Study of $B_s^0 \rightarrow D_s^{(*)-} \mu^+ \nu_\mu$ decays at the LHCb experiment

Author: Lino Ferreira Lopes

February 22, 2017

Abstract

The study of the semileptonic decays $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^{*-} \mu^+ \nu_\mu$ is reported using a data sample corresponding to an integrated luminosity of 3 fb$^{-1}$, collected by the LHCb experiment in proton-proton collisions at centre-of-mass energies of 7 and 8 TeV. The $B^0 \rightarrow D^- \mu^+ \nu_\mu$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ decays reconstructed in the same dataset are used as normalisation channels. After an investigation of possible ways to cope with the partial reconstruction of the decay candidates, the composition of the data sample is determined by a fit to the approximated mass of the candidates. The total efficiency is split into five contributions which are estimated either from simulations or calibration samples, when possible. The preliminary values of the branching ratios are found to be:

$\mathcal{B}(B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu) = (2.03 \pm 0.05{\rm (stat)} \pm 0.14{\rm (BR)} \pm 0.12(f_s/f_d))\%$,

$\mathcal{B}(B_s^0 \rightarrow D_s^{*-} \mu^+ \nu_\mu) = (4.35 \pm 0.16{\rm (stat)} \pm 0.22{\rm (BR)} \pm 0.25(f_s/f_d))\%$,

where the first uncertainty is statistical, the second and the third are due to the uncertainties of the branching ratios and $f_s/f_d$ taken as input from external measurements, respectively.
1 Introduction

The standard model (SM) of particle physics is a theory describing the interaction between matter and three of the four fundamental forces, namely the electromagnetic, weak and strong interactions. It does not encompass gravity. It has been developed during the twentieth century and the last fundamental particle predicted is the Higgs boson, whose discovery was announced in July 2012 by the ATLAS and CMS experiments at the Large Hadron Collider at CERN [1, 2]. The SM has been very successful and most of its predictions have been verified to a high level of accuracy. Nonetheless, it is not the definitive theory since there are multiple phenomena which are not explained in this framework, e.g. neutrino oscillations, the problem of the mass hierarchy, the strong CP problem, the origin of dark energy and dark matter [3]. For these reasons, the SM is widely regarded as a low energy effective field theory which could be encompassed in a more complete theory. A large part of the current research is focused on the search and description of new physics which is to extend the SM.

In the domain of heavy-flavour physics, decays of hadrons containing a $b$ or a $c$ quark are studied. These decays are useful e.g. to make precision tests of $CP$ violation or to search for new physics which affects transitions suppressed in the SM. Another property of the SM currently under investigation is that of lepton universality. This property states that the coupling between the electroweak gauge bosons, i.e. the $\gamma$, $W^\pm$ and $Z^0$, and the leptons is independent of the leptons’ flavour. A possible way to test this property is to study semileptonic decays and to compute ratios of branching fractions:

$$R(D^{(*)}) = \frac{B(B^0 \to D^{(*)-} \tau^+ \nu_\tau)}{B(B^0 \to D^{(*)-} \mu^+ \nu_\mu)}.$$  \hfill (1)

These $B^0$ decays have been studied intensively by the $B$-factories, i.e. the Babar and Belle experiments, and also by LHCb. The current determinations of $R(D^{(*)})$ are presented in Tab. 1. The combined measurements display a 3.9$\sigma$ tension with SM expectations, see Fig. 1. This demand additional investigations to understand if this discrepancy is due to genuine new physics effect; for instance complementary measurements for the $B^0_s$ meson (and $\Lambda_b$ baryon), such as:

$$R(D_s^{(*)}) = \frac{B(B^0_s \to D_s^{(*)-} \tau^+ \nu_\tau)}{B(B^0_s \to D_s^{(*)-} \mu^+ \nu_\mu)}.$$ \hfill (2)

could help in shed some light. While $B^0$ mesons cannot be produced at $B$-factories running at the $\Upsilon(4S)$ energy (and Belle collected only a small sample at $\Upsilon(5S)$ energy), they are abundantly produced in hadron collisions, and only more recently their investigation has started with the aim of mirroring the same level of knowledge reached for the $B^0$ mesons decays.

The aim of this project is to study for the first time the $B^0_s \to D_s^- \mu^+ \nu_\mu$ and $B^0_s \to D_s^{*-} \mu^+ \nu_\mu$ decays in details. These decays will serve as the reference mode to perform lepton-flavour universality tests in the $B_s$ sector in the next future.

\footnote{Charge conjugation is implied throughout this report, unless otherwise specified.}
Table 1: Current determinations of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ by the Babar [4, 5], Belle [6, 7, 8] and LHCb [9] experiments. The SM predictions of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ are taken from Ref. [10] and [11], respectively.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$\mathcal{R}(D)$</th>
<th>$\mathcal{R}(D^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BaBar</td>
<td>$0.440 \pm 0.058 \pm 0.042$</td>
<td>$0.332 \pm 0.024 \pm 0.018$</td>
</tr>
<tr>
<td></td>
<td>$0.375 \pm 0.064 \pm 0.026$</td>
<td>$0.293 \pm 0.038 \pm 0.015$</td>
</tr>
<tr>
<td>Belle</td>
<td>--</td>
<td>$0.302 \pm 0.030 \pm 0.011$</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>$0.276 \pm 0.034^{+0.029}_{-0.026}$</td>
</tr>
<tr>
<td>LHCb</td>
<td>--</td>
<td>$0.336 \pm 0.027 \pm 0.030$</td>
</tr>
<tr>
<td>Average</td>
<td>$0.403 \pm 0.040 \pm 0.024$</td>
<td>$0.310 \pm 0.015 \pm 0.008$</td>
</tr>
<tr>
<td>SM</td>
<td>$0.300 \pm 0.008$</td>
<td>$0.252 \pm 0.003$</td>
</tr>
</tbody>
</table>

