A precise measurement of the $B^0$ meson oscillation frequency
A precise measurement of the $B^0$ meson oscillation frequency

LHCb Collaboration

CERN, 1211 Geneva 23, Switzerland

Received: 13 April 2016 / Accepted: 4 July 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract  The oscillation frequency, $\Delta m_d$, of $B^0$ mesons is measured using semileptonic decays with a $D^-$ or $D^{*-}$ meson in the final state. The data sample corresponds to 3.0 fb$^{-1}$ of $p p$ collisions, collected by the LHCb experiment at centre-of-mass energies $\sqrt{s} = 7$ and 8 TeV. A combination of the two decay modes gives $\Delta m_d = (505.0 \pm 2.1 \pm 1.0)$ ns$^{-1}$, where the first uncertainty is statistical and the second is systematic. This is the most precise single measurement of this parameter. It is consistent with the current world average and has similar precision.

1 Introduction

Flavour oscillation, or mixing, of neutral meson systems gives mass eigenstates that are different from flavour eigenstates. In the $B^0$–$\bar{B}^0$ system, the mass difference between mass eigenstates, $\Delta m_d$, is directly related to the square of the product of the CKM matrix elements $V_{td}$ and $V_{tb}^*$, and is therefore sensitive to fundamental parameters of the Standard Model, as well as to non-perturbative strong-interaction effects and the square of the top quark mass [1]. Measurements of mixing of neutral $B$ mesons were published for the first time by UA1 [2] and ARGUS [3]. Measurements of $B^0$–$\bar{B}^0$ mixing have been performed by CLEO [4], experiments at LEP and SLC [5], experiments at the Tevatron [6,7], the $B$ Factories experiments [8,9] and, most recently, at LHCb [10–12]. The combined world average value for the mass difference, $\Delta m_d = (510 \pm 3)$ ns$^{-1}$, has a relative precision of 0.6% [13]. This paper reports a measurement of $\Delta m_d$ based on $B^0 \to D^-(\mu^+\nu_\mu)X$ and $B^0 \to D^{*-}(\mu^-\bar{\nu}_\mu)X$ decays, where $X$ indicates any additional particles that are not reconstructed. The data sample used for this measurement was collected at LHCb during LHC Run 1 at $\sqrt{s} = 7(8)$ TeV in 2011 (2012), corresponding to integrated luminosities of 1.0 (2.0) fb$^{-1}$.

The relatively high branching fraction for semileptonic decays of $B^0$ mesons, along with the highly efficient lepton identification and flavour tagging capabilities at LHCb, results in abundant samples of $B^0 \to D^{(*)}\mu^+\nu_\mu X$ decays, where the flavour of the $B^0$ meson at the time of production and decay can be inferred. In addition, the decay time $t$ of $B^0$ mesons can be determined with adequate resolution, even though the decay is not fully reconstructed, because of the potential presence of undetected particles. It is therefore possible to precisely measure $\Delta m_d$ as the frequency of matter-antimatter oscillations in a time-dependent analysis of the decay rates of unmixed and mixed events,

$$N^{\text{unmix}}(t) \equiv N(B^0 \to D^{(*)}\mu^+\nu_\mu X)(t) \propto e^{-\Gamma_d t} \times [1 + \cos(\Delta m_d t)] ,$$

$$N^{\text{mix}}(t) \equiv N(B^0 \to \bar{B}^0 \to D^{(*)}\mu^-\bar{\nu}_\mu X)(t) \propto e^{-\Gamma_d t} \times [1 - \cos(\Delta m_d t)] ,$$

where the state assignment is based on the flavours of the $B^0$ meson at production and decay, which may be the same (unmixed) or opposite (mixed). In Eq. 1, $\Gamma_d = 1/\tau_{B^0}$ is the decay width of the $B^0$ meson, $\tau_{B^0}$ being its lifetime. Also, in Eq. 1 the difference in the decay widths of the mass eigenstates, $\Delta \Gamma_d$, and $CP$ violation in mixing are neglected, due to their negligible impact on the results. The flavour asymmetry between unmixed and mixed events is

$$A(t) = \frac{N^{\text{unmix}}(t) - N^{\text{mix}}(t)}{N^{\text{unmix}}(t) + N^{\text{mix}}(t)} = \cos(\Delta m_d t) .$$

