Measurement of $CP$ violation in $B^0 \to D^{*\pm} D^{\mp}$ decays

The LHCb collaboration

E-mail: vecchi@fe.infn.it

ABSTRACT: The decay-time-dependent $CP$ asymmetry in $B^0 \to D^{*\pm} D^{\mp}$ decays is measured using a data set corresponding to an integrated luminosity of 9 fb$^{-1}$ recorded by the LHCb detector in proton-proton collisions at centre-of-mass energies of 7, 8 and 13 TeV. The $CP$ parameters are measured as

\[ S_{D^*D} = -0.861 \pm 0.077 \text{ (stat)} \pm 0.019 \text{ (syst)}, \]
\[ \Delta S_{D^*D} = 0.019 \pm 0.075 \text{ (stat)} \pm 0.012 \text{ (syst)}, \]
\[ C_{D^*D} = -0.059 \pm 0.092 \text{ (stat)} \pm 0.020 \text{ (syst)}, \]
\[ \Delta C_{D^*D} = -0.031 \pm 0.092 \text{ (stat)} \pm 0.016 \text{ (syst)}, \]
\[ A_{D^*D} = 0.008 \pm 0.014 \text{ (stat)} \pm 0.006 \text{ (syst)}. \]

The analysis provides the most precise single measurement of $CP$ violation in this decay channel to date. All parameters are consistent with their current world average values.

KEYWORDS: B physics, CP violation, Electroweak interaction, Flavor physics, Hadron-Hadron scattering (experiments)

ArXiv ePrint: 1912.03723
1 Introduction

In the hadronic sector of the Standard Model (SM), CP violation originates from an irreducible complex phase in the Cabibbo-Kobayashi-Maskawa matrix that describes the mixing of the quark mass eigenstates into weak-interaction eigenstates [1, 2]. Interference caused by a weak-phase difference between the $B^0$-$\bar{B}^0$ oscillation and the decay amplitudes leads to a CP asymmetry in the decay-time distributions of $B^0$ and $\bar{B}^0$ mesons. Decays involving $b \to c\bar{c}s$ tree transitions at leading order, such as $B^0 \to J/\psi K^0_S$, are sensitive to the weak phase $2\beta$, where $\beta \equiv \text{arg}[-(V_{cd}V_{cb}^*)/(V_{td}V_{tb}^*)]$ is one of the angles of the Unitarity Triangle. Measurements of this phase were performed by several experiments using different channels [3]. The same phase appears in $b \to c\bar{c}d$ transitions, which contribute to $B^0 \to D^{*\pm}D^{\mp}$ decays, when the leading-order colour-favoured tree diagram is considered. However, $B^0 \to D^{*\pm}D^{\mp}$ decays can also proceed through several other decay diagrams, that include penguin, $W$-exchange and annihilation topologies, where additional contributions to CP violation both from the SM and new physics (NP) may arise. Tests of the SM have been performed by relating CP asymmetries and branching fractions of different decay modes of neutral and charged beauty mesons to two charm mesons [4, 5].

---

1Charge-conjugated processes are implicitly included in the following, unless specified.
Each of the $D^{∗+}D^{−}$ and $D^{∗−}D^{+}$ final states are accessible from both $B^0$ and $B^0$ mesons. The time-dependent decay rates for the four configurations of initial $B$ flavour and final states can be written as

$$\frac{d\Gamma_{B^0 f}(t)}{dt} = \frac{e^{-t/\tau_d}}{8\tau_d} (1 + A_{ff}) \left[ 1 + S_f \sin(\Delta m_d t) - C_f \cos(\Delta m_d t) \right],$$
$$\frac{d\Gamma_{B^0 \bar{f}}(t)}{dt} = \frac{e^{-t/\tau_d}}{8\tau_d} (1 - A_{\bar{f}f}) \left[ 1 - S_f \sin(\Delta m_d t) + C_f \cos(\Delta m_d t) \right],$$ \hspace{1cm} (1.1)

$$\frac{d\Gamma_{B^0 \bar{f}}(t)}{dt} = \frac{e^{-t/\tau_d}}{8\tau_d} (1 - A_{\bar{f}f}) \left[ 1 + S_f \sin(\Delta m_d t) - C_f \cos(\Delta m_d t) \right],$$
$$\frac{d\Gamma_{B^0 f}(t)}{dt} = \frac{e^{-t/\tau_d}}{8\tau_d} (1 + A_{ff}) \left[ 1 - S_f \sin(\Delta m_d t) + C_f \cos(\Delta m_d t) \right],$$

where $f = D^{∗+}D^{−}$ and $\bar{f} = D^{∗−}D^{+}$. The parameter $A_{ff}$ represents the overall asymmetry in the production of the $f$ and $\bar{f}$ final states and is defined as

$$A_{ff} = \frac{\left| |A_f|^2 + |A_\bar{f}|^2 \right| - \left( |A_f|^2 + |A_\bar{f}|^2 \right)}{\left( |A_f|^2 + |A_\bar{f}|^2 \right)} + \left( |A_f|^2 + |A_\bar{f}|^2 \right),$$ \hspace{1cm} (1.2)

with $A_f$ ($A_\bar{f}$) and $\bar{A}_f$ ($\bar{A}_\bar{f}$) indicating the amplitudes of the decay of a $B^0$ and a $\bar{B}^0$ meson to final state $f$ ($\bar{f}$). Here, $\tau_d$ is the $B^0$ lifetime and $\Delta m_d$ is the mass difference of the two $B^0$ mass eigenstates, which are assumed to have the same decay width $[3]$. Introducing $q$ and $p$ to describe the relation between the mass and flavour eigenstates, $|B_{H LL}| = p|B^0| \pm q|\bar{B}^0|$, the parameters $S_f$ and $C_f$ are defined as

$$S_f = \frac{2\tau_m \lambda_f}{1 + |\lambda_f|^2}, \quad C_f = \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2}, \quad \lambda_f = \frac{q}{p} \bar{A}_f,$$ \hspace{1cm} (1.3)

with analogous definitions holding for $S_{\bar{f}}$, and $C_{\bar{f}}$. By combining these parameters, the CP observables for $B^0 \rightarrow D^{∗±}D^{±}$ decays can be defined as $[3]$

$$S_{D^∗D} = \frac{1}{2} (S_f + S_{\bar{f}}),$$
$$\Delta S_{D^∗D} = \frac{1}{2} (S_f - S_{\bar{f}}),$$
$$C_{D^∗D} = \frac{1}{2} (C_f + C_{\bar{f}}),$$
$$\Delta C_{D^∗D} = \frac{1}{2} (C_f - C_{\bar{f}}),$$
$$A_{D^∗D} = A_{ff}.$$ \hspace{1cm} (1.4)

In absence of CP violation, $S_{D^∗D}$ and $C_{D^∗D}$ vanish. While $\Delta S_{D^∗D}$ is related to the relative strong phase between the decay amplitudes, the parameter $\Delta C_{D^∗D}$ is a measure of how flavour specific the decay mode is. For a flavour-specific decay only one final state is accessible for each flavour of the decaying neutral $B$ meson, $\Delta C_{D^∗D} = \pm 1$ and no CP violation in the interference between decays with and without mixing is possible. Decays with
$\Delta C_{D^*D} = 0$ present the highest sensitivity to mixing-induced CP violation. In the case of $B^0 \to D^{*\pm}D^{\mp}$ decays, if the contribution of higher-order SM processes and NP are negligible, the amplitudes for $B^0 \to D^{*+}D^-$ and $B^0 \to D^{*-}D^+$ have the same hadronic phase and magnitude. As a result, $A_{D^*D}, C_{D^*D}, \Delta C_{D^*D}$ and $\Delta S_{D^*D}$ vanish and $S_{D^*D} = \sin(2\beta)$.