Figure 1: Combination plot of the latest $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ measurements. The average value shows a 3.9\textsigma tension with SM prediction. Image taken from [12].
2 Motivation

The exclusive branching ratios \( \mathcal{B}(B^0_s \to D_s^- \mu^+\nu_\mu) \) and \( \mathcal{B}(B^0_s \to D_s^{*-} \mu^+\nu_\mu) \) are not yet measured. Currently only semi-inclusive branching ratios have been determined and their measurements have relatively poor precision [3]:

\[
\begin{align*}
\mathcal{B}(B^0_s \to D_s^- \mu^+\nu_\mu) &= (8.1 \pm 1.3)\% \\
\mathcal{B}(B^0_s \to D_s^{*-} \mu^+\nu_\mu) &= (5.4 \pm 1.1)\%
\end{align*}
\]

This project is the first step towards the measurement of the exclusive branching ratios mentioned above. In order to do so, we will measure the signal yields of the following decays:

\[
\begin{align*}
B^0_s &\to D_s^- (\to \phi \to K^+K^-) \pi^- ) \mu^+\nu_\mu, \\
B_s^0 &\to D_s^{*-} (\to D_s^- (\to \phi \to K^+K^-) \pi^- ) X) \mu^+\nu_\mu,
\end{align*}
\]

with respect to the similar \( B^0 \) decays, used as normalization:

\[
\begin{align*}
B^0 &\to D^- (\to \phi \to K^+K^-) \pi^- ) \mu^+\nu_\mu, \\
B^0 &\to D^{*-} (\to D^- (\to \phi \to K^+K^-) \pi^- ) X) \mu^+\nu_\mu,
\end{align*}
\]

where \( X \) represents both a \( \gamma \) and a \( \pi^0 \). We use the \( B^0 \) decays as normalization decays because we reconstruct them in the same final state and thus numerator and denominator are very similar, allowing the suppression of the systematic uncertainties. The Feynman diagram of the \( B^0_s \to D_s^- \mu^+\nu_\mu \) decays is shown in Fig. 2. We can then compute the following yields ratios:

\[
\begin{align*}
\frac{N_{B_s}}{N_B} &= \frac{f_s \mathcal{B}(B_s \to D_s^- \mu^+\nu_\mu) \mathcal{B}(D_s^- \to \phi\pi^-) \epsilon_s}{f_d \mathcal{B}(B \to D^- \mu^+\nu_\mu) \mathcal{B}(D^- \to \phi\pi^-) \epsilon_d} \\
\frac{N_{B_s}^*}{N_B^*} &= \frac{f_s \mathcal{B}(B_s \to D_s^{*-} \mu^+\nu_\mu) \mathcal{B}(D_s^{*-} \to D_s^- X) \mathcal{B}(D_s^- \to \phi\pi^-) \epsilon_s^*}{f_d \mathcal{B}(B \to D^{*-} \mu^+\nu_\mu) \mathcal{B}(D^{*-} \to D^- X) \mathcal{B}(D^- \to \phi\pi^-) \epsilon_d^*}
\end{align*}
\]

where \( N_{B_s} \) is the yield of Eq. (5) and \( N_B \) is the yield of Eq. (7); \( N_{B_s}^* \) and \( N_B^* \) are their excited counterparts, i.e. Eqs. (6) and (8); \( f_s \) and \( f_d \) are the fragmentation fractions, i.e. the probability that a \( b \) quark will hadronize into a meson containing an \( s \) or a \( d \) quark, respectively; \( \epsilon_s \) and \( \epsilon_d \) are the total efficiencies for the numerator and the denominator decays, respectively; \( \epsilon_s^* \) and \( \epsilon_d^* \) are their excited counterparts.
We will measure the events yields ratios in data and the efficiencies ratios in Monte Carlo simulation, or using control sample of data when possible. The value of the other parameters are taken from external measurements, as reported in Tab. 2.

We observe that the normalization branching ratio and $D_s(\rightarrow)$ branching ratios contribute to a total relative uncertainty of 7.3% in the case of $B_s \rightarrow D_s \mu \nu$ and 5.6% in the case of $B_s \rightarrow D_s^* \mu \nu$. With the dataset collected by the LHCB experiment during 2011 and 2012 (known as Run 1 data), we expect to measure $\frac{N_{B_s}}{N_B}$ and $\frac{N_{B_s}^*}{N_B}$ with $\mathcal{O}(3\%)$ and $\mathcal{O}(4\%)$ relative uncertainty, respectively. We aim at controlling the efficiencies with a relative uncertainty below 5%. We could then extract a measurement of the exclusive decays’ branching ratios with a $\mathcal{O}(10\%)$ relative uncertainty.