A description of the LHCb detector and the datasets used in this measurement is given in Sect. 2. Section 3 presents the selection criteria, the flavour tagging algorithms, and the method chosen to reconstruct the $B^0$ decay time. The fitting strategy and results are described in Sect. 4. A summary of the systematic uncertainties is given in Sect. 5, and conclusions are reported in Sect. 6.
2 Detector and simulation

The LHCb detector [14,15] 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/$c$. The minimum distance of a track to a primary 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 (RICH) detectors. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic 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 [16], 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. Candidate events are first required to pass the hardware trigger, which selects muons with a transverse momentum $p_T > 1.48 \text{ GeV}/c$ in the 7 TeV data or $p_T > 1.76 \text{ GeV}/c$ in the 8 TeV data. The software trigger requires a two-, three- or four-track secondary vertex, where one of the tracks is identified as a muon, with a significant displacement from the primary $pp$ interaction vertices. At least one charged particle must have a transverse momentum $p_T > 1.7 \text{ GeV}/c$ and be inconsistent with originating from a PV. As it will be explained later, the software trigger introduction introduces a bias on the $\Delta m_d$ measurement, which is corrected for. A multivariate algorithm [17] is used for the identification of secondary vertices consistent with the decay of a $b$ hadron.

The method chosen to reconstruct the $B^0$ decay time relies on Monte Carlo simulation. Simulation is also used to estimate the main background sources and to verify the fit model. In the simulation, $pp$ collisions are generated using PYTHIA [18,19] with a specific LHCb configuration [20]. Decays of hadronic particles are described by EVTGEN [21], in which final-state radiation is generated using PHOTOS [22]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [23,24] as described in Ref. [25]. Large samples of mixtures of semileptonic decays resulting in a $D^-$ or a $D^{*-}$ meson in the final state were simulated and the assumptions used to build these samples are assessed in the evaluation of systematic uncertainties.

3 Event selection

For charged particles used to reconstruct signal candidates, requirements are imposed on track quality, momentum, transverse momentum, and impact parameter with respect to any PV. Tracks are required to be identified as muons, kaons or pions. The charm mesons are reconstructed through the $D^+ \to K^+\pi^+\pi^-$ decay, or through the $D^{*-} \to \bar{D}^0\pi^-$, $\bar{D}^0 \to K^+\pi^-$ decay chain. The masses of the reconstructed $D^-$ and $\bar{D}^0$ mesons should be within 70 MeV/$c^2$ and 40 MeV/$c^2$ of their known values [13], while the mass difference between the reconstructed $D^{*-}$ and $\bar{D}^0$ mesons should lie between 140 MeV/$c^2$ and 155 MeV/$c^2$. For $D^-$ and $\bar{D}^0$ candidates, the scalar sum of the $p_T$ of the daughter tracks should be above 1800 MeV/$c$. A good quality vertex fit is required for the $D^-$, $\bar{D}^0$, and $D^{*-}$ candidates, and for the $D^{(s)*}\mu^+$ combinations. When more than one combination is found in an event, the one with the smallest vertex $\chi^2$ (hereafter referred to as the $B$ candidate) is chosen. The reconstructed vertices of $D^-$, $\bar{D}^0$, and $B$ candidates are required to be significantly displaced from their associated PV, where the associated PV is that which has the smallest $\chi^2$ increase when adding the candidate. For $D^-$ and $\bar{D}^0$ candidates, a large IP with respect to the associated PV is required in order to suppress charm mesons promptly produced in $pp$ collisions. The momentum of the $B$ candidate, and its flight direction measured using the PV and the $B$ vertex positions, are required to be aligned. These selection criteria reduce to the per-mille level or lower the contribution of $D^{(*)}\pi^-\pi^0$ decays where the charmed meson originates from the PV. The invariant mass of the $B$ candidate is required to be in the range $[3.0, 5.2] \text{ GeV}/c^2$.

