{f}} \equiv -S_f$  is required.

<sup>1163</sup> **5.3.1 Fit validation**

<sup>1164</sup> **Check of nuisance parameters**

<sup>1165</sup> The values of the nuisance parameters obtained in the fit (production/detection asymme-  
<sup>1166</sup> tries, flavour tagging calibrations) are compared with available external measurements.

The production asymmetry  $A_P$  is compared with the LHCb measurement of Ref. [75]. 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  $\varepsilon_i = \frac{f_i}{\sum_i f_i}$  in each bin  $i$  of  $B^0 \rightarrow D^\mp\pi^\pm$  data:

$$A_P = \sum_i \varepsilon_i A_{P,i}, \quad (5.11)$$

where  $f_i$  is the number of signal candidates. This yields

$$A_P = -0.0100 \pm 0.0047 \text{ (stat)} \pm 0.0004 \text{ (syst)}. \quad (5.12)$$

<sup>1167</sup> This value is compatible within  $0.65 \sigma$  with that obtained from the  $B^0 \rightarrow D^\mp\pi^\pm$  decay-  
<sup>1168</sup> time fit.

<sup>1169</sup> The detection asymmetry  $A_D$  is again obtained from Ref. [75], and is measured using both  
<sup>1170</sup>  $B^0 \rightarrow J/\psi K^{*0}$  to be  $0.0098 \pm 0.0046$  and  $0.0056 \pm 0.0030$  for 2011 and 2012, respectively,  
<sup>1171</sup> and  $B_s^0 \rightarrow D_s^\pm \pi^\mp$  to be  $0.0143 \pm 0.0086$  and  $0.0103 \pm 0.0058$  for 2011 and 2012, respectively.  
<sup>1172</sup> The central value fitted on  $B^0 \rightarrow D^\mp\pi^\pm$  are in agreement with this set of results.

<sup>1173</sup> The values of the parameters of the tagging calibrations are compared with those found  
<sup>1174</sup> in the control samples as described in Sec. 3.2.1 and 3.2.2. The strategy presented in  
<sup>1175</sup> Sec. 3.2 is followed, and no perfect portability of the calibrations is assumed a priori,

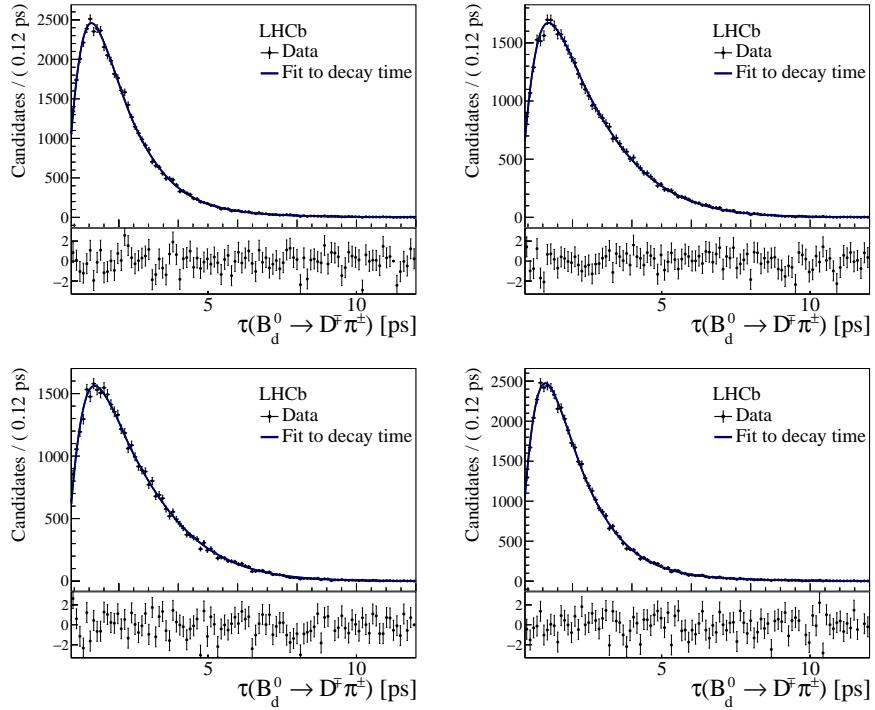


Figure 5.7 – Decay-time distributions of the *sWeighted* data samples for (top left)  $B^0 \rightarrow D^- \pi^+$ , (top right)  $\bar{B}^0 \rightarrow D^- \pi^+$ , (bottom left)  $B^0 \rightarrow D^+ \pi^-$  and (bottom right)  $\bar{B}^0 \rightarrow D^+ \pi^-$  for OS inclusively tagged candidates.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

as Sec. 3.2.1 has shown this not to be the case. However, the values of the parameters found in the signal fit are expected to be in the same ballpark interval of those from the control channels. A full comparison that takes into account the correlation between the parameters with a  $\chi^2$  test which uses the covariance matrices of the fit to data (reduced to the tagging parameters only), and the fit of the calibration from the control channel, is performed. The discrepancy (corresponding to the  $\chi^2$  minimum) is around  $0.91\sigma$  for the OS tagger, and  $0.29\sigma$  for the SS tagger. The parameters which present the largest disagreement are  $\Delta p_3^{OS}$  for the OS tagger and the  $\Delta p_0^{SS}$  for the SS tagger.

#### 1184 Fits in data subsamples

1185 A check of the stability of the results against the different data taking conditions is  
 1186 performed by repeating the fit in four subsamples of the data, namely data taken with  
 1187 magnet “up” and “down” polarities, and data taken in 2011 and in 2012. The *sWeights*  
 1188 for each subsample are obtained via the mass fits described in Sec. 4.3.4. The detailed  
 1189 results of these time fits are reported in Appendix I. A comparison between the fitted  
 1190 values for  $S_f$  and  $S_{\bar{f}}$  obtained in each subsample is shown in Fig. 5.10. In all cases, the

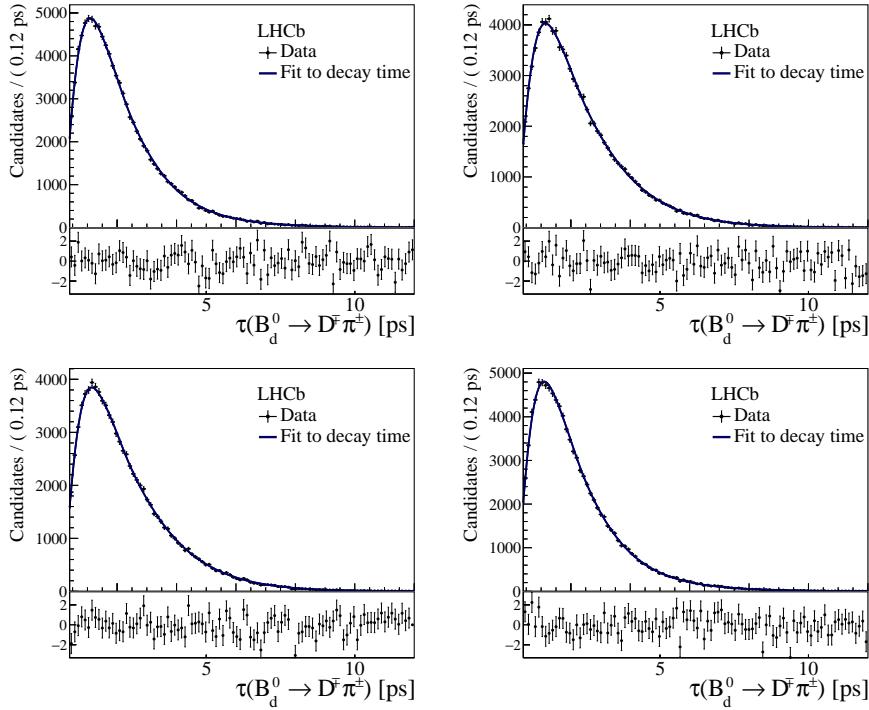


Figure 5.8 – Decay-time distributions of the  $sWeighted$  data sample for (top left)  $B^0 \rightarrow D^- \pi^+$ , (top right)  $\bar{B}^0 \rightarrow D^- \pi^+$ , (bottom left)  $B^0 \rightarrow D^+ \pi^-$  and (bottom right)  $\bar{B}^0 \rightarrow D^+ \pi^-$  for SS inclusively tagged candidates.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

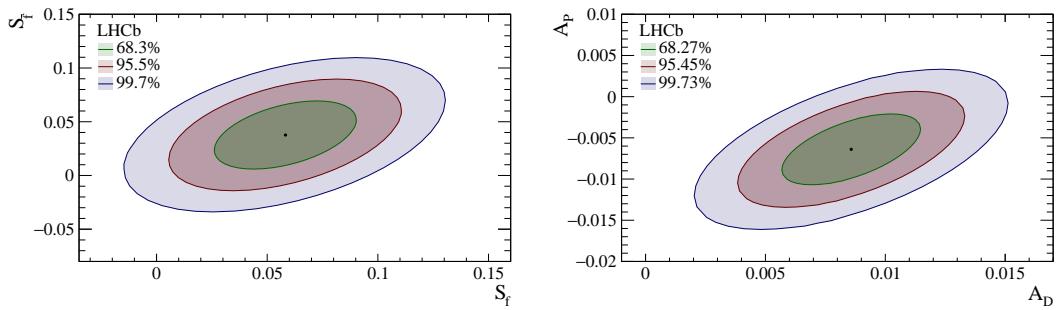


Figure 5.9 – Contour plots for  $(S_f, S_{\bar{f}})$  (left) and  $(A_P, A_D)$  showing the one, two and three sigma contours. The shown uncertainties include the full statistical uncertainty and the systematic uncertainty due to the Gaussian constraints on the mixing frequency  $\Delta m$  and the 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

### 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
$S_f$	$0.058 \pm 0.021$	Stat. uncert. when fitting w/o Gauss-const. and PIDk syst., 0.0198
$S_{\bar{f}}$	$0.038 \pm 0.021$	Stat. uncert. 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. 5.12)
$A_D$	$0.0086 \pm 0.0019$	Compare with [75] (Sec. 5.3.1)
$\Gamma$	$0.6587 \pm 0.0017 \text{ ps}^{-1}$	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)
$\varepsilon_{\text{tag}}^{\text{OS}}$	$0.93046 \pm 0.00040$	fraction of SS tagged candidates (relative to tagged candidates only)
$v_1$	$0.3192 \pm 0.0062$	time acceptance coefficients
$v_2$	$0.494 \pm 0.010$	
$v_3$	$0.793 \pm 0.016$	
$v_4$	$0.994 \pm 0.019$	
$v_5$	$1.093 \pm 0.021$	
$v_6$	$1.117 \pm 0.021$	
$v_7$	$1.140 \pm 0.021$	
$v_8$	$1.175 \pm 0.019$	
$v_9$	$1.154 \pm 0.026$	

<sup>1191</sup> parameters are in agreement and have good  $p$ -values; the smallest being the one between

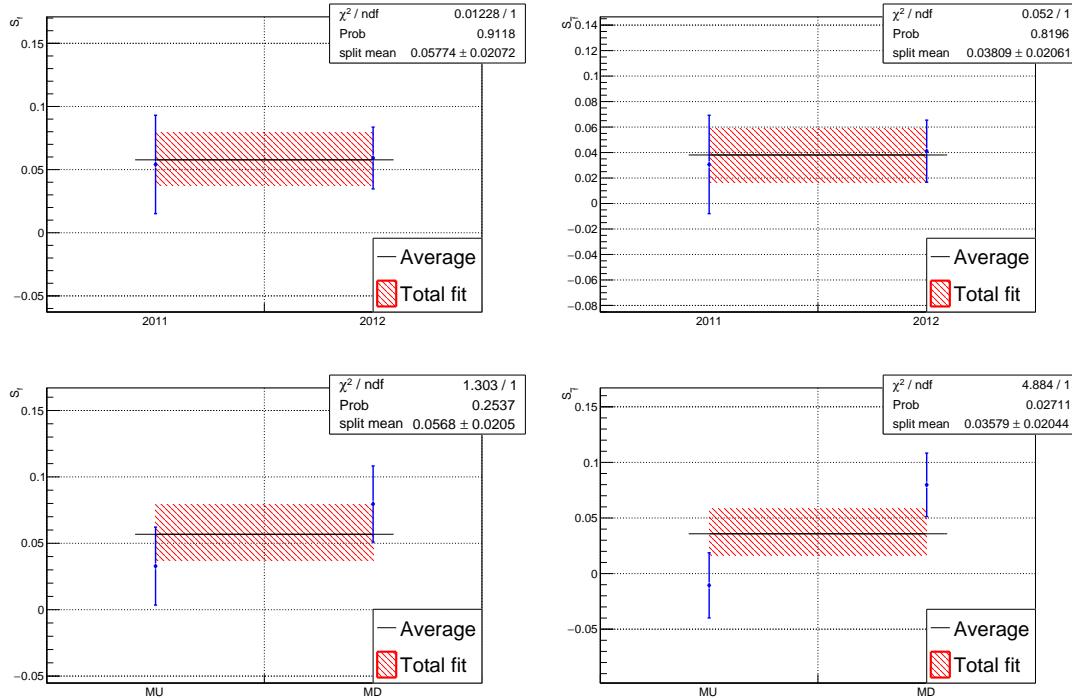


Figure 5.10 – Fitted values of  $S_f$  (left) and  $S_{\bar{f}}$  (right) as a function of the data-taking year (top) and magnet polarity (bottom). The red hatched band shows the values obtained from the nominal fit of the full sample.

1192 the values of  $S_{\bar{f}}$  from the magnet polarity splits (2.7%). In addition, the average between  
 1193 the fitted values in each split (black line in Fig. 5.10) is always very close to the central  
 1194 value from the nominal fit (red hatched band).

1195 The stability of the results against the tagging algorithm adopted in the fit are also  
 1196 checked. In this case, the data sample with sWeights obtained from the nominal mass fit  
 1197 (Sec. 4.3) is split in three independent subsamples according to the tagging decision:

- 1198 • candidates tagged exclusively by the OS tagger;  
 1199 • candidates tagged exclusively by the SS tagger;  
 1200 • candidates tagged by both OS and SS taggers simultaneously.

1201 The values of  $S_f$  and  $S_{\bar{f}}$  obtained in these subsamples are compared in Fig. 5.11. All  
 1202 values are compatible. More details are given in Appendix I. Given the difference of the  
 1203 tagging algorithms and their calibrations, the stability of the results in this test provides  
 1204 additional confidence on the strategy adopted of floating the calibration parameters in  
 1205 the fit.

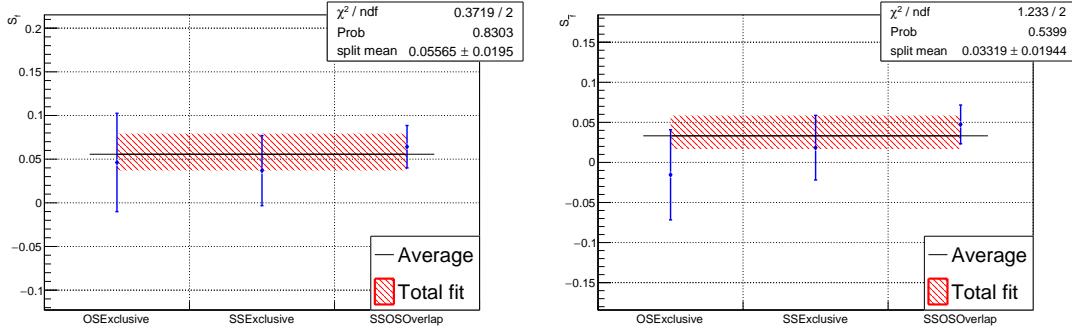


Figure 5.11 – Fitted values of  $S_f$  (left) and  $S_{\bar{f}}$  (right) when candidates tagged exclusively by OS or SS, or both simultaneously are considered. The red hatched band shows the values obtained from the nominal fit of the full sample.

1206 The stability of the results against the  $B^0$  kinematics and global properties of the event is  
1207 tested. More specifically, the decay time fit is repeated in bins of the following variables:

- 1208 • transverse momentum of the  $B^0$  (4 bins);
- 1209 • number of reconstructed primary vertices (3 bins);
- 1210 • number of reconstructed tracks (3 bins);
- 1211 • difference in pseudorapidity ( $\Delta\eta$ ) between  $D$  meson and bachelor pion (4 bins).

1212 The motivation for these test is that flavour tagging calibration parameters depend on  
1213 the above observables; as a consequence, the fitted values for the  $S_f$  and  $S_{\bar{f}}$  coefficients  
1214 might also show a significant trend in these variables because of the correlation with  
1215 the flavour tagging calibrations. Moreover, the difference in pseudorapidity is sensitive  
1216 to potential misalignments in the detectors which might affect the measured value of  
1217  $CP$  asymmetries. The values of  $S_f$  and  $S_{\bar{f}}$  obtained in these subsamples are compared in  
1218 Fig. 5.12, whereas more details are given in Appendix I. All values are compatible, and  
1219 no significant dependence of  $S_f$  and  $S_{\bar{f}}$  on the studied variables is observed.

1220 Finally, the time fit is repeated separately for  $B^0 \rightarrow D^- \pi^+$  candidates TOS on L0Hadron  
1221 and all the other candidates. The values of  $S_f$  and  $S_{\bar{f}}$  obtained in these subsamples are  
1222 compared in Fig. 5.13. All values are compatible, and no significant dependence of  $S_f$   
1223 and  $S_{\bar{f}}$  is observed. More details can be found in Appendix I.

#### 1224 Time fits to bootstrapped Monte Carlo samples

1225 The fit is also validated using Monte Carlo simulation. The  $B^0 \rightarrow D^\mp \pi^\pm$  simulated  
1226 sample is *bootstrapped* [76] (*i.e.*, resampled allowing repetition of the same candidates)

## Chapter 5. Measurement of $CP$ violation in $B^0 \rightarrow D^\mp \pi^\pm$ decays

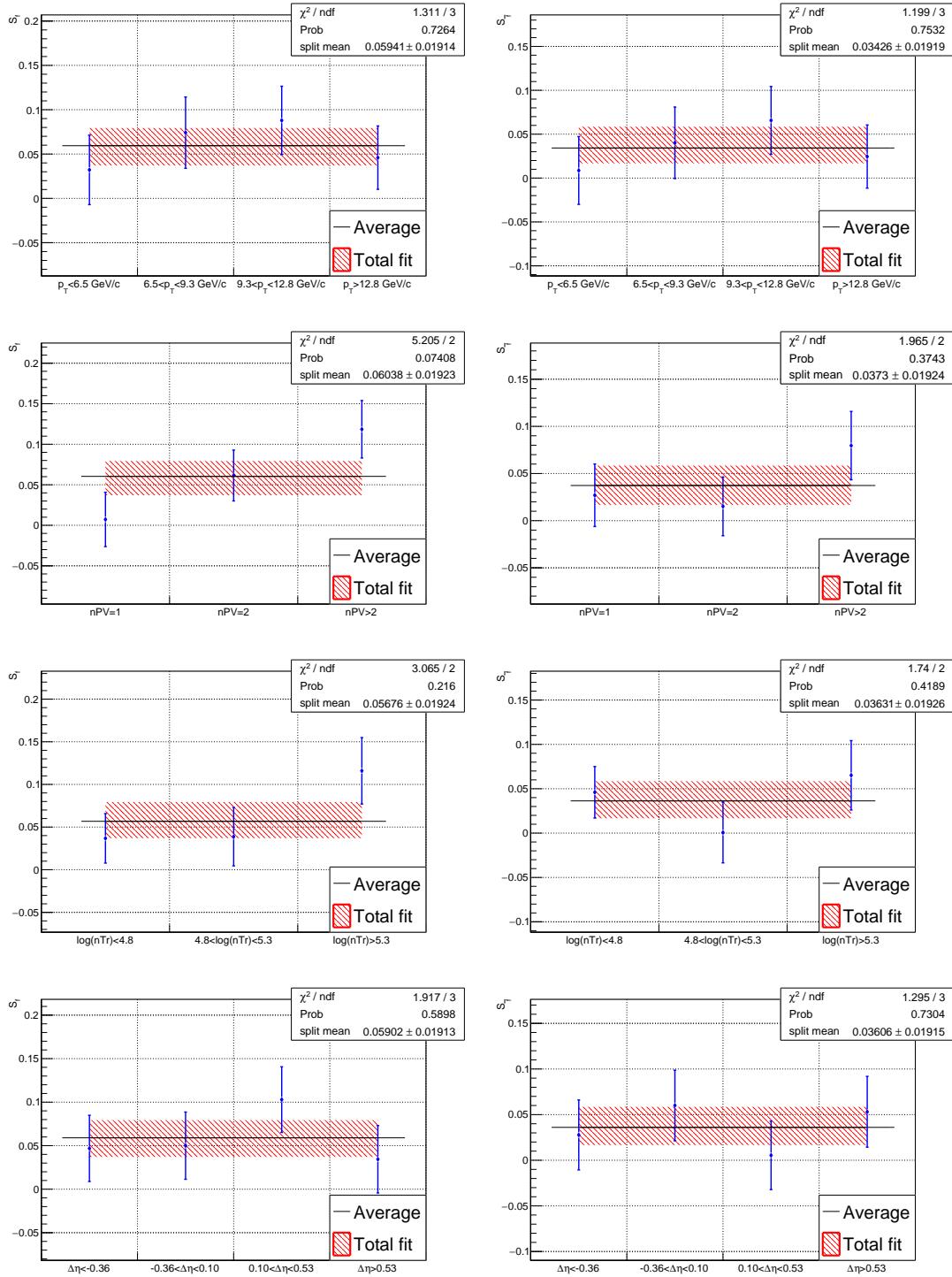


Figure 5.12 – Fitted values of  $S_f$  (left) and  $S_{\bar{f}}$  (right) when the decay time fit is performed in bins of (from top to bottom) the transverse momentum of the  $B^0$ , number of primary vertices, number of tracks and difference in pseudorapidity between the  $D$  meson and the bachelor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

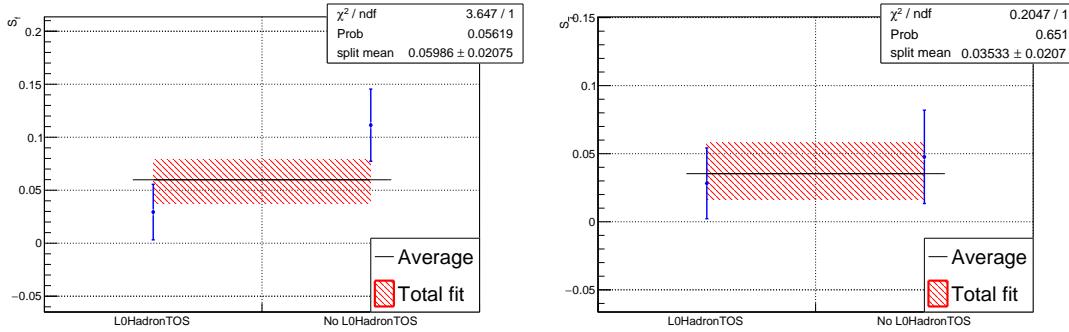


Figure 5.13 – Fitted values of  $S_f$  (left) and  $S_{\bar{f}}$  (right) when the decay time fit is performed separately for candidates TOS on L0Hadron and all the other candidates. The red hatched band shows the values obtained from the nominal fit of the full sample.