On the other hand, if the ratios $\frac{\mathcal{B}(B_s \rightarrow D_s^{(*)} \mu \nu)}{\mathcal{B}(B \rightarrow D^{(*)} \mu \nu)}$ can be predicted theoretically (e.g. from lattice Quantum Chromodynamics (QCD)) with a relative uncertainty of a few percent, we could then extract a measurement of $f_s/f_d$ competitive with current determination. The current determination is based on two measurements, one using inclusive semileptonic decays [13], and another using hadronic decays [14]:

\[
(f_s/f_d)_{\text{semileptonic}} = 0.268 \pm 0.008(\text{stat})^{+0.022}_{-0.026}(\text{syst}) \quad (11)
\]

\[
(f_s/f_d)_{\text{hadronic}} = 0.238 \pm 0.004(\text{stat}) \pm 0.015(\text{syst}) \pm 0.021(\text{theo}) \quad (12)
\]

Our decays have more statistics than the measurement of Eq. (11). The method to extract $f_s/f_d$ from semileptonic decays used in Ref. [13] is different from the approach adopted here, and exploit only 3 pb$^{-1}$ of the Run 1 dataset. Our decays are also theoretically cleaner than the ones used in Ref. [14], the $B^0_s \rightarrow D_s^- \pi^+$ and $B^0 \rightarrow D^- K^+$ decays. Note also that these measurements are done at a centre-of-mass energy $\sqrt{s} = 7$ TeV. We can add a measurement at $\sqrt{s} = 8$ TeV.

This project is divided in three parts:

1. Investigate some variables that allow discrimination between the two decays reconstructed inclusively, i.e. in the final state $K^+ K^- \pi^- \mu^+$, as we do not
reconstruct the neutrino nor the $\gamma$ or the $\pi^0$ coming from the decay of the $D^{*-}_{(s)}$ meson. This is the subject of section 4.

2. Make the first step towards the measurement of the exclusive branching ratio by counting the yields of the $B^0_s$ and $B^0$ decay modes using one of the variables identified previously. This is described in section 5.

3. Estimate the efficiencies. This is described in section 6.

In section 7 we outline the possible source of systematic uncertainties affecting the measurements, and in section 8 we draw the conclusions of this work.

## 3 The LHCb experiment

The LHCb detector is described in great detail in Ref. [16]. In this section we summarize the main features of interest for our analysis.

The LHCb experiment is a dedicated heavy flavour physics experiment at the Large Hadron Collider (LHC) at CERN in Geneva, Switzerland. Its main goal is to search for indirect evidence of new physics in $CP$ violation and rare decays of beauty and charm hadrons, by looking for the effects of new particles in processes that are precisely predicted in the Standard Model (SM) and by utilising the distinctive flavour structure of the SM with no tree-level flavour-changing neutral currents. The LHCb detector, sketched in figure 3, is a single-arm forward spectrometer covering the pseudo-rapidity range $2 < \eta < 5$. This choice of geometry comes from the fact that at high energies, the $b$ quarks are mainly produced in a cone along the $z$-axis, i.e. the proton beam axis. The detectors sub-systems are the following

**Vertex Locator (VELO):** Surrounds the interaction region. Its purpose is to measure the positions of the interaction vertices with high precision. It is composed of 42 silicon disks, able to measure the polar position $(r, \phi)$ of the protons’ interaction
points, the primary vertexes (PV), and of the decay vertexes of particles produced in the primary interaction (so-called secondary vertexes, SV). The VELO provides a track finding efficiency above 98%. The PV resolution is 13 $\mu$m in the transverse plane and 71 $\mu$m along the beam axis. The impact parameter (IP) of a track is the minimal distance between the track and the PV. Since the $B^0$ and $D^-$ mesons are long lived particles, their decay vertex are displaced from the PV. The IP of the $D^-$ daughters are relatively large on average and thus a requirement on the IP of these particles will be effective at removing prompt backgrounds. The resolution on the IP decreases linearly as a function of the transverse momentum, $p_T$, of the particles. The fitted resolution $\sigma$ using 2012 data is given by $\sigma = 11.6 + 23.4/p_T$. The resolution on the IP is smaller than 35 $\mu$m for particles with $p_T > 1$ GeV/c [17].

Ring Imaging Cherenkov Detector (RICH): There are two of them. Their purpose is to identify charged hadrons in the momentum range $2-100$ GeV/c. Using the Cherenkov effect, the velocity $\beta c$ of the particles can be found by measuring the light-cone’s aperture angle $\theta$:

$$\cos \theta = \frac{1}{\beta n},$$

(13)

where $n$ is the refractive index of the medium. By combining the velocity with the momentum and the electric charge measurements coming from the trackers and the magnet, the identity of the particle is retrieved. The first RICH, filled with $C_4F_{10}$ and using as additional radiator a small piece of silica aerogel (in Run 1 only) is placed just after the VELO and covers a momentum range of 2-60 GeV/c. The second RICH, filled with CF$_4$, is placed after the tracking stations and covers a momentum range of 15-100 GeV/c. In order to produce Cherenkov light in the $C_4F_{10}$ medium, the kaons require a minimal momentum of 9.3 GeV/c [18]. If the momentum of a kaon is below this threshold, it can only be identified as a kaon by the absence of Cherenkov light. The typical pion-kaon misidentification rate is around 10% for a kaon identification efficiency of 95% for the particle identification requirement $\Delta \log(K - \pi) > 0$ (explained in further detail in section 9.5), when averaging over the 2-100 GeV/c range. If the requirement is tightened to $\Delta \log(K - \pi) > 5$, the mis-identification rate falls to ~3% for a ~10% efficiency loss in kaon identification. Figure 35 in section 9.5 shows the distribution of the efficiency and mis-identification rate for kaons and pions under a ProbNN requirement.