Backgrounds from $B \to J/\psi X$ decays, where one of the muons from the $J/\psi \to \mu^+\mu^-$ decay is correctly identified and the other misidentified as a pion and used to reconstruct a $D^{(*)}\pi^-$, are suppressed by applying a veto around the $J/\psi$ mass. Similarly, a veto around the $\Lambda_c^+\pi^-$ mass is applied to suppress semileptonic decays of the $\Lambda_c^0$ baryon, in which the proton of the subsequent $\Lambda_c^+\pi^-$ decay into $pK^-\pi^+$ is misidentified as a pion.

The dominant background is due to $B^+ \to D^{(*)}\mu^+\nu_\mu X$ decays, where additional particles coming from the decay of higher charm resonances, or from multi-body decays of $B^+$ mesons, are neglected. The fractions of $B^+$ decays in the $D^-$ and $D^{*-}$ samples are expected to be 13 and 10%, based on the branching fractions of signal and background, with uncertainties at the 10% level. This background is reduced by using a multivariate discriminant based on a boosted decision.
tree (BDT) algorithm [26, 27], which exploits information on the B candidate, kinematics of the higher charm resonances and isolation criteria for tracks and composite candidates in the B decay chain. Training of the BDT classifier is carried out using simulation samples of $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ signal and $B^+ \rightarrow D^*+ \mu^+ \nu_\mu X$ background. The variables used as input for the BDT classifier are described in the Appendix. Only candidates with BDT output larger than $-0.12 \, (\sim -0.16)$ are selected in the 2011 (2012) data sample for the $B^0 \rightarrow D^{-} \mu^+ \nu_\mu X$ mode. The BDT output is required to be larger than $-0.3$ in both 2011 and 2012 data samples for the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ mode. The impact of this requirement on signal efficiency and background retention can be seen in Fig. 3. The background from $B^\pm$ decays is reduced by 70% in both modes. Combinatorial background is evaluated by using reconstructed candidates in the $D^{(*)-}$ signal mass sidebands. Backgrounds due to decays of $B^0_s$ and $\Lambda_b^0$ into similar final states to those of the signal are studied through simulations.

The BDT output is defined as $t = (M_{BDT} \cdot L)/(p_{rec} \cdot c/k)$, where $M_{BDT}$ is the mass of the $B^0$, taken from Ref. [13], $L$ is the measured decay length and $p_{rec}$ is the magnitude of the visible momentum, measured from the $D^{(*)-}$ meson and the muon. The correction factor $k$ is determined from simulation by dividing the visible $B^0$ momentum by its true value and taking the average, $k = (p_{rec}/p_{true})$. This correction represents the dominant source of uncertainty in the determination of the decay time of the $B^0$ meson for $t > 1.5 \text{ ps}$. Since the $k$-factor depends strongly on the decay kinematics, it is parametrised by a fourth-order polynomial as a function of the visible mass of the $B^0$ candidate.

The $B^0$ flavour at production is determined by using information from the other $b$ hadron present in the event. The decision of flavour tagging algorithms [28] based on the charge of leptons, kaons and of an inclusively reconstructed detached vertex, is used for the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ channel. In the $B^0 \rightarrow D^{-} \mu^+ \nu_\mu X$ channel, which is subject to a larger $B^\pm$ background contamination, the decision of the tagging algorithm based on the detached vertex is excluded in order to avoid spurious background asymmetries. The statistical uncertainty on $\Delta m_d$ decreases as $T^{-1/2}$ where the tagging power is defined as $T = \varepsilon_l (1 - 2 \omega)^2$, where $\varepsilon_l$ is the tagging efficiency and $\omega$ is the mistag rate. To increase the statistical precision, the events are grouped into four tagging categories of predicted mistag probability $\eta$, defined by $\eta \in [0, 0.25], [0.25, 0.33], [0.33, 0.41], [0.41, 0.47]$. The mistag probability $\eta$ is evaluated for each $B$ candidate from event and taggers properties and was calibrated on data using control samples [28]. The average mistag rates for signal and background are taken as free parameters when fitting for $\Delta m_d$. The combined tagging power [28] for the $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ mode is $(2.38 \pm 0.05)\%$ and $(2.46 \pm 0.04)\%$ in 2011 and 2012. For the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ mode, the tagging power in 2011 and 2012 is $(2.55 \pm 0.07)\%$ and $(2.32 \pm 0.04)\%$.

