Measurement of $CP$ asymmetry in $B^0_s \to D^\mp_s K^\pm$ decays

The LHCb collaboration

Abstract
Reported are measurements of the time-dependent $CP$ violating observables in $B^0_s \to D^\pm_s K^\pm$ decays using a dataset corresponding to 3.0 fb$^{-1}$ of $pp$ collisions recorded with the LHCb detector in Run 1 of the LHC. We measure $C_f = 0.735 \pm 0.142 \pm 0.048$, $S_f = -0.518 \pm 0.202 \pm 0.073$, $S_T = -0.496 \pm 0.197 \pm 0.071$, $A^{\Delta \Gamma}_f = 0.395 \pm 0.277 \pm 0.122$, $A^{\Delta \Gamma}_T = 0.314 \pm 0.274 \pm 0.107$, where the uncertainties are statistical and systematic, respectively. These observables are used together with a recent measurement of the $B^0_s$ mixing phase $-2\beta_s$ to obtain a measurement of the CKM angle $\gamma$. 
1 Introduction

Time-dependent analyses of tree-level $B_s^0 \rightarrow D_s^{\mp} \pi^\pm$ and $B_s^0 \rightarrow D_s^{\mp} K^\pm$ decays\(^1\) are sensitive to the Cabibbo-Kobayashi-Maskawa (CKM) unitarity triangle angle $\gamma \equiv \arg(-V_{td}V_{tb}^*/V_{ts}V_{tb})$ through CP violation in the interference of mixing and decay amplitudes $\lambda^3$. A comparison between the value of $\gamma$ measured from tree-level processes with the value of $\gamma$ and other unitarity triangle parameters measured in loop-level processes provides a powerful consistency check of the Standard Model (SM) picture of CP violation.

Due to the interference between mixing and decay amplitudes, the physical CP-violating observables in these decays are functions of a combination of $\lambda^3$ and $\gamma$. Due to the interference between mixing and decay amplitudes, the physical CP-violating observables in these decays are functions of a combination of $\lambda^3$ and $\gamma$. Due to the interference between mixing and decay amplitudes, the physical CP-violating observables in these decays are functions of a combination of $\lambda^3$ and $\gamma$.

The leading-order Feynman diagrams contributing to the interference of decay and mixing in $B_s^0 \rightarrow D_s^{\mp} K^\pm$ are shown in Fig. 1. In contrast to $B^0 \rightarrow D^{(*)\mp} \pi^\pm$ decays, here both the $B_s^0 \rightarrow D_s^+ K^+$ ($b \rightarrow cs\bar{u}$) and $B_s^0 \rightarrow D_s^- K^-$ ($b \rightarrow u\bar{c}s$) amplitudes are of the same order in the sine of the Cabibbo angle $\lambda = 0.2252 \pm 0.0007$\(^1\), and the amplitude ratio of the interfering diagrams is approximately $|V_{ub}V_{cb}/V_{tb}V_{us}| \approx 0.4$. Moreover, the decay-width difference in the $B_s^0$ system, $\Delta\Gamma_s$, is nonzero\(^1\), which, compared to the $B^0$ case, gives access to two additional CP-violating observables which are coefficients of in time-dependent hyperbolical terms. This in turn allows a determination of $\gamma - 2\beta_s$ from the sinusoidal and hyperbolic terms in the decay-time evolution, up to a two-fold ambiguity.

This note presents an updated measurement of the CP-violating observables in $B_s^0 \rightarrow D_s^{\mp} K^\pm$ decays using a data set corresponding to 1.0 (2.0) fb\(^{-1}\) of pp collisions recorded with the LHCb detector at $\sqrt{s} = 7 (8)$ TeV in 2011 (2012), and the most precise determination of $\gamma - 2\beta_s$ in these decays.

Figure 1: Feynman diagrams for $B_s^0 \rightarrow D_s^+ K^-$ without (left) and with (right) $B_s^0$ mixing.

The formalism is unchanged with respect to the previous analysis in Ref. \(^1\) and is

\(^1\)Inclusion of charge-conjugate modes is implied throughout except where explicitly stated.
briefly summarised here for convenience. The decay-time-dependent decay rates of the initially produced flavour eigenstates \(|B_s^0(t = 0)\) and \(|\bar{B}_s^0(t = 0)\) are proportional to
\[
\frac{d\Gamma_{B_s^0 \rightarrow f}(t)}{dt} \propto e^{-\Gamma_s t} \left[ \cosh \left( \frac{\Delta \Gamma_s t}{2} \right) + A_f^\Delta \sinh \left( \frac{\Delta \Gamma_s t}{2} \right) + C_f \cos (\Delta m_s t) - S_f \sin (\Delta m_s t) \right], \tag{1}
\]
\[
\frac{d\Gamma_{\bar{B}_s^0 \rightarrow f}(t)}{dt} \propto e^{-\Gamma_s t} \left[ \cosh \left( \frac{\Delta \Gamma_s t}{2} \right) + A_f^\Delta \sinh \left( \frac{\Delta \Gamma_s t}{2} \right) - C_f \cos (\Delta m_s t) + S_f \sin (\Delta m_s t) \right], \tag{2}
\]
where \(\Gamma_s\) is the average \(B_s^0\) decay width, and \(\Delta \Gamma_s\) is the positive \cite{15} decay-width difference between the heavy and light mass eigenstates in the \(B_s^0\) system. Similar equations can be written for the \(CP\)-conjugate decays replacing \(S_f\) by \(S_f^\dag\), and \(A_f^\Delta\) by \(A_f^{\Delta F}\), while \(C_f = -C_f^\tau\) under the assumption of no \(CP\) violation in either the decay or mixing amplitudes. The \(CP\) observables are related to the magnitude of the amplitude ratio \(r_{D,s} \equiv |\lambda_{D,K}| = |A(\bar{B}_s^0 \rightarrow D_s^- K^+)/A(B_s^0 \rightarrow D_s^- K^+)|\), the strong phase difference between the favoured and suppressed amplitudes \(\delta\), and the weak phase difference \(\gamma - 2\beta_s\), by the following equations:
\[
C_f = \frac{1 - r_{D,s}^2}{1 + r_{D,s}^2},
\]
\[
A_f^{\Delta F} = -\frac{2 r_{D,s} \cos (\delta - (\gamma - 2\beta_s))}{1 + r_{D,s}^2}, \quad A_f^{\Delta F} = \frac{-2 r_{D,s} \cos (\delta + (\gamma - 2\beta_s))}{1 + r_{D,s}^2},
\]
\[
S_f = \frac{2 r_{D,s} \sin (\delta - (\gamma - 2\beta_s))}{1 + r_{D,s}^2}, \quad S_f = \frac{-2 r_{D,s} \sin (\delta + (\gamma - 2\beta_s))}{1 + r_{D,s}^2}. \tag{3}
\]