Tracker Turicensis (TT) and Tracking Stations (T1-T3): Together with the VELO they compose the tracking system. TT is upstream of the magnet, while the three tracking stations T1-T3 are downstream of the magnet. TT uses silicon microstrip sensors while T1-T3 use straw tubes. Charged particles require a minimum momentum of 1.5 GeV/c in order to reach T1-T3. The hit resolution is 52.6 $\mu$m for the TT and 50.3 $\mu$m for the IT. The tracking efficiency, i.e. the probability of reconstructing the trajectory of a charged particle which has passed through the full tracking system, is greater than 96% on average, for particles in the momentum range 5 to 200 GeV/c.

Magnet: Dipole magnet with an integrated field of about 4 Tm, placed between TT and T1-T3. It is used to deflect charged particles and therefore measure their electric charge and momentum. The relative momentum resolution $\delta p/p$ ranges from about 0.5%, for particles with momentum below 20 GeV/c, up to 0.8% for 100
GeV/c particles.

**Calorimeters:** There are two of them. The electromagnetic (ECAL) and the hadronic (HCAL) calorimeters. The HCAL is placed after the ECAL which is placed after the second RICH. Their purpose is to measure the energy and positions of electrons, photons and hadrons. When a particle hits the scintillating pads of the detector, it produces secondary particles which will excite the medium and create an amount of light proportional to the energy of the incoming particles. The calorimeters permit to detect neutral particles like the photons or the neutral pion. The ECAL is made of sampling/lead structure with a total thickness of 25 X₀. The HCAL is made from iron and scintillating tiles with a thickness of 5.6 nuclear interaction lengths.

**Muon chambers (M1-M5):** The role of the muon detection system is to identify muons. It also contributes to the L0 level of the trigger, which is explained in further details in section 6.3. The muon station M1 is placed before the calorimeters, while M2-M5 are placed at the end of the detector. Iron absorbers with 80 cm thickness are placed between the Stations M2 to M5 in order to select penetrating muons. In order for a muon to traverse all the stations, a minimal momentum of approximately 6 GeV/c is required. The full system is composed of 1380 chambers and covers a total area of 435 m². The muon identification efficiency is between 95% and 98% for an hadron mis-identification rate at the level of 1% [19].

### 4 Dealing with partial reconstruction

In order to extract the yields ratios of Eqs. (9) and (10), we want to fit a distribution which allows us to disentangle the signal components and to distinguish them from the backgrounds. Thus the study presented here.

The differential decay rate of \( B^0 \rightarrow D^- \mu^+ \nu_\mu \) can be parametrised as a function of the momentum transfer squared, \( q^2 \), and one helicity angle: \( \theta_l \). The momentum transfer squared \( q^2 = (p_l + p_\nu)^2 \) is the invariant mass squared of the virtual \( W \) boson; \( \theta_l \) is the angle between the momentum of the lepton (in our case the muon) in the virtual \( W \) frame, relatively to the momentum of the \( W \) in the \( B^0 \) frame. These two momenta form a plane. The spin-parity of the \( D^- \) and \( D^{*-} \) mesons are \( J^P = 0^- \) and \( J^P = 1^- \), respectively. In the case of \( B^0 \rightarrow D^{*-} \mu^+ \nu_\mu \), because of the higher spin-value of the excited charmed meson, the decay rate is described by two additional helicity angles, \( \theta_V \) and \( \chi \): \( \theta_V \) is the angle between the momentum of the \( D^- \) in the \( D^{*-} \) frame and the momentum of the \( D^{*-} \) in the \( B^0 \) frame. Again, these two momenta form a plane. Finally, \( \chi \) is the angle between the two previously defined planes. The helicity angles are sketched in Fig. 4. The differential decay rate as a function of \( q^2 \) and the three helicity angles for \( B^0 \rightarrow D^{*-} \mu^+ \nu_\mu \) is given in Ref. [20]. Measuring the decay rate as a function of these four variables grants more experimental discrimination between both decays and also allows us to measure the form factors and thus distinguish between different decay models, see Fig. 5. The from factors are phenomenological parameters that describe the hadronic current of the transition amplitude. They are not measured yet, and can be compared with expectations to validate calculations from lattice QCD. If we do not fully reconstruct
4 DEALING WITH PARTIAL RECONSTRUCTION

Figure 4: Kinematics describing the decay $B^0 \to D^{*-} \mu^+ \nu_\mu$.