Theoretical models, based on QCD factorization and heavy quark symmetry, estimate the contribution of penguin amplitudes in $B^0 \to D^{*\pm}D^{\mp}$ to be up to a few percent [6, 7].

By combining eqs. (1.1) and (1.4) the decay rate can be rewritten as

$$\frac{d\Gamma(t)}{dt} = \frac{e^{-t/\tau_d}}{8\tau_d}(1 + rA_{D^*D})$$

$$\times \left[ 1 - d(S_{D^*D} + r\Delta S_{D^*D})\sin(\Delta mt) + d(C_{D^*D} + r\Delta C_{D^*D})\cos(\Delta mt) \right],$$

where $d$ takes values $+1 (-1)$ for mesons whose initial flavour is $B^0 (\bar{B}^0)$ and $r$ takes values $+1 (-1)$ for the final states $f$ ($\bar{f}$).

This paper reports the first measurement of CP violation in $B^0 \to D^{*\pm}D^{\mp}$ decays at the LHCb experiment. The measurement is based on a sample of $pp$ collision data corresponding to integrated luminosities of 1 and 2 fb$^{-1}$ at centre-of-mass energies of 7 and 8 TeV (referred to as Run 1) and of 6 fb$^{-1}$ at 13 TeV (Run 2), recorded by the LHCb experiment between 2011 and 2018. Previous measurements with this $B^0$ decay mode have been performed by the BaBar [8] and Belle experiments [9].

In this analysis the $B^0 \to D^{*\pm}D^{\mp}$ candidates are reconstructed through the subsequent decays $D^- \to K^+\pi^-\pi^-$ and $D^{*+} \to D^0\pi^+$. For the $D^0$ meson, the $D^0 \to K^-\pi^+\pi^-\pi^-$ and $D^0 \to K^-\pi^+$ decay modes are used. The analysis proceeds as follows: $B^0 \to D^{*\pm}D^{\mp}$ candidates, reconstructed in the two $D^0$ decay modes and the two data-taking periods, are selected and analysed separately, as outlined in section 3. The signal contribution is determined in each of the four samples with fits to the $B^0$ mass distributions, as described in section 4. A key ingredient for measurements of CP violation in time-dependent analyses is the determination of the flavour of the neutral $B$ mesons by means of tagging algorithms, described in section 5. The evaluation of instrumental asymmetries that affect the measurement of the overall $CP$ charge asymmetry $A_{D^*D}$ is discussed in section 6. A simultaneous fit to the $B^0$ decay-time distributions of the four samples is performed to determine the $CP$ parameters, as described in section 7. The estimate of the systematic uncertainties is presented in section 8 and finally, conclusions are drawn in section 9.

2 Detector and simulation

The LHCb detector [10, 11] is a single-arm forward spectrometer covering the pseudorapidity range $2-5$, designed to study particles containing $b$ or $c$ quarks. The detector includes a high-precision tracking system consisting of a silicon-strip vertex detector surrounding the $pp$ interaction region [12], a large-area silicon-strip detector located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors and straw drift tubes [13, 14] placed downstream of the magnet. The tracking system provides a measurement of the momentum, $p$, of charged particles with a relative uncertainty...
that varies from 0.5% at low momentum to 1.0% at 200 GeV/$c$. During the data taking, the polarity of the magnetic field was periodically reversed to reduce the residual detection asymmetries that affect the determination of charge asymmetries. The minimum distance of a track to a primary collision vertex (PV), the impact parameter (IP), is measured with a resolution of $(15 + 29/p_T) \mu m$, where $p_T$ is the component of the momentum transverse to the beam, in GeV/$c$. Different types of charged hadrons are distinguished using information from two ring-imaging Cherenkov detectors [15]. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers [16].

Simulated data samples are used to model the effects of the detector acceptance and the imposed selection requirements. Samples of signal decays are produced in order to determine inputs for the analysis, such as the parametrisation of the mass distribution and the decay-time resolution model. Multibody $D^0$ and $D^+$ decays are modelled in the simulation according to the previously measured resonant structures [17, 18]. Samples of the most relevant background from partially reconstructed and misidentified $B$ meson decays, as well as specific $B$ decays useful for studies related to flavour tagging, are also produced. In the simulation, $pp$ collisions are generated using PYTHIA [19, 20] with a specific LHCb configuration [21]. Decays of unstable particles are described by EVTGEN [22], in which final-state radiation is generated using PHOTOS [23]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [24, 25] as described in ref. [26].

3 Selection

The online event selection is performed by a trigger [27], which consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage, which applies a full event reconstruction. At the hardware-trigger stage, events are required to have a muon with high $p_T$ or a hadron, photon or electron with high transverse energy in the calorimeters. The software trigger requires a two-, three- or four-track secondary vertex with a large sum of the transverse momenta of the tracks and a significant displacement from the primary $pp$ interaction vertices. A multivariate algorithm [28] is used for the identification of secondary vertices consistent with the decay of a $b$ hadron.

In the offline selection, the $D^{*+}$, $D^0$ and $D^-$ candidates are reconstructed through their decays into the selected final-state particles whose tracks are required to have good quality, exceed threshold values on $p$ and $p_T$ and satisfy loose particle identification (PID) criteria, mostly relying on the Cherenkov detectors information. These tracks are also required to have a $\chi^2_{IP}$ value with respect to any PV greater than four, where $\chi^2_{IP}$ is defined as the difference in the vertex-fit $\chi^2$ of a given PV reconstructed with and without the particle being considered. The distance of closest approach between all possible combinations of particles forming a common vertex should be smaller than 0.5 mm and the vertex should be downstream of the PV. The invariant mass of $D^0$ ($D^-$) meson candidates is required to lie within $\pm 40 \text{ MeV}/c^2$ ($\pm 50 \text{ MeV}/c^2$) of the known value [18], while the difference
of the $D^{*+}$ and $D^0$ invariant masses is required to be smaller than 150 MeV/$c^2$. These windows correspond to about $\pm 5$ times the mass resolutions. Candidate $B^0$ mesons are reconstructed from $D^{*+}$ and $D^-$ meson candidates that form a common vertex. The scalar sum of transverse momenta of the all final-state particles should exceed 5000 MeV/$c$ and the momentum direction of the $B^0$ meson should point to its associated PV. If more than one PV is reconstructed in an event, the associated PV is that with respect to which the signal $B^0$ candidate has the smallest $\chi^2_{IP}$.

Background can be due to the misidentification of one hadron in the charged $D$ decay chain. To suppress $A^0_1 \rightarrow \Lambda^+_c D_s^{*+}$ with $\Lambda^+_c \rightarrow pK^+\pi^-$ and $B^0 \rightarrow D_s^+ D^{*-}$ with $D_s^+ \rightarrow K^+K^-\pi^+$ decays, mass vetoes are applied. The pion from the $D^+ \rightarrow K^-\pi^+\pi^+$ reconstructed decay with the higher $p_T$ is assumed to be a proton (kaon) and the candidate is rejected if the recomputed invariant mass is within $\pm 25$ MeV/$c^2$ of the known $\Lambda^+_c$ ($D_s^+$) mass and the PID requirement for the alternative particle assignment is satisfied. To further reduce background contributions of $\phi \rightarrow K^+K^-$ from $D_s^+$ decays, the kaon mass hypothesis is assigned to the pion with the higher $p_T$ in the $D^- \rightarrow K^+\pi^-\pi^-$ reconstructed decay and the candidate $D^-$ is rejected if the invariant mass of the kaon pair is compatible with the $\phi$ mass within $\pm 10$ MeV/$c^2$. These vetoes reduce the $\Lambda^+_c$ and $\phi$ background contributions to a negligible level. The background due to $B^0 \rightarrow D^+_s D^{*-}$ decay is only partially suppressed by the $D_s^+$ veto which includes only loose selection criteria to retain high signal efficiency.