4 Fit strategy and results

The fit proceeds as follows. First, $D^{(*)-}$ mesons originating from semileptonic $B^0$ or $B^+$ decays are separated from the background coming from combinations of tracks not associated to a charm meson decay, by a fit to the invariant mass distributions of the selected candidates. This fit assigns to each event a covariance-weighted quantity $sWeight$, which is used in the subsequent fits to subtract statistically the contribution of the background by means of the $sPlot$ procedure [29]. Then, the contribution of $D^{(*)-}$ from $B^\pm$ decays is determined in a fit to the distributions of the BDT classifier output weighted by signal $sWeights$. Next, a cut is applied on the BDT output in order to suppress the $B^+$ background, the mass distributions are fitted again, and new $sWeights$ are determined. Finally, the oscillation frequency $\Delta m_d$ is determined by a fit to the decay time distribution of unmixed and mixed candidates, weighted for the signal $sWeights$ determined in the previous step.

An extended binned maximum likelihood fit to the data distributions is performed for each stage, simultaneously for the four tagging categories defined above. Data samples collected in 2011 and 2012 are treated separately.

Figure 1 shows the results of the fits to the $D^-$ candidate mass distributions for $B^0 \rightarrow D^{-} \mu^+ \nu_\mu X$ candidates. In these fits, the distributions of $D^-$ from $B^0$ and $B^+$ decays are summed as they are described by the same probability density function (PDF): the sum of two Gaussian functions and a Crystal Ball function [30]. The yields corresponding to the $D^- \pi^-$ peak are $(5.30 \pm 0.02) \times 10^5$ and $(1.393 \pm 0.003) \times 10^5$ in 2011 and 2012 data, respectively. The combinatorial background, which contributes typically 6% under the $D^-$ peak, is modelled with an exponential distribution.

For the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ samples, a simultaneous fit to the distributions of $K^+\pi^-$ invariant mass, $m_{K^+\pi^-}$, and the invariant mass difference of $K^+\pi^-\pi^-$ and $K^+\pi^+\pi^-$ combinations, $\delta m = m_{K^+\pi^-\pi^-} - m_{K^+\pi^+\pi^-}$, is performed. Three different components are considered: the signal $D^*$ from $B^0$ or $B^+$ decays and two background sources. The PDF for the mass distributions of $D^*$ from $B$ decays is defined by the sum of two Gaussian functions and a Crystal Ball function in the $m_{K^+\pi^-}$ mass projection and by two Gaussian functions and a Johnson function [31] in the $\delta m$ mass projection. Background candidates containing a $T^{0}$ originating from a $b$ hadron decay without an intermediate $D^*$ resonance, which contribute about 15% in the full $\delta m$ mass range, are described by the same distribution as that of the signal for $m_{K^+\pi^-}$, and by an empirical function based on a phase-space distribution.
Fig. 1 Distribution of $m_{K\pi\pi}$ for the $B^0 \to D^- \mu^+ \nu_\mu X$ candidates in (left) 2011 and (right) 2012 data. Projections of the fit function are superimposed (blue continuous line) for the full PDF and its components: (red dashed line) signal $D'$ from $B^0$ or $B^+$ decays and (filled yellow area) combinatorial background.

Fig. 2 Distributions of (top) $m_{K\pi}$ and (bottom) $\delta m$ for $B^0 \to D^{*-} \mu^+ \nu_\mu X$ candidates in (left) 2011 and (right) 2012 data. Projections of the fit function are superimposed for (blue continuous line) the full PDF and its components: (red dashed line) signal $D^{*-}$ from $B^0$ or $B^+$ decays, (black dashed-dotted line) $\bar{D}^0$ from $B$ and (filled yellow area) combinatorial backgrounds.

for $\delta m$. A combinatorial background component which contributes typically 0.8% under the $D^*$ peak is modelled with an exponential distribution for $m_{K^+\pi^-}$ and the same empirical distribution for $\delta m$ as used for the $\bar{D}^0$ background. All parameters that describe signal and background shapes are allowed to vary freely in the invariant mass fits. The results of the 2011 and 2012 fits for these parameters are compatible within the statistical uncertainties. Figure 2 shows the results of the fit to the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ samples, projected onto the two mass observables. The yields corresponding to the $D^*$ peak are $(2.514 \pm 0.006) \times 10^5$ and $(5.776 \pm 0.009) \times 10^5$ in 2011 and 2012 data.