the $D^*$ decay, we cannot define the $\theta_V$ and $\chi$ angles. In this case the decay rate as function of $q^2$ and $\cos(\theta_l)$ still gives access to the form factors measurements, and in the heavy quark effective theory is given by [21]:

$$
\frac{d\Gamma(B \to Dl\nu)}{dq^2} = \frac{G_F^2|V_{cb}|^2}{24\pi^3} |p_D|^3 |F_1(q^2)|^2,
$$

where $G_F$ is the Fermi constant, $V_{cb}$ is the CKM matrix element describing the $b \to c$ transition, $p_D$ is the momentum of the $D$ meson in the $B$ rest-frame and $F_1(q^2)$ is a form factor. In the case of the $B^0 \to D^- \mu^+ \nu_\mu$ decay, the virtual $W$ has a defined helicity and therefore no polarization amplitude. In the case of the $B^0 \to D^{*-} \mu^+ \nu_\mu$ decay, we need to introduce the angular dependence of the different polarizations:

$$
\frac{d^2\Gamma(B \to D^*l\nu)}{dq^2 d\cos(\theta_l)} = \frac{G_F^2|V_{cb}|^2}{768\pi^3} |p_D|^2 \frac{m_B^2}{q^2} \left( (1 + \cos(\theta_l))^2 |H_+|^2 + (1 - \cos(\theta_l))^2 |H_-|^2 + 2 \sin^2(\theta_l) |H_0|^2 \right),
$$

where $\theta_l$ is replaced by $\pi - \theta_l$ for the decay $\bar{B}^0 \to D^{*-} \mu^- \bar{\nu}_\mu$, $m_B$ is the $B$ mass, and $|H_+|$, $|H_-|$, $|H_0|$ are invariant helicity amplitudes related to the form factors, corresponding to transverse and longitudinal polarizations, respectively.

4.1 Experimental challenges

The major experimental challenge of the semileptonic decays that we consider is the fact that we cannot reconstruct all final state particles. Indeed, we cannot reconstruct the neutrino and, in the case where a $D_s^*$ is present, an additional particle which is either a $\gamma$ or a $\pi^0$. This experimental situation is represented in Fig. 6. Since the final state is the same regardless of the decay, the question arises as how to distinguish both decays. In order to be able to distinguish them, one needs to find a good discriminating variable such that its distribution is different enough to allow the disentanglement of both components.

Monte Carlo (MC) simulations are used to model the signal decays of interest, as well as different backgrounds present in data, allowing the study of the sample composition explained in further detail in section 5.4. Details on the simulated sample are described in that section and appendix 9.3. In this section we used simulated...
4.1 Experimental challenges

Figure 5: The left plot shows the distribution of $q^2$ in MC for the decays $B^0 \rightarrow D^- \mu^+ \nu_\mu$ (in blue) and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ (in red). We see that the distribution allows the separation of both components. The right plot shows the distribution of $q^2$ in MC for the decay $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ for two different sets of form factors used in the generations of the decays. We see that the distribution exhibits discrimination power between different phenomenological decay models.

candidates of the decays $B^0 \rightarrow D^- \mu^+ \nu_\mu$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$. In simulation the properties of particles that we cannot reconstruct in data, such as the neutrino, are known and therefore we can study the distribution of variables which are otherwise inaccessible.

A variable that we can use to discriminate the two signal components is the $B(s)$ visible mass, i.e. $m(D(s)\mu) = m(K\bar{K}\pi\mu)$. The distribution of this variable in MC simulations is presented in Fig. 7. We notice that both distributions are relatively far from the $B^0$ mass value $m_{B^0} = (5279.62 \pm 0.15)$ MeV$^2$ [3]. This is expected since we do not reconstruct the neutrino and thus lose some energy in the computation of the invariant mass. In the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ distribution we miss an additional particle on top of the neutrino and therefore the distribution is even more biased relatively to the true mass. The distributions also display a relatively poor resolution for the same reasons. Nonetheless, the $B$ visible mass distributions still allow discrimination between both decays and thus $m(D\mu)$ is an interesting variable as a tool to disentangle both decay channels.

While $m(D\mu)$ is an interesting variable, it is possible to find a better one. We consider for this purpose the $B^0_{(s)}$ corrected mass:

$$m_{\text{corr}} = \sqrt{m(D(s)\mu)^2 + p_T^2}$$

where $m(D(s)\mu)$ is the $B^0_{(s)}$ visible mass and $p_T$ is the norm of the transverse momentum of the $D(s)\mu$ system relatively to the $B^0_{(s)}$ flight direction. If we consider the $B^0_{(s)} \rightarrow D^-_{(s)} \mu^+ \nu_\mu$ decay (see Fig. 6) since the total transverse momentum relatively to the $B^0_{(s)}$ flight direction is zero by definition, then the transverse momentum relatively to the $B^0_{(s)}$ flight direction of the $D(s)\mu$ system is opposite to the one from the neutrino. Their norms are thus equals. The $B^0_{(s)}$ corrected mass is thus the $B^0_{(s)}$ invariant mass, where the longitudinal part of the neutrino’s momentum, which cannot be measured nor constrained, is neglected. The full computation is given in section 9.1 of the appendix.

\footnote{In this report we use the convention $c = 1$ unless otherwise specified.}
4 DEALING WITH PARTIAL RECONSTRUCTION

Figure 6: Experimental situation (left) of a $B_{(s)}^0 \rightarrow D_{(s)}^{(*)} \mu^+ \nu_\mu$ decay. We miss at least the neutrino and in the case where an excited charmed meson is produced we miss an additional particle, which is a $\gamma$ or a $\pi^0$. The final state measured is the same in both cases: $K^+ K^- \pi^- \mu^+$. Using kinematical constraints (right), we are able to retrieve some information on the neutrino, namely its transverse momentum relatively to the $B_{(s)}^0$ flight direction. The proxy variables $p_\perp(D_{(s)})$ and $p_\perp(\mu)$ are also indicated.