Single-charm $B$ decays such as $B^0 \rightarrow D^{*-}h^-h^+h^+$, where the three hadrons are not produced in a $D^+$ decay, but directly originate from the $B^0$ decay, are another potential source of background. To reject $B^0 \rightarrow D^{*-}\pi^-\pi^+\pi^+$ decays with a pion misidentified as a kaon, the $D^{*-}\pi^-\pi^+\pi^+$ invariant mass is calculated with the pion mass assigned to the kaon. The candidate is rejected if the mass is within $\pm 40$ MeV/$c^2$ of the known $B^0$ mass, and either the kaon candidate has a high probability to be a pion or the $\chi^2$ of the flight distance of the $D^+$ with respect to the $B^0$ decay vertex is less than four. Background arising from possible $B^0 \rightarrow D^{*-}\pi^-\pi^+\pi^+$ decays is suppressed by rejecting candidates with the $D^{*-}\pi^-\pi^+\pi^+$ mass within $\pm 25$ MeV/$c^2$ of the known $B^0$ mass and the $D^+ \rightarrow D_s^+ \pi^+$ vertex reconstructed upstream the $B^0$ decay vertex, or the $\chi^2$ of the flight distance of the $D^+$ candidate with respect to the $B^0$ decay vertex smaller than two. This veto is applied only to $D^0$ decaying into $K^-\pi^-\pi^+\pi^+$, as no excess is observed for the $K^-\pi^+$ sample. The invariant-mass distribution of signal candidates is not significantly modified by the vetoes.

In order to separate further the $B^0 \rightarrow D^{*\pm}D^\mp$ signal candidates from the combinatorial background, a boosted decision tree (BDT) utilising the AdaBoost method [29, 30] implemented in the TMVA toolkit [31, 32] is used. To train the BDT, simulated $B^0 \rightarrow D^{*\pm}D^\mp$ candidates are used as a proxy for signal, whereas candidates in data with invariant mass in the upper sideband are used as proxy for background. The upper sideband is defined as $5400 < m_{D^{*\pm}D^\mp} < 6000$ MeV/$c^2$ ($5600 < m_{D^{*\pm}D^\mp} < 6000$ MeV/$c^2$) for the $D^0$ final states $K^-\pi^+$ and $K^-\pi^-\pi^+\pi^+$. A two-folding procedure is applied to avoid overtraining for the $K^-\pi^+$ candidates in the mass range $5400 < m_{D^{*\pm}D^\mp} < 5600$ MeV/$c^2$, which are both used for the BDT training and the mass fit. Separate classifiers are trained for each final state and data taking period. Various kinematic and topological quantities are used in the BDT to exploit the features of the signal decay in order to distinguish it from
background, namely the transverse momentum of the $B^0, D^{*+}, D^- \text{ mesons}$ and of the $D^-$ and $D^0 \text{ decay products}$; the decay-time significance of the $D^- \text{ and } D^0 \text{ mesons}$, and their flight distance $\chi^2$ with respect to the associated PV; the $B^0 \chi^2_D$; the angles formed by each of the $D^-$ decay products and the $D^-$ direction; the angle formed by the pion from the $D^0 \text{ decay}$ and the $D^0 \text{ direction}$, and the angle formed by the $D^{*+}$ and the $B^0 \text{ mesons}$.

The PID probabilities of the kaon and pion in the final state are also used in the BDT.

The requirement on the output of each BDT classifiers is chosen to minimise the uncertainties on the $CP$ parameters $S_{D^*D}$ and $C_{D^*D}$.

A kinematic fit to the $B^0 \text{ decay chain}$ with constraints on the masses of all the charm mesons and on the PV is performed to significantly improve ($\sim 60\%$) the resolution on the invariant mass of the $B^0 \text{ candidate}$ [33], while the $B^0$ decay time is calculated using the same fit with only the PV constraint in order to avoid correlations of the decay time with the invariant mass. Candidate $B^0 \text{ mesons}$ are retained if their invariant mass and decay time are in the ranges $5000–5600 \text{ MeV/c}^2$ and $0.3–10.3 \text{ ps}$, respectively, where the lower boundary of the decay time is set to reduce promptly produced background. After the selection a fraction of events below $5\%$ have multiple candidates. In these cases a single candidate is randomly selected with negligible change in the final result.

### 4 Mass fit

An unbinned extended maximum-likelihood fit is performed on the invariant-mass distribution of the selected $B^0 \text{ candidates}$, on each of the four data samples independently. The sPlot technique [34] is employed to determine per-candidate weights that are used for background subtraction in the subsequent decay-time fit. The model describing the $D^{*\pm}D^\mp$ mass distribution consists, in addition to the signal $B^0 \to D^{*\pm}D^\mp$, of the following background components: $B^0_s \text{ decays}$ to the same $D^{*\pm}D^\mp \text{ final states}$; $B^0 \to D_s^{+}D^{*-}$ decays with a misidentified kaon that pass the selection; $B^0$ and $B^0_s \text{ decays}$ to $D^{*+}D^{*-}$ with one of the excited charm mesons decaying into a charged $D$ meson and an additional unreconstructed neutral pion or photon; and combinatorial background. The distribution of the reconstructed mass of the $B^0 \text{ signal component}$ is parametrised with the sum of two (three) Crystal Ball functions [35] for the $K^-\pi^-\pi^+\pi^+$ ($K^-\pi^+$) $D^0 \text{ final state}$, with common mean but different width and tail parameters. The mass model for the $B^0_s \text{ decays}$ is the same as for the $B^0 \text{ decays}$, but the peak position is shifted by the difference between the known values of the $B^0_s$ and $B^0 \text{ mesons}$ [18]. The mass distribution of $B^0 \to D_s^{+}D^{*-}$ decays is described with the sum of two Crystal Ball functions with a common mean, which is floated in the data fit, while the remaining parameters are fixed from simulation. The mass distribution for the partially reconstructed $B^0 \to D^{*+}D^{*-}$ and $B^0_s \to D^{*+}D^{*-}$ background contributions are described by a combination of functions corresponding to pure longitudinal and transverse polarizations of the two $D^{*\pm}$ mesons. The relative fraction of the two possible contributions are floated in the fit, while the shapes, that are a double peak in case of longitudinal polarization and a single broad one for transverse polarization, are fixed to the evaluated on simulated samples. The relative fraction of the two contributions in $B^0_s \to D^{*+}D^{*-}$ and $B^0 \to D^{*+}D^{*-}$ decays are assumed to be the same. The distribution of the reconstructed mass for the combinatorial background component...
is modelled by an exponential function. The results of the fits to the four data samples are shown in figure 1 with the partial contribution of each component overlaid; the resulting signal yields are 469 ± 28 (1570 ± 48) and 856 ± 32 (3265 ± 61) for the $D^0$ final states $K^-\pi^-\pi^+\pi^+$ and $K^-\pi^+$ in Run 1 (Run 2), respectively.