The sinusoidal \(CP\) observables can only be measured using signal candidates in which the initial flavour of the \(B_s^0\) meson is determined, a process known as “flavour tagging”. By contrast, all signal candidates provide sensitivity to the hyperbolic \(CP\) observables. The analysis strategy largely follows that described in Ref. \cite{14}. The kinematically similar mode \(B_s^0 \rightarrow D_s^- \pi^+\) is used as a control channel which helps in the determination of the decay-time-dependent efficiency and flavour tagging performance. A two-stage likelihood fit is performed: first the signal and background are separated and the so-called \textit{sWeights} \cite{16} are obtained; second, these \textit{sWeights} are used to statistically subtract the background, and the decay-time-dependent \(CP\) observables defined in Eq. 3 are measured by fitting the signal Probability Density Function (PDF) to the decay-time distribution \cite{17}.

2 Detector and software

The LHCb detector \cite{18,19} is a single-arm forward spectrometer covering the pseudorapidity range \(2 < \eta < 5\), designed for the study of 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, 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 placed downstream of
the magnet. The tracking system provides a measurement of momentum, \(p\), of charged
particles with a relative uncertainty that varies from 0.5% at low momentum to 1.0%
at 200 GeV (units in which \(c = \hbar = 1\) are used throughout). The minimum distance of a
track to a primary vertex (PV), the impact parameter, 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. Different types of charged hadrons are distinguished using information from two
ring-imaging Cherenkov detectors. Photons, electrons and hadrons are identified by a
calorimeter system consisting of scintillating-pad and preshower detectors, an electromag-
netic calorimeter and a hadronic calorimeter. Muons are identified by a system composed
of alternating layers of iron and multiwire proportional chambers.

The online event selection is performed by a trigger \[20\], 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. The software trigger requires a two- or
three-track secondary vertex with a large sum of the transverse momenta of the charged
particles and a significant displacement from the primary \(pp\) interaction vertices (PVs).
The secondary vertex must either be classified as being consistent with the decay of a \(b\)
hadron by a multivariate algorithm \[21\], or in the case of a two-track vertex, be identified
as compatible with a \(\phi(1020) \rightarrow K^+K^-\) decay.

Simulations of \(pp\) collisions are generated using \textsc{Pythia} \[22\] with a specific LHCb
configuration \[23\]. Decays of hadronic particles are described by \textsc{EvtGen} \[24\], in
which final-state radiation is generated using \textsc{Photos} \[25\]. The interaction of the
generated particles with the detector, and its response, are implemented using the
\textsc{Geant4} toolkit \[26\] as described in Ref. \[27\].

## 3 Event selection

The event selection forms \(D_s^- \rightarrow K^-K^+\pi^-\), \(D_s^- \rightarrow K^-\pi^+\pi^-\), and \(D_s^- \rightarrow \pi^-\pi^+\pi^-\) candidates from reconstructed charged particles. These \(D_s^-\) candidates are subsequently combined with a fourth particle, referred to as the “companion”, to form \(B_s^0 \rightarrow D_s^\pm K^\mp\) and \(B_s^0 \rightarrow D_s^-\pi^+\) candidates. The decay-time and \(B_s^0\) mass resolutions are improved by
performing a kinematic fit \[28\] in which the \(B_s^0\) candidate is constrained to originate
from its associated PV, and the \(B_s^0\) invariant mass and decay-time are computed with a
constraint on the \(D_s^-\) invariant mass.

The \(B_s^0 \rightarrow D_s^-\pi^+\) mode is used for the optimisation of the selection and for studying
and constraining physics backgrounds to the \(B_s^0 \rightarrow D_s^\pm K^\mp\) decay. The \(B_s^0 \rightarrow D_s^\pm K^\mp\) and \(B_s^0 \rightarrow D_s^-\pi^+\) candidates are required to be matched to the secondary vertex candidates
found in the software trigger. Subsequently, a preselection is applied to the \(B_s^0 \rightarrow D_s^\pm K^\mp\) and \(B_s^0 \rightarrow D_s^-\pi^+\) candidates using a similar multivariate displaced vertex algorithm to
the trigger selection, but with offline-quality reconstruction.