1227 1000 times. Each bootstrapped sample contains the same signal yield as obtained from  
 1228 the nominal mass fit (Table 4.15), corrected for the *sWeights* dilution factor of Eq. 5.6 to  
 1229 have the same effective yield as in the data time fit.

1230 Each sample is then fitted using exactly the same strategy as described in Sec. 5.3. The  
 1231 Gaussian-constrained parameters, *i.e.*  $\Gamma$  and  $\Delta m$ , are treated in the following way. For  
 1232 each fit, a random value of  $\Gamma$  and  $\Delta m$  is drawn from two Gaussian distributions having  
 1233 mean and standard deviation equal to the centre and the width of the Gaussian constraints.  
 1234 The mean of each Gaussian constraints is taken as the Monte Carlo generation value  
 1235 (Appendix M), whereas the standard deviation is the same used for the constraints in the  
 1236 data fit (Eq. 5.7). This gives  $1.519 \pm 0.004$  ps and  $0.5100 \pm 0.0023$  ps $^{-1}$  for  $\tau$  and  $\Delta m$   
 1237 respectively. Then, these random values are used as the central value of the Gaussian  
 1238 constraints in that fit. This allows fluctuations of the  $\Gamma$  and  $\Delta m$  measurements, and  
 1239 avoids underestimation of the fitted uncertainties.

1240 The distributions of the fitted value, uncertainty, pull and residual<sup>3</sup> of  $S_f$  and  $S_{\bar{f}}$  are  
 1241 shown in Fig. 5.14. Other fitted parameters are reported in Appendix K. Each of these  
 1242 distribution is fitted with a Gaussian PDF. The width of the fitted pull distributions  
 1243 are close to unity, meaning that the uncertainty coming from the maximum likelihood  
 1244 optimisation is correctly estimated. The mean value of the distribution of the uncertainties  
 1245 of each parameter is close to the value of the uncertainty found in the fit to data. The  
 1246 on-average better precision found in the fit to MC is due to the higher tagging performance  
 1247 of the simulation.

1248 The distribution of the residuals of the  $S_f$  parameter shows a mean of  $0.0071 \pm 0.0006$ ,  
 1249 corresponding to one third of the statistical uncertainty of the fit to data; for  $S_{\bar{f}}$ , the

<sup>3</sup>The residual is defined as fitted value minus generated value, whereas the pull is the residual divided by the fitted uncertainty.

1250 mean is  $-0.0013 \pm 0.0006$ , which corresponds to about 6% of the statistical uncertainty  
1251 of the fit to data.

1252 Several configurations are implemented to test the bootstrap study and its results, and  
1253 to try to address the origin of these biases. The fits to the bootstrapped samples are  
1254 repeated in the following different configurations:

- 1255 • using the true flavour of the  $B^0$  candidate instead of the tagging decision and mistag  
1256 probability. No biases are found on  $S_f$  and  $S_{\bar{f}}$ ;
- 1257 • using the true flavour of the  $B^0$  candidate instead of the tagging decision and mistag  
1258 probability, but reducing the signal yield to that equivalent to a sample of perfectly  
1259 tagged candidates given the tagging power on data. No biases are found on  $S_f$  and  
1260  $S_{\bar{f}}$ ;
- 1261 • using a *toy* (or *cheated*) tagger, as explained in Appendix L. No biases are found  
1262 on  $S_f$  and  $S_{\bar{f}}$ ;
- 1263 • using the calibration parameters obtained in the signal MC sample using the true  
1264 flavour information (see Sec. 3.2.1 and Sec. 3.2.2). No biases are found on  $S_f$  and  
1265  $S_{\bar{f}}$ ;
- 1266 • fixing the calibration parameters to the values obtained from the MC samples of  
1267 the control channels. Biases of the order  $1\sigma$  on  $S_f$  and  $S_{\bar{f}}$  are found;
- 1268 • applying Gaussian-constraints on the calibration parameters using the values ob-  
1269 tained from the MC samples of the control channels. Biases of the order of half the  
1270 statistical uncertainty of  $S_f$  and  $S_{\bar{f}}$  are found;

1271 This study confirms that the strategy of floating the calibration parameters in the fit is  
1272 the optimal choice. Other than the biases related to the flavour-tagging calibrations, the  
1273 origin of the small bias observed on the  $S_f$  parameter in the nominal configuration could  
1274 not be clarified. To confirm this bias, the study is repeated by fitting additional 1000  
1275 bootstrapped samples using an independent fitter. The mean of the distribution of the  
1276 residuals in this second study is confirmed to be of the same size, namely  $0.0064 \pm 0.0007$   
1277 for  $S_f$  and  $-0.0024 \pm 0.0007$  for  $S_{\bar{f}}$ . Hence, the weighted average of the small residuals  
1278 on  $S_f$  ( $0.0068 \pm 0.0005$ ) and  $S_{\bar{f}}$  ( $-0.0018 \pm 0.0005$ ) of both studies are considered as  
1279 systematic uncertainties. As described in Sec J, the correlation between the systematic  
1280 uncertainties on  $S_f$  and  $S_{\bar{f}}$  associated to the fit biases reported here is 0.4.

### 5.3. Decay-time fit to data

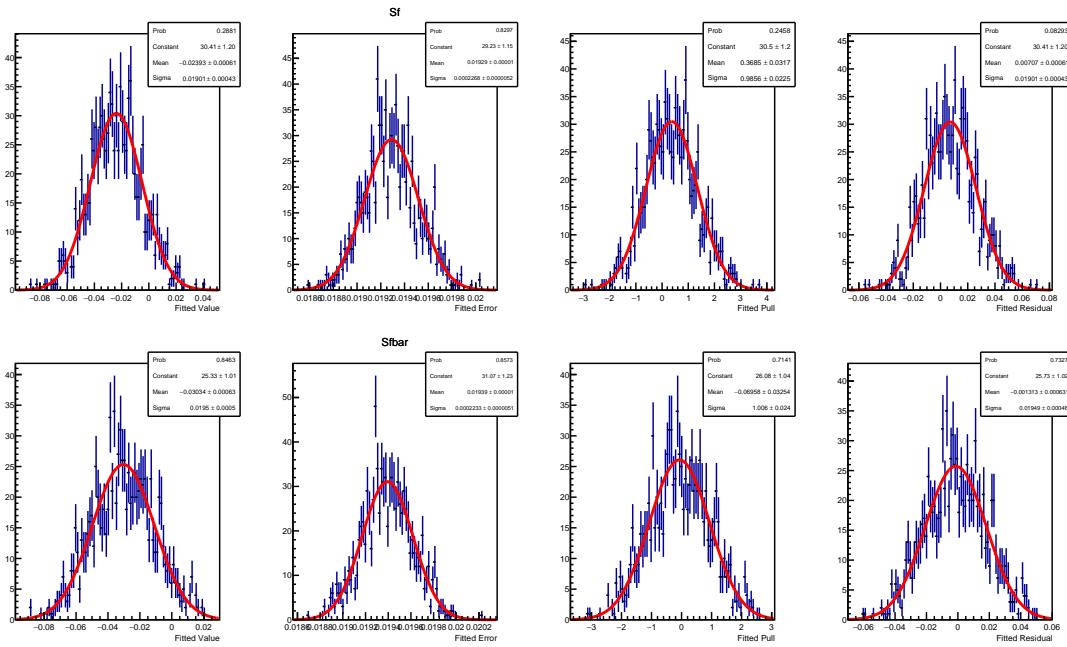


Figure 5.14 – Distributions of the fitted value, error, pull and residual for  $S_f$  (top) and  $S_{\bar{f}}$  (bottom) obtained from fits to bootstrapped Monte Carlo samples. Each distribution is fitted with a Gaussian PDF.

---

## 1281 5.4 Systematics

1282 The systematic uncertainties are listed in Table 5.5 in decreasing order of their size. They  
 1283 total 0.011 and 0.007 for  $S_f$  and  $S_{\bar{f}}$ , respectively. A description of each systematic effect  
 1284 is given in the following subsections. The “fit biases” are the residuals observed in the  
 1285 Monte Carlo bootstrap study discussed in Sec. 5.3.1.

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 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
resolution model	0.0012	0.0008
total systematic uncertainty	0.0111	0.0073
statistical uncertainties	0.0198	0.0199

### 1286 5.4.1 Systematic uncertainties from Gaussian constraints

1287 Systematic uncertainties due to external measurements used in the PDF are accounted  
 1288 for through Gaussian constraints in the likelihood. These parameters are the mixing  
 1289 frequency,  $\Delta m$ , and the  $B^0$  lifetime,  $\tau$ . The fit has been repeated by fixing the Gaussian-  
 1290 constrained parameters to their central values, in order to not propagate the uncertainty  
 1291 of these parameters to the statistical uncertainties of the fit. The statistical uncertainties  
 1292 of  $S_f$  and  $S_{\bar{f}}$  with  $\Delta m$  fixed are 0.0198 and 0.0199, respectively, whereas the statistical  
 1293 correlation is 0.6. Considering the difference in quadrature between the uncertainty  
 1294 from the nominal fit and that from this fit, the systematic uncertainty due to  $\Delta m$  are  
 1295 0.0073 and 0.0061 for  $S_f$  and  $S_{\bar{f}}$ , respectively. The correlation between the systematic  
 1296 uncertainties due to  $\Delta m$  on  $S_f$  and  $S_{\bar{f}}$  is  $-1$ , as described in Sec. J. The fit with  $\tau$  fixed  
 1297 shows that the systematics uncertainty due to this parameter is negligible.

1298 Systematic uncertainty associated with the PIDK efficiencies (Table 4.10) are taken into  
 1299 account in the mass fit by means of Gaussian constraints on these parameters (Sec. 4.3.2).  
 1300 The mass fit is repeated by neglecting these uncertainties in the Gaussian constraints.  
 1301 Then, the time fit is performed with this new set of *sWeights*. The difference in quadrature  
 1302 between the uncertainty from this fit and that from the nominal fit gives the systematic

1303 due to the binning scheme in the PIDK resampling, which is 0.0008 for both  $S_f$  and  $S_{\bar{f}}$ .

1304 **5.4.2 Systematic uncertainties estimated with pseudoexperiments**

1305 When computing the systematic uncertainties with pseudoexperiments (or *toys*), a sample  
1306 with the same size as the data is generated by sampling the PDF with parameters fixed  
1307 to the value found in the data fit. The values of  $S_f$  and  $S_{\bar{f}}$  are fixed to those used in the  
1308 generation of the Monte Carlo sample (Appendix M). In the generation of the samples  
1309 the PDF is modified to consider alternative models according to the source of systematic  
1310 uncertainty under investigation. The generated sample is then fitted with the nominal  
1311 model. For each parameter, the mean of the distribution of the residuals from 1000 toys  
1312 is taken as the systematic uncertainty. If the mean is consistent with zero, the error on  
1313 the mean is taken instead. The systematic uncertainties estimated with this toy-based  
1314 method are the following:

- 1315 • the flavour tagging calibration model;
- 1316 • the acceptance model;
- 1317 • the resolution model;
- 1318 • the assumption  $C_f = -C_{\bar{f}} = 1$ ;
- 1319 • the assumption  $\Delta\Gamma = 0$ .

1320 **Flavour tagging calibration model**

1321 Toys are generated using for the SS calibration the nominal model with a first order  
1322 polynomial, and for the OS the model is reduced by one degree as compared to the  
1323 nominal one. In the fit, the calibration models of both taggers are increased by one degree  
1324 compared to what was used in the generation step. The distribution of the residuals  
1325 of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.15. The residuals are not compatible with zero and  
1326 therefore they are assigned as systematic uncertainties.

1327 **Flavour tagging efficiency asymmetries**

1328 Toys are generated with the flavour tagging asymmetries set to their estimate from  
1329 simulation minus their uncertainty, namely  $-0.14\%$  and  $-0.13\%$  for the OS and SS  
1330 tagger, respectively. The distributions of the residuals of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.16.  
1331 The residuals are not compatible with zero and therefore they are assigned as systematic  
1332 uncertainties.

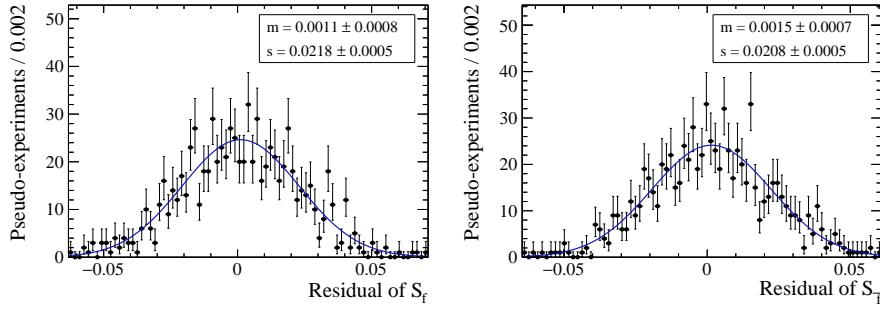


Figure 5.15 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the tagging calibration models.

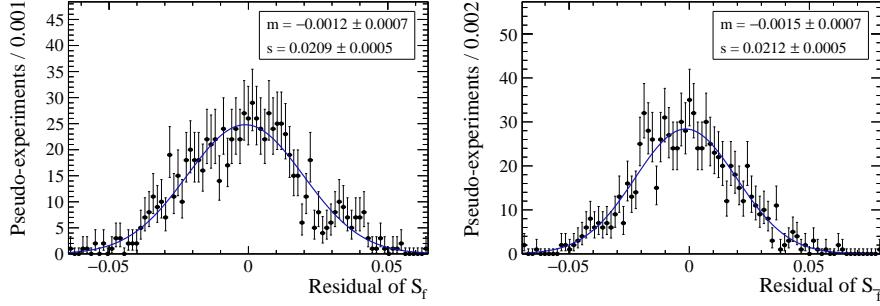


Figure 5.16 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the assumption on the flavour tagging efficiency asymmetry.

### <sup>1333</sup> Acceptance model

<sup>1334</sup> The acceptance model is modified in the generation by considering new knots for the  
<sup>1335</sup> spline function, namely at 0.4, 0.45, 0.8, 1.3, 2.5, 6.0, and 12.0 ps. The distribution of the  
<sup>1336</sup> residuals of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.17. Residuals consistent with zero are found  
<sup>1337</sup> and therefore the uncertainty on the residuals is assigned as systematic uncertainty.

### <sup>1338</sup> Decay time resolution

<sup>1339</sup> Toys are generated with time resolutions 20 fs larger and 20 fs smaller than the nominal  
<sup>1340</sup> value of 55 fs. The distributions of the fitted value of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.18.  
<sup>1341</sup> The largest residual is considered as overall systematic uncertainty.

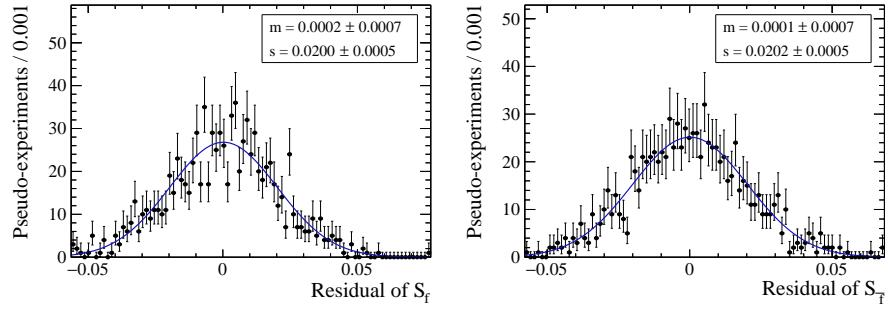


Figure 5.17 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the acceptance model.

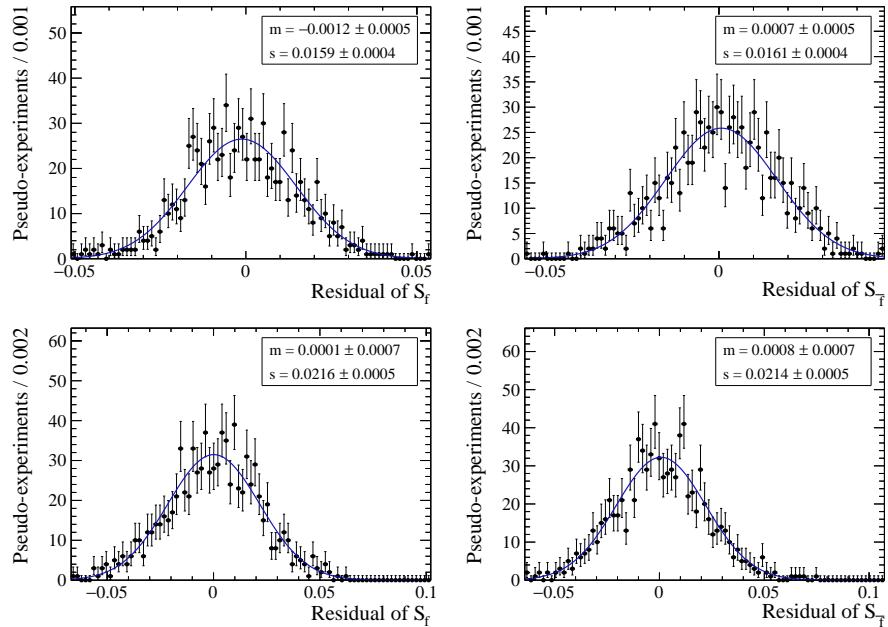


Figure 5.18 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the resolution model. Top: 75 fs resolution model. Bottom: 35 fs resolution model.

**1342 Fixed  $C_f$**

**1343** Toys are generated with  $C_f$  set to the average of the measurements by Belle and BaBar mi-  
**1344** nus the largest uncertainty among the two measurements, namely 0.993 [26, 27]. The  
**1345** distributions of the residuals of  $S_f$  and  $S_{\bar{f}}$  are shown in Fig. 5.19. Residuals consistent  
**1346** with zero are found, therefore the uncertainty on the residuals is assigned as systematic  
 uncertainty.

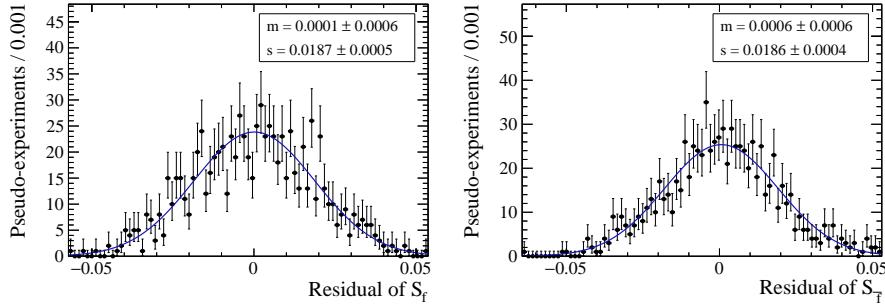


Figure 5.19 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the assumption  $C_f = -C_{\bar{f}} = 1$ .

**1347**

**1348 Fixed  $\Delta\Gamma$**

**1349** Toys are generated with  $\Delta\Gamma$  set to the world average value plus its uncertainty, namely  
**1350**  $0.0079 \text{ ps}^{-1}$  [77]. Moreover, the  $D_f$  and  $D_{\bar{f}}$  coefficients (defined in Eqs. 1.35-1.38) have  
**1351** been fixed to their expected values of  $-0.0103$  and  $-0.0155$ , the same used in the Monte  
**1352** Carlo production of the  $B^0 \rightarrow D^- \pi^+$  sample (Appendix M). The distribution of the  
**1353** 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.

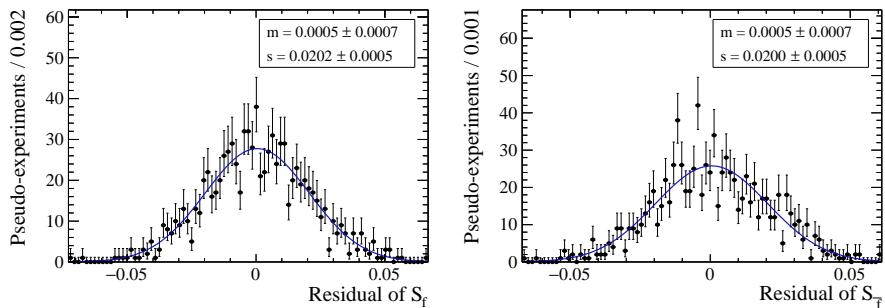


Figure 5.20 – Distribution of  $S_f$  (left) and  $S_{\bar{f}}$  (right) residuals for the determination of the systematic uncertainty due to the assumption  $\Delta\Gamma = 0$ .