## 5 Flavour tagging

Measurements of $CP$ violation in decay-time-dependent analyses of $B^0$ meson decays require the determination of the production flavour of the $B^0$ meson. Methods to infer the
initial flavour of a reconstructed candidate, i.e. whether it contained a $b$ or a $\bar{b}$ quark at production, are referred to as flavour tagging. Two classes of algorithms are used. The opposite side (OS) tagger exploits the fact that $b$ and $\bar{b}$ quarks are almost exclusively produced in pairs in $pp$ collisions, allowing the flavour of the signal $B^0$ candidate to be inferred from the flavour of the other $b$ hadron in the event. The OS tagger combines information on the charge of the muon or electron from semileptonic $b$ decays, the charge of the kaon from the $b \rightarrow c \rightarrow s$ decay chain, the charge of a reconstructed secondary charm hadron and the charges of the tracks that form the secondary vertex of the other $b$-hadron decay, combined into a weighted average, with weights depending on the transverse momenta of the tracks [36, 37]. The same-side (SS) tagger exploits the production of correlated protons or pions in the hadronization of the $\bar{b}$ ($b$) quark that forms the signal $B^0$ ($\bar{B}^0$) candidate, with its initial flavour identified by the charge of the particle [38].

Each tagging algorithm provides a flavour-tagging decision, $d$, and an estimate, $\eta$, of the probability that the decision is incorrect (mistag) for a reconstructed $B^0$ candidate. The tagging decision takes the value $\pm 1$ for tagged $B$ candidates and 0 if no decision on initial flavour can be assigned (untagged). The mistag probability varies in the range from 0 to 0.5 for tagged candidates and is equal to 0.5 for untagged candidates.

Each tagging algorithm is implemented as a BDT that is trained and optimised using large data samples of $B^+ \rightarrow J/\psi K^+$ decays for the OS and of $B^0 \rightarrow D^- \pi^+$ decays for the SS taggers, respectively [39]. The mistag probability for each tagger is given by the output of the BDT, which is calibrated using dedicated data control channels to relate $\eta$ to the true mistag probability, $\omega$. The performance of the flavour tagging is measured by the tagging power, $\varepsilon_{\text{tag}} D^2$, where $\varepsilon_{\text{tag}}$ is the fraction of tagged candidates and $D = 1 - 2\omega$ represents the dilution induced on the oscillation amplitude. The tagging power represents the effective loss in signal yield compared to a perfectly tagged sample.

5.1 Calibration of the tagging output

Flavour tagging algorithms are calibrated using control samples of flavour-specific $B$ decays, separately in Run 1 and Run 2 data. At first approximation, the measured mistag fraction $\omega$ in the control channel can be expressed as a linear function of the predicted mistag estimate $\eta$ as

$$\omega(\eta) = p_0 + p_1 (\eta - \langle \eta \rangle),$$

(5.1)

where the use of the arithmetic mean $\langle \eta \rangle$ of the $\eta$ distribution decorrelates the $p_0$ and $p_1$ parameters. A perfect calibration of the taggers would result in $p_0 = \langle \eta \rangle$ and $p_1 = 1$. A signal candidate tagged with decision $d$ and a calibrated mistag $\omega > 0.5$ corresponds to an opposite tagging decision $d' = -d$ with a mistag probability $\omega' = 1 - \omega$.

The performance of the flavour taggers, $\varepsilon_{\text{tag}}$ and $\omega$, may depend on the initial flavour of the neutral $B$ meson. Charged decay products, like the $K^+$ and $K^-$ which are used by the OS kaon tagger, can have significantly different interaction rates with the detector material and therefore lead to different reconstruction efficiencies. To account for these
asymmetries eq. (5.1) is modified as

$$\bar{\alpha}(\eta) = \left( p_0 \pm \frac{\delta p_0}{2} \right) + \left( p_1 \pm \frac{\delta p_1}{2} \right) (\eta - \langle \eta \rangle), \quad (5.2)$$

where the mistag fractions $\omega$ and $\bar{\alpha}$ for an initial $B^0$ and $\bar{B}^0$ correspond to the plus and minus sign, respectively, and $\delta p_{0,1} = p_{0,1}^P - p_{0,1}^s$. Similarly, the tagging asymmetry, $A_{tag}$, is defined as

$$A_{tag} = \frac{\varepsilon_{tag}^P - \varepsilon_{tag}^s}{\varepsilon_{tag}^P + \varepsilon_{tag}^s}. \quad (5.3)$$

For this analysis, the calibration of the OS and SS taggers is performed using samples of $B^0 \to D_s^+ D^-$ and $B^0 \to D_s^+ D^{*-}$ flavour-specific decays which have similar topology, kinematics and event characteristics as the signal decay. Candidate $B^0$ mesons in the two decay modes are reconstructed using the decays $D_s^+ \to K^- K^+ \pi^+$, $D^- \to K^+ \pi^− \pi^−$ and $D^{*-} \to \pi^- \bar{D}^0$, followed by $\bar{D}^0 \to K^+ \pi^−$ or $\bar{D}^0 \to K^+ \pi^− \pi^+ \pi^−$. These decays are selected using the same trigger requirements, and similar kinematic and geometric criteria, to those applied for the signal selection. Requirements are applied on the identification of the final-state particles as well as on the invariant mass of $D_s^+$, $D^{*-}$ and $\bar{D}^0$ mesons. Candidate $B^0$ mesons are retained if their invariant mass and decay time are in the ranges $5220 - 5500$ MeV/c$^2$ and $0.3 - 10.3$ ps, respectively. In case multiple candidates are selected in the same event, the candidate with the highest $p_T$ is retained.

A fit to the invariant-mass distribution of the $B^0$ candidates is performed on each sample independently to separate the signal from the background contribution, which consists of random combination of particles. The signal is modelled by the sum of two Gaussian functions with a common mean, while the background shape is described by an exponential function. In the case of the $B^0 \to D_s^+ D^-$ control channel, an additional contribution due to $B_s^0 \to D_s^- D^+$ decays is also needed. It is modelled by the same function used for the $B^0$ decays except for the mean value, that is shifted by the known difference of the $B_s^0$ and $B^0$ masses [18]. Signal yields of 11,400 (39,200) $B^0 \to D_s^+ D^{*-}$ decays and 24,900 (102,900) $B^0 \to D_s^+ D^-$ decays, in the Run 1 (Run 2) data sample, are found.

The mass fit determines for each selected candidate a weight that is used to statistically subtract the background contribution, using the sPlot technique. The tagging calibration parameters are determined from an unbinned maximum-likelihood fit to the weighted distribution of the decay time $t$, the final state $r$, with $r = +1 (-1)$ for $D_s^+ D^{(*)^-}$, tag decisions $\bar{d} = (d_{OS}, d_{SS})$ and probabilities $\bar{\eta} = (\eta_{OS}, \eta_{SS})$, according to the following PDF

$$\mathcal{P}(t, r, \bar{d} | \bar{\eta}) = \epsilon(t) \cdot \left[ \mathcal{F}(t', r, \bar{d} | \bar{\eta}) \otimes \mathcal{R}(t - t') \right]. \quad (5.4)$$

Here, $\mathcal{R}(t - t')$ is the decay-time resolution, which is modelled using simulation. The decay-time efficiency, $\epsilon(t)$, is parametrised using a cubic spline model [40], with parameters determined from the fit to the data. The number of spline coefficients and the knot positions are very similar among the $B^0 \to D_s^+ D^-$ and $B^0 \to D_s^+ D^{*-}$ decays, as well as among Run 1 and Run 2 samples. The signal PDF, $\mathcal{F}$, describes the decay-time distribution of flavour-specific $B^0$ decays; it is based on eqs. (7.2) and (7.3) of section 7, with appropriate values
asymmetry of the $B^0$ detection asymmetry of the control channel, that are determined from the fit to data. The tagging efficiencies and their asymmetry, on the $B^0$ values [18]. In addition, the PDF depends on the tagging parameters of eq. (5.2), on the coefficients corresponding to flavour-specific $B^0$ decays. It depends on the $\tau_d$ and $\Delta m_d$ parameters, that describe the $B^0$ decay and mixing, which are fixed to their known values [18]. In addition, the PDF depends on the tagging parameters of eq. (5.2), on the tagging efficiencies and their asymmetry, on the $B^0$ production asymmetry and on the detection asymmetry of the control channel, that are determined from the fit to data. The asymmetry of the $B^0$ meson production is defined as $A_{\text{prod}} = (\sigma_{\overline{B}^0} - \sigma_{B^0})/(\sigma_{\overline{B}^0} + \sigma_{B^0})$, where $\sigma_{\overline{B}^0}$ and $\sigma_{B^0}$ are the production rates of $\overline{B}^0$ and $B^0$ mesons.