A selection using the gradient boosted decision tree (BDTG) \[29\] implementation in the
\textsc{TMVA} software package \[30\] further suppresses combinatorial backgrounds. The BDTG
is trained on data using the \(B_s^0 \rightarrow D_s^-\pi^+\) decay sample, which is purified with respect to

\[2\] The \(\chi^2\) of the \(B_s^0\) candidate is computed with respect to each PV in the event as the change in the
PV-fit \(\chi^2\), when the \(B_s^0\) is or is not included in the vertex fit. The PV which gives the smallest IP \(\chi^2\) is
associated to the \(B_s^0\) candidate.
the previous preselection exploiting particle identification (PID) information from the Cherenkov detectors. Since all channels in this analysis are kinematically similar, and since no PID information is used as input to the BDTG, the resulting BDTG performs similarly well on the other $D_s^-$ decay modes. The optimal working point is chosen to maximise the expected sensitivity to the $CP$-violating observables in $B^0_s \rightarrow D_s^\mp K^\pm$ decays. In addition, the $B^0_s$ and $D_s^-$ candidates are required to be within $m(B^0_s) \in [5300, 5800] \text{ MeV}/c^2$ and $m(D^-) \in [1930, 2015] \text{ MeV}/c^2$, respectively.

Finally, a combination of PID information and kinematic vetoes is used to distinguish the different $D_s^-$ final states from each other and from cross-feed backgrounds such as $B^0 \rightarrow D^- K^+$ or $\Lambda_b^0 \rightarrow \Lambda^- K^+$. The selection structure and most criteria are identical to those used in Ref. [14]; the specific values of certain PID selection criteria were changed to perform optimally with the latest detector reconstruction.

Fig. 2 shows the relevant mass distributions for candidates passing the PID and BDTG selections. In addition to the previous requirements, a loose PID requirement is imposed on the companion track to complete the selection. After all selection requirements, fewer than 1% of retained events contain more than one signal candidate. All candidates are used in the subsequent analysis.

4 Signal and background shapes

Signal and background are statistically separated with a three-dimensional (“multivariate”) fit. The dimensions are the $B^0_s$ invariant mass, $D_s^-$ invariant mass, and the log-likelihood difference between the kaon and pion hypotheses for the companion ($L(K/\pi)$). The signal and background shapes are obtained using a mixture of data-driven approaches and simulation. The simulated events need to be corrected for kinematic differences between simulation and data, as well as for the efficiency of the PID selection requirements. The decay mode $B^0 \rightarrow D^- \pi^+$, which can be selected with very high purity without the use of any PID requirements and is kinematically very similar to the $B^0_s$ signals, is used to obtain kinematic distributions in data for this weighting. The PID efficiencies are obtained as a function of particle momentum and event occupancy using dedicated data-driven calibration samples of pions, kaons, and protons [31].

In order to model the tails due to radiative signal decays and non-Gaussian reconstruction effects, the signal shape in the $B^0_s$ invariant mass is the sum of two Crystal Ball functions [32] with common mean and power-law tails on both sides. The signal shapes are determined separately for $B^0_s \rightarrow D_s^\mp K^\pm$ and $B^0_s \rightarrow D_s^- \pi^+$ from simulated candidates, and fixed in the multivariate fit with two exceptions. The common mean of the Crystal Ball functions floats for both $B^0_s \rightarrow D_s^- \pi^+$ and $B^0_s \rightarrow D_s^\mp K^\pm$, while a scale factor which accounts for data-simulation differences in the signal width floats in the $B^0_s \rightarrow D_s^- \pi^+$ fit and is subsequently fixed to the measured value in the $B^0_s \rightarrow D_s^\mp K^\pm$ fit. The functional form of the combinatorial background is taken from the upper $B^0_s$ sideband $> 5800 \text{ MeV}/c^2$, with its parameters left free to vary in the subsequent multivariate fit. The combinatorial background is considered independently for each $D_s^-$ mode and parameterised by either an exponential function or by a combination of an exponential and a constant function. The shapes of the fully or partially reconstructed backgrounds are fixed from simulated events using a non-parametric kernel estimation method (KEYS, [33]). Exceptions to this are the $B^0 \rightarrow D^- \pi^+$ background in the $B^0_s \rightarrow D_s^- \pi^+$ fit and the $B^0_s \rightarrow D_s^- \pi^+$ background.
in the $B^0_s \rightarrow D^{\pm}_s K^\mp$ fit, which are described with KEYS shapes obtained from data with the same method used in Ref. [14].

Similarly to the $B^0_s$ invariant mass, the $D_s^-$ invariant mass is a sum of two Crystal Ball functions with common mean and oppositely oriented tails. The signal shapes are determined separately for each $D_s^-$ decay mode from simulated events that have the full selection chain applied to them. As with the $B^0_s$ invariant mass signal shape, only the common mean and the width scale factor float in the fits; the $B^0_s$ and $D_s^-$ scale factors are different. The combinatorial background consists of both random combinations of tracks, which do not peak in the $D_s^-$ invariant mass, and, in some $D_s^-$ decay modes, backgrounds that contain a true $D_s^-$, and a random companion track. It is parameterised separately for each $D_s^-$ decay mode either by an exponential function or by the sum of an exponential
function and the signal $D_s^-$ shape. The fully and partially reconstructed backgrounds which contain a correctly reconstructed $D_s^-$ candidate ($B_s^0 \rightarrow D_s^- K^\pm$ and $B^0 \rightarrow D_s^- \pi^+$ as backgrounds in the $B_s^0 \rightarrow D_s^- \pi^+$ fit; $B^0 \rightarrow D_s^- K^+$ and $B_s^0 \rightarrow D_s^- \pi^+$ as backgrounds in the $B_s^0 \rightarrow D_s^\mp K^\pm$ fit) are assumed to have the same invariant mass distribution as the signal. For other backgrounds, the shapes are KEYS templates taken from simulated events.