Figure 7: $B^0$ visible (left) and corrected (right) mass distributions for both signal types in MC. We observe some discrimination power in both cases but the $m_{corr}$ distributions are less biased and have a better resolution than for $m(D_{(s)}\mu)$.

The MC distribution of the $B^0$ corrected mass is plotted in Fig. 7. We observe that those distributions are less biased and have a better resolution than the $B^0$ visible mass distributions. These distributions keep some discriminating power between both decay signals and also accentuate more the difference between signal and background, relatively to the visible mass. The corrected mass distribution is a standard tool for similar analysis where a particle is not reconstructed. It is a powerful variable which has been used intensively in other analyses [22, 23], and proved to be of great interest. Nevertheless, we want to investigate whether we can find a better variable.

4.2 Attempt to improve the $B_{(s)}^0$ mass determination

The idea to define a new variable which could potentially improve the $B_{(s)}^0$ mass determination is the following: using MC simulation, we can compute the ratio
4.2 Attempt to improve the \( B^0_{(s)} \) mass determination

Figure 8: Distribution of \( p_\perp(D) \) for \( B^0 \rightarrow D^- \mu^+ \nu_\mu \) and \( B^0 \rightarrow D^{*-} \mu^+ \nu_\mu \) (noted Cat1 and Cat2, respectively). The sum of both distributions is also shown in green.

\[ r = \frac{m(D(\perp)\mu)}{m(D(\perp)\mu)} \]

and find some “proxy” variable \( X \) which is correlated with \( r \). We can then produce a two-dimensional histogram of \( r \) as a function of \( X \), do a profile and fit the profile with a polynomial function. This function \( r(X) \) gives the value of \( r \) for a given \( X \). It should be noted that \( r(X) \) does not provide a perfect relationship between both variables. Indeed, when profiling the 2D histogram, we are doing the average of \( r \) in bins of \( X \). Therefore the value \( r(X_0) \), returned by the polynomial function \( r(X) \) applied at a certain value \( X_0 \) of the proxy variable, is only the average value of all \( r \) falling in the particular bin containing \( X_0 \). The choice of the proxy variable is based on the correlation displayed with \( r \). If there is no correlation between them then \( r(X) \) is constant and brings no information whatsoever. Now given \( r(X) \) we are able to correct candidate-by-candidate the value of the \( B^0_{(s)} \) visible mass in data:

\[
m_{\text{fit}} = r(X) \cdot m(D_{(s)}\mu),
\]

where \( X \) is the value of the proxy variable for the given candidate and \( m_{\text{fit}} \) is the modified visible mass. We choose as proxy variable \( p_\perp(D) \), the momentum of the \( D_{(s)} \) transverse to the \( B^0_{(s)} \) flight direction (see Fig. 6). To keep a good correlation between \( r \) and \( p_\perp(D) \), we select candidates with \( p_\perp(D) \) in the range 500-2500 MeV. This results in an efficiency of 96.48%. Figure 8 shows the distribution of \( p_\perp(D) \) for both decay signals.

The dependence of \( r \) on \( p_\perp(D) \) is computed only on the \( B^0 \rightarrow D^- \mu^+ \nu_\mu \) decay channel in MC, but applied to all data candidates (Fig. 9). Figure 10 shows the \( m_{\text{fit}} \) distributions for \( B^0 \rightarrow D^- \mu^+ \nu_\mu \), \( B^0 \rightarrow D^{*-} \mu^+ \nu_\mu \) and \( B^0 \rightarrow D^{**} \mu^+ \nu_\mu \) denoted Category 1,2 and 3, respectively. Category 3 represents decays where the \( B^0 \) meson decays to higher resonances than \( D^{**} \), e.g. \( D_1(2420) \) or \( D'_1(2430) \) (the full details are in appendix 9.3). The full composition is shown in table 25, in section 5.4. We observe that we have lost all discrimination power between \( B^0 \rightarrow D^- \mu^+ \nu_\mu \) and \( B^0 \rightarrow D^{*-} \mu^+ \nu_\mu \). The reason why these two distributions are no longer disentangled, relatively to the \( B^0 \) visible or corrected mass distributions, is the following. The \( B^0 \) visible mass distributions peaks at lower values for \( B^0 \rightarrow D^{*-} \mu^+ \nu_\mu \) than for \( B^0 \rightarrow D^- \mu^+ \nu_\mu \), as seen in Fig. 7. However, as explained previously, \( r \) is computed on the \( B^0 \rightarrow D^- \mu^+ \nu_\mu \) decay but applied to all candidates. For the \( B^0 \rightarrow D^{*-} \mu^+ \nu_\mu \) candidates, this value of \( r \) is actually overestimated relatively to the true value (Fig. 9). Then when computing \( m_{\text{fit}} = r \cdot m(D\mu) \) for \( B^0 \rightarrow D^{*-} \mu^+ \nu_\mu \) candidates, the lower values of the visible mass, relatively to the \( B^0 \rightarrow D^- \mu^+ \nu_\mu \) distribution, are
compensated by the overestimated values of $r$, thus resulting in two distributions which are centred at almost the same value. We attempt at producing a variable