The fit is performed independently on each sample. For each data-taking period, the parameters of the tagging calibration and of the production asymmetries are consistent between the two channels, and the results are combined according to their statistical uncertainties.

Different sources of systematic uncertainty are considered. Systematic uncertainties on the calibration parameters are determined by repeating the fit to data with modified assumptions. The deviations of the fit results from the nominal values are assigned as systematic uncertainties. The largest uncertainty is related to the determination of the signal weights from the fit to the mass distribution. An alternative determination of such weights is performed fitting the two-dimensional invariant-mass distribution of the $D_s^+$ and $\overline{D}^0$ ($D^{*-}$) candidates. To account for possible effects related to the decay-time efficiency model, an analytic function is considered instead of the cubic spline function. The uncertainties related to the input values of the $B^0$ decay and mixing properties are determined repeating the fit with inputs varied within their uncertainties. Finally, uncertainties related to differences between data and simulation concerning the time resolution model are neglected given their insignificant impact on the tagging parameters for variations up to 50%. The resulting tagging parameters are listed in table 1.

The portability of the calibration from the control channels to the signal decay channel is assessed with simulation. Both OS and SS taggers show compatible results among the different decay modes. A deviation from linearity of the SS tagger calibration is observed.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Run 1</th>
<th>Run 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{\text{prod}}$</td>
<td>$-0.011 \pm 0.008 \pm 0.003$</td>
<td>$0.004 \pm 0.005 \pm 0.002$</td>
</tr>
<tr>
<td>$A_{\text{tag}}^{\text{OS}}$</td>
<td>$0.008 \pm 0.008 \pm 0.005$</td>
<td>$0.001 \pm 0.004 \pm 0.009$</td>
</tr>
<tr>
<td>$\delta_{0}^{\text{OS}}$</td>
<td>$0.022 \pm 0.007 \pm 0.013$</td>
<td>$0.009 \pm 0.004 \pm 0.003$</td>
</tr>
<tr>
<td>$\delta_{1}^{\text{OS}}$</td>
<td>$0.20 \pm 0.06 \pm 0.11$</td>
<td>$0.02 \pm 0.03 \pm 0.01$</td>
</tr>
<tr>
<td>$p_{0}^{\text{OS}} - \langle \eta^{\text{OS}} \rangle$</td>
<td>$0.037 \pm 0.005 \pm 0.006$</td>
<td>$0.032 \pm 0.003 \pm 0.005$</td>
</tr>
<tr>
<td>$p_{1}^{\text{OS}}$</td>
<td>$0.95 \pm 0.04 \pm 0.10$</td>
<td>$0.87 \pm 0.02 \pm 0.02$</td>
</tr>
<tr>
<td>$A_{\text{tag}}^{\text{SS}}$</td>
<td>$-0.006 \pm 0.004 \pm 0.005$</td>
<td>$-0.001 \pm 0.001 \pm 0.002$</td>
</tr>
<tr>
<td>$\delta_{0}^{\text{SS}}$</td>
<td>$-0.004 \pm 0.005 \pm 0.009$</td>
<td>$0.001 \pm 0.003 \pm 0.003$</td>
</tr>
<tr>
<td>$\delta_{1}^{\text{SS}}$</td>
<td>$-0.03 \pm 0.08 \pm 0.06$</td>
<td>$-0.11 \pm 0.04 \pm 0.02$</td>
</tr>
<tr>
<td>$p_{0}^{\text{SS}} - \langle \eta^{\text{SS}} \rangle$</td>
<td>$0.007 \pm 0.004 \pm 0.006$</td>
<td>$-0.002 \pm 0.002 \pm 0.002$</td>
</tr>
<tr>
<td>$p_{1}^{\text{SS}}$</td>
<td>$0.94 \pm 0.06 \pm 0.07$</td>
<td>$0.88 \pm 0.03 \pm 0.03$</td>
</tr>
</tbody>
</table>

Table 1. Flavour-tagging parameters obtained as a weighted average of the values measured in the two control channels. The quoted uncertainties are statistical and systematic.
Table 2. Tagging efficiency and tagging power for $B^0 \to D^{*\pm} D^{\mp}$ signal candidates in the four data samples, computed using the event-by-event predicted mistag $\eta$ and the calibration parameters obtained from control channels. The quoted uncertainties are statistical only.

5.2 Tagging results

Approximately 40% of the tagged candidates in the signal decay samples are tagged by both the OS and the SS algorithms. Since the algorithms select different samples of charged particles and hence are uncorrelated, the two tagging results are combined taking into account both decisions and their corresponding estimates of $\eta$ as detailed in ref. [41]. The combined estimated mistag probability and the corresponding uncertainty are obtained by combining the individual calibrations for the OS and SS taggers and propagating their uncertainties in the decay-time fit. The effective tagging power and efficiency for signal candidates tagged by one or both of the OS and SS algorithms are given in table 2.

6 Instrumental asymmetries

The overall asymmetry that is measured in the decay-time fit has to be corrected for instrumental asymmetries in order to determine the physical parameter $A_{D^{*}D}$. These asymmetries affect reconstruction, detection and particle identification efficiencies and are related to the different interaction cross-section with matter and different detection and identification efficiencies of positive and negative pions and kaons. The $B^0 \to D^{*\pm} D^{\mp}$ decay is charge symmetric, however since all instrumental efficiencies depend on momenta, and the $p$ and $p_T$ spectra of kaons and pions in the $D^{*\pm}$ and $D^{\mp}$ decays are observed to be slightly different, the cancellation is not expected to be complete.
be smaller in Run 1 than in Run 2, the difference of the kinematic distributions of the final state particles of the signal, as described in the following. In order to account for the PID asymmetry, the same kaon and pion identification requirements used in the \( B^0 \rightarrow D^{\pm} D^\mp \) selection are applied to the prompt \( D^- \) meson sample. The additional asymmetries induced by the use of PID variables in vetoes and in the BDT employed for the selection are considered as a source of systematic uncertainty on the asymmetry \( A_{\text{det}}^{D^-} \). A sample of about 6 million prompt \( D^- \rightarrow K^+ \pi^- \pi^- \) decays collected in Run 2 is used and divided into four subsamples: for each \( D^0 \) decay mode of the \( B^0 \rightarrow D^{\pm} D^\mp \) signal, one sample is used to measure the \( D^{\pm} \) asymmetry, the other is used to determine the \( D^- \) asymmetry. In each subsample a fit to the \( K^+ \pi^- \pi^- \) invariant mass distribution is performed to determine the weights to be used to subtract background. The signal mass model is parametrised by two Crystal Ball functions with common mean and the background is modelled by an exponential function. For the \( D^0 \rightarrow K^- \pi^+ \) decay mode, one of the prompt \( D^- \rightarrow K^+ \pi^- \pi^- \) subsamples is weighted to match the spectra of the final state particles in the signal \( D^{\pm} \) decay and the other is weighted to match the spectra of the final state particles in the signal \( D^- \) decay [42]. The same procedure is used for the \( D^0 \rightarrow K^- \pi^- \pi^+ \pi^+ \) decay mode, but not considering two of the pions with opposite charge from the \( D^0 \) decay, since the \( \pi^- \pi^+ \) pair is found to contribute negligibly to additional asymmetries. Finally, the asymmetries \( A_{\text{det}}^{D^\pm} \) and \( A_{\text{det}}^{D^-} \) are calculated from the \( D^+ \) and \( D^- \) yields obtained by fitting the mass distributions of each weighted prompt \( D^+ \) and \( D^- \) meson samples and subtracted. The results are reported in table 3 and \( A_{\text{det}}^{D^\pm-D^+} \) values are consistent with zero.