The PDFs describing the $L(K/\pi)$ distributions of pions and kaons are obtained from dedicated $D^{*+}$ calibration samples, while those for protons are obtained using a calibration sample of $\Lambda_c^+ \rightarrow pK^-\pi^+$ decays. These samples are weighted to match the signal kinematic and event track multiplicity distributions in the same way as the simulated events. The weighting is done separately for each signal and background component, as well as for each magnet polarity. The shapes for each magnet polarity and year of data taking are described by binned histograms and combined according to the integrated luminosity in each sample. The signal companion $L(K/\pi)$ shape is obtained separately for each $D_s^-$ decay mode to account for small kinematic differences between them. The combinatorial background companion $L(K/\pi)$ shape consists of a mixture of pions, protons, and kaons, whose relative normalisation is left floating in the multivariate fit. The companion $L(K/\pi)$ shape for fully or partially reconstructed backgrounds is obtained by weighting the PID calibration samples to match the event distributions of simulated events, for each background type.

5 Multivariate fit to $B_s^0 \rightarrow D_s^- \pi^+$ and $B_s^0 \rightarrow D_s^\mp K^\pm$

The total PDF for the multivariate fit is built from the product of the PDFs in each of the three fit dimensions, since correlations between the fitting variables are measured to be small in simulation. The product is built for each fit component and $D_s^-$ decay mode, and these products then added to form the final PDF. A systematic uncertainty is assigned for the impact of residual correlations. Almost all background yields are left free to float, except where the expected yield is below 2% of the signal yield. These are $B^0 \rightarrow D^- K^+$, $B^0 \rightarrow D^- \pi^+$, $\Lambda_b^0 \rightarrow \Lambda_c^- K^+$, and $\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+$ for the $B_s^0 \rightarrow D_s^\mp K^\pm$ fit, and $B^0 \rightarrow D^- \pi^+$, $\Lambda_b^0 \rightarrow \Lambda_c^- \pi^+$, and $B_s^0 \rightarrow D_s^\mp K^\pm$ for the $B_s^0 \rightarrow D_s^- \pi^+$ fit. Such background yields are fixed from known branching fractions and relative efficiencies measured using simulated events.

The multivariate fit results in a signal yield of $96942 \pm 345 B_s^0 \rightarrow D_s^- \pi^+$ and $5955 \pm 90 B_s^0 \rightarrow D_s^\mp K^\pm$ decays, where uncertainties are statistical only. The results are shown in Fig. 3 for both $B_s^0 \rightarrow D_s^- \pi^+$ and $B_s^0 \rightarrow D_s^\mp K^\pm$, summed over all $D_s^-$ decay modes. The multivariate fit is validated and found to be unbiased using large samples of data-like pseudo-experiments.

6 Decay-time resolution and acceptance

Due to the fast $B_s^0 - \overline{B_s^0}$ oscillations, any mismodelling of the decay-time resolution potentially biases the CP observables. The signal decay-time PDF is convolved with a Gaussian resolution function that has a different width for each candidate, making use of the per-candidate decay-time uncertainty estimated by the decay-time kinematic fit. The per-candidate decay-time uncertainty $\sigma_t$ is compared to the measured decay-time resolution, and a correction factor is determined. This calibration is performed using prompt $D_s^-$ mesons combined with a random track to give a sample of “fake $B^0_w$” candidates with a known lifetime of zero. The spread of the observed decay times follows the shape of
Figure 3: The multivariate fit to the (left) $B^0_s \rightarrow D^- s \pi^+$ and (right) $B^0_s \rightarrow D^0 K^\pm s$ candidates for all $D^- s$ decay modes combined. From top to bottom: distributions of candidates in $B^0_s$ invariant mass, $D^- s$ invariant mass, absolute value of companion $L(K/\pi)$. The blue solid line represents the sum of the fit components. The dashed red line represents the signal, while the other shaded areas represent different background components.
a double Gaussian distribution, where the negative decay times are used to determine the resolution. The resulting two widths are combined into a single effective resolution, by calculating the corresponding effective dilution, \( D = f_1 e^{-\sigma_1^2 \Delta m_s^2/2} + (1 - f_1) e^{-\sigma_2^2 \Delta m_s^2/2} \), where \( \sigma_{1,2} \) are the widths, and \( f_1 \) and \( (1 - f_1) \) are the fractions of the two Gaussian components. The effective decay-time resolution depends on the per-candidate decay-time uncertainty as \( \sigma(\sigma_i) = 10.3 \text{ fs} + 1.28\sigma_i \). The uncertainty on the decay-time resolution is dominated by the uncertainty on the modelling of the observed decay times of the “fake \( B_s^0 \)” candidates. Modelling the spread by a single Gaussian or by taking only the central Gaussian from the double Gaussian fit, results in the correction \( \sigma(\sigma_i) = 1.77\sigma_i \) and \( \sigma(\sigma_i) = 1.24\sigma_i \), respectively, which are used to estimate the uncertainty on the measured \( CP \) observables.

The measured decay-time resolution on “fake \( B_s^0 \)” candidates can be used for true \( B_s^0 \) candidates, as the measured decay-time resolution does not significantly depend on the transverse momentum of the companion particle, which is the main kinematic difference between the samples. In addition, simulation shows that the “fake \( B_s^0 \)” and signal \( B_s^0 \) samples have similar correction factors, varying in the range 1.19 – 1.27.