**1354**

1355 5.4.3 Systematics related to the background subtraction

1356 Systematic uncertainties can arise from the choice of the mass fit strategy adopted to  
 1357 calculate *sWeights* (Sec. 4.3).

1358 Fit B, used to compute the *sWeights*, is repeated in the full mass window ( $[5090, 6000]$  MeV/ $c^2$ )  
 1359 instead of the narrow signal region ( $[5220, 5600]$  MeV/ $c^2$ ). In this way, the resulting sample  
 1360 is enriched in background events. The aim of this test is to estimate how much background  
 1361 events (with negative *sWeights*) affect the result for  $S_f$  and  $S_{\bar{f}}$  in the final decay time  
 1362 fit. The fitted total background yield in this new mass fit configuration is  $199767 \pm 481$ ,  
 1363 compared to  $34102 \pm 299$  in the nominal fit configuration (Table 4.15). The projection  
 1364 of the PDF used for Fit B in the wide mass range is shown in Fig. 5.21. A decay time  
 1365 fit is then performed on the resulting sample following the same strategy as reported  
 1366 in Sec. 5.3. The correlated disagreement<sup>4</sup> between the result of this fit and that of the  
 1367 nominal fit is  $2.3 \sigma$  and  $1.8 \sigma$  for  $S_f$  and  $S_{\bar{f}}$  respectively. Because of this discrepancy, the  
 1368 difference between the newly obtained  $S_f$  and  $S_{\bar{f}}$  coefficients, and the nominal values is  
 1369 taken as systematic uncertainty, yielding 0.0042 and 0.0023 for  $S_f$  and  $S_{\bar{f}}$  respectively.  
 1370 The correlation between two systematic uncertainties on  $S_f$  and  $S_{\bar{f}}$  is estimated to be  
 1371 0.7, as shown in Sec. J.

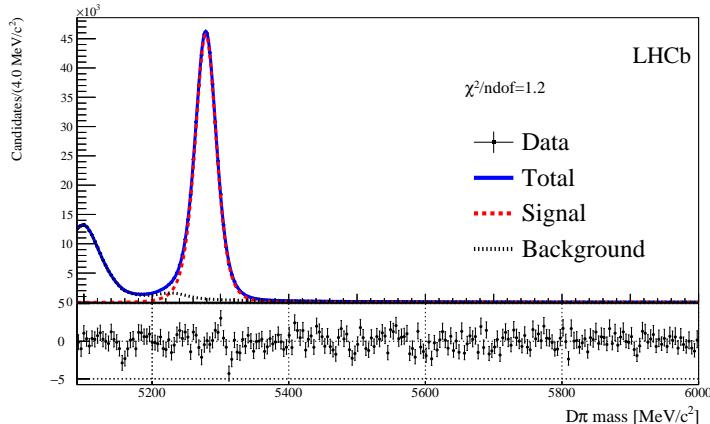


Figure 5.21 –  $D^\mp\pi^\pm$  mass distribution of the  $\pi$  sample with the results of Fit B with large mass window superimposed.

1372 Another test is made by repeating the mass fit with a different strategy:

- 1373 • a PIDK < 0 cut (instead of PIDK < 5) is applied on the pion PID in order to define  
 1374 the pion sample;
- 1375 • both Fit A and Fit B are performed in the narrow signal region ( $[5220, 5600]$  MeV/ $c^2$ );

<sup>4</sup>Defined as the difference between the fit results divided by the difference in quadrature between the fitted uncertainties.

## Chapter 5. Measurement of $CP$ violation in $B^0 \rightarrow D^\mp\pi^\pm$ decays

- during Fit A, only the pion sample is considered (no simultaneous fit in kaon and pion samples is performed);
- only  $B^0 \rightarrow D^\mp K^\pm$  and combinatorial background are considered, whereas all the other physical background are neglected;
- the  $B^0 \rightarrow D^\mp K^\pm$  yield is Gaussian constrained to be  $0.0101 \pm 0.0012$  of the signal yield, based on the selection efficiencies (including the  $\text{PIDK} < 0$  cut) found on Monte Carlo.

The signal and total background yield obtained in this fit are  $406818 \pm 674$  and  $23938 \pm 266$  respectively. The projection of the PDF used for Fit A and Fit B in this configuration is shown in Fig. 5.22. A decay time fit is performed on the resulting sample with *sWeights* by following the same strategy as reported in Sec. 5.3. The correlated discrepancy between 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. Given the good level of agreement, and the fact that systematic uncertainties on the PID efficiencies are already considered, no further systematics are assigned.

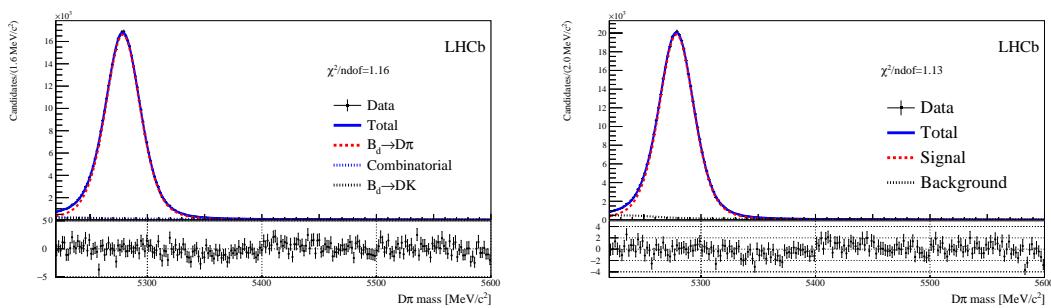


Figure 5.22 –  $D^\mp\pi^\pm$  mass distribution of the alternative  $\pi$  sample defined by the cut  $\text{PIDK} < 0$  on the bachelor pion with the result of Fit A (left) and Fit B (right) superimposed.

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

- the amount of combinatorial and  $B^0 \rightarrow D^\mp K^\pm$  backgrounds in the signal region is very small, and their presence doesn't affect significantly the fitted  $S_f$  and  $S_{\bar{f}}$  coefficients as these are compatible with the nominal fit result.
- the  $B^0 \rightarrow D^\mp\pi^\pm$  decay time and invariant mass are uncorrelated observables, so that it is possible to adopt the *sWeights* technique to unfold signal and background.

1401 In fact, *sWeights* are not affecting the fitted  $S_f$  and  $S_{\bar{f}}$  coefficients as they are  
1402 compatible with the nominal fit with *sWeights*.

- 1403 • any systematics due to a wrong modelling of signal and/or background PDF is  
1404 negligible, since the fitted value obtained from the nominal fit (with *sWeights*) and  
1405 this alternative fit (with no mass fit at all) are compatible.

Table 5.6 – 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	$6.8 \pm 2.9 \pm 1.2$	$3.1 \pm 3.0 \pm 1.2$
Babar	$-2.3 \pm 4.8 \pm 1.4$	$4.3 \pm 4.8 \pm 1.4$
This analysis	$5.8 \pm 2.0 \pm 1.1$	$3.8 \pm 2.0 \pm 0.7$

## 1406 5.5 Summary and interpretation of the result

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

$$S_f = 0.058 \pm 0.020(\text{stat}) \pm 0.011(\text{syst}) \quad (5.13)$$

$$S_{\bar{f}} = 0.038 \pm 0.020(\text{stat}) \pm 0.007(\text{syst}). \quad (5.14)$$

1407 The statistical and systematic correlations between  $S_f$  and  $S_{\bar{f}}$  are 60% and -41%,  
 1408 respectively. These values are in agreement with and more precise than the measurements  
 1409 from Belle and Babar [31, 28]. A direct comparison is shown in Table 5.6.

The measurement of  $S_f$  and  $S_{\bar{f}}$  is interpreted in terms of the angle  $(2\beta + \gamma)$ , the ratio of amplitudes  $r_{D\pi}$ , and the strong phase  $\delta$ , using a similar procedure to that of [78]. By taking external measurements of  $\beta$  and  $r_{D\pi}$ , we derive confidence intervals for  $\gamma$  and  $\delta$ . The values  $\beta = (22.2 \pm 0.7)^\circ$  from [77] and  $r_{D\pi} = (1.82 \pm 0.12 \pm 0.36(\text{SU}(3)))\%$  are used as inputs. The latter is calculated from the branching fraction of  $B^0 \rightarrow D_s^\mp\pi^\pm$  decays, assuming SU(3) symmetry, following the same relation of Refs. [27, 26]:

$$r_{D\pi} = \tan \theta_c \frac{f_{D^+}}{f_{D_s}} \sqrt{\frac{\mathcal{B}(B^0 \rightarrow D_s^\mp\pi^\pm)}{\mathcal{B}(B^0 \rightarrow D^\mp\pi^\pm)}}, \quad (5.15)$$

1410 where  $\tan \theta_c = 0.23101 \pm 0.00032$  is the tangent of the Cabibbo angle from Ref. [19],  
 1411 and the ratio of decay constants  $\frac{f_{D_s}}{f_{D^+}} = 1.173 \pm 0.003$  and the branching fractions  
 1412  $\mathcal{B}(B^0 \rightarrow D_s^\mp\pi^\pm) = (2.16 \pm 0.26) \times 10^{-5}$  and  $\mathcal{B}(B^0 \rightarrow D^\mp\pi^\pm) = (2.52 \pm 0.13) \times 10^{-3}$  are  
 1413 taken from Ref. [20]. An additional 20% relative error is added to account for uncertainties  
 1414 due to possible non-factorizable SU(3)-breaking effects, as reported in Ref. [79].

1415 Using the PLUGIN method as described in Ref. [78],  $\gamma$  is determined to be in the interval  
 1416  $[5, 86]^\circ \cup [185, 266]^\circ$  and  $\delta$  to be in the interval  $[-41, 41]^\circ \cup [140, 220]^\circ$ , both at the 68%  
 1417 C.L. These intervals are illustrated in Fig. 5.23. In Fig. 5.24  $\gamma$  and  $\delta$  are shown in the 2  
 1418 dimensional plane. The interval of  $|\sin(2\beta + \gamma)|$  is determined to be  $[0.77, 1.0]$  at the 68%  
 1419 C.L. as shown in Fig. 5.25. The intervals for  $\gamma$ ,  $\delta$  and  $|\sin(2\beta + \gamma)|$  are also determined  
 1420 by assuming a SU(3)-breaking uncertainty of 0%, 20% and 100% on the value of  $r_{D\pi}$ .  
 1421 These are presented in Fig. 5.26 and Fig. 5.27.

## 5.5. Summary and interpretation of the result

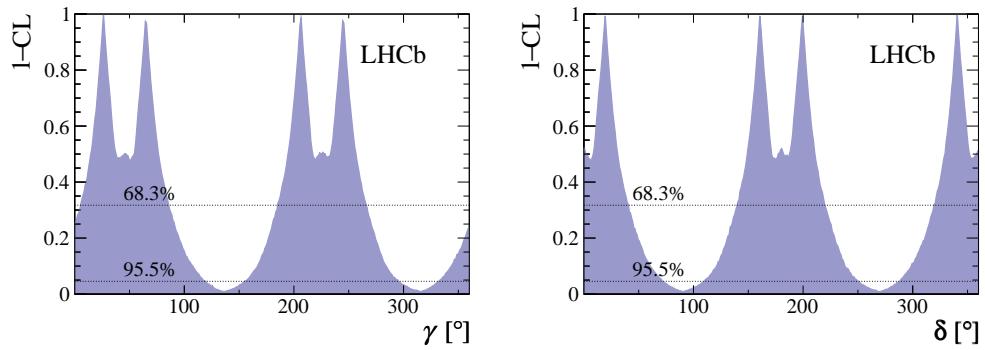


Figure 5.23 – 1–CL distributions of  $\gamma$  and strong phase  $\delta$  obtained using the measured values of  $S_f$  and  $S_{\bar{f}}$ .

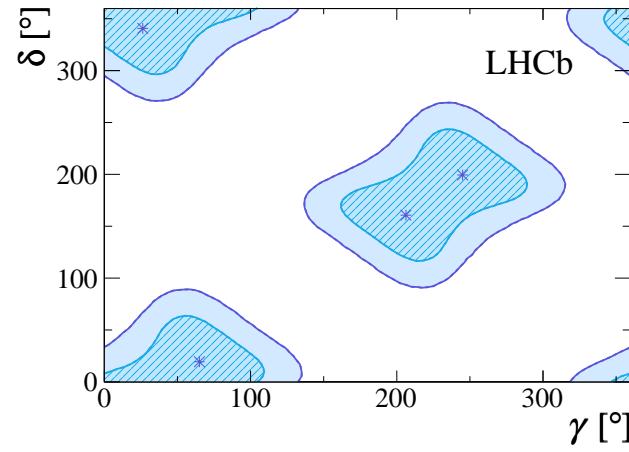


Figure 5.24 – 1–CL distributions of  $\gamma$  and strong phase  $\delta$  obtained using the measured values of  $S_f$  and  $S_{\bar{f}}$ .

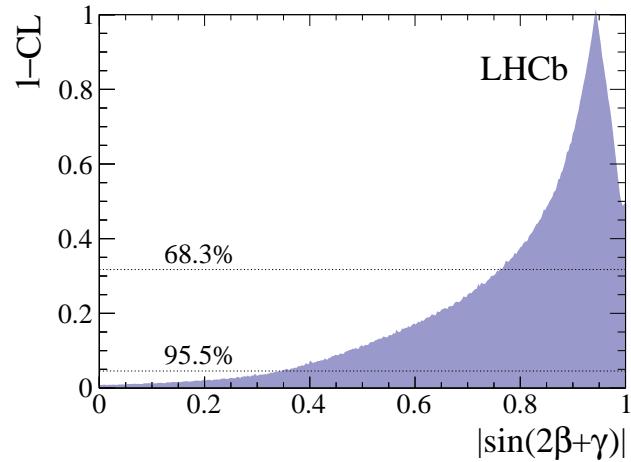


Figure 5.25 – 1–CL distribution of  $|\sin(2\beta + \gamma)|$  using the measured values of  $S_f$  and  $S_{\bar{f}}$ .

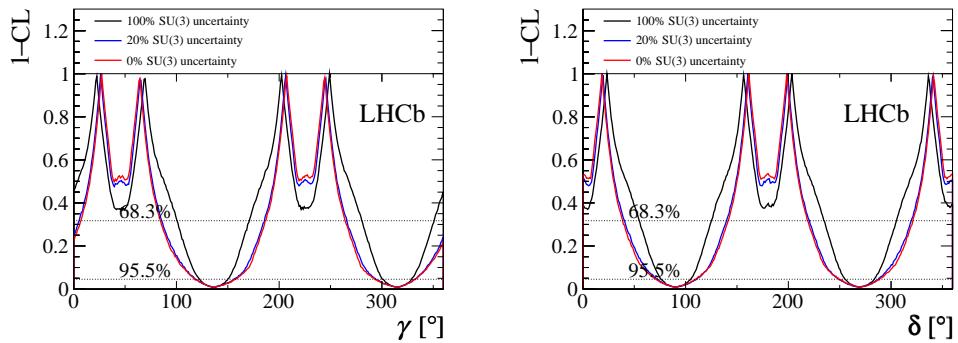


Figure 5.26 – 1–CL distribution of (left)  $\gamma$  and (right)  $\delta$  for assumptions of 0 %, 20 % and 100 % for the SU(3) uncertainty on the parameter  $r_{D\pi}$ .

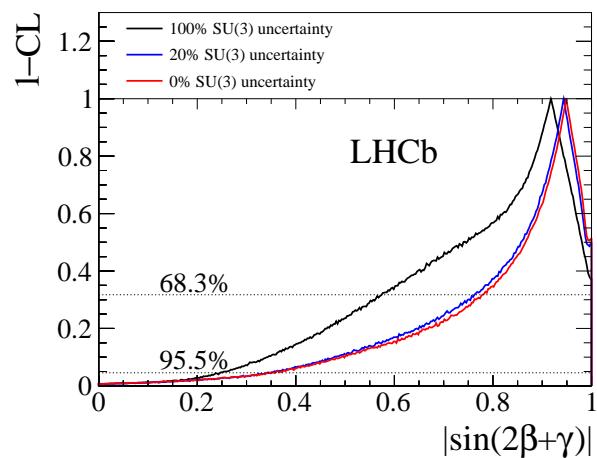


Figure 5.27 – 1–CL distribution of  $|\sin(2\beta + \gamma)|$  for assumptions of 0 %, 20 % and 100 % for the SU(3) uncertainty on the parameter  $r_{D\pi}$ .

## Appendices

**A Opposite-side tagging studies**

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

A fit to the mass distribution of  $B^\pm$  candidates is done to calculate *sWeights*, used in the subsequent steps of the analysis to subtract the backgrounds surviving the selection. A two-steps procedure similar to that adopted for the  $B^0 \rightarrow D^\mp\pi^\pm$  analysis (“Fit A” in a wide mass window to account for all backgrounds, and “Fit B” in a subset to calculate the weights, as described in Sec. 4.3) is used for the fit of the  $B^+ \rightarrow D^0\pi^+$  candidates. The projection of the total PDF on the  $\pi$  sample and the  $K$  sample (“Fit A”) is shown in Fig. 28, as well as the projection of the total PDF in the reduced sample (“Fit B”). The  $\pi$  sample and  $K$  sample are defined by the PID requirement on the companion track,  $\text{PIDK} < 5$  and  $\text{PIDK} > 5$ , respectively.

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

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

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

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

All the PDFs listed above are defined in Appendix F.

The values for the fitted parameters floated in the fit are reported in Table 7 for Fit A and Table 8 for Fit B. The naming convention for each parameter is similar to the one used in Sec. 4.3.2.

## A. Opposite-side tagging studies

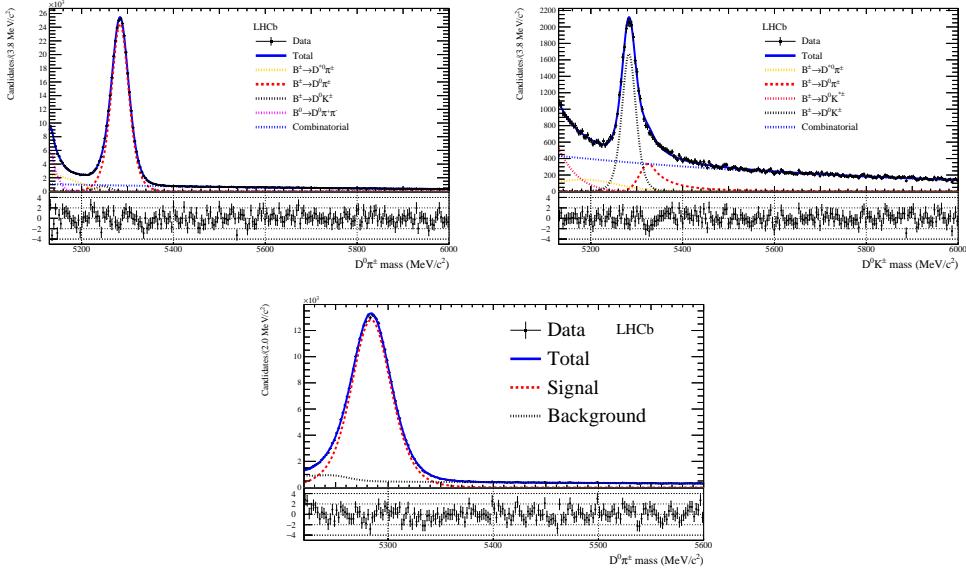


Figure 28 – Top left:  $D^0\pi^\pm$  mass distribution of the  $\pi$  sample. Top right:  $D^0K^\pm$  mass distribution of the  $K$  sample. (in the two samples). The result of the simultaneous fit (Fit A) to both samples is superimposed. Bottom:  $D^\mp\pi^\pm$  mass distribution of the  $\pi$  sample with the results of Fit B superimposed.

### 1451 A.2 Reweighting of $B^\pm \rightarrow D^0\pi^\pm$ to $B^0 \rightarrow D^\mp\pi^\pm$

1452 In order to improve the OS calibration portability, a multi-dimensional reweighting of the  
 1453 *sWeighted*  $B^\pm \rightarrow D^0\pi^\pm$  distributions is made to match the  $B^0 \rightarrow D^\mp\pi^\pm$  kinematics.

1454 The reweighting is made in two steps. In the first step, the variables considered in the  
 1455 reweighting are the transverse momentum, the pseudo-rapidity  $\eta$  and the decay time  $\tau_B$  of  
 1456 the  $B$  candidate, as well as the number of tracks and the number of primary vertices of the  
 1457 events. A BDT-based approach is followed in order to cope with the high dimensionality  
 1458 of the space as well as with the correlations among variables [80]. A comparison between  
 1459 weighted and unweighted distributions is provided in Figs. 29 and 30.

1460 In the second step, a new weight is computed by comparing the two-dimensional distribu-  
 1461 tions of the  $D$  meson decay time and HLT2 trigger composition between  $B^\pm \rightarrow D^0\pi^\pm$   
 1462 and  $B^0 \rightarrow D^\mp\pi^\pm$  after *sWeights* and the weights from the first step are applied. The  
 1463 HLT2 composition observable is a categorical variable which describes which HLT2 trigger  
 1464 line has been fired by the  $B$  candidate:

- 1465     • `Hlt2Topo2BodyBBDTDecision` only (value 0);
- 1466     • `Hlt2Topo3BodyBBDTDecision` or `Hlt2Topo4BodyBBDTDecision` only (value 1);
- 1467     • overlap of the first two categories (value 2).