The fraction of $B^+$ background in data, $\alpha_{B^+}$, is determined with good precision by fitting the distribution of the BDT classifier, where templates for signal and $B^+$ background are obtained from simulation. Fits are performed separately in tagging categories for 2011 and 2012 data, giving fractions of $B^+$ of 6 and 3% on average for the $B^0 \to D^- \mu^+ \nu_\mu X$ and the $B^0 \to D^{*-} \mu^+ \nu_\mu X$ modes with relative variation of the order of 10% between samples. The results of the fits to 2012 data for both modes are given in Fig. 3. Limited knowledge of the exclusive decays used to build the simulation templates leads to systematic uncertainties of 0.5 and 0.4% on the $B^+$ fractions for $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$. In the decay time fit, the $B^+$ fractions are kept fixed. The statistical and systematic uncertainties on $\alpha_{B^+}$ lead to a systematic uncertainty on $\Delta m_{J}$, which is reported in Sect. 5.

The oscillation frequency $\Delta m_{J}$ is determined from a binned maximum likelihood fit to the distribution of the $B^0$ decay time $t$ of candidates classified as mixed ($q = -1$) or
The mistag fractions for signal and $B^+$ background are determined for each tagging category directly from the tagged time-dependent fit to data.

The fitting strategy is validated with simulation. A bias is observed in the $\Delta m_d$ value, due to a correlation between the decay time and its resolution, which is not taken into account when parameterizing the signal shape. Simulation shows that this correlation is introduced by the requirements of the software trigger and offline selection on the impact parameters of $D^-$ and $D^{*0}$ with respect to the PV. Values for this bias, of up to $4 \text{ ns}^{-1}$ with a 10% uncertainty, are determined for each mode and for each year by fitting the true and corrected time distributions and taking the differences between the resulting values of $\Delta m_d$. The uncertainty on the bias is treated as a systematic uncertainty on $\Delta m_d$.

The values of $\Delta m_d$, obtained from the time-dependent fit and corrected for the fit bias, are reported in Table 1. Systematic uncertainties are discussed below. The four independent $\Delta m_d$ values are compatible within statistical uncertainties. Figure 4 shows the fit projections for the decay time distributions for the candidates in the category with lowest mistag rate in 2012 data. The time-dependent asymmetries for the $B^0 \to D^- \mu^+\nu_X$ and $B^0 \to D^{*-}\mu^+\nu_X$ modes in 2011 and 2012 data are shown in Figs. 5 and 6. Fits are also performed in subsamples of different track multiplicity, number of primary vertices, magnet polarity, run periods, and muon charges. Statistically compatible results are obtained in all cases. A combination of the two $\Delta m_d$ determinations, including systematic uncertainties, is given in Sect. 6.

5 Systematic uncertainties

The contribution of each source of systematic uncertainty is evaluated by using a large number of parameterized simula-
Table 1 Results for $\Delta m_d$ measured in each mode for 2011 and 2012 data separately, for the total sample, and for the combination of the two modes. The quoted uncertainties for the separate samples are statistical only. For the total samples and the combination, they refer to statistical and total systematic uncertainties, respectively.

<table>
<thead>
<tr>
<th>Mode</th>
<th>2011 sample $\Delta m_d$ (ns$^{-1}$)</th>
<th>2012 sample $\Delta m_d$ (ns$^{-1}$)</th>
<th>Total sample $\Delta m_d$ (ns$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^0 \rightarrow D^+ \mu^+\nu_\mu X$</td>
<td>506.2 ± 5.1</td>
<td>505.2 ± 3.1</td>
<td>505.5 ± 2.7 ± 1.1</td>
</tr>
<tr>
<td>$B^0 \rightarrow D^{*-} \mu^+\nu_\mu X$</td>
<td>497.5 ± 6.1</td>
<td>508.3 ± 4.0</td>
<td>504.4 ± 3.4 ± 1.0</td>
</tr>
<tr>
<td>Combination</td>
<td></td>
<td></td>
<td>505.0 ± 2.1 ± 1.0</td>
</tr>
</tbody>
</table>