$$m_{\text{fit}}^{\text{corr}} = r' \cdot m_{\text{corr}};$$

with $r' = \frac{m(D_{\mu\nu})}{m_{\text{corr}}}$. This variable is similar to $m_{\text{fit}}$ where the visible mass is replaced by the corrected mass. Figure 11 shows the two-dimensional histogram $r'(p_\perp(D))$ with the fit to its profile, while figures 10 and 12 show the distributions of $m_{\text{fit}}^{\text{corr}}$ for the three previously defined categories and a comparison between the corrected mass distribution, $m_{\text{fit}}$ and $m_{\text{corr}}^{\text{corr}}$. Despite that the $m_{\text{fit}}^{\text{corr}}$ distribution is less biased and presents a better resolution than both $m_{\text{corr}}$ and $m_{\text{fit}}$, we cannot distinguish between $B^0 \rightarrow D^-\mu^+\nu_\mu$ and $B^0 \rightarrow D^*^-\mu^+\nu_\mu$ decays. Therefore, the corrected mass is still preferable and we shall use it in the rest of the analysis.
4.3 Approximation of $q^2$

Since we are unable to fully reconstruct $q^2$, we want to find a good approximation of its value. By using kinematic constraints similarly to those used to define $m_{\text{corr}}$, assuming that we only miss the neutrino, it is possible to extract an equation for the longitudinal component, along the $B^0_{(s)}$ flight direction, of the neutrino. This is the
standard method to estimate $q^2$. The trouble is that this is a second order equation, thus yielding two solutions. See section 9.2 for the computation. There is no way to know which of these two solutions is the correct one. Therefore, when computing the $q^2$ using the neutrino’s momentum obtained in this way, we are left with a 2-fold ambiguity yielding solutions $q^2_1$ and $q^2_2$. We noticed that the distribution of $q^2_2$ is less biased and has a better resolution than the distribution of $q^2_1$. Therefore, we will compare our results with $q^2_2$. Another problem is that, due to the experimental resolution, some of the solutions are complex-valued and thus cannot be used. This results in a loss of statistics. To be precise, only $\mathcal{O}(80)\%$ of the candidates produce a real-valued $q^2$ in $B^0 \to D^- \mu^+ X$ data.

We want to find an approximation for $q^2$ using a similar idea as previously. We find some variable correlated with the true $q^2$, known in MC simulation; we fit the relationship between both variables and then we apply the fit function to the proxy variable in data and we obtain a value $q^2_{\text{fit}}$. Again we choose $p_\perp(D)$ as our proxy variable. We fit separately $B^0 \to D^- \mu^+ \nu_\mu$ and $B^0 \to D^+ \mu^+ \nu_\mu$ decays. Figure 13 shows the 2D histograms for $B^0 \to D^- \mu^+ \nu_\mu$ and $B^0 \to D^+ \mu^+ \nu_\mu$ decays, in data. Figure 14 shows the momentum distributions for the true value ($q^2_{\text{true}}$), $q^2_{\text{fit}}$ and $q^2_2$ in the $B^0 \to D^- \mu^+ \nu_\mu$ and $B^0 \to D^+ \mu^+ \nu_\mu$ modes. It should be noted that in these two plots, the ranges of $q^2_{\text{true}}$ and $q^2_2$ are limited between the minimum and maximum $q^2_{\text{fit}}$ values. Notice that there are less events in the $q^2_2$ distribution than in the $q^2_{\text{true}}$ distribution because of the complex-valued $q^2_2$ solutions. Figure 15 shows the resolution of the $q^2_{\text{fit}}$ and $q^2_2$ distributions relatively to the true distribution in the $B^0 \to D^- \mu^+ \nu_\mu$ and $B^0 \to D^+ \mu^+ \nu_\mu$ modes. We see that in both cases the distribution for $q^2_{\text{fit}}$ is less biased than for $q^2_2$. The resolution is better for $q^2_{\text{fit}}$ in the $B^0 \to D^- \mu^+ \nu_\mu$ mode while it is a bit worse compared to $q^2_2$ in the $B^0 \to D^+ \mu^+ \nu_\mu$ mode.

4.4 Approximation for $\cos(\theta_l)$

We have seen that the method using proxy variables displaying a correlation with the observable of interest yields promising results for the momentum transfer squared. Therefore, we apply the same procedure for $\cos(\theta_l)$.

This time we choose as proxy variable the transverse momentum of the muon, relatively to the $B^0$ flight direction, $p_\perp(\mu)$. Again, multiple proxy variables have been tested, including $p_\perp(D)$, and $p_\perp(\mu)$ was the one displaying the best correlation with $\cos(\theta_l)$. Figure 16 shows the 2D histogram and the fit to its profile. Figure 17 shows the true and approximated $\cos(\theta_l)$ distributions and the resolution of the fitted $\cos(\theta_l)$. In the left plot of this figure, the true distribution is also limited between the minimum and maximum $\cos(\theta_l)_{\text{fit}}$ values. We observe that the resolution is relatively poor. Currently other investigations are being pursued by the semileptonic working group at LHCb but are not yet conclusive. Our method could be compared to their results.