## Appendix . Appendices

---

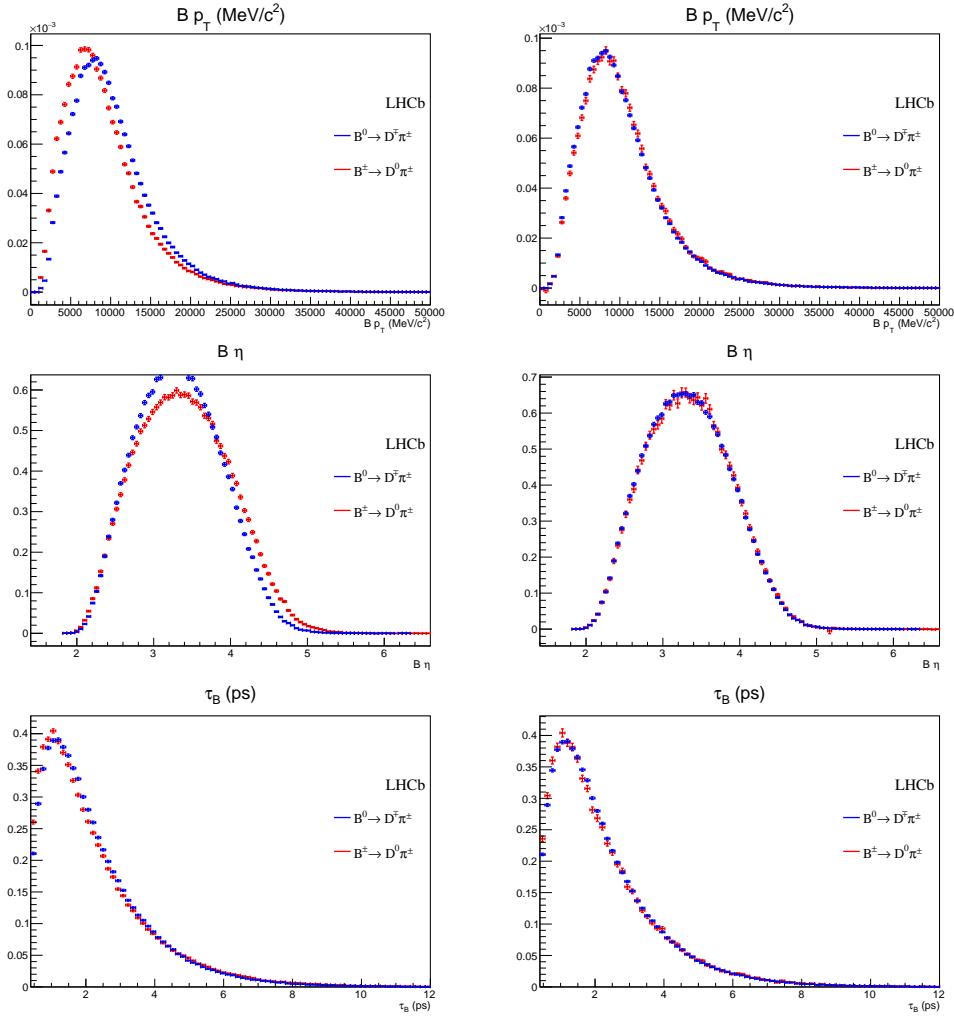


Figure 29 – Normalised *sWeighted* distributions of the transverse momentum, the pseudo-rapidity  $\eta$  and the decay time  $\tau_B$  of the  $B^0$  and  $B^\pm$  mesons. Left: unweighted distributions. Right: distributions after reweighting  $B^\pm \rightarrow D^0\pi^\pm$ .

## A. Opposite-side tagging studies

Table 7 – Results of the  $B^\pm \rightarrow D^0\pi^\pm$  mass fit (Fit A).

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

<sup>1468</sup> This reweighting is done separately from the first one in order to avoid a too fine partition  
<sup>1469</sup> of the samples, which would give a very low statistics in less populated bins. The reason  
<sup>1470</sup> why HLT2 trigger and  $D$  decay time are reweighted simultaneously is that these two  
<sup>1471</sup> observables are correlated. The result of this second reweighting is shown in Fig. 31.

### <sup>1472</sup> A.3 GOF tests for OS calibration on $B^\pm \rightarrow D^0\pi^\pm$ data

<sup>1473</sup> The number of free parameters (10) used in the GLM model for the OS calibration  
<sup>1474</sup> (Sec. 3.2.1) is the minimum number to obtain satisfactory GOF metrics. The GOF tests  
<sup>1475</sup> are performed automatically by the EPM; the metrics include the Pearson  $\chi^2$ , the deviance

Table 8 – Results of the  $B^\pm \rightarrow D^0\pi^\pm$  mass fit (Fit B)

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

## Appendix . Appendices

---

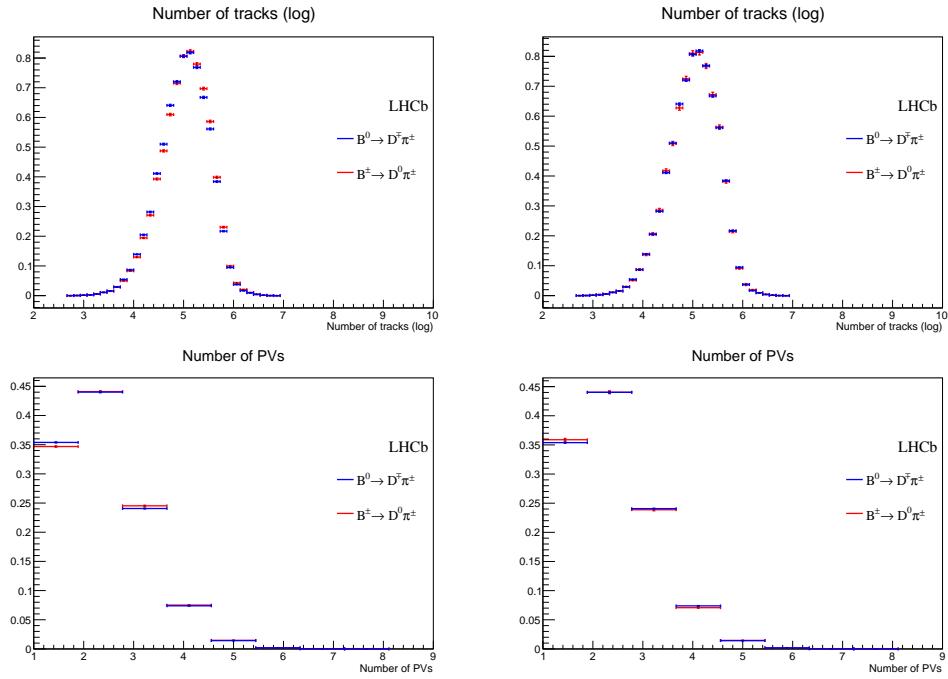


Figure 30 – Normalised *sWeighted* distributions of the number of tracks and PVs in a  $B^0$  or  $B^\pm$  event. Left: unweighted distribution. Right: distributions after reweighting  $B^\pm \rightarrow D^0\pi^\pm$ .

<sup>1476</sup>  $G^2$ , the Cressie-Read ( $CR$ ) metric and the le Cessie-van Houwelingen-Copas-Hosmer  
<sup>1477</sup> metric ( $S$ ), all described in Ref. [54].

<sup>1478</sup> All these tests return a normally distributed score: this means that the score is equal  
<sup>1479</sup> to the distance (measured in standard deviations) from the perfect case, which is a null  
<sup>1480</sup> score. A comparison between the GOF scores obtained for the nominal calibration (10  
<sup>1481</sup> free parameters) and a simplified model (8 free parameters) is shown in Table 9. In a  
<sup>1482</sup> simplified model, all scores are more than  $\sim 3$  standard deviations away from a perfect  
<sup>1483</sup> fit, whereas the scores for the nominal model are  $\sim 2$  standard deviations at most. For  
<sup>1484</sup> this reason, 10 free parameters are chosen, and the model cannot be simplified further.

Table 9 – 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

## A. Opposite-side tagging studies

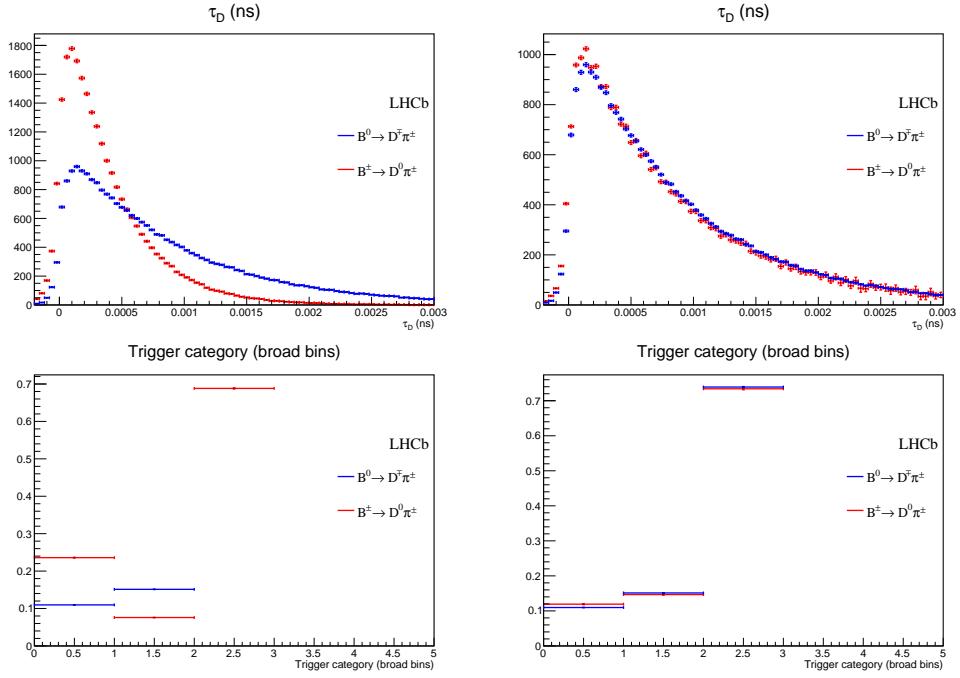


Figure 31 – Normalised  $sWeighted$  distributions of the  $D^{\mp}$  and  $D^0$  mesons decay time and HLT2 trigger composition, where the weight obtained from the first reweighting step is also applied. Left: unweighted distributions. Right: distributions after reweighting  $B^{\pm} \rightarrow D^0\pi^{\pm}$ .

## **B Opposite-side electron optimisation**

The correlation between the predicted mistag  $\eta$  and the input features is shown in Figs. 32, 33 and 34 for the Run 1 new, Run 2 B2CC and Run 2 B2OC implementations of the OSe tagger, respectively. This correlation, known as *partial dependence*, allows to check the impact of each feature on the classifier, in addition to the F-score (Sec. 3.3.4).

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

- For a given set of maximum depth and number of trees values, the training set is bootstrapped 10 times. Each bootstrapped sample is then divided in three exclusive subsamples.
- The first subsample is used to train a BDT. The BDT is then transformed into a mistag probability, and a calibration is performed on the second subsample (a simple second order logistic function is used). Finally, the calibration is applied on the third sample, where the per-event tagging power is computed. The ROC AUC is also obtained as additional performance metric.
- The procedure at the previous point is repeated by permutating the 3 samples. This means that, in total, there are  $3 \times 10 = 30$  approximately independent estimations of the BDT performance for each set of hyperparameters. The average tagging power and ROC AUC values are finally computed over the 30 estimations, together with the standard error on the mean.

The result for the Run 1 new OSe algorithm is shown in Fig. 35. The performance is weakly dependent on the hyperparameters. For this reason, the maximum depth and the number of trees are fixed to 3 and 300 respectively, in order to maintain the size of the algorithm sufficiently small.

## B. Opposite-side electron optimisation

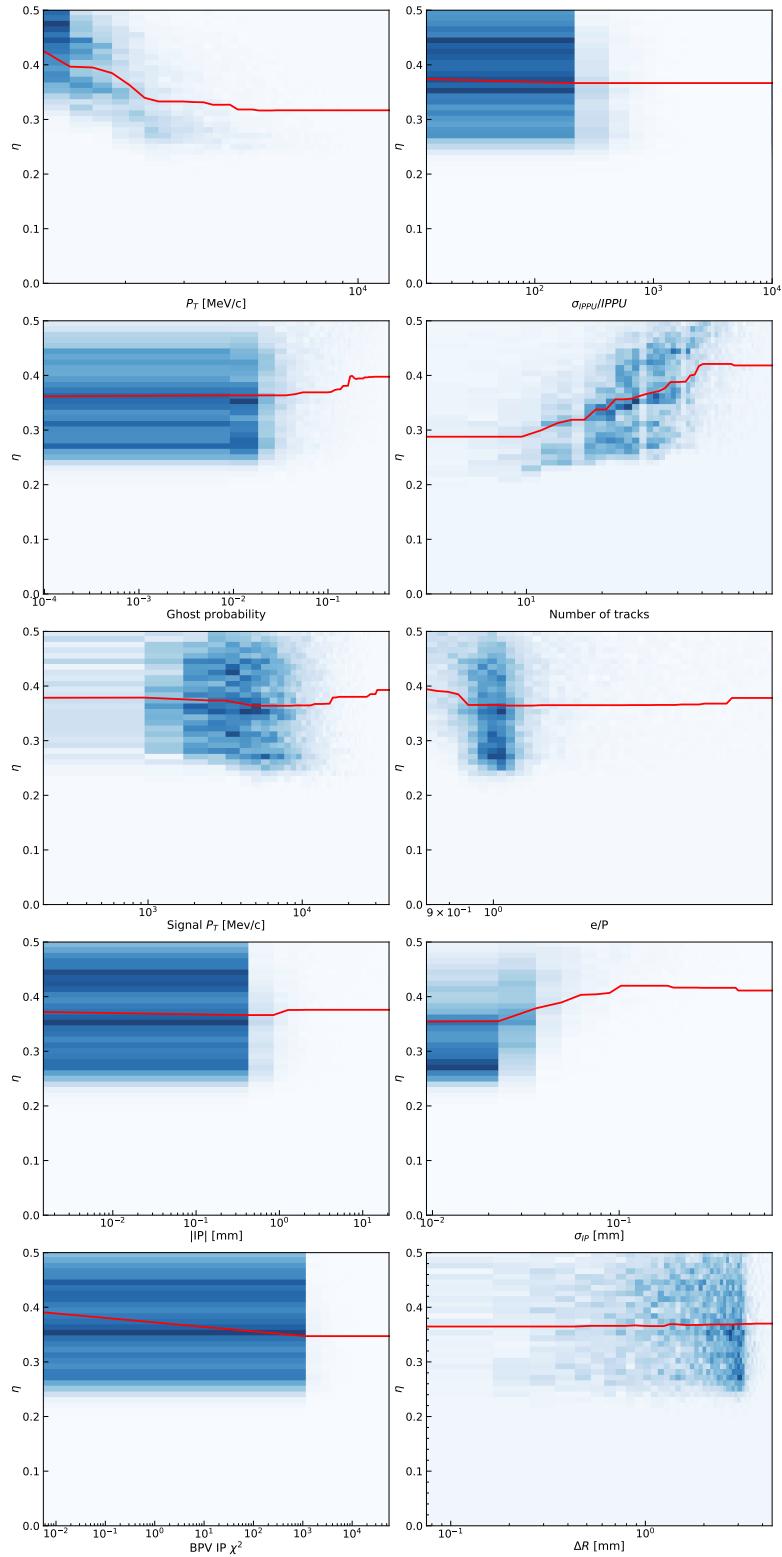


Figure 32 – Partial dependence of the predicted mistag  $\eta$  (OSe Run 1 tagger) for each feature used as BDT input, marginalised over any other feature. The blue two-dimensional distributions represent the *sWeighted* data, whereas the red line shows the average  $\eta$  for each feature bin.

## Appendix . Appendices

---

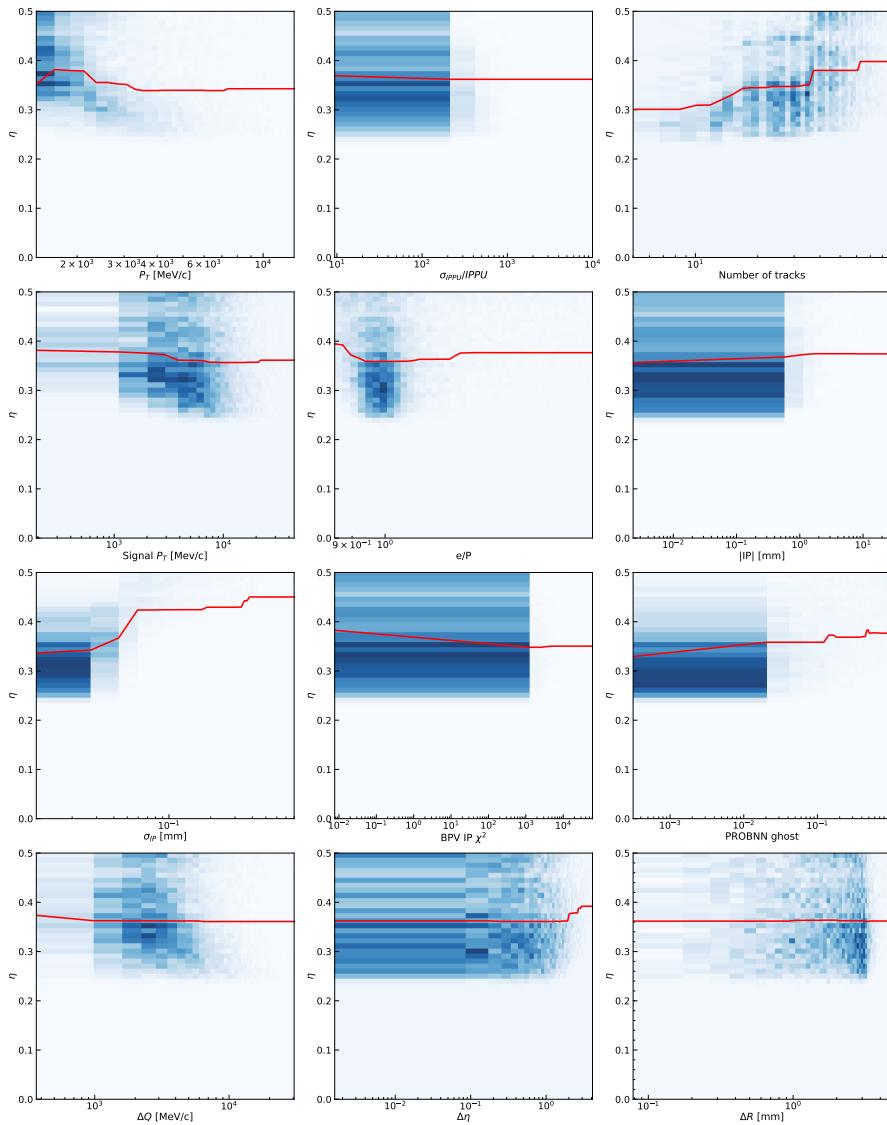


Figure 33 – Partial dependence of the predicted mistag  $\eta$  (OSe Run 2 B2CC tagger) for each feature used as BDT input, marginalised over any other feature. The blue two-dimensional distributions represent the *sWeighted* data, whereas the red line shows the average  $\eta$  for each feature bin.

## B. Opposite-side electron optimisation

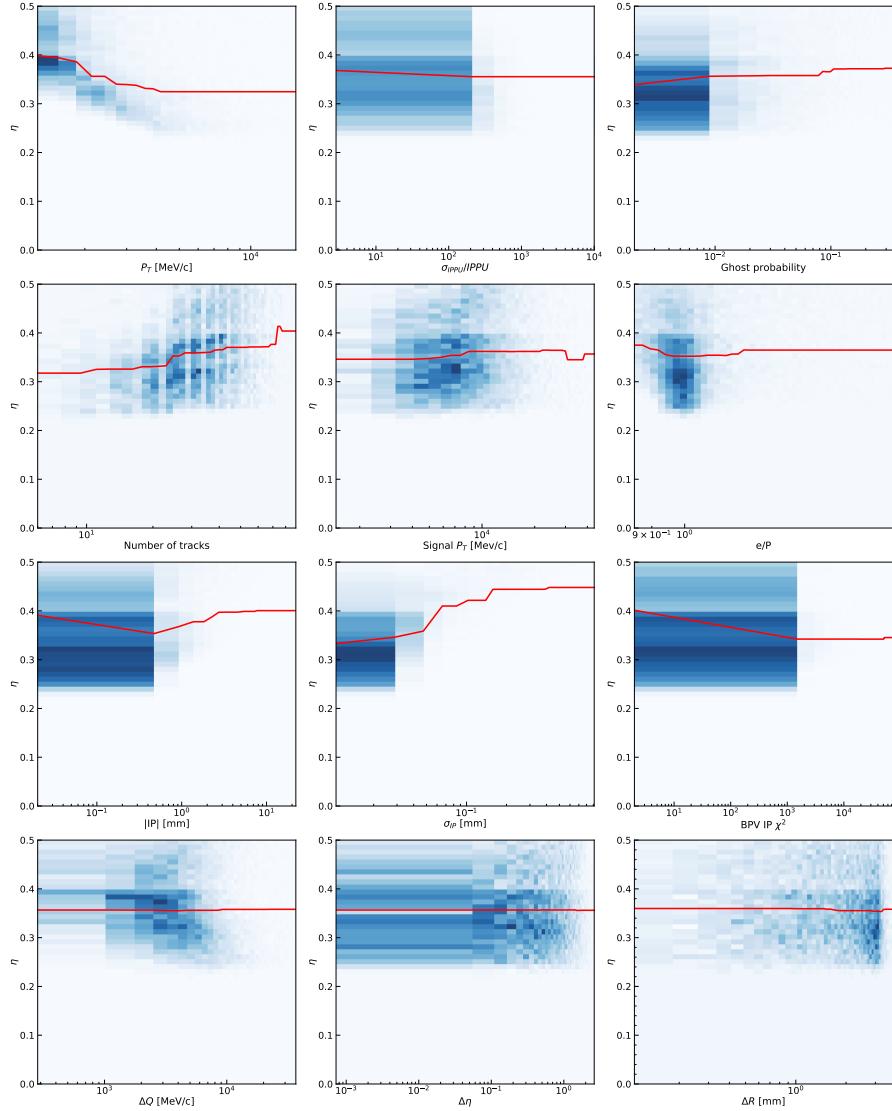


Figure 34 – Partial dependence of the predicted mistag  $\eta$  (OSe Run 2 B2OC tagger) for each feature used as BDT input, marginalised over any other feature. The blue two-dimensional distributions represent the *sWeighted* data, whereas the red line shows the average  $\eta$  for each feature bin.