The decay-time acceptance of \( B_s^0 \rightarrow D^{\pm}_s K^{\mp} \) candidates cannot be left free in the fit because its shape is heavily correlated with the \( CP \) observables. However, in the case of \( B_s^0 \rightarrow D^-_s \pi^+ \), the acceptance can be measured by fixing \( \Gamma_s \) and leaving the acceptance parameters free in the fit. The decay-time acceptance in the \( B_s^0 \rightarrow D^{\pm}_s K^{\mp} \) fit is fixed to that found in the \( B_s^0 \rightarrow D^-_s \pi^+ \) data fit, after correcting for differences between the two decay modes using fully simulated events, the latter being weighted to match the kinematic distributions observed in data. The mixture of \( D^-_s \) decay modes in these simulated samples is weighted to match the yields observed in data.

In all cases, the acceptance is described using segments of smooth polynomial functions (“splines”), which can be implemented in an analytic way in the decay-time fit [34]. The spline boundaries (“knots”) are chosen in an ad hoc fashion to model reliably the features of the acceptance shape, and placed at 0.5, 1.0, 1.5, 2.0, 3.0, 12.0 ps. The decay-time fit to the \( B_s^0 \rightarrow D^-_s \pi^+ \) data is an \textit{sWeighted} fit using the signal PDF from Sec. 4, with \( C_f \) fixed to 1, \( S_f, S_T, A_f^{\Delta T}, \) and \( A_T^{\Delta T} \) all fixed to zero, and the knot magnitudes and \( \Delta m_s \) left free. The decay-time fit to the \( B_s^0 \rightarrow D^-_s \pi^+ \) data together with the measured decay-time acceptance is shown in Fig. 4.

7 Flavour Tagging

The identification of the initial flavour of the \( B_s^0 \) is performed by means of two flavour-tagging algorithms, referred to as the opposite-side (OS) and same-side (SS) taggers [35,36]. While the former exploits the pair-wise production of \( b \) quarks, the latter makes use of the hadronisation remnants of the signal quark. The tagging algorithms use a neural network trained on simulated events (SS tagger) or data (OS tagger) to compute the probability \( \eta \) to assign an incorrect flavour, or “mistag”, to the \( B_s^0 \) candidate. This estimated mistag \( \eta \) is treated as a per-candidate variable, thus adding an observable to the fit. Due to variations in the properties of tagging tracks for different channels and to data-MC discrepancies, the predicted mistag probability \( \eta \) requires to be calibrated using flavour specific, and therefore self-tagging, decays to represent the true mistag rate \( \omega \). The statistical uncertainty on \( C_f, S_f, \) and \( S_T \) scales with \( 1/\sqrt{\varepsilon_{\text{eff}}} \), defined as \( \varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1-2\omega)^2 \).
Figure 4: Result of the fit to the decay-time distribution of $B^0_s \rightarrow D_s^- \pi^+$ candidates, which is used to measure the decay-time acceptance in $B^0_s \rightarrow D_s^\mp K^\pm$ decays. The red solid curve is the measured decay-time acceptance.

Table 1: Calibration parameters of the OS and SS taggers determined from $B^0_s \rightarrow D_s^- \pi^+$ decays. For $p_0$ and $p_1$ the first listed uncertainty is statistical and the second systematic; other parameters are not used in the fit but provided for reference only and their uncertainties, where listed, are statistical. For a perfectly calibrated tagger one expects $p_1 = 1$ and $p_0 - \langle \eta \rangle = 0$.

<table>
<thead>
<tr>
<th>Tagger</th>
<th>$p_0$</th>
<th>$p_1$</th>
<th>$\langle \eta \rangle$</th>
<th>$\varepsilon_{\text{tag}}$ [%]</th>
<th>$\varepsilon_{\text{eff}}$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>$0.377 \pm 0.007 \pm 0.001$</td>
<td>$1.12 \pm 0.08 \pm 0.01$</td>
<td>$0.370$</td>
<td>$37.15 \pm 0.17$</td>
<td>$3.55 \pm 0.33$</td>
</tr>
<tr>
<td>SS</td>
<td>$0.441 \pm 0.005 \pm 0.000$</td>
<td>$1.09 \pm 0.08 \pm 0.01$</td>
<td>$0.437$</td>
<td>$63.93 \pm 0.17$</td>
<td>$1.92 \pm 0.22$</td>
</tr>
</tbody>
</table>

Therefore, the tagging algorithms are tuned for maximum effective tagging power $\varepsilon_{\text{eff}}$.

Both the OS and SS taggers are calibrated using a decay-time-dependent fit to the $B^0_s \rightarrow D_s^- \pi^+$ control channel in which $\Delta m_s$ is fixed to its world-average value. A linear model is used as a calibration function,

\[
\omega(\eta) = p_0 + p_1 \cdot (\eta - \langle \eta \rangle),
\]

where $\langle \eta \rangle$ is fixed to the average of the estimated mistag probability $\eta$. The use of $B^0_s \rightarrow D_s^- \pi^+$ to calibrate both taggers maximizes the portability of this calibration to $B^0_s \rightarrow D_s^\mp K^\pm$ in the sense of minimized associated systematic uncertainties. The compatibility between the calibrations in $B^0_s \rightarrow D_s^- \pi^+$ and $B^0_s \rightarrow D_s^\mp K^\pm$ decays is verified using simulated events. The measured tagging parameters are listed in Tab. 1. Fig. 5 shows the linear dependence between the measured mistag $\omega$ and the estimated mistag probability for the OS and SS taggers. The tagging calibration can depend on the initial $B_s^0$ flavour due to the different interaction cross-sections of $K^+$ and $K^-$ with matter.
These asymmetries have been measured in previous analyses \cite{36}, and a cross-check using the $B^0_s \rightarrow D^- \pi^+$ control mode gives consistent results. In the nominal $B^0_s \rightarrow D^\pm_s K^\mp$ decay-time fit the tagging asymmetries are set to zero and influence of this choice is evaluated by means of pseudo-experiments.