The following sources of systematic uncertainties on \( A_{\text{det}} \) are considered: asymmetries due to the particle identification and hadronic hardware-trigger requirements, variation of the prompt \( D^- \rightarrow K^+ \pi^- \pi^- \) selection criteria and of the mass-fit model and imperfect cancellation of the prompt \( D^\pm \) production asymmetry. Since the differences between the kinematic distributions of the final state particles of \( D^{\pm} \) and \( D^\mp \) decays are observed to be smaller in Run 1 than in Run 2, the difference of \( D^{\pm} \) and \( D^\mp \) asymmetries in Run 1 is assumed to be zero with the same statistical and systematic uncertainties as in Run 2.

<table>
<thead>
<tr>
<th>Final state</th>
<th>( A_{\text{det}}^{D^\pm} )</th>
<th>( A_{\text{det}}^{D^-} )</th>
<th>( A_{\text{det}}^{D^\pm-D^+} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D^0 \rightarrow K^- \pi^+ )</td>
<td>0.0169 ± 0.0036</td>
<td>0.0158 ± 0.0018</td>
<td>0.0011 ± 0.0040 ± 0.0030</td>
</tr>
<tr>
<td>( D^0 \rightarrow K^- \pi^- \pi^+ \pi^+ )</td>
<td>0.0146 ± 0.0022</td>
<td>0.0138 ± 0.0015</td>
<td>0.0009 ± 0.0026 ± 0.0051</td>
</tr>
</tbody>
</table>

Table 3. Instrumental asymmetries of \( D^\pm, D^{\pm} \) and their combination for \( B^0 \rightarrow D^{\pm} D^\mp \) decays obtained from prompt \( D \) meson decays. The quoted uncertainties are statistical, in the last column statistical and systematic.

To a good level of approximation, the asymmetry in \( B^0 \rightarrow D^{\pm} D^\mp \) decays, denoted with \( A_{\text{det}} \), can be related to the \( D^\pm \) and \( D^- \) asymmetries, \( A_{\text{det}}^{D^\pm} \) and \( A_{\text{det}}^{D^-} \), as

\[
A_{\text{det}}^{D^\pm-D^+} \simeq A_{\text{det}}^{D^\pm} - A_{\text{det}}^{D^-}.
\]
7 Decay-time fit

A decay-time fit is performed simultaneously on the four data samples to measure the CP coefficients $S_{D^*D}$, $\Delta S_{D^*D}$, $C_{D^*D}$, $\Delta C_{D^*D}$ and $A_{D^*D}$. The weights determined with the mass fit are used to subtract the background, using the sPlot technique. Before performing the fit, a blinding transformation is applied on the CP parameters that is removed only at the end of the analysis.

The PDF describing the measured $B^0$ decay-time distribution and tag decisions $\tilde{d} = (d_{OS}, d_{SS})$, given the mistag probability estimates $\tilde{\eta} = (\eta_{OS}, \eta_{SS})$, can be expressed as

$$P(t, r, \tilde{d} | \tilde{\eta}) = \epsilon(t) \cdot \left( \mathcal{F}(t', r, \tilde{d} | \tilde{\eta}) \otimes R(t - t') \right),$$

(7.1)

where $\mathcal{P}(t', r, \tilde{d} | \tilde{\eta})$ is the PDF describing the distribution of the true decay time, $t'$, $R(t - t')$ is the decay-time resolution function, while $\epsilon(t)$ describes the decay-time efficiency. The PDF describing the $B^0$ decay-time distribution has the same form as eq. (1.5) but with effective coefficients

$$\mathcal{F}(t', r, \tilde{d} | \tilde{\eta}) = Ne^{-t'/\tau_s}(1 + r A_{\text{raw}}) \left[ C_{\text{cosh}}^{\text{eff}} - C_{\text{sin}}^{\text{eff}} \sin(\Delta m_{d} t') + C_{\text{cos}}^{\text{eff}} \cos(\Delta m_{d} t') \right].$$

(7.2)

These coefficients, that depend on the final state variable $r = \pm 1$, can be expressed as

$$C_{\text{cosh}}^{\text{eff}} = (S_{D^*D} + r \Delta S_{D^*D}) (\Delta^- - A_{\text{prod}} \Delta^+),$$
$$C_{\text{sin}}^{\text{eff}} = (C_{D^*D} + r \Delta C_{D^*D}) (\Delta^- - A_{\text{prod}} \Delta^+),$$
$$C_{\text{cos}}^{\text{eff}} = (\Delta^+ - A_{\text{prod}} \Delta^-),$$

(7.3)

where the factors $\Delta^\pm$ contain the dependence on the mistag fraction, the tagging decisions and efficiencies and the asymmetries in the tagging efficiencies of the OS and SS taggers. The parameter $A_{\text{raw}}$ is the sum of the CP asymmetry, $A_{D^*D}$, and the detection asymmetry, $A_{D^*D}^{\text{det}}$. Since the instrumental asymmetries depend on the final state and on the data-taking period, four different parameters are used in the fit.

The decay-time resolution model is determined from simulation and fixed in the fit to data with an effective resolution of 60 fs. The same set of parameters describes well both Run 1 and Run 2 data, with small differences between the two $D^0$ final-state samples. Due to the low $B^0$-$\bar{B}^0$ oscillation frequency, the decay-time resolution has a very small influence on the relatively low CP parameters.

The selection and reconstruction efficiency depends on the $B^0$ decay time due to displacement requirements made on the signal final-state particles and a decrease in the reconstruction efficiency for tracks with large impact parameter with respect to the beam line [43]. The decay-time efficiency is modelled by a cubic spline function with five knots [40]. The knot positions are chosen from a fit to simulated signal candidates at $(0.3, 0.5, 2.7, 6.3, 10.3)$ ps. The spline coefficients are determined from the fit to data where the $B^0$ lifetime is fixed to its known value [18]. The same parameters are assigned to both $K^-\pi^+$, $K^-\pi^-\pi^+\pi^+$ final-state samples, while different values are used for Run 1 and Run 2 data, as suggested by simulation studies. The mass difference $\Delta m_{d}$ is fixed to its known value [18].
The average mistag values for OS and SS tags are fixed to the values calculated in each corresponding signal sample. All the tagging parameters, as reported in table 1, are introduced in the fit through Gaussian constraints, in order to account for their associated uncertainties. The tagging efficiencies are free to vary in the fit. The OS and SS taggers are combined in the fit. The production asymmetries are also constrained in the fit to the values measured in the flavour-tagging control channels. The result of the production asymmetry in Run 1 is in agreement with the dedicated LHCb measurement [44], when considering the kinematics of the signal data sample.