5.1 Background from $B^+$

The fraction of $B^+$ background is estimated from data with a very small statistical uncertainty. A variation, within their uncertainties, of the branching fractions of semileptonic $B^0$ decays resulting in a $D^{*-}$ or $D^-$ in the final state gives systematic uncertainties on the $B^+$ fractions of 0.5 and 0.4% for $B^0 \rightarrow D^+ \mu^+\nu_\mu X$ and $B^0 \rightarrow D^{*-} \mu^+\nu_\mu X$. The resulting uncertainty on $\Delta m_d$ is 0.1 ns$^{-1}$ in $B^0 \rightarrow D^+ \mu^+\nu_\mu X$ and is negligible for $B^0 \rightarrow D^{*-} \mu^+\nu_\mu X$. In the default fit, the decay time acceptance ratio of the $B^0$ and the $B^+$ components is taken from simulation. The time acceptance is to a large extent due to the cut on the $D^0$ impact parameter. A possible systematic effect due to an incorrect determination of the acceptance ratio from simulation is estimated by fitting events, generated with the default signal and background acceptances, with an acceptance ratio determined by using a tighter $D^0$ IP cut than the default. This gives an uncertainty of 0.4 ns$^{-1}$ on both decay modes. The above systematic uncertainties are considered as uncorrelated between the two channels.

5.2 Other backgrounds

The impact of the knowledge of backgrounds due to semileptonic $B^0$ decays with $D^{(*)-}$ in the final state is estimated by varying their contributions within the uncertainties on their branching fractions. This effect has a negligible impact on $\Delta m_d$ for both channels. For the $B^0 \rightarrow D^+ \mu^+\nu_\mu X$ channel, there is an additional contribution from $B^0 \rightarrow D_{s}^{*-} \mu^+\nu_\mu$ decays, where a kaon in the $D_{s}^- \rightarrow K^-K^+\pi^-$ decay is misidentified as a pion, which gives an 8% contribution due to $D_s^-$ peaking under the $D^-$ mass. A difference in $\Delta m_d$ of 0.5 ns$^{-1}$ is observed.

The $\Lambda_{c}^{0} \rightarrow nD^{*-} \mu^+\nu_\mu$ decay has not been observed. However, because of the similar final state, it can be mistaken for $B^+$ background, since neither of them exhibits oscillatory behaviour. Dedicated simulated samples are generated by assuming colour suppression with respect to signal, and are used to estimate a signal contamination of 0.2% from $\Lambda_{c}^{0}$ decays, with 100% uncertainty, which gives a negligible effect on $\Delta m_d$.

Small contributions from $B \rightarrow D^{(*)-}D_s^+ X$ decays, with the $D_s^+$ decaying semileptonically give an uncertainty of 0.2 ns$^{-1}$ on $\Delta m_d$ in the $B^0 \rightarrow D^- \mu^+\nu_\mu X$ mode, and a negligible effect for the $B^0 \rightarrow D^{*-} \mu^+\nu_\mu X$ mode.

Fig. 4 Decay time distributions for (left) $B^0 \rightarrow D^+ \mu^+\nu_\mu X$ and (right) $B^0 \rightarrow D^{*-} \mu^+\nu_\mu X$ in the category with lowest mistag in 2012 data.
5.3 The $k$-factor

Two main sources of systematic uncertainty are related to the $k$-factor. The first, due to possible differences in the $B$ momentum spectrum between simulation and data, is studied by comparing the $B$ momentum in $B^+ \to J/\psi K^+$ decays in data and simulation, and reweighting signal simulation to estimate the effect on the $k$-factor distribution and therefore on $\Delta m_d$. The systematic uncertainties on $\Delta m_d$ from this effect for $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ are 0.3 ns$^{-1}$ and 0.5 ns$^{-1}$. The second source, related to the uncertainties on the measurements of the branching fractions for the exclusive modes which are used to build the simulated samples, is evaluated by varying the branching fractions of exclusive decays one at a time by one standard deviation, and reweighting the corresponding $k$-factor distribution. An uncertainty of 0.4 ns$^{-1}$ is obtained for both $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$ channels. The systematic uncertainties from the $k$-factor correction are taken to be correlated between the two channels.