Two types of systematic uncertainties affect the tagging calibration. The dominant systematic uncertainty comes from the uncertainty on the decay-time resolution used in the fit to $B^0_s \rightarrow D^- \pi^+$ data. In order to avoid double counting systematic uncertainties from the resolution on the $CP$ observables we exclude the systematic uncertainty of the resolution when constraining the tagging calibration parameters in the $B^0_s \rightarrow D^\pm_s K^\mp$ decay-time fit. Instead, the tagging calibration is redone varying the decay-time resolution, and subsequently the combination of the varied decay-time resolution and varied tagging calibration is applied to the $B^0_s \rightarrow D^- \pi^+$ data. This combined systematic uncertainty is discussed further in Sec. 9 and is not part of the systematic uncertainties listed in Tab. 1.

Other systematic uncertainties are evaluated for changes in the background PDFs, different calibrations for 2011 and 2012 data sets, varying the decay-time acceptance and other fixed parameters in the $B^0_s \rightarrow D^- \pi^+$ fit. In addition, a binned fit is performed where instead of using $\eta$ as a per-event observable, the sample is split in bins of predicted $\eta$ and the average calibrated mistag is measured in each bin. The sum in quadrature of all these systematic uncertainties is an order of magnitude smaller than either the statistical uncertainty or the dependence on the decay-time resolution. The overall tagging performance for the OS and SS combination is reported in Tab. 2 considering three categories of tagged candidates: OS only, SS only and both OS and SS.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Measured mistag rate against the average predicted mistag rate for the (left) OS and (right) SS taggers in $B^0_s \rightarrow D^- \pi^+$ decays. The error bars represent only the statistical uncertainties. The solid curve is the linear fit to the data points, the shaded area defines the 95\% confidence level region of the calibration function (statistical only).}
\end{figure}
Table 2: The Flavour Tagging performances for only OS tagged, only SS tagged and both OS and SS tagged events for $B^0_s \rightarrow D^-_s \pi^+$, where uncertainties are statistical.

<table>
<thead>
<tr>
<th></th>
<th>$\varepsilon_{\text{tag}}$ [%]</th>
<th>$\varepsilon_{\text{eff}}$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS only</td>
<td>11.95 ± 0.11</td>
<td>1.28 ± 0.12</td>
</tr>
<tr>
<td>SS only</td>
<td>34.47 ± 0.15</td>
<td>1.19 ± 0.14</td>
</tr>
<tr>
<td>both OS-SS</td>
<td>19.27 ± 0.13</td>
<td>2.51 ± 0.18</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>65.69 ± 0.23</td>
<td>4.98 ± 0.26</td>
</tr>
</tbody>
</table>

8 Decays-time fit to $B^0_s \rightarrow D^+_s K^\pm$

An unbinned maximum likelihood fit is performed to measure the $CP$ observables defined in Eq. 3 using the $sWeights$ obtained in the multivariate fit to statistically subtract the backgrounds. The following physics parameters are fixed in the fit

\[
\Delta m_s = (17.757 \pm 0.021) \text{ps}^{-1},
\]

\[
\Gamma_s = (0.6643 \pm 0.0020) \text{ps}^{-1},
\]

\[
\Delta \Gamma_s = (0.083 \pm 0.006) \text{ps}^{-1},
\]

\[
\rho(\Gamma_s, \Delta \Gamma_s) = -0.239,
\]

\[
A_{\text{prod}}(B^0_s) = (1.1 \pm 2.7)\%,
\]

\[
A_{\text{det}}(K\pi) = (1 \pm 1)\%.
\]

based on HFAG \cite{37} world average of $B^0_s$ oscillation frequency and decay widths, and LHCb measurements of the $B^0_s$ production asymmetry \cite{38} and the $K^+\pi^-$ detection asymmetry \cite{39}. The signal production asymmetry is fixed to zero because the fast $B^0_s$ oscillations wash out any initial asymmetry and make its effect on the $CP$ observables negligible. The tagging calibration parameters are constrained to the values obtained from the control channel. The decay-time PDF is convolved with a single Gaussian representing the per-candidate decay-time resolution, and multiplied by the decay-time acceptance described in Sec. \cite{6}.

The measured $CP$-violating observables are given in Tab. \cite{3} and the correlations of their statistical uncertainties are given in Tab. \cite{4}. The fit to the decay-time distribution is shown in Fig. \cite{6} together with the folded asymmetry plots for $D^+_s K^-$ and $D^-_s K^+$ final states. The folded asymmetry plots show the difference in the rates of $B^0_s$ and $\bar{B}^0_s$ tagged $D^+_s K^-$ and $D^-_s K^+$ candidates, plotted in slices of $2\pi/\Delta m_s$, where the $sWeights$ obtained with the multivariate fit have been used to subtract the background. The central values of the $CP$ observables are used to define the plotted asymmetry, while the normalization is floated and found to be in good agreement with the expected dilution due to mistag and decay-time resolution.