The $CP$ observables resulting from the decay-time fit are
\[
S_{D^+D} = -0.861 \pm 0.077,
\]
\[
\Delta S_{D^+D} = 0.019 \pm 0.075,
\]
\[
C_{D^+D} = -0.059 \pm 0.092,
\]
\[
\Delta C_{D^+D} = -0.031 \pm 0.092.
\]

After subtracting the detection asymmetries to each of the four $A_{raw}$ values the weighted mean is calculated as $A_{D^+D} = 0.008 \pm 0.014$. The quoted uncertainties include contributions due to the size of the samples and due to the external parameters constrained in the fit. The correlations among the $CP$ parameters are

<table>
<thead>
<tr>
<th></th>
<th>$S_{D^+D}$</th>
<th>$\Delta S_{D^+D}$</th>
<th>$C_{D^+D}$</th>
<th>$\Delta C_{D^+D}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{D^+D}$</td>
<td>1</td>
<td>0.07</td>
<td>0.44</td>
<td>0.05</td>
</tr>
<tr>
<td>$\Delta S_{D^+D}$</td>
<td>1</td>
<td>0.04</td>
<td>0.46</td>
<td>0.05</td>
</tr>
<tr>
<td>$C_{D^+D}$</td>
<td>1</td>
<td>0.04</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>$\Delta C_{D^+D}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The four $A_{raw}$ parameters are almost uncorrelated among each other, with the largest correlation coefficient being $10^{-4}$. Their correlation with all the other fit parameters is also small, between 0.1% and 1%. The spline parameters have large correlations among them but have correlations between 0.1% and 1% with the $CP$ parameters.

Figure 2 shows the decay-time distribution of the full $B^0 \rightarrow D^*\pm D^\mp$ data sample, where the fitted PDF is overlaid. For illustration, figure 3 shows the asymmetry between $B^0$ and $B^0$ signal yields as a function of the decay time, separately for $D^*+D^-$ and $D^*-D^+$ final states, with the corresponding fit functions overlaid.

8 Systematic uncertainties

Several cross checks of the analysis and possible sources of systematic uncertainties of the results are considered in the following and summarised in table 4. As a first validation, the decay-time fit is performed on a simulated sample of signal decays corresponding to 36 times the data sample size. The resulting $CP$ parameters are compatible with the generation values within less than two standard deviations. In order to test if the likelihood estimate for the $CP$ parameter values are accurate, the same mass and decay-time models as used for
Figure 2. Decay-time distribution of the $B^0 \to D^{\pm} D^{\mp}$ signal candidates, summed over all data samples, where the background contribution is subtracted by means of the *sPlot* technique. The projection of the PDF is represented by the full line.

Figure 3. Asymmetry between $B^0$ and $\bar{B}^0$ signal yields as a function of the decay time, for (blue full dot) $D^{\ast+} D^-$ and (red empty dot) $D^{\ast-} D^+$ signal candidates with a non null tagging decision. The background contribution is subtracted by means of the *sPlot* technique. The corresponding projections of the PDF are represented by the blue dashed (red continuous) line. The $B^0$ flavour is determined by the combination of all flavour-tagging algorithms.

data are fitted to 2000 pseudoexperiments. For each pseudoexperiment, four samples are generated corresponding to the four data samples. In each sample, the input parameters for the signal component are the same as in the fit to data, except for the production and tagging asymmetries, which are set to zero for convenience. The $CP$ parameters, the
tagging efficiencies and the coefficients of the splines describing the time acceptance are taken from the best fit result to data. All the parameters that are constrained in the fit are set to values randomly generated according to the constraints applied. Each background contribution is generated with a specific time dependence, which in some cases accounts for \( CP \) violation. No bias is found for the \( CP \) parameters nor for the raw asymmetries \( A_{\text{raw}} \). The uncertainty on the mean value of the bias is taken as systematic uncertainty on the parameter.

In order to cross-check the statistical uncertainty obtained from the fit to data, a bootstrapping procedure is used [45]. In this frequentist model-independent approach a new data sample is generated by drawing candidates from the nominal data sample until the number of candidates matches that of the original one (the same event can be drawn multiple times). The nominal fit to the decay time is executed, the fit result is stored and the distribution of the residuals with respect to the starting values of the parameters is analysed. Given that the standard deviation of the Gaussian fits of the residuals agree with the mean values of the uncertainty from the fit it can be assumed that the uncertainty of the nominal fit is accurate.

Cross checks on the stability of the result are performed by dividing the data sample into categories, according to the \( D^0 \) final state, the tagging algorithm (OS and SS) and the magnet polarity. No evidence of bias is found, as all variations of the fit parameters obtained in the splittings are smaller than two standard deviations.

<table>
<thead>
<tr>
<th>Source</th>
<th>( \Delta C_{D^0} )</th>
<th>( C_{D^0} )</th>
<th>( \Delta S_{D^0} )</th>
<th>( S_{D^0} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit bias</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>Mass model</td>
<td>0.006</td>
<td>0.014</td>
<td>0.003</td>
<td>0.011</td>
</tr>
<tr>
<td>( \Delta m_d, \tau_d, \Delta \Gamma_d )</td>
<td>0.001</td>
<td>0.003</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Decay-time resolution</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
</tr>
<tr>
<td>Decay-time acceptance</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
</tr>
<tr>
<td>Flavour tagging</td>
<td>0.015</td>
<td>0.014</td>
<td>0.012</td>
<td>0.015</td>
</tr>
<tr>
<td>Total syst. uncertainty</td>
<td>0.016</td>
<td>0.020</td>
<td>0.012</td>
<td>0.019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>( A_{\text{raw}}^{K_{\pi\pi}, \text{Run1}} )</th>
<th>( A_{\text{raw}}^{K_{\pi\pi}, \text{Run2}} )</th>
<th>( A_{\text{raw}}^{K_{\pi}, \text{Run1}} )</th>
<th>( A_{\text{raw}}^{K_{\pi}, \text{Run2}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit bias</td>
<td>0.0013</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0004</td>
</tr>
<tr>
<td>Mass model</td>
<td>0.0025</td>
<td>0.0024</td>
<td>0.0021</td>
<td>0.0016</td>
</tr>
<tr>
<td>( \Delta m_d, \tau_d, \Delta \Gamma_d )</td>
<td>0.0003</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0001</td>
</tr>
<tr>
<td>Decay-time resolution</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Decay-time acceptance</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0001</td>
</tr>
<tr>
<td>Flavour tagging</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Total syst. uncertainty</td>
<td>0.0028</td>
<td>0.0025</td>
<td>0.0023</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

Table 4. Summary of the systematic uncertainties. The total systematic uncertainties are computed as quadratic sums of individual contributions.
Specific studies are performed to estimate the effect of an inaccurate determination of the mass model, the decay-time model, the flavour-tagging performance and variations of the input parameters to the decay-time fit. Four systematic effects on the mass model, described below, are considered and the sum in quadrature of each systematic uncertainty is taken as uncertainty on the mass model reported in table 4. The signal model is changed to two (three) Crystal Ball functions for the mass distribution of $K^-\pi^+ (K^-\pi^-\pi^+\pi^+)$ final state or to two Crystal Ball functions for both final states. The model for the combinatorial background is changed to a second-order polynomial. These systematic uncertainties are evaluated using pseudoexperiments, with a procedure that is used also in the rest of the paper for other sources of uncertainty. In each pseudoexperiment, the mass fit to each of the four samples is performed with the nominal model and then repeated with the alternative model. Results of the subsequent decay-time fit are compared to those obtained with the nominal fit and the distribution of their difference is built. The systematic uncertainty is defined as the sum in quadrature of the average and root mean square of the distribution.

Possible differences between data and simulation in the mass resolution are considered. The mass fit to data is repeated with the parameters describing the mass resolution of the signal and of the $B^0_s\rightarrow D^{\ast\pm}D^{\mp}$ and $B^0\rightarrow D^{\ast+}D^{*-}$ background contributions (widths of the Crystal Ball functions, fixed to the values determined from simulation in the nominal model) multiplied by a common, free scaling factor. The decay-time fit to data is repeated and the variation of the $CP$ parameters is taken as systematic uncertainty.