The systematic uncertainties on $\Delta m_d$ from the finite number of events in the simulation sample used to compute the $k$-factor corrections are 0.3 and 0.4 ns$^{-1}$ ($B^0 \to D^- \mu^+ \nu_\mu X$) and 0.2 and 0.3 ns$^{-1}$ ($B^0 \to D^{*-} \mu^+ \nu_\mu X$) for the 2011 and 2012 samples, respectively.

5.4 Other systematic uncertainties

Possible differences between data and simulation in the resolution on the $B^0$ flight distance are evaluated by using the results of a study reported in Ref. [33], and scaling the widths of the triple Gaussian function by a factor 1.5 with respect to the default. Uncertainties of 0.3 ns$^{-1}$ and 0.5 ns$^{-1}$ on $\Delta m_d$ are obtained for $B^0 \to D^- \mu^+ \nu_\mu X$ and $B^0 \to D^{*-} \mu^+ \nu_\mu X$. Both channels are affected by the same discrepancy between data and simulation; thus these systematic uncertainties are taken as correlated.

Since all parameters are allowed to vary freely in the invariant mass fits, the uncertainties from the invariant mass model are small. As a cross-check, when the fits are repeated...
using the sWeights determined without splitting the mass fits in tagging categories, negligible variation in $\Delta m_d$, is found. Signal and background mistag probabilities are free parameters in the fit, and therefore no systematic uncertainty is associated to them.

Asymmetries in the production of neutral and charged $B$ mesons, in tagging efficiency and mistag probabilities, and in the reconstruction of the final state are neglected in the $\Delta m_d$ fits. Also, the $B^0$ semileptonic $CP$ asymmetry $a_{CP}^0$ is assumed to be zero. The systematic uncertainty on $\Delta m_d$ arising from these assumptions is studied using parameterized simulations with the asymmetries set to zero, to their measured values, and to random variations from their central values within the uncertainties [34]. The resulting uncertainty on $\Delta m_d$ is found to be negligible.

The bias in $\Delta m_d$ from the correlation between the decay time and its resolution is determined using the simulation. The dependence of $\Delta m_d$ on possible differences between data and simulation has already been considered above by varying the composition of the simulation sample used to construct the $k$-factor distribution. Since the bias is related to the cut on the $D$ meson IP with respect to the PV, the fits are repeated with a $k$-factor distribution obtained with a tighter cut on the IP, and the difference with respect to the default is taken as the systematic uncertainty. The systematic uncertainties (0.5 and 0.3 ns$^{-1}$ for $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$, respectively) related to the bias are considered as uncorrelated between the channels, as they are determined from different simulation samples and the time-biasing cuts, responsible for the systematic uncertainty on the bias, are different for the two channels.

The knowledge of the length scale of the LHCb experiment is limited by the uncertainties from the metrology measurements of the silicon-strip vertex detector. This was evaluated in the context of the $\Delta m_s$ measurement and found to be 0.022% [33]. This translates into an uncertainty on $\Delta m_d$ of 0.1 ns$^{-1}$. The uncertainty on the knowledge of the momentum scale is determined by reconstructing the masses of various particles and is found to be 0.03% [35]. This uncertainty results in a 0.2 ns$^{-1}$ uncertainty in $\Delta m_d$ in both modes.

Both uncertainties are considered correlated across the two channels.

Effects due to the choice of the binning scheme and fitting ranges are found to be negligible.

### 6 Summary and conclusion

A combined value of $\Delta m_d$ is obtained as a weighted average of the four measurements performed in $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ in the years 2011 and 2012. First, the 2011 and 2012 results for each decay mode are averaged according to their statistical uncertainties. The combined results are shown in the last column of Table 1. Then, the resulting $\Delta m_d$ values of each mode are averaged taking account of statistical and uncorrelated systematic uncertainties. The combined systematic uncertainty is added in quadrature to the resulting uncertainty. The combined result is shown in the last row of Table 1.