9 Systematic uncertainties

Systematic uncertainties arise from the fixed parameters $\Delta m_s$, $\Gamma_s$, and $\Delta \Gamma_s$, and from the limited knowledge of the detection asymmetry, decay-time resolution, and acceptance. These uncertainties are estimated using large sets of simulated pseudoexperiments, which are generated using the central values of the above fixed parameters and later fitted.
Table 3: Fitted values of the CP observables to the $B_0^s \to D^\pm K^\mp$ decay-time distribution where the first uncertainty is statistical, the second is systematic.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_f$</td>
<td>$0.735 \pm 0.142 \pm 0.048$</td>
</tr>
<tr>
<td>$A_f^{\Delta \Gamma}$</td>
<td>$0.395 \pm 0.277 \pm 0.122$</td>
</tr>
<tr>
<td>$A_f^{\Delta \Gamma}$</td>
<td>$0.314 \pm 0.274 \pm 0.107$</td>
</tr>
<tr>
<td>$S_f$</td>
<td>$-0.518 \pm 0.202 \pm 0.073$</td>
</tr>
<tr>
<td>$S_f$</td>
<td>$-0.496 \pm 0.197 \pm 0.071$</td>
</tr>
</tbody>
</table>

Table 4: Statistical correlation matrix of the CP parameters. Other fit parameters have negligible correlations with the CP parameters and are omitted for brevity.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$C_f$</th>
<th>$A_f^{\Delta \Gamma}$</th>
<th>$A_f^{\Delta \Gamma}$</th>
<th>$S_f$</th>
<th>$S_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_f$</td>
<td>1.00</td>
<td>0.09</td>
<td>0.08</td>
<td>0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td>$A_f^{\Delta \Gamma}$</td>
<td>0.09</td>
<td>1.00</td>
<td>0.51</td>
<td>-0.07</td>
<td>-0.01</td>
</tr>
<tr>
<td>$A_f^{\Delta \Gamma}$</td>
<td>0.08</td>
<td>0.51</td>
<td>1.00</td>
<td>-0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>$S_f$</td>
<td>0.01</td>
<td>-0.07</td>
<td>-0.03</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$S_f$</td>
<td>-0.06</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

varying these fixed parameters within their uncertainties. They are processed by both the multivariate fit in order to obtain the sWeights, and subsequently by the decay-time fit. Before they are used to evaluate systematic uncertainties, the pseudoexperiments are processed with the nominal fit configuration, and the fitted observables are found to be unbiased for both the multivariate and decay-time fits. Furthermore, when these fitted observables are used to obtain $\gamma$, the result is found to be unbiased with good coverage.

The fitted values of the observables are subsequently compared between the nominal fit and the systematic fit, in which each fixed parameter is varied according to its uncertainty. A distribution is formed by normalising the resulting differences to the uncertainties measured in the nominal fit, and the mean and width of this distribution are added in quadrature and assigned as the total systematic uncertainty. These pseudoexperiments are also used to compute the systematic covariance matrix due to each source of uncertainty.

The total systematic covariance matrix is obtained by adding the individual covariance matrices. The resulting systematic uncertainties and their correlations are shown in Tab. 5 and 6. The contributions from $\Gamma_s$, $\Delta \Gamma_s$, and the acceptance are listed independently for comparison to convey a feeling for their relative importance. When computing the total, however, the correlations between these two, as well as between them and the acceptance parameters, are accounted for, and the full systematic uncertainty which enters into the total is listed as “acceptance, $\Gamma_s$, $\Delta \Gamma_s$”.

The result is cross-checked by splitting the sample into subsets according to the two magnet polarities, the year of data taking, the $B_0^s$ momentum, and the BDTG response. In addition, simulated signal and background events are fitted in order to check for systematic effects due to neglecting correlations between the different variables in the signal and background PDFs. No bias is found, and only the sensitivity of the closure test performed to the fully simulated signal events is assigned as a systematic uncertainty.

A number of other possible systematic effects are studied, but found to be negligible. These include possible production asymmetries, missing or imperfectly modelled back-
Figure 6: Result of the decay-time fit to the $B_s^0 \rightarrow D_s^0 K^\pm$ candidates. The folded asymmetry plots for (bottom left) $D_s^+ K^-$, and (bottom right) $D_s^- K^+$ are also shown.