## Appendix . Appendices

---

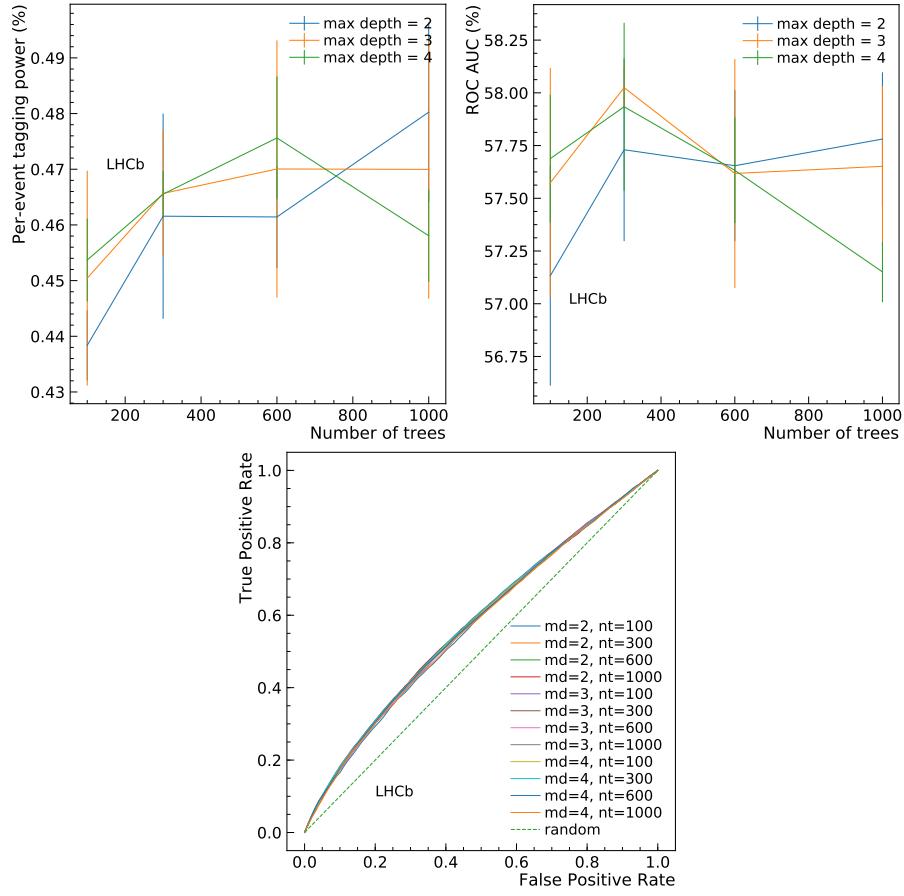


Figure 35 – Hyperparameter tuning for the BDT classifier used for the mistag estimation by the Run 1 new version of the OSe tagger (cross-validation). The tagging power (top left), ROC AUC (top right), and ROC curves (bottom) for each set of hyperparameters are shown.

1510 

## C Selection studies

 1511 

### C.1 BDT input features

1512 The distributions of the input features for the BDT are shown in Figs. 36, 37 and 38.

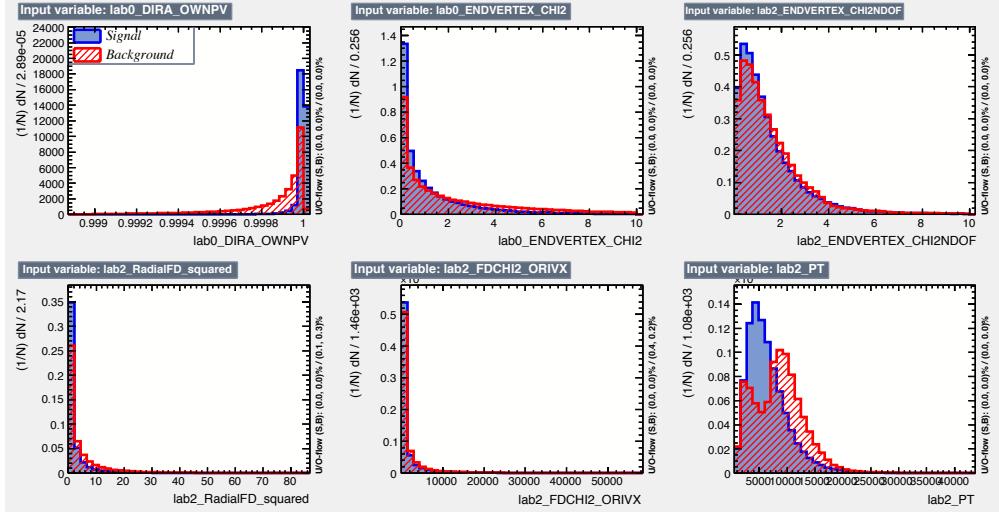


Figure 36 – Input features used in the BDT training. From top left to bottom right: cosine of the direction angle of the  $B^0$ ,  $\chi^2$  of the  $B^0$  vertex,  $\chi^2/\text{ndof}$  of the  $D^\mp$  vertex,  $D^\mp$  radial flight distance,  $D^\mp$  flight distance  $\chi^2$  with respect to the  $B^0$  vertex and transverse momentum of the  $D^\mp$ .

 1513 

### C.2 Multiple candidates

 1514 In Table 10, a summary of the multiple candidates left after the stripping selection is  
 1515 given, while Table 11 reports the number of multiple candidates after both stripping and  
 1516 offline selection.

## Appendix . Appendices

---

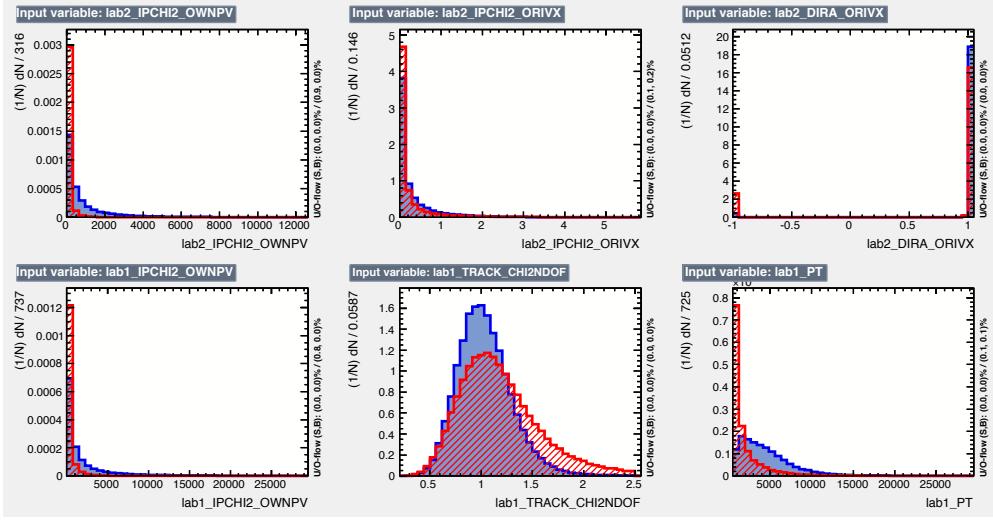


Figure 37 – Input features used in the BDT training. From top left to bottom right:  $D^\mp$   $\text{IP}\chi^2$  with respect to the associated PV and the  $B^0$  vertex, cosine of the direction angle of the  $D^\mp$ , the  $\text{IP}\chi^2$  with respect to the associated PV of the bachelor pion, track  $\chi^2/\text{ndof}$  of the bachelor pion and the transverse momentum of the bachelor pion.

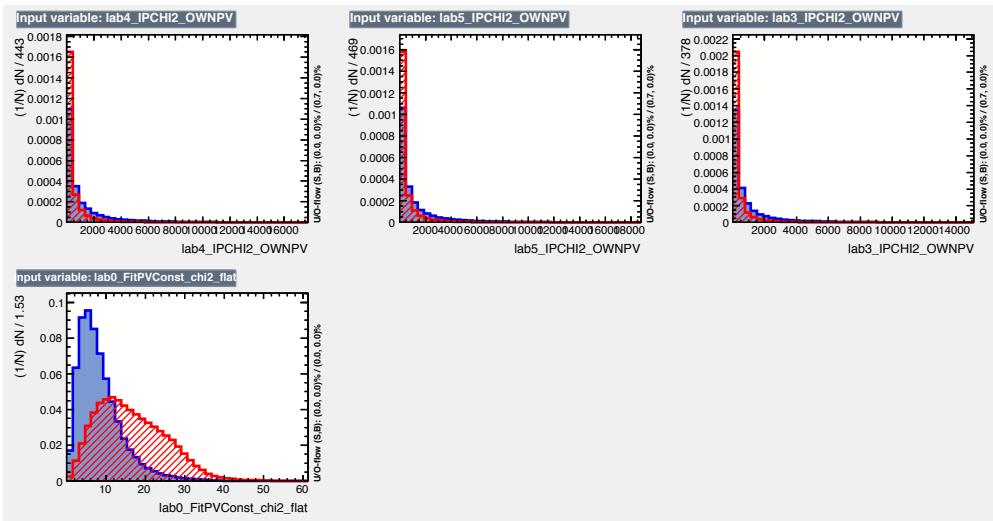


Figure 38 – Input features used in the BDT training. From top left to bottom right:  $\text{IP}\chi^2$  of the associated primary vertex of the three  $D^\mp$  daughters and the  $\chi^2$  of the decay tree fit with PV constraint.

### C. Selection studies

Table 10 – Statistical information on multiple  $B^0 \rightarrow D^\mp\pi^\pm$  candidates left after the stripping selection.

	2011	2012		
fraction of $B^0$ candidates that are not unique in a given event	18.3 %	19.5 %		
fraction of events with multiple $B^0$ candidates	9.0 %	9.6 %		
	# $B^0$ cands	#events	# $B^0$ 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	
fraction of $B^0$ candidates to be discarded to maintain one candidate per event	10.1 %	11.0 %		

## Appendix . Appendices

---

Table 11 – Statistical information on multiple  $B^0 \rightarrow D^\mp\pi^\pm$  candidates left after both stripping and offline selection.

	2011	2012	
fraction of $B^0$ candidates pairs that are not unique in an event	0.8 %	0.8 %	
fraction of events with multiple $B^0$ candidates	0.4 %	0.4 %	
	# $B^0$ cands	#events	# $B^0$ cands
	1	483074	1
	2	1886	2
	3	38	3
	4	4	4
	5	1	5
fraction of $B^0$ candidates to be discarded to maintain one candidate per event	0.4 %	0.4 %	

<sup>1517</sup> **D Particle identification plots**

<sup>1518</sup> The  $p$ ,  $\eta$  distributions and the PIDK efficiency, misidentification rates before the resampling in bins of  $p$ ,  $\eta$  are shown in Figs. 39 and 40 for the bachelor particle of the signal,  
<sup>1519</sup> and in Fig. 41 for the  $D^\mp$  daughters.

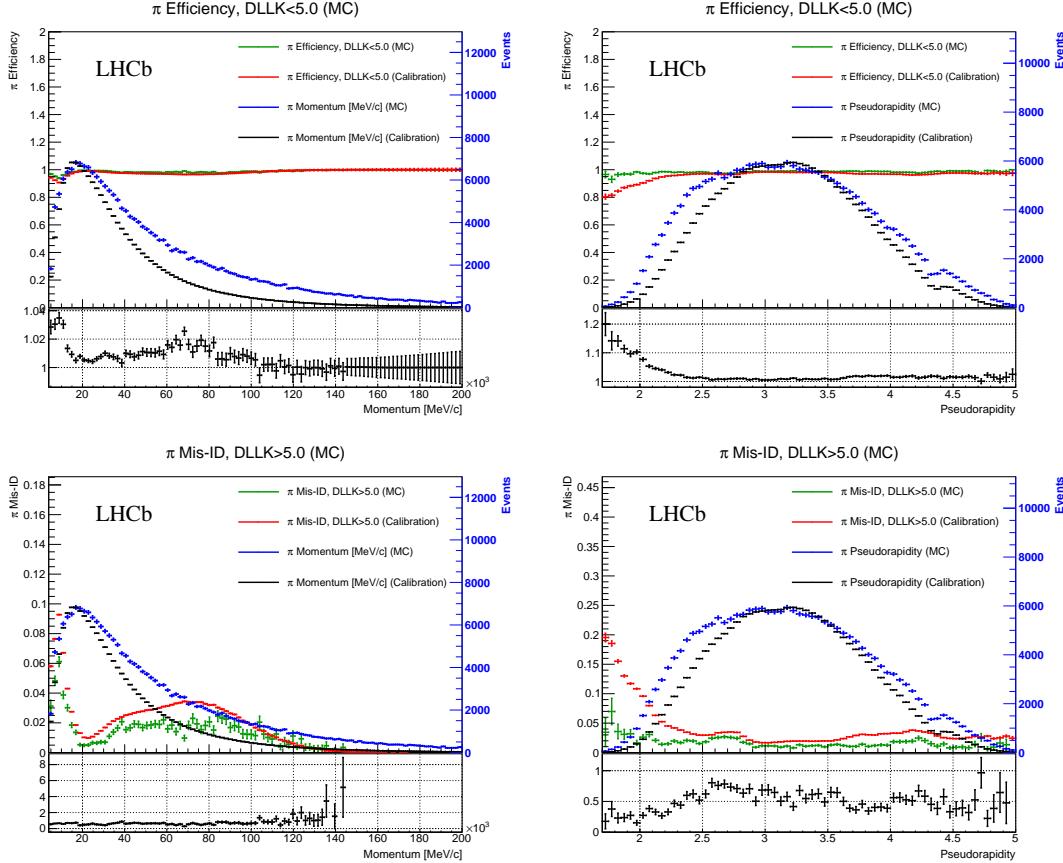


Figure 39 – PIDK efficiency (top) and misidentification rate (bottom) for bachelor pions as a function of momentum  $p$  (left) and pseudorapidity  $\eta$  (right), both for  $B^0 \rightarrow D^\mp \pi^\pm$  signal MC (green) and calibration mode (red). The superimposed histograms show the  $p$  and  $\eta$  distributions of the MC signal (blue) and calibration (black) samples. The ratio of the efficiency or misidentification rate between the Monte Carlo and calibration samples is shown in the lower pad (black).

<sup>1520</sup>

## Appendix . Appendices

---

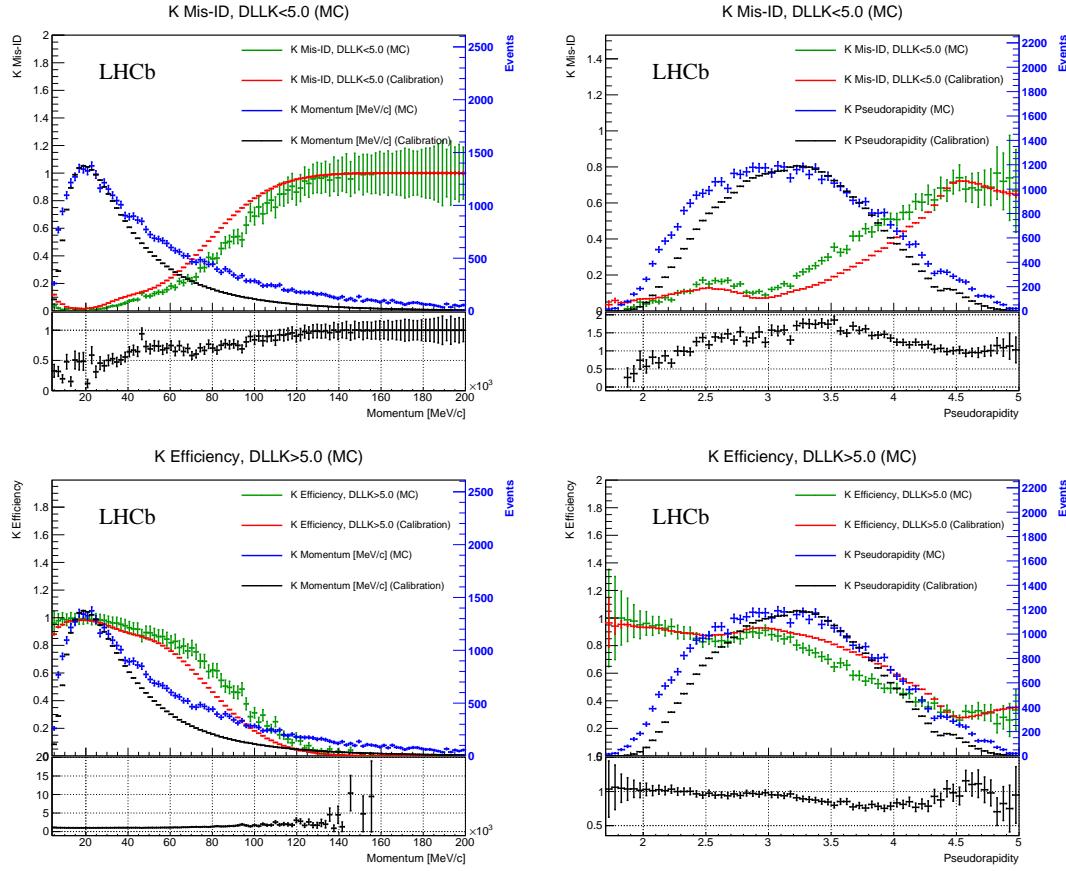


Figure 40 – PIDK misidentification rate (top) and efficiency (bottom) for bachelor kaons as a function of momentum  $p$  (left) and pseudorapidity  $\eta$  (right), both for  $B^0 \rightarrow D^\mp K^\pm$  MC (green) and calibration mode (red). The superimposed histograms show the  $p$  and  $\eta$  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 Monte Carlo and calibration samples is shown in the lower pad (black).

## D. Particle identification plots

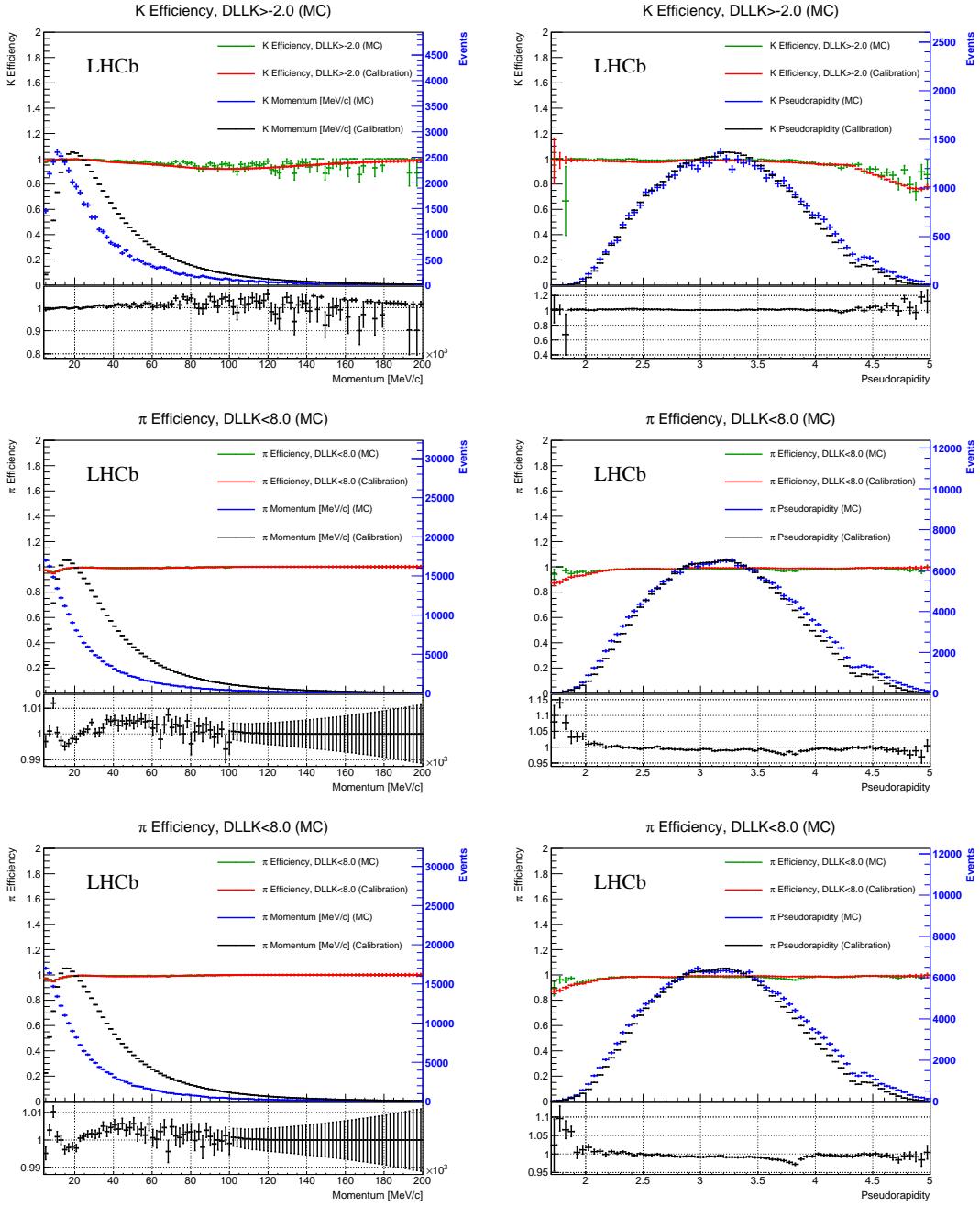


Figure 41 – PIDK efficiency for the kaon (top) and the two pions (middle, bottom) produced in the  $D^\mp$  decay as a function of momentum  $p$  (left) and pseudorapidity  $\eta$  (right), both for  $B^0 \rightarrow D^\mp \pi^\pm$  signal MC (green) and calibration mode (red). The superimposed histograms show the  $p$  and  $\eta$  distributions of the MC signal (blue) and calibration (black) samples. The ratio of the efficiency between the Monte Carlo and calibration samples is shown in the lower pad (black).