A possible contribution to the mass distribution due to background from $B^0_s\rightarrow D_s^{+}D_s^{-}$ decays is also considered. The same shape and parameters for the mass PDF as for the $B^0\rightarrow D_s^{+}D_s^{-}$ component is taken, but the mean is shifted by the difference of the known values of the $B^0_s$ and $B^0$ mesons [18]. No significant yield for this decay mode as well as no significant variation in the signal yield is found, therefore no systematic uncertainty is associated to this background.

To assess the systematic uncertainty on the decay-time efficiency description, pseudoexperiments are generated with alternative positions of the spline knots, (0.3, 1.3, 2.2, 6.3, 10.3) ps, chosen using simulation to provide a good alternative fit. The systematic uncertainty is found to be negligible, as expected from the small correlation between the decay-time efficiency and the $CP$ parameters.

The validity of using the same decay-time efficiency function for the two $D^0$ final state samples is tested by producing 2500 data sets with the bootstrap method and fitting each sample once with the nominal PDF and once with different acceptances for the two final states. Due to the limited amount of data candidates, only 32% of the fits with separate acceptances converge with a good covariance matrix quality, however the systematic uncertainty evaluated from those fits is found to be negligible. The systematic uncertainty on the $CP$ parameters due to variation of the decay-time resolution, assessed with pseudoexperiments, is found to negligible.

The calibration parameters of flavour tagging are constrained in the decay-time fit within their statistical and systematic uncertainties, therefore their variation is included in the statistical uncertainty of the $CP$ parameters. Two additional sources of uncertainties related to the tagging parameters, as mentioned in section 5, are considered. The first
test concerns a small deviation from a perfect linear calibration of the mistag probability of the SS tagger observed in signal simulated data. The second test accounts for the effect of a slight dependence of the OS tagger mistag probability on the $B^0$ decay time. Pseudoexperiments are used to evaluate the variations of the $CP$ parameters with respect to the unbiased case. The systematic uncertainties corresponding to the two tests are summed in quadrature and reported in table 4.

The systematic uncertainty related to the decay-time fit input parameters ($\tau_d$, $\Delta m_d$ and $\Delta \Gamma_d$) is determined by varying each parameter according to its uncertainty [3]. A test is done to check the impact on the $CP$ parameters of the assumption of no $CP$ violation in the $B^0 \rightarrow D^*_s D^{*-}$ background. Pseudoexperiments with a charge asymmetry of 10% included at generation for the $B^0 \rightarrow D^*_s D^{*-}$ component are studied. No bias is found on the $CP$ parameters, as a consequence no systematic uncertainty is assigned due to this source.

The systematic uncertainty on $A_{D^* D}$ is evaluated as the weighted average of the quadratic sum of the uncertainties on the detection asymmetry, from table 3 and the uncertainties on the raw asymmetries, from table 4.

9 Results and conclusion

A flavour-tagged decay-time-dependent analysis of $B^0 \rightarrow D^{**} D^\mp$ decays is performed using $pp$ collision data collected by the LHCb experiment between 2011 and 2018, corresponding to an integrated luminosity of about 9 fb$^{-1}$. The $D^-$ meson is reconstructed as $K^+\pi^-\pi^-$, while the $D^{**}$ meson is reconstructed as $D^{**} \rightarrow D^0\pi^+$, where the $D^0$ meson final states $K^-\pi^+$ and $K^-\pi^+\pi^-\pi^-$ are considered. In total, about 6,160 signal decays are selected.

Opposite-side and same-side tagging algorithms are used to determine the flavour of the $B^0$ mesons at production, with a total tagging power of 5.6 to 7.1%. The following $CP$ parameters are measured

$$ S_{D^* D} = -0.861 \pm 0.077 \text{ (stat)} \pm 0.019 \text{ (syst)}, $$

$$ \Delta S_{D^* D} = 0.019 \pm 0.075 \text{ (stat)} \pm 0.012 \text{ (syst)}, $$

$$ C_{D^* D} = -0.059 \pm 0.092 \text{ (stat)} \pm 0.020 \text{ (syst)}, $$

$$ \Delta C_{D^* D} = -0.031 \pm 0.092 \text{ (stat)} \pm 0.016 \text{ (syst)}, $$

$$ A_{D^* D} = 0.008 \pm 0.014 \text{ (stat)} \pm 0.006 \text{ (syst)}. $$

The largest statistical correlations are found between the $S_{D^* D}$ and $C_{D^* D}$ parameters and the $\Delta S_{D^* D}$ and $\Delta C_{D^* D}$. They amount to $\rho(S_{D^* D}, C_{D^* D}) = 0.44$ and $\rho(\Delta S_{D^* D}, \Delta C_{D^* D}) = 0.46$, respectively.

This measurement using $B^0 \rightarrow D^{**} D^\mp$ decays excludes the hypothesis of $CP$ conservation at more than 10 standard deviations, obtained using Wilk’s theorem [46]. This result is the most precise single measurement of the $CP$ parameters in $B^0 \rightarrow D^{**} D^\mp$ decays and it is compatible with previous measurements by the Belle [9] and BaBar [8] experiments. The precision of $\Delta C_{D^* D}$ and $C_{D^* D}$ parameters is comparable with that of previous measurements, while for $S_{D^* D}$, $\Delta S_{D^* D}$ and $A_{D^* D}$, this measurement improve significantly the precision of the current world average [3].
Acknowledgments

We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at the LHCb institutes. We acknowledge support from CERN and from the national agencies: CAPES, CNPq, FAPERJ and FINEP (Brazil); MOST and NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); NWO (Netherlands); MNiSW and NCN (Poland); MEN/IFA (Romania); MSHE (Russia); MinECo (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); DOE NP and NSF (U.S.A.). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (Netherlands), PIC (Spain), GridPP (United Kingdom), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland) and OSC (U.S.A.). We are indebted to the communities behind the multiple open-source software packages on which we depend. Individual groups or members have received support from AvH Foundation (Germany); EPLANET, Marie Skłodowska-Curie Actions and ERC (European Union); ANR, Labex P2IO and OCEVU, and Région Auvergne-Rhône-Alpes (France); Key Research Program of Frontier Sciences of CAS, CAS PIFI, and the Thousand Talents Program (China); RFBR, RSF and Yandex LLC (Russia); GVA, XuntaGal and GENCAT (Spain); the Royal Society and the Leverhulme Trust (United Kingdom).

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


T. Skwarnicki, *A study of the radiative cascade transitions between the $\gamma'$ and $\gamma$ resonances*, Ph.D. thesis, Institute of Nuclear Physics, Krakow, Poland (1986) [DESY-F31-86-02].


D. Fazzini, *Flavour tagging in the LHCb experiment*, in the proceedings of the 6th Large Hadron Collider Physics Conference (LHCP 2018), June 4–9, Bologna, Italy (2018).


LHCb collaboration, *Measurement of $C P$ violation in $B^0 \rightarrow D^\mp \pi^\pm$ decays*, *JHEP* 06 (2018) 084 [arXiv:1805.03448] [arXiv].


LHCb collaboration, *Measurements of the $B^+, B^0, B^0_s$ meson and $\Lambda^0_b$ baryon lifetimes*, *JHEP* 04 (2014) 114 [arXiv:1402.2554] [arXiv].

LHCb collaboration, *Measurement of $B^0, B^0_s, B^+$ and $\Lambda^0_b$ production asymmetries in 7 and 8 TeV proton-proton collisions*, *Phys. Lett.* B 774 (2017) 139 [arXiv:1703.08464] [arXiv].