grounds, and fixed signal shape parameters in the multivariate fit. Potential systematic effects due to fixed background yields are evaluated by generating pseudoexperiments with the nominal value for these yields, and fitting back with the yields fixed to twice or half their nominal value. No significant bias is observed and no systematic uncertainty assigned. The decay-time fit is repeated adding one or two additional spline functions to the decay-time acceptance description and no significant change in the fit result is observed. Randomly removing multiple candidates (instead of keeping them all) the multivariate and decay-time fits are repeated without observing a significant change in the fit result. The asymmetries of the tagging are neglected in the nominal fit. The impact of their uncertainties on the CP parameters is evaluated by fitting the nominal pseudoexperiments with the tagging asymmetries parameters fixed to random parameters generated according to the covariance matrices. The results are compared to the nominal fit results, no significant deviation is observed and no systematic is assigned. No systematic
Table 5: Total systematic uncertainties, relative to the statistical uncertainty. †The daggered contributions \((\Gamma_s, \Delta \Gamma_s, \text{acceptance})\) are shown separately only for comparison. The phrase “MC ratio” refers to the ratio of \(B^0_s \to D^\mp_s K^\pm\) and \(B^0_s \to D_s^-\pi^+\) decay-time acceptances measured in simulated events.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(C_f)</th>
<th>(A^{\Delta \Gamma}_f)</th>
<th>(A^{\Delta \Gamma}_S)</th>
<th>(S_f)</th>
<th>(S_T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection asymmetry</td>
<td>0.01</td>
<td>0.23</td>
<td>0.26</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>(\Delta m_s)</td>
<td>0.06</td>
<td>0.01</td>
<td>0.01</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>Tagging and scale factor</td>
<td>0.15</td>
<td>0.06</td>
<td>0.06</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>Correlation among observables</td>
<td>0.27</td>
<td>0.25</td>
<td>0.18</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td>Closure test</td>
<td>0.12</td>
<td>0.19</td>
<td>0.19</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>(\Gamma^\dagger_s)</td>
<td>0.02</td>
<td>0.16</td>
<td>0.18</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>(\Delta \Gamma^\dagger_s)</td>
<td>0.01</td>
<td>0.07</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Acceptance, MC ratio†</td>
<td>0.04</td>
<td>0.09</td>
<td>0.10</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Acceptance, data fit†</td>
<td>0.07</td>
<td>0.18</td>
<td>0.20</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Acceptance, (\Gamma_s, \Delta \Gamma_s)</td>
<td>0.07</td>
<td>0.19</td>
<td>0.06</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.34</td>
<td>0.44</td>
<td>0.39</td>
<td>0.36</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 6: Total systematic correlations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(C_f)</th>
<th>(A^{\Delta \Gamma}_f)</th>
<th>(A^{\Delta \Gamma}_S)</th>
<th>(S_f)</th>
<th>(S_T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C_f)</td>
<td>1.00</td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>(A^{\Delta \Gamma}_f)</td>
<td>0.02</td>
<td>1.00</td>
<td>-0.34</td>
<td>0.04</td>
<td>-0.01</td>
</tr>
<tr>
<td>(A^{\Delta \Gamma}_S)</td>
<td>0.06</td>
<td>-0.34</td>
<td>1.00</td>
<td>-0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>(S_f)</td>
<td>0.02</td>
<td>0.04</td>
<td>-0.03</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>(S_T)</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

uncertainty is attributed to the imperfect knowledge of the momentum and longitudinal scale of the detector since both effects are taken into account by the systematic uncertainty in \(\Delta m_s\).

10 Interpretation and conclusion

The measurement of the \(CP\)-sensitive parameters is interpreted in terms of \(\gamma - 2\beta_s\) and subsequently \(\gamma\) following the same procedure as in Refs. [14, 40]. The value of \(\beta_s\) is constrained to the LHCb measurement from \(B^0_s \to J/\psi K^+K^-\) and \(B^0_s \to J/\psi\pi^+\pi^-\) decays, \(\phi_s = (-0.01 \pm 0.039)\) rad [13]. Neglecting penguin pollution and assuming no BSM contribution in these decays, \(\phi_s = -2\beta_s\). The resulting confidence intervals are,

\[
\gamma = (127^{+17}_{-22})^\circ, \quad \delta = (358^{+15}_{-16})^\circ, \quad r_{D_s K} = 0.37^{+0.10}_{-0.09}, \quad (68.3\%\text{CL}) \quad (11)
\]

\[
\gamma = (127^{+33}_{-50})^\circ, \quad \delta = (358^{+31}_{-33})^\circ, \quad r_{D_s K} = 0.37^{+0.19}_{-0.19}, \quad (95.4\%\text{CL}) \quad (12)
\]

where the intervals for the angles are expressed modulo 180°. The quoted uncertainties are the sum in quadrature of the statistical and systematic components. A compatibility of 2.15\(\sigma\) is found between the presented measurement and the LHCb \(\gamma\) average [40] computed excluding the previous \(B^0_s \to D^\pm_s K^\pm\) measurement (Ref. [14]). The fit probability is 95.1\%,
indicating excellent internal consistency between the measured CP observables. A non-zero value of $\gamma - 2\beta_s$ would constitute evidence for CP violation in $B^0_s \to D^\pm_s K^\pm$. We test for this by repeating the frequentist fit to the measured CP observables, including both statistical and systematic uncertainties, but force $\gamma - 2\beta_s$ to be equal to 0. The observed change in the fit log-likelihood indicates $3.6\sigma$ evidence for CP violation in $B^0_s \to D^\pm_s K^\mp$. Fig. 7 shows the measured CP-violating observables in the complex plane. Fig. 8 shows the $1-\text{CL}$ curve for $\gamma$, $\delta$, and $r_{D_sK}$, as well as the two-dimensional contours of the profile likelihood.

Figure 7: The CP-violating observables for the (left) 1 fb$^{-1}$ and (right) this analysis. Correlations between observables are ignored when making this plot.
Figure 8: Graph showing $1 - \text{CL}$ for $\gamma$, together with the central value and the 68.3% CL interval as obtained from the frequentist method described in the text (top). Graph showing $1 - \text{CL}$ for $\delta$ (middle left) and $r_{D,K}$ (middle right). Profile likelihood contours of $r_{D,K}$ vs. $\gamma$ (bottom left), and $\delta$ vs. $\gamma$ (bottom right). The contours are the 1σ (2σ) profile likelihood contours, where $\Delta \chi^2 = 1$ ($\Delta \chi^2 = 4$), corresponding to 39% CL (86% CL) in Gaussian approximation. The markers denote the best-fit values.
References


[8] Belle collaboration, F. J. Ronga et al., *Measurement of CP violation in \(B^0 \to D^{*}\pi^\pm\) and \(B^0 \to D^\pm\rho^\pm\) decays*, Phys. Rev. **D84** (2011) 021101, [arXiv:1102.0888](https://arxiv.org/abs/1102.0888).