## 1521 E Invariant mass fit studies

1522 The small correlation between the  $B^0$  invariant mass and decay time is shown by comparing  
 1523 the distribution of the decay time in bins of the invariant mass after applying the full  
 1524 selection. This is done separately for signal and background. For the signal distribution  
 1525 simulated data is used and the decay time is shown in six bins of the invariant mass.  
 1526 (Fig. 42). In order to account for the combinatorial background, the upper mass sideband  
 1527 is chosen as a proxy. Fig. 43 shows the decay time in four bins of the invariant mass. The  
 1528 physics background contribution in the signal region is considered to be small enough,  
 1529 so that even a large correlation does not matter. Given the small differences for all  
 1530 distributions, the correlations between decay time and invariant mass is assumed to be  
 1531 small enough to justify the use of the invariant mass in the *sPlot* [66] technique for  
 1532 disentangling signal from background.

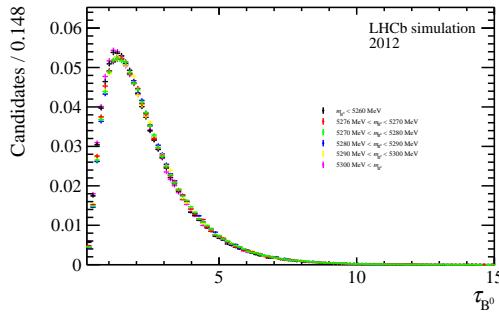


Figure 42 – Normalised signal decay time distributions in six bins of the invariant mass.

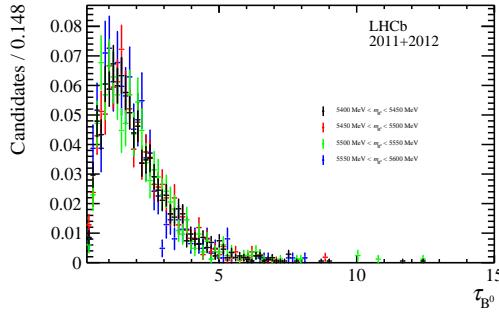


Figure 43 – Upper mass side band decay time distribution divided in four bins of the invariant mass. The shapes are shown normalised.

<sub>1533</sub> **F PDF definitions**

<sub>1534</sub> Throughout the text, each parameter  $p$  defined inside a PDF used for the mass fit is  
<sub>1535</sub> labelled as  $p_c^s$ , where  $s = \pi, K$  indicates the sample and  $c = B^0 \rightarrow D^\mp \pi^\pm, B^0 \rightarrow D^\mp \rho^\pm \dots$   
<sub>1536</sub> indicates the component. For sake of clarity, the  $s$  and  $c$  labels are dropped in the equations  
<sub>1537</sub> that follow. The mass *observable* is always indicated as  $m$ . The  $\propto$  symbol indicates that  
<sub>1538</sub> all PDFs are defined up to a normalisation constant, which depends on the interval chosen  
<sub>1539</sub> 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 the following parameters:

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

the single-sided Crystal Ball function [69] is expressed as follows:

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

- **Double-sided Hypatia function**

Having defined:

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

and its first derivative with respect to  $m, h'$ , then the double-sided Hypatia function  $H$  [67] is expressed as follows:

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

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

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

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

• **Single-sided Hypatia function**

A single-sided Hypatia function is obtained from a double-sided Hypatia function in 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 following 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 [68] is expressed as follows:

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

1543 **G Signal PDF for the decay-time fit**

1544 The PDF describing the  $B^0$  decay time distribution can be written in the most general  
1545 case as follows:

$$P(t|\eta_{OS}, \eta_{SS}, d_{OS}, d_{SS}) = \frac{1}{4\tau} e^{-\frac{t}{\tau}} \left[ C_{\cosh}^{\text{eff}} \cosh \left( \frac{\Delta\Gamma t}{2} \right) + C_{\sinh}^{\text{eff}} \sinh \left( \frac{\Delta\Gamma t}{2} \right) \right. \\ \left. + C_{\cos}^{\text{eff}} \cos(\Delta m t) + C_{\sin}^{\text{eff}} \sin(\Delta m t) \right]. \quad (16)$$

1546 The four “effective”  $CP$  coefficients inside Eq. 16 depend on the final state ( $f = D^+ \pi^-$  or  
1547  $f = D^- \pi^+$ ), the tagging decision ( $d_i = -1, 0, +1$ , for  $i = \text{OS, SS}$ ), the mistag and the  
1548 tagging efficiency of the OS and SS taggers, and the asymmetries (production, detection  
1549 and tagging efficiency). In the ideal case (no asymmetries, zero mistag, 100% tagging  
1550 efficiency), these effective coefficients become the ‘physical’ coefficients already introduced  
1551 in Sec. 1.4.2.

1552 The OS and SS taggers are combined on the fly during the time fit. All the steps to build  
1553 the final PDF are described in details below.

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

$$\varepsilon_b^i = \varepsilon_{\text{tag}}^i - \frac{1}{2}\Delta\varepsilon^i, \quad (17)$$

$$\varepsilon_{\bar{b}}^i = \varepsilon_{\text{tag}}^i + \frac{1}{2}\Delta\varepsilon^i, \quad (18)$$

1554 where  $\varepsilon_b^i$  and  $\varepsilon_{\bar{b}}^i$  are the two different tagging efficiencies for  $B^0$  and  $\bar{B}^0$ . Moreover, the  
1555 OS and SS mistag are calibrated taking into account asymmetries between  $B^0$  and  $\bar{B}^0$   
1556 according to Eq. 3.10. We will refer to them here as  $\omega_i^b$  and  $\omega_i^{\bar{b}}$  (for  $i = \text{OS, SS}$ ). It’s now  
1557 convenient to define the quantity  $\Delta^\pm(d_{OS}, d_{SS})$  in the following way according to the OS  
1558 and SS tagging decisions.

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

$$\Delta^\pm = (1 - \varepsilon_b^{\text{OS}} - \varepsilon_b^{\text{SS}} + \varepsilon_b^{\text{OS}} \varepsilon_b^{\text{SS}}) \pm (1 - \varepsilon_{\bar{b}}^{\text{OS}} - \varepsilon_{\bar{b}}^{\text{SS}} + \varepsilon_{\bar{b}}^{\text{OS}} \varepsilon_{\bar{b}}^{\text{SS}}). \quad (19)$$

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

$$\Delta^\pm = \frac{1}{2}\varepsilon_b^j \left[ 1 - \varepsilon_b^i + d_j \left( 1 - \varepsilon_b^i - 2\omega_j^b + 2\omega_j^b \varepsilon_b^i \right) \right] \\ \pm \frac{1}{2}\varepsilon_{\bar{b}}^j \left[ 1 - \varepsilon_{\bar{b}}^i + d_j \left( 1 - \varepsilon_{\bar{b}}^i - 2\omega_j^{\bar{b}} + 2\omega_j^{\bar{b}} \varepsilon_{\bar{b}}^i \right) \right]. \quad (20)$$

## Appendix . Appendices

---

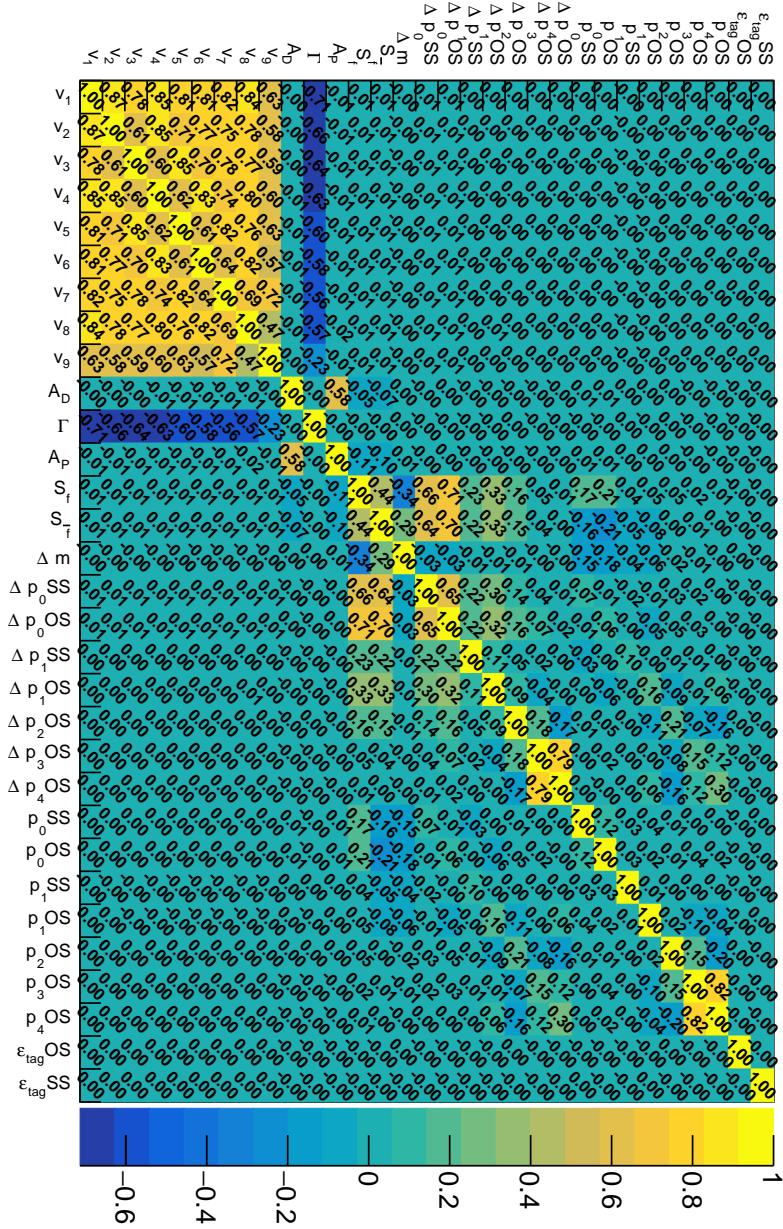
- If  $d_{\text{OS}} = d_{\text{SS}} = 1$  ( $B^0$  candidates tagged by both taggers):

$$\begin{aligned} \Delta^\pm &= \frac{1}{4} \varepsilon_b^{\text{SS}} \varepsilon_b^{\text{OS}} \left[ 1 + d_{\text{SS}} \left( 1 - 2\omega_{\text{SS}}^b \right) + d_{\text{OS}} \left( 1 - 2\omega_{\text{OS}}^b \right) \right. \\ &\quad \left. + d_{\text{OS}} d_{\text{SS}} \left( 1 - 2\omega_{\text{SS}}^b - 2\omega_{\text{OS}}^b + 4\omega_{\text{SS}}^b \omega_{\text{OS}}^b \right) \right] \\ &\pm \frac{1}{4} \varepsilon_b^{\text{SS}} \varepsilon_b^{\text{OS}} \left[ 1 + d_{\text{SS}} \left( 1 - 2\omega_{\text{SS}}^{\bar{b}} \right) + d_{\text{OS}} \left( 1 - 2\omega_{\text{OS}}^{\bar{b}} \right) \right. \\ &\quad \left. + d_{\text{OS}} d_{\text{SS}} \left( 1 - 2\omega_{\text{SS}}^{\bar{b}} - 2\omega_{\text{OS}}^{\bar{b}} + 4\omega_{\text{SS}}^{\bar{b}} \omega_{\text{OS}}^{\bar{b}} \right) \right]. \end{aligned} \quad (21)$$

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

$$C_{\sin, \cos}^{\text{eff}} = \begin{cases} (1 + A_D) C_{\sin, \cos}^{\text{phys}} (\Delta^- - A_P \Delta^+) & \text{for } f = D^- \pi^+ \\ -(1 - A_D) C_{\sin, \cos}^{\text{phys}} (\Delta^- - A_P \Delta^+) & \text{for } f = D^+ \pi^- \end{cases} \quad (22)$$

$$C_{\sinh, \cosh}^{\text{eff}} = \begin{cases} (1 + A_D) C_{\sinh}^{\text{phys}} (\Delta^+ - A_P \Delta^-) & \text{for } f = D^- \pi^+ \\ (1 - A_D) C_{\sinh}^{\text{phys}} (\Delta^+ - A_P \Delta^-) & \text{for } f = D^+ \pi^- \end{cases} \quad (23)$$

<sup>1559</sup> H Correlation matrix of decay-time fit of  $B^0 \rightarrow D^\mp \pi^\pm$ 

 Figure 44 – Correlation matrix from the decay-time fit of  $B^0 \rightarrow D^\mp \pi^\pm$ .

---

 1560 I Decay-time fits of  $B^0 \rightarrow D^\mp \pi^\pm$  data subsamples

Table 12 – Results of the decay-time fits to the 2011, 2012, magnet-up and down subsamples.

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

## I. Decay-time fits of $B^0 \rightarrow D^\mp \pi^\pm$ data subsamples

---

Table 13 – Results of the decay-time fits when only OS tagged, SS tagged or OS and SS candidates are considered exclusively.

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

## Appendix . Appendices

---

Table 14 – Results of the decay-time fits in bins of  $B^0$  transverse momentum ( GeV/c).

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

## I. Decay-time fits of $B^0 \rightarrow D^\mp \pi^\pm$ data subsamples

Table 15 – Results of the decay-time fits in bins of number of primary vertices (PV).

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

## Appendix . Appendices

---

Table 16 – Results of the decay-time fits in bins of the logarithm of the number of tracks  $\ln N_{\text{tr}}$ .

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

## I. Decay-time fits of $B^0 \rightarrow D^\mp \pi^\pm$ data subsamples

---

Table 17 – 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	$\Delta\eta \in (-\infty, -0.36)$	$\Delta\eta \in [-0.36, 0.10)$	$\Delta\eta \in [0.10, 0.53)$	$\Delta\eta \in [0.53, +\infty)$
$S_f$	$0.047 \pm 0.038$	$0.050 \pm 0.039$	$0.103 \pm 0.038$	$0.034 \pm 0.039$
$S_{\bar{f}}$	$0.028 \pm 0.038$	$0.060 \pm 0.039$	$0.006 \pm 0.038$	$0.054 \pm 0.039$
$A_P$	$-0.0092 \pm 0.0053$	$-0.0121 \pm 0.0054$	$0.0004 \pm 0.0053$	$-0.0051 \pm 0.0053$
$A_D$	$0.0072 \pm 0.0036$	$0.0043 \pm 0.0037$	$0.0118 \pm 0.0036$	$0.0106 \pm 0.0035$
$\Gamma$	$0.6587 \pm 0.0017$	$0.6587 \pm 0.0017$	$0.6588 \pm 0.0017$	$0.6588 \pm 0.0017$
$\Delta m$	$0.5052 \pm 0.0023$	$0.5047 \pm 0.0023$	$0.5046 \pm 0.0023$	$0.5059 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.127 \pm 0.039$	$-0.145 \pm 0.039$	$-0.157 \pm 0.039$	$-0.188 \pm 0.041$
$p_1^{\text{OS}}$	$-0.032 \pm 0.049$	$-0.033 \pm 0.044$	$-0.044 \pm 0.045$	$-0.010 \pm 0.045$
$p_2^{\text{OS}}$	$-0.037 \pm 0.023$	$0.011 \pm 0.017$	$0.003 \pm 0.018$	$-0.016 \pm 0.015$
$p_3^{\text{OS}}$	$-0.01 \pm 0.32$	$-0.12 \pm 0.22$	$0.01 \pm 0.23$	$-0.95 \pm 0.20$
$p_4^{\text{OS}}$	$-0.7 \pm 2.0$	$0.16 \pm 0.92$	$0.66 \pm 0.96$	$-2.93 \pm 0.91$
$\Delta p_0^{\text{OS}}$	$-0.105 \pm 0.092$	$-0.082 \pm 0.093$	$0.001 \pm 0.089$	$-0.140 \pm 0.093$
$\Delta p_1^{\text{OS}}$	$0.185 \pm 0.074$	$0.276 \pm 0.069$	$0.029 \pm 0.067$	$0.082 \pm 0.067$
$\Delta p_2^{\text{OS}}$	$-0.021 \pm 0.037$	$-0.054 \pm 0.026$	$0.007 \pm 0.025$	$-0.032 \pm 0.022$
$\Delta p_3^{\text{OS}}$	$0.16 \pm 0.79$	$-0.31 \pm 0.30$	$-0.10 \pm 0.31$	$-0.56 \pm 0.28$
$\Delta p_4^{\text{OS}}$	$-2.5 \pm 2.9$	$-0.1 \pm 1.3$	$-0.4 \pm 1.3$	$-1.2 \pm 1.3$
$p_0^{\text{SS}}$	$-0.061 \pm 0.040$	$-0.009 \pm 0.038$	$-0.029 \pm 0.040$	$-0.071 \pm 0.041$
$p_1^{\text{SS}}$	$0.018 \pm 0.044$	$-0.060 \pm 0.039$	$0.055 \pm 0.043$	$-0.054 \pm 0.044$
$\Delta p_0^{\text{SS}}$	$-0.143 \pm 0.083$	$-0.121 \pm 0.082$	$0.024 \pm 0.081$	$-0.098 \pm 0.084$
$\Delta p_1^{\text{SS}}$	$0.157 \pm 0.066$	$0.033 \pm 0.058$	$-0.038 \pm 0.063$	$0.017 \pm 0.063$
$\varepsilon_{tag}^{\text{OS}}$	$0.4370 \pm 0.0015$	$0.4344 \pm 0.0015$	$0.4346 \pm 0.0015$	$0.4236 \pm 0.0014$
$\varepsilon_{tag}^{\text{SS}}$	$0.93563 \pm 0.00072$	$0.92611 \pm 0.00077$	$0.92349 \pm 0.00078$	$0.93669 \pm 0.00071$
$v_1$	$0.2796 \pm 0.0082$	$0.377 \pm 0.011$	$0.396 \pm 0.012$	$0.2429 \pm 0.0072$
$v_2$	$0.439 \pm 0.015$	$0.565 \pm 0.019$	$0.591 \pm 0.020$	$0.404 \pm 0.013$
$v_3$	$0.768 \pm 0.025$	$0.872 \pm 0.029$	$0.879 \pm 0.030$	$0.679 \pm 0.022$
$v_4$	$0.882 \pm 0.028$	$1.061 \pm 0.034$	$1.105 \pm 0.035$	$0.952 \pm 0.029$
$v_5$	$1.078 \pm 0.033$	$1.152 \pm 0.037$	$1.180 \pm 0.038$	$0.982 \pm 0.031$
$v_6$	$1.009 \pm 0.031$	$1.181 \pm 0.037$	$1.202 \pm 0.038$	$1.095 \pm 0.033$
$v_7$	$1.112 \pm 0.034$	$1.171 \pm 0.037$	$1.208 \pm 0.038$	$1.083 \pm 0.033$
$v_8$	$1.095 \pm 0.029$	$1.222 \pm 0.034$	$1.200 \pm 0.034$	$1.191 \pm 0.031$
$v_9$	$1.130 \pm 0.046$	$1.128 \pm 0.048$	$1.312 \pm 0.052$	$1.065 \pm 0.045$

## Appendix . Appendices

---

Table 18 – Results of the decay-time fits performed for candidates TOS on L0HadronTOS and all the other candidates.