In conclusion, the oscillation frequency, $\Delta m_d$, in the $B^0 \rightarrow \bar{B}^0$ system is measured in semileptonic $B^0$ decays using data collected in 2011 and 2012 at LHCb. The decays $B^0 \rightarrow D^- \mu^+ \nu_\mu X$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu X$ are used, where the $D$ mesons are reconstructed in Cabibbo-favoured decays $D^- \rightarrow K^+ \pi^- \pi^-$ and $D^{*-} \rightarrow \bar{B}^0 \pi^-$, with $\bar{B}^0 \rightarrow K^+ \pi^-$. A combined $\Delta m_d$ measurement is obtained,

$$\Delta m_d = (505.0 \pm 2.1 \text{ (stat)} \pm 1.0 \text{ (syst)}) \text{ ns}^{-1},$$

which is compatible with previous LHCb results and the world average [13]. This is the most precise single measurement of this quantity, with a total uncertainty similar to the current world average.

### 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); NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); FOM and NWO (The Netherlands); MNiSW and NCN (Poland); MEN/IFA (Romania); MinES and FANO (Russia); MinECo (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United King-
The variables used as input for the BDT classifier are the following:

- Visible mass of the $B$ candidate, $m_B \equiv m(D^{(*)-}\mu^+)$
- Corrected mass [36], defined as $m_{corr} = \sqrt{m_B^2 + p_T(B)^2}$
- Impact parameter, $\text{IP}(\pi, D)$, with respect to the decay vertex of the $D^{-}(\bar{D}^0)$, of the track with the smallest impact parameter with respect to the $B$ candidate
- Smallest vertex $\chi^2$ of the combination of the $D^- (D^{*-})$ with any other track, and the invariant mass of this combination
- Cone isolation $I = \frac{p_T(B)}{p_T(B) + \sum p_T()}$, where the sum is computed over tracks which satisfy $\sqrt{\delta\eta_i^2 + \delta\phi_i^2} < 1$, $\delta\eta_i$ and $\delta\phi_i$ being the difference in pseudorapidity and in polar angle $\phi$ between the track and the $B$ candidate
- Track isolation variables, used to discriminate tracks originating from the $B$ vertex from those originating elsewhere:
  - Number of nearby tracks [37], computed for each track in the $B$ decay chain
  - The output of an isolation BDT [37] estimated for the $B$ candidate
  - A second isolation BDT, similar to the previous, which exploits a different training strategy and additional variables, computed for tracks originating from $D^- (\bar{D}^0)$ decays, those coming from the $B$ decay, and all tracks in the decay chain.

The TMVA package [38], used to train and test the classifier, ranks the input variables according to their discriminating power between signal and background.

### A.2 Distributions of the $k$-factor

Figure 7 shows distributions of the $k$-factor as a function of the visible mass of the $B$ candidate, as obtained with samples of simulated signal events. In each plot, the average $k$-factor and the result of a polynomial fit are also shown.
Departamento de Fisica, Universidad Nacional de Colombia, Bogota, Colombia, associated to
Institut für Physik, Universität Rostock, Rostock, Germany, associated to
National Research Centre Kurchatov Institute, Moscow, Russia, associated to
Yandex School of Data Analysis, Moscow, Russia, associated to
Instituto de Fisica Corpuscular (IFIC), Universitat de Valencia-CSIC, Valencia, Spain, associated to
Van Swinderen Institute, University of Groningen, Groningen, The Netherlands, associated to

a Universidade Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil
b P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia
c Università di Bari, Bari, Italy
d Università di Bologna, Bologna, Italy
e Università di Cagliari, Cagliari, Italy
f Università di Ferrara, Ferrara, Italy
g Università di Urbino, Urbino, Italy
h Università di Modena e Reggio Emilia, Modena, Italy
i Università di Genova, Genova, Italy
j Università di Milano Bicocca, Milan, Italy
k Università di Roma Tor Vergata, Rome, Italy
l Università di Roma La Sapienza, Rome, Italy
m Università della Basilicata, Potenza, Italy
n AGH-University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland
o LIFAELS, La Salle, Universitat Ramon Llull, Barcelona, Spain
p Hanoi University of Science, Hanoi, Viet Nam
q Università di Padova, Padua, Italy
r Università di Pisa, Pisa, Italy
s Scuola Normale Superiore, Pisa, Italy
t Università degli Studi di Milano, Milan, Italy
† Deceased