Parameter	L0HadronTOS	Other candidates
$S_f$	$0.029 \pm 0.026$	$0.111 \pm 0.034$
$S_{\bar{f}}$	$0.029 \pm 0.026$	$0.048 \pm 0.034$
$A_P$	$-0.0092 \pm 0.0035$	$-0.0002 \pm 0.0049$
$A_D$	$0.0079 \pm 0.0024$	$0.0093 \pm 0.0032$
$\Gamma$	$0.6588 \pm 0.0017$	$0.6587 \pm 0.0017$
$\Delta m$	$0.5053 \pm 0.0023$	$0.5049 \pm 0.0023$
$p_0^{\text{OS}}$	$-0.196 \pm 0.027$	$-0.083 \pm 0.034$
$p_1^{\text{OS}}$	$-0.066 \pm 0.030$	$-0.005 \pm 0.039$
$p_2^{\text{OS}}$	$-0.006 \pm 0.012$	$-0.010 \pm 0.015$
$p_3^{\text{OS}}$	$-0.20 \pm 0.16$	$-0.46 \pm 0.18$
$p_4^{\text{OS}}$	$-0.66 \pm 0.79$	$-0.70 \pm 0.80$
$\Delta p_0^{\text{OS}}$	$-0.145 \pm 0.060$	$0.044 \pm 0.084$
$\Delta p_1^{\text{OS}}$	$0.122 \pm 0.046$	$0.151 \pm 0.059$
$\Delta p_2^{\text{OS}}$	$-0.032 \pm 0.017$	$-0.012 \pm 0.020$
$\Delta p_3^{\text{OS}}$	$-0.34 \pm 0.25$	$0.04 \pm 0.25$
$\Delta p_4^{\text{OS}}$	$-1.8 \pm 1.3$	$0.7 \pm 1.1$
$p_0^{\text{SS}}$	$0.013 \pm 0.024$	$-0.229 \pm 0.046$
$p_1^{\text{SS}}$	$-0.024 \pm 0.025$	$-0.029 \pm 0.050$
$\Delta p_0^{\text{SS}}$	$-0.108 \pm 0.053$	$-0.061 \pm 0.083$
$\Delta p_1^{\text{SS}}$	$0.012 \pm 0.038$	$0.100 \pm 0.070$
$\varepsilon_{tag}^{\text{OS}}$	$0.40498 \pm 0.00094$	$0.4834 \pm 0.0014$
$\varepsilon_{tag}^{\text{SS}}$	$0.93331 \pm 0.00048$	$0.92502 \pm 0.00071$
$v_1$	$0.4111 \pm 0.0092$	$0.1889 \pm 0.0052$
$v_2$	$0.608 \pm 0.015$	$0.331 \pm 0.010$
$v_3$	$0.909 \pm 0.022$	$0.633 \pm 0.019$
$v_4$	$1.087 \pm 0.025$	$0.863 \pm 0.024$
$v_5$	$1.174 \pm 0.027$	$0.981 \pm 0.028$
$v_6$	$1.188 \pm 0.027$	$1.018 \pm 0.028$
$v_7$	$1.200 \pm 0.027$	$1.051 \pm 0.029$
$v_8$	$1.234 \pm 0.024$	$1.095 \pm 0.026$
$v_9$	$1.218 \pm 0.034$	$1.063 \pm 0.039$

## 1561 J Correlation between systematic uncertainties

1562 As reported in Table 5.5, the main systematic uncertainties on  $S_f$  and  $S_{\bar{f}}$  are the ones due  
1563 to the constraint on  $\Delta m$ , the fit biases, and the background subtraction. The correlations  
1564 between the systematics uncertainties on  $S_f$  and  $S_{\bar{f}}$  due to these sources are described in  
1565 details in this appendix. The correlation of the total systematic resulting from these three  
1566 contributions is  $-0.41$ . The correlation between other sources of systematics is neglected.

### 1567 J.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. 5.4.1. This correlation  $\rho^{\Delta m}(S_f, S_{\bar{f}})$  is simply computed from the difference in covariance between the two fit results:

$$\rho^{\Delta m}(S_f, S_{\bar{f}}) = \frac{\sigma_{S_f}^{\text{nominal}} \sigma_{S_{\bar{f}}}^{\text{nominal}} \rho^{\text{nominal}}(S_f, S_{\bar{f}}) - \sigma_{S_f}^{\Delta m \text{ fixed}} \sigma_{S_{\bar{f}}}^{\Delta m \text{ fixed}} \rho^{\Delta m \text{ fixed}}(S_f, S_{\bar{f}})}{\sqrt{(\sigma_{S_f}^{\text{nominal}})^2 - (\sigma_{S_f}^{\Delta m \text{ fixed}})^2} \sqrt{(\sigma_{S_{\bar{f}}}^{\text{nominal}})^2 - (\sigma_{S_{\bar{f}}}^{\Delta m \text{ fixed}})^2}} \quad (24)$$

1568 The value of  $\rho^{\Delta m}(S_f, S_{\bar{f}})$  so obtained is  $-1$ , meaning that  $S_f$  and  $S_{\bar{f}}$  are fully anticorrelated  
1569 because of the uncertainties on  $\Delta m$ .

### 1570 J.2 Correlation of systematics due to fit biases

1571 As described in Sec. 5.3.1, a bias on  $S_f$  and  $S_{\bar{f}}$  is observed from bootstrapped MC samples,  
1572 and the size of this bias is assigned as systematic uncertainty. The associated correlation is  
1573 estimated by computing the correlation between  $(S_f^{\text{fit}} - S_f^{\text{gen}})$  and  $(S_{\bar{f}}^{\text{fit}} - S_{\bar{f}}^{\text{gen}})$  obtained on  
1574 the same set of bootstrapped MC samples, as shown in Fig. 45. The resulting correlation  
1575 is  $0.4$ .

### 1576 J.3 Correlation of systematics due to background subtraction

1577 A systematic due to *sWeighting* and background subtraction is assigned by repeating Fit  
1578 B in a wider mass range, as described in Sec. 5.4.3. In order to estimate the correlation  
1579 of this systematic uncertainty between  $S_f$  and  $S_{\bar{f}}$ , the data sample is bootstrapped in a  
1580 similar way as done for the Monte Carlo (Sec. 5.3.1). Then, *sWeights* are obtained twice  
1581 on each sample, once with the nominal strategy, and once by selecting a wide mass range  
1582 in Fit B. Finally, the time fit is performed on each sample using both the sets of *sWeights*  
1583 computed in the previous step. The correlation is estimated from the two-dimensional  
1584 distribution of  $(S_f^{\text{nominal}} - S_f^{\text{wide mass}})$  versus  $(S_f^{\text{nominal}} - S_f^{\text{wide mass}})$ , which is shown in  
1585 Fig. 46. The resulting correlation is  $0.7$ .

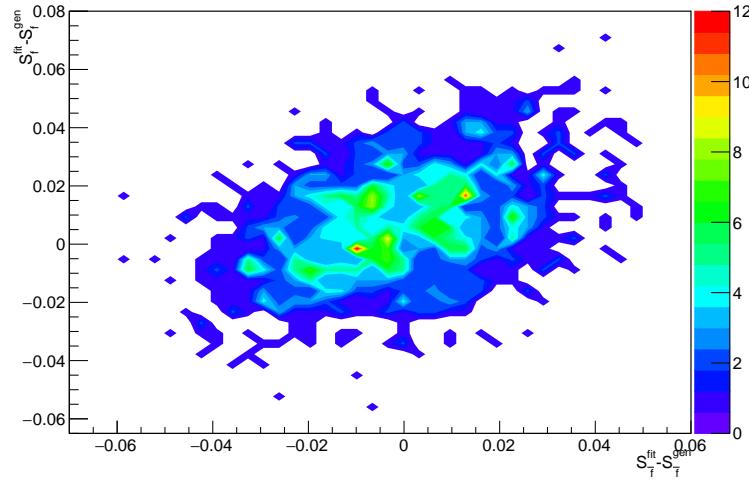


Figure 45 – Two-dimensional distribution of  $(S_f^{\text{fit}} - S_f^{\text{gen}})$  versus  $(S_{\bar{f}}^{\text{fit}} - S_{\bar{f}}^{\text{gen}})$  obtained from the fits to bootstrapped MC samples (Sec. 5.3.1).

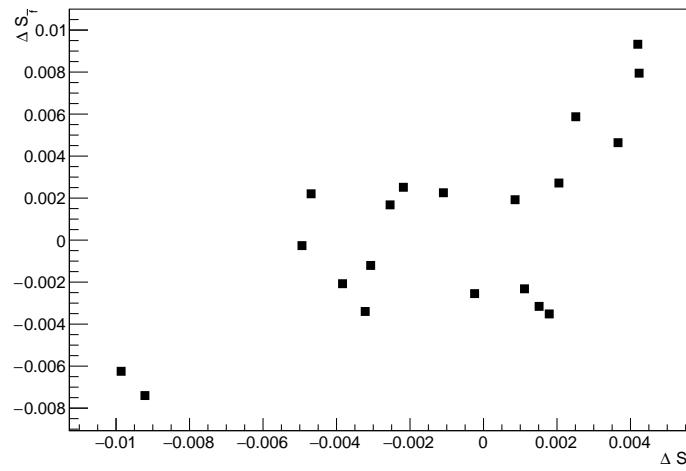


Figure 46 – Two-dimensional distribution of  $S_{\bar{f}}^{\text{nominal}} - S_{\bar{f}}^{\text{wide mass}}$  ( $\Delta S_{\bar{f}}$ ) versus  $S_f^{\text{nominal}} - S_f^{\text{wide mass}}$  ( $\Delta S_f$ ) obtained from fits to bootstrapped data samples, where Fit B is performed with the nominal strategy or with a wider mass range for each sample.

## K. Decay-time fit validation with bootstrapping

---

### 1586 K Decay-time fit validation with bootstrapping

1587 In this appendix, further results of the Monte Carlo time fit validation are shown in  
 1588 addition to Fig. 5.14. The results for production, detection asymmetry, and acceptance  
 1589 are not shown because no independent reference value for these parameters exist in Monte  
 1590 Carlo.

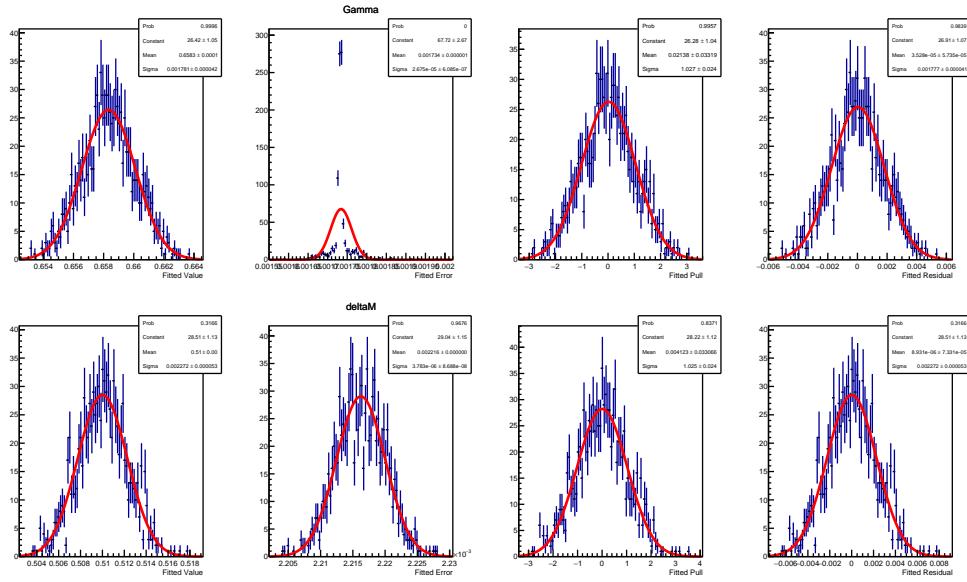


Figure 47 – Distributions of the fitted value, error, pull and residual for  $\Gamma$  (top) and  $\Delta m$  (bottom). Each distribution is fitted with a Gaussian PDF. Pulls and residuals are computed by taking the Monte Carlo generation value as reference.

## Appendix . Appendices

---

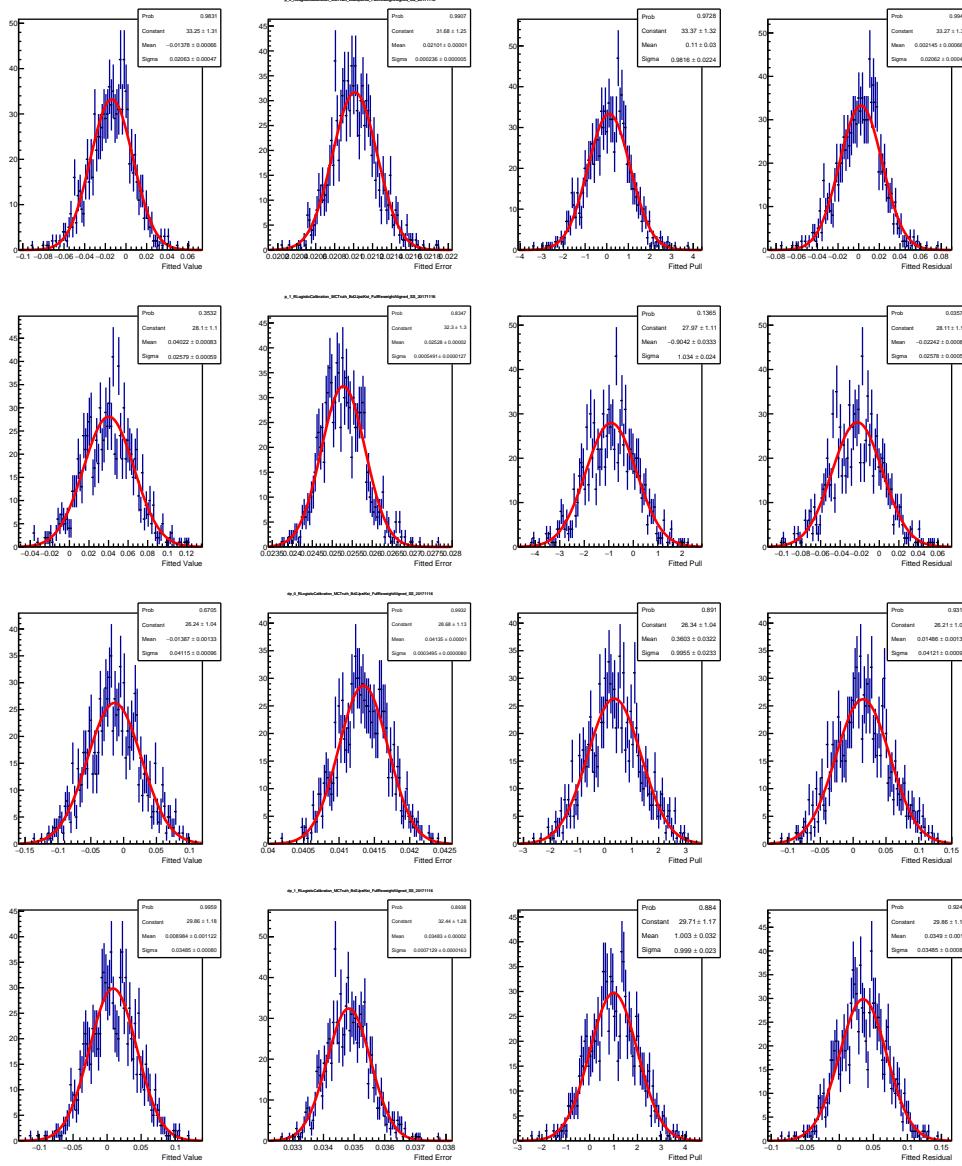


Figure 48 – Distributions of the fitted value, error, pull and residual for the SS tagger calibration parameters  $p_0^{\text{SS}}$ ,  $p_1^{\text{SS}}$ ,  $\Delta p_0^{\text{SS}}$ , and  $\Delta p_1^{\text{SS}}$  (from top to bottom). Each distribution is fitted with a Gaussian PDF. Pulls and residuals are computed by taking the values found on the  $B^0 \rightarrow J/\psi K^{*0}$  Monte Carlo calibration as reference.

## K. Decay-time fit validation with bootstrapping

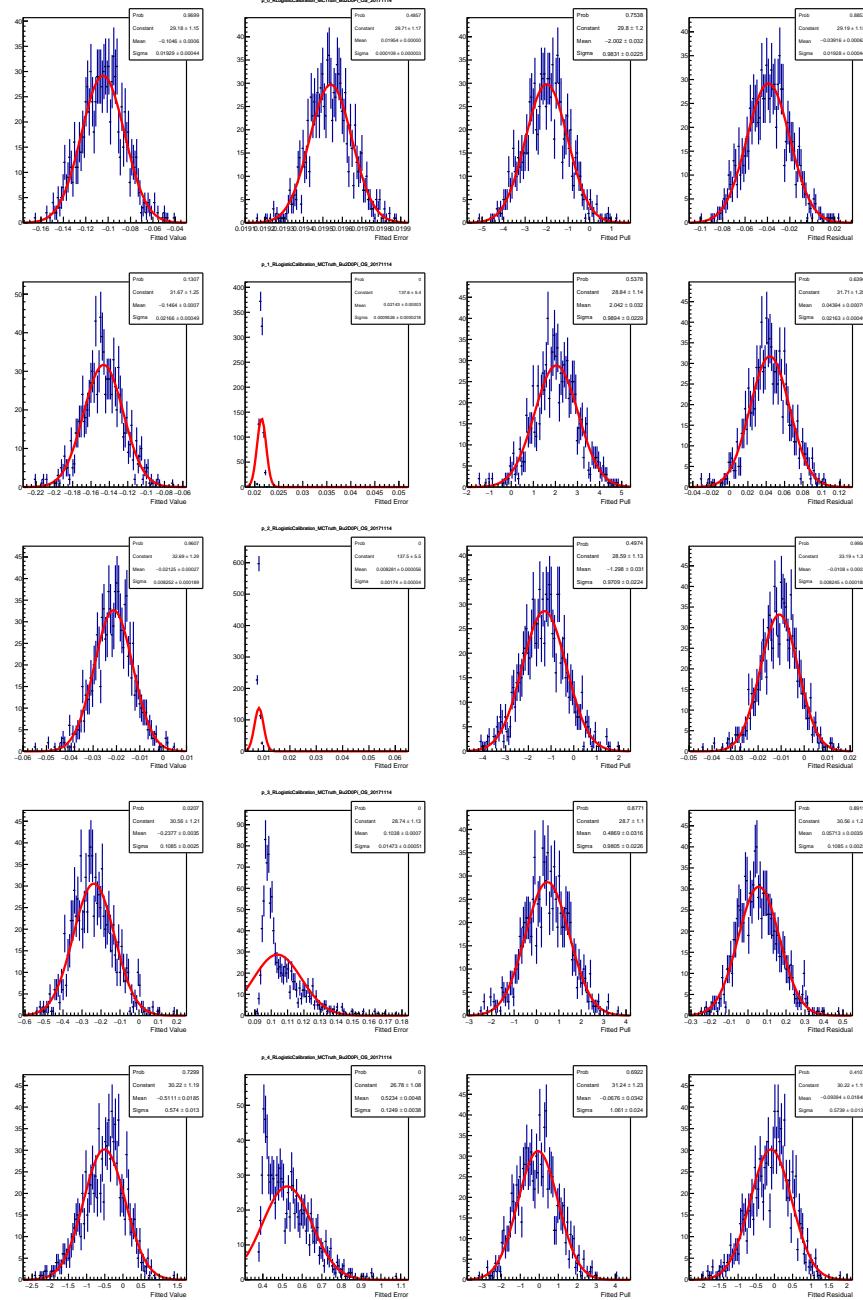


Figure 49 – Distributions of the fitted value, error, pull and residual for the OS tagger calibration parameters  $p_0^{\text{OS}}$ ,  $p_1^{\text{OS}}$ ,  $p_2^{\text{OS}}$ ,  $p_3^{\text{OS}}$ , and  $p_4^{\text{OS}}$  (from top to bottom). Each distribution is fitted with a Gaussian PDF. Pulls and residuals are computed by taking the values found on the  $B^+ \rightarrow D^0\pi^+$  Monte Carlo calibration as reference.

## Appendix . Appendices

---

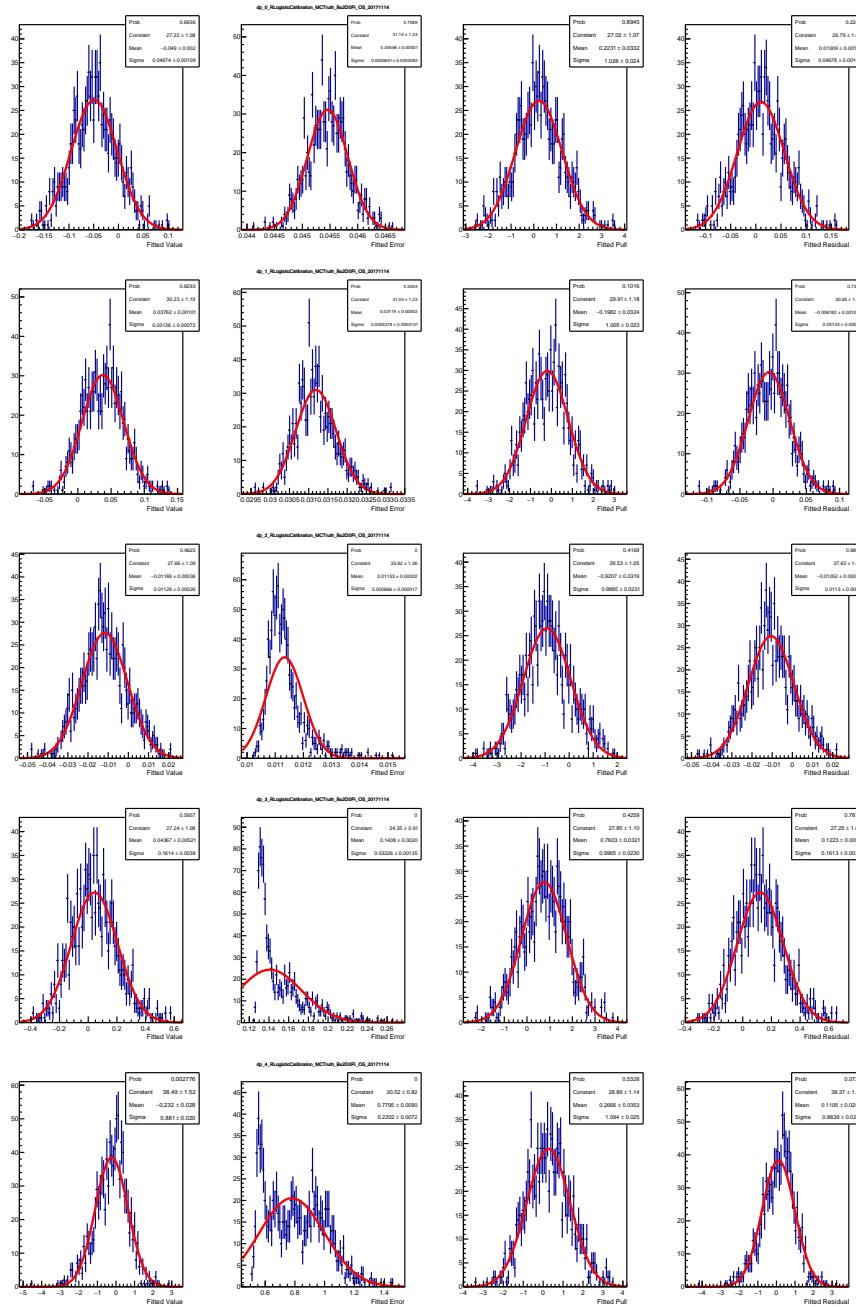


Figure 50 – Distributions of the fitted value, error, pull and residual for the OS tagger calibration parameters  $\Delta p_0^{\text{OS}}$ ,  $\Delta p_1^{\text{OS}}$ ,  $\Delta p_2^{\text{OS}}$ ,  $\Delta p_3^{\text{OS}}$ , and  $\Delta p_4^{\text{OS}}$  (from top to bottom). Each distribution is fitted with a Gaussian PDF. Pulls and residuals are computed by taking the values found on the  $B^+ \rightarrow D^0\pi^+$  Monte Carlo calibration as reference.

<sub>1591</sub> **L Test of the decay-time fit via a toy tagger**

The *toy tagger* used to perform the test mentioned in Sec. 5.3.1 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  $d_i$  is flipped (and thus made wrong) according to the following criterion:

$$d_i \rightarrow \begin{cases} -d_i & \text{if } r_i \leq \eta_i \\ d_i & \text{otherwise} \end{cases}. \quad (25)$$

<sub>1592</sub> During the time fit, the mistag calibration is simply taken as a linear function (Eq. 3.9)  
<sub>1593</sub> with  $p_0 = \langle \eta \rangle = 0.370029$  (taken from the adopted template) and  $p_1 = 1$ , which means  
<sub>1594</sub>  $\omega = \eta$  for all candidates. In fact, the per-event mistag  $\eta$  is the true mistag  $\omega$  probability by  
<sub>1595</sub> construction. In this way, it is possible to test the time fit with a per-event mistag without  
<sub>1596</sub> relying on any approximation or uncertainty coming from the calibration procedure.  
<sub>1597</sub> Moreover, the tagging efficiency is 100% by construction.

<sub>1598</sub> The distributions of the fitted value, error, pull and residual for the relevant parameters are  
<sub>1599</sub> shown in Fig. 51. Each distribution is fitted with a Gaussian PDF. All pull distributions  
<sub>1600</sub> have means compatible with 0 and widths compatible with 1, meaning that the maximum  
<sub>1601</sub> likelihood estimation of the parameters is unbiased and returns correct uncertainties.

## Appendix . Appendices

---

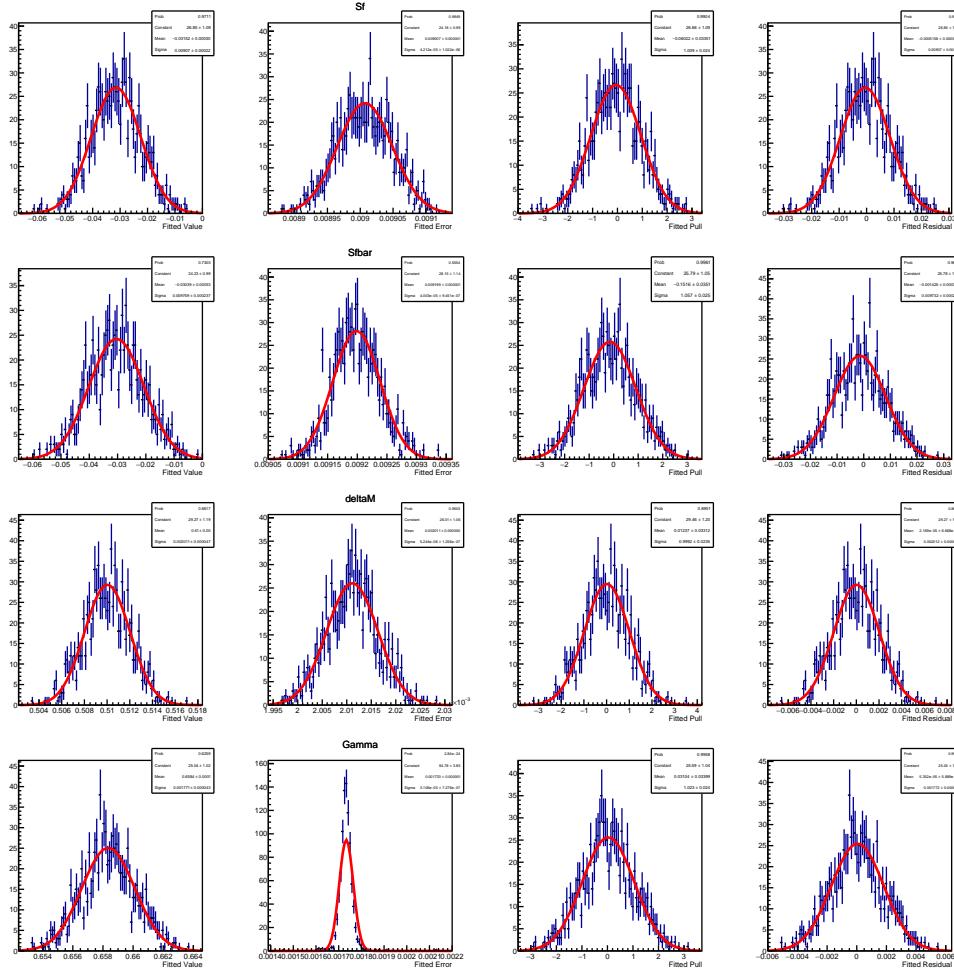


Figure 51 – Distributions of the fitted value, error, pull and residual for the main parameters ( $S_f$ ,  $S_{\bar{f}}$ ,  $\Delta m$ , and  $\Gamma$ , from top to bottom) fitted on bootstrapped Monte Carlo samples with a toy tagger. Each distribution is fitted with a Gaussian PDF. Pulls and residuals are computed by taking the Monte Carlo generation value as reference.

<sup>1602</sup> M Inputs for  $B^0 \rightarrow D^\pm \pi^\mp$  simulation

 Table 19 – 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

- 1604 [1] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367.
- 1605 [2] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.
- 1606 [3] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.
- 1607 [4] <http://www-f9.ijs.si/%7Elubelj/SM.pdf>.
- 1608 [5] E. Fermi, *Sulla quantizzazione del gas perfetto monoatomico*, Rendiconti Lincei **3** (1926) 145.
- 1610 [6] P. A. M. Dirac, *On the theory of quantum mechanics*, Proceedings of the Royal Society A **112** (**762**) (1926) 661.
- 1611 [7] S. Bose, *Plancks Gesetz und Lichtquantenhypothese*, Z. Phys. **26** (1924) 178.
- 1613 [8] LHCb collaboration, R. Aaij *et al.*, *Observation of the resonant character of the*  
1614 *Z(4430)<sup>-</sup> state*, Phys. Rev. Lett. **112** (2014) 222002, [arXiv:1404.1903](https://arxiv.org/abs/1404.1903).
- 1615 [9] LHCb collaboration, R. Aaij *et al.*, *Observation of J/ψp resonances consistent with*  
1616 *pentaquark states in Λ\_b^0 → J/ψpK^- decays*, Phys. Rev. Lett. **115** (2015) 072001,  
1617 [arXiv:1507.03414](https://arxiv.org/abs/1507.03414).
- 1618 [10] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,  
1619 Phys. Rev. Lett. **13** (1964) 321.
- 1620 [11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett.  
1621 **13** (1964) 508.
- 1622 [12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and*  
1623 *Massless Particles*, Phys. Rev. Lett. **13** (1964) 585.
- 1624 [13] The ATLAS collaboration, *Observation of a new particle in the search for the*  
1625 *Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716**  
1626 (2012) 1, [arXiv:1207.7214](https://arxiv.org/abs/1207.7214).
- 1627 [14] The CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the*  
1628 *CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, [arXiv:1207.7235](https://arxiv.org/abs/1207.7235).

## Bibliography

---

- 1629 [15] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **178** (1963)  
1630 531.
- 1631 [16] M. Kobayashi and T. Maskawa, *CP violation in the renormalizable theory of weak  
1632 interaction*, Progr. Theor. Phys. **49** (1973) 652.
- 1633 [17] L. L. Chau and W. Y. Keung, *Comments on the parametrization of the Kobayashi-  
1634 Maskawa matrix*, Phys. Rev. Lett. **53** (1984) 1802.
- 1635 [18] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa matrix*, Progr. Theor.  
1636 Phys. **49** (1973) 652.
- 1637 [19] CKMfitter group, J. Charles *et al.*, *Current status of the Standard  
1638 Model CKM fit and constraints on  $\Delta F = 2$  new physics*, Phys. Rev. **D91**  
1639 (2015) 073007, arXiv:1501.05013, updated results and plots available at  
1640 <http://ckmfitter.in2p3.fr/>.
- 1641 [20] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys.  
1642 **C40** (2016) 100001, and 2017 update.
- 1643 [21] I. Bigi and A. I. Sanda, *CP violation*, Cambridge Monographs on Particle Physics,  
1644 Nuclear Physics and Cosmology, 2009.
- 1645 [22] R. Fleischer, *New strategies to obtain insights into CP violation through  
1646  $B_{(s)} \rightarrow D_{(s)}^\pm K^\mp$ ,  $D_{(s)}^{*\pm} K^\mp$ , ... and  $B_{(d)} \rightarrow D^\pm \pi^\mp$ ,  $D^{*\pm} \pi^\mp$ , ... decays*, Nucl. Phys.  
1647 **B671** (2003) 459, arXiv:hep-ph/0304027.
- 1648 [23] R. Aleksan, I. Dunietz, and B. Kayser, *Determining the CP violating phase  $\gamma$* , Z.  
1649 Phys. **C54** (1992) 653.
- 1650 [24] I. Dunietz and R. G. Sachs, *Asymmetry between inclusive charmed and anticharmed  
1651 modes in  $B^0$ ,  $\bar{B}^0$  decay as a measure of CP violation*, Phys. Rev. D **37** (1988) 3186,  
1652 Erratum ibid. **D39** (1989) 3515.
- 1653 [25] LHCb collaboration, R. Aaij *et al.*, *Measurement of the time-dependent CP asymme-  
1654 tries in  $B_s^0 \rightarrow J/\psi K_s^0$* , JHEP **06** (2015) 131, arXiv:1503.07055.
- 1655 [26] BaBar collaboration, B. Aubert *et al.*, *Measurement of the branching fractions of  
1656 the rare decays  $B^0 \rightarrow D_s^{(*)+} \pi^-$ ,  $B^0 \rightarrow D_s^{(*)+} \rho^-$ , and  $B^0 \rightarrow D_s^{(*)-} K^{(*)+}$* , Phys. Rev.  
1657 **D78** (2008) 032005, arXiv:0803.4296.
- 1658 [27] Belle collaboration, A. Das *et al.*, *Measurements of branching fractions for  
1659  $B^0 \rightarrow D_s^+ \pi^-$  and  $\bar{B}^0 \rightarrow D_s^+ K^-$* , Phys. Rev. **D82** (2010) 051103, arXiv:1007.4619.
- 1660 [28] BaBar collaboration, B. Aubert *et al.*, *Measurement of time-dependent CP asymme-  
1661 tries in  $B^0 \rightarrow D^{(*)\pm} \pi^\mp$  and  $B^0 \rightarrow D^\pm \rho^\mp$  decays*, Phys. Rev. D **73** (2006) 111101,  
1662 arXiv:hep-ex/0602049.

- 1663 [29] BaBar collaboration, B. Aubert *et al.*, *Measurement of time-dependent CP-violating  
1664 asymmetries and constraints on  $\sin(2\beta + \gamma)$  with partial reconstruction of  $B \rightarrow$   
1665  $D^{*\mp}\pi^\pm$  decays*, Phys. Rev. **D71** (2005) 112003, [arXiv:hep-ex/0504035](https://arxiv.org/abs/hep-ex/0504035).
- 1666 [30] Belle collaboration, S. Bahinipati *et al.*, *Measurements of time-dependent CP asymmetries in  $B \rightarrow D^{*\mp}\pi^\pm$  decays using a partial reconstruction technique*, Phys. Rev.  
1667 **D84** (2011) 021101, [arXiv:1102.0888](https://arxiv.org/abs/1102.0888).
- 1669 [31] Belle, F. J. Ronga *et al.*, *Measurement of CP violation in  $B^0 \rightarrow D^{*-}\pi^+$  and  
1670  $B^0 \rightarrow D^-\pi^+$  decays*, Phys. Rev. **D73** (2006) 092003, [arXiv:hep-ex/0604013](https://arxiv.org/abs/hep-ex/0604013).
- 1671 [32] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP asymmetry in  $B_s^0 \rightarrow D_s^\mp K^\pm$   
1672 decays*, [arXiv:1712.07428](https://arxiv.org/abs/1712.07428), submitted to JHEP.
- 1673 [33] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007,  
1674 [arXiv:1405.7808](https://arxiv.org/abs/1405.7808).
- 1675 [34] LHCb collaboration, R. Antunes Nobrega *et al.*, *LHCb reoptimized detector design and  
1676 performance: Technical Design Report*, CERN-LHCC-2003-030. LHCb-TDR-009.
- 1677 [35] LHCb collaboration, A. Franca Barbosa *et al.*, *LHCb inner tracker: Technical Design  
1678 Report*, CERN-LHCC-2002-029. LHCb-TDR-008.
- 1679 [36] LHCb collaboration, P. R. Barbosa Marinho *et al.*, *LHCb outer tracker: Technical  
1680 Design Report*, CERN-LHCC-2001-024. LHCb-TDR-006.
- 1681 [37] V. Battista, F. Blanc, M. Martinelli, and M. Tobin, *A study of spillover clusters and  
1682 ghost tracks in the Silicon Tracker with 25 ns bunch spacing*, CERN-LHCb-INT-  
1683 2016-010.
- 1684 [38] A. Powell *et al.*, *Particle identification at LHCb*, Proceedings of ICHEP2010 **020**  
1685 (2010), LHCb-PROC-2011-008.
- 1686 [39] LHCb collaboration, S. Amato *et al.*, *LHCb calorimeters: Technical Design Report*,  
1687 CERN-LHCC-2000-036.
- 1688 [40] A. A. Alves Jr *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013)  
1689 P02022.
- 1690 [41] LHCb collaboration, R. Antunes Nobrega *et al.*, *LHCb trigger system: Technical  
1691 Design Report*, CERN-LHCC-2003-031.
- 1692 [42] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*,  
1693 Comput. Phys. Commun. **178** (2008) 852, [arXiv:0710.3820](https://arxiv.org/abs/0710.3820).
- 1694 [43] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth.  
1695 **A462** (2001) 152.

## Bibliography

---

- 1696 [44] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections*  
1697 *in  $Z$  and  $W$  decays*, Eur. Phys. J. **C45** (2006) 97, [arXiv:hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026).
- 1698 [45] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl.  
1699 Instrum. Meth. **A506** (2003) 250.
- 1700 [46] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE  
1701 Trans. Nucl. Sci. **53** (2006) 270.
- 1702 [47] D. Fazzini, *Development of "same side" flavour tagging algorithms for measurements*  
1703 *of flavour oscillations and CP violation in the  $B^0$  mesons system*, CERN-THESIS-  
1704 2015-040.
- 1705 [48] LHCb collaboration, R. Aaij *et al.*, *New algorithms to tag the flavour of  $B^0$  mesons*  
1706 *using pions and protons*, Eur. Phys. J. (2016), no. C77 [arXiv:1610.06019](https://arxiv.org/abs/1610.06019).
- 1707 [49] LHCb collaboration, R. Aaij *et al.*, *Opposite-side flavour tagging of  $B$  mesons at the*  
1708 *LHCb experiment*, Eur. Phys. J. **C72** (2012) 2022, [arXiv:1202.4979](https://arxiv.org/abs/1202.4979).
- 1709 [50] LHCb collaboration, R. Aaij *et al.*,  *$B$  flavour tagging using charm decays at the*  
1710 *LHCb experiment*, JINST **10** (2015) P10005, [arXiv:1507.07892](https://arxiv.org/abs/1507.07892).
- 1711 [51] M. Grabalosa, *Flavour Tagging developments within the LHCb experiment*, PhD  
1712 thesis, Barcelona U., 2012-03-28.
- 1713 [52] LHCb collaboration, *Optimization and calibration of the LHCb flavour tagging*  
1714 *performance using 2010 data*, LHCb-CONF-2011-003.
- 1715 [53] V. Battista,  *$b$ -flavour tagging in  $pp$  collisions at LHCb*, Il Nuovo Cimento **39C** (2016)  
1716 335.
- 1717 [54] J. Wimberley, *Calibration flavor tagging algorithms with binomial regression*, LHCb-  
1718 INT-2017-002.
- 1719 [55] *scikit-optimize python package*, <https://github.com/scikit-optimize>.
- 1720 [56] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*,  
1721 [arXiv:1603.02754](https://arxiv.org/abs/1603.02754).
- 1722 [57] L. Breiman, *Bagging predictors*, Machine Learning **24** (1996) 123.
- 1723 [58] LHCb collaboration, *A single track HLT1 trigger*, LHCb-PUB-2011-003.
- 1724 [59] LHCb collaboration, *The LHCb inclusive  $B$  triggers*, LHCb-INT-2011-030.
- 1725 [60] W. D. Hulsbergen *et al.*, *Decay chain fitting with a Kalman filter*, Nucl. Instrum.  
1726 Meth. **A552** (2005) 566.
- 1727 [61] P. Koppenburg, *Statistical biases in measurements with multiple candidates*,  
1728 [arXiv:1703.01128](https://arxiv.org/abs/1703.01128).

- 1729 [62] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and*  
1730 *regression trees*, Wadsworth international group, Belmont, California, USA, 1984.
- 1731 [63] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neu-*  
1732 *ral networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577,  
1733 arXiv:physics/0408124.
- 1734 [64] A. Hoecker *et al.*, *TMVA: Toolkit for multivariate data analysis*, PoS **ACAT** (2007)  
1735 040, arXiv:physics/0703039.
- 1736 [65] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning*  
1737 *and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.
- 1738 [66] M. Pivk and F. R. L. Diberder, *sPlot: a statistical tool to unfold data distributions*,  
1739 Nucl. Instrum. Meth. **A555** (2005) 356.
- 1740 [67] D. M. Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*,  
1741 Nucl. Instrum. Meth. **A764** (2014) 150.
- 1742 [68] N. L. Johnson, *Systems of frequency curves generated by methods of translation*,  
1743 Biometrika **36** (1949) 149.
- 1744 [69] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime*  
1745 *and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986,  
1746 DESY-F31-86-02.
- 1747 [70] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP-violating phase  $\phi_s$  in*  
1748  $\bar{B}_s^0 \rightarrow D_s^+ D_s^-$  *decays*, Phys. Rev. Lett. **113** (2014) 211801, arXiv:1409.4619.
- 1749 [71] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of CP violation in*  
1750  $B_s^0 \rightarrow J/\psi K^+ K^-$  *decays*, Phys. Rev. Lett. **114** (2015) 041801, arXiv:1411.3104.
- 1751 [72] M. Karbach, G. Raven, and M. Schiller, *Decay time integrals in neutral meson mixing*  
1752 *and their efficient evaluation*, arXiv:1407.0748.
- 1753 [73] Y. Xie, *sFit: a method for background subtraction in maximum likelihood fit*,  
1754 arXiv:0905.0724.
- 1755 [74] LHCb collaboration, R. Aaij *et al.*, *A precise measurement of the  $B^0$  meson oscillation*  
1756 *frequency*, Eur. Phys. J. **C76** (2016) 412, arXiv:1604.03475.
- 1757 [75] LHCb collaboration, R. Aaij *et al.*, *Measurement of  $B^0$ ,  $B_s^0$ ,  $B^+$  and  $\Lambda_b^0$  produc-*  
1758 *tion asymmetries in 7 and 8 TeV pp collisions*, Phys. Lett. **B774** (2017) 139,  
1759 arXiv:1703.08464.
- 1760 [76] P. Good, *Permutation, parametric and bootstrap tests of hypotheses*, Springer Series  
1761 in Statistics, 2004.

## Bibliography

---

- 1762 [77] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-*  
1763 *hadron, and  $\tau$ -lepton properties as of summer 2016*, Eur. Phys. J. **C77**  
1764 (2017) 895, arXiv:1612.07233, updated results and plots available at  
1765 <http://www.slac.stanford.edu/xorg/hflav/>.
- 1766 [78] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle  $\gamma$  from a*  
1767 *combination of tree-level LHCb analyses*, LHCb-PAPER-2016-032, in preparation.
- 1768 [79] K. De Bruyn *et al.*, *Exploring  $B_s \rightarrow D_s^{(*)\pm} K^\mp$  decays in the presence of a sizable*  
1769 *width difference  $\Delta\Gamma_s$* , Nucl. Phys. **B868** (2013) 351, arXiv:1208.6463.
- 1770 [80] A. Rogozhnikov, *hep\_ml*, [https://arogozhnikov.github.io/hep\\_ml/](https://arogozhnikov.github.io/hep_ml/).