As a partial conclusion on the investigation of these proxy variables, we have seen that this approximation method gives promising results concerning $q^2$. For
cos(θ₁) the results are less conclusive. Unfortunately, the mass variables $m_{\text{fit}}$ and $m_{\text{corr}}^{m_{\text{fit}}}$ are not useful as discriminating tools between the two signal modes.

5 Measuring the signal yields in data

We now proceed to the first step in the study of the exclusive decays: the measurement of the integrated branching ratios. Indeed, due to the more challenging condition imposed by the partial reconstruction, we do not attempt in this project to measure the differential decay rates, but focus only on the determination of the exclusive total branching ration of the two decay modes. We will count the signal yield of $B_s^0 \to D^-_s \mu^+ \nu_\mu$ and $B_s^0 \to D_s^- \mu^+ \nu_\mu$ through a fit to the $B_s^0$ corrected mass which has been shown in the previous section to provide a discrimination between the two signal modes. The strategy is the following. First, a selection is applied in order to increase the signal purity. Then, we measure the yields of the signal by selecting the $D_s^-$ peak in the $K K \pi$ mass distribution and fitting the $B_s^0$ corrected mass. The signals and physics backgrounds are described by simulated mass templates, while the combinatorial background is derived from data. The method is validated with an high statistic sample of $B^0 \to D^- (\to K^+ \pi^- \pi^-) \mu^+ \nu_\mu X$ decays, and with the reference decays $B^0 \to D^- (\to K^+ K^- \pi^-) \mu^+ \nu_\mu X$, whose composition
5 MEASURING THE SIGNAL YIELDS IN DATA

Figure 14: Distributions of $q^2$ in the $B_0^0 \rightarrow D^- \mu^+ \nu_\mu$ (left) and $B_0^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ (right) modes for the true MC value in blue, $q_{fit}^2$ in red and $q^2$ in green.

Figure 15: Comparison between the resolutions of the $q^2$ distributions in the $B_0^0 \rightarrow D^- \mu^+ \nu_\mu$ (left) and $B_0^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ (right) modes for $q_{fit}^2$ in blue and $q^2$ in red. In both cases, the distribution for $q_{fit}^2$ is less biased than for $q^2$. It presents a better standard deviation in the $B_0^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ case and a slightly worse in the $B_0^0 \rightarrow D^- \mu^+ \nu_\mu$.

is well known. Once the method has been validated, it is applied to the $B_s^0$ samples.

5.1 Selection

We are using the full Run 1 data set collected by the LHCb experiment. This corresponds approximatively to 3 fb$^{-1}$ of integrated luminosity. The selection used is inherited from another analysis [22], which is concerned with the measurement of the $B_s^0$ lifetime. The candidates $B_{(s)}^0 \rightarrow D_{(s)}(\rightarrow K^+K^-\pi^-)\mu^+\nu_\mu X$ decays are reconstructed from the line b2DsPhiPiMuXB2DMuNuX of Stripping20r{0,1}p3, while the $B_0^0 \rightarrow D^- (\rightarrow K^+\pi^-\pi^-)\mu^+\nu_\mu X$ candidates from the line b2DpMuXB2DMuNuX of Stripping20r{0,1}p0. The offline selection consists of tighter cuts relatively to the stripping and of mass vetoes. By “stripping” we mean a first step of the reconstruction and selection of the decay candidates of interest. The stripping and offline selections are presented in tables 3 and 4, respectively. Note that $p_T(X)$ is the transverse momentum of particle $X$ relatively to the $z$-axis. This is not to be confused with $p_L(X)$ defined in section 4.2.

The signal modes are characterised by a $D_{(s)}$ and a muon whose charge are of opposite sign (OS). The stripping lines also reconstruct events where the muon and the charmed meson candidates have the same charge. This same sign (SS) data is useful to model combinatorial background formed by a real $D_{(s)}$ meson and a random muon, as well as random $K^+K^-\pi^-\mu^+$ combinations. The signal candidates
5.1 Selection

Figure 16: Two-dimensional histogram of the true $\cos(\theta_{l})$ as a function of $p_{\perp}(\mu)$ (left) and profile of the histogram (right) for $B^{0} \rightarrow D^{*-}\mu^{+}\nu_{\mu}$. The profile is fitted with a polynomial of order 4. The range of $p_{\perp}(\mu)$ from 500 MeV to 2500 MeV results in a loss of 1.17% of candidates, in MC.

Figure 17: Distributions of the true (blue) and fitted (red) $\cos(\theta_{l})$ (left) and resolution of the fitted $\cos(\theta_{l})$ distribution (right), in the $B^{0} \rightarrow D^{*-}\mu^{+}\nu_{\mu}$ mode.

are formed by requiring a $D^{-}$ mass comprised between 1850 and 1890 MeV, for both $K^{+}K^{-}\pi^{-}$ and $K^{+}\pi^{-}\pi^{-}$ modes; while we require a $D_{s}^{-}$ mass in the range 1940-2000 MeV. The $D_{s}^{-}(s)$ mass distributions after the stripping selection are presented in figure 18 for the $K^{+}K^{-}\pi^{-}$ and $K^{+}\pi^{-}\pi^{-}$ modes\textsuperscript{3}. Some same-sign candidates are peaking at the $D_{s}^{-}(s)$ mass. These candidates are fully reconstructed $D_{s}^{-}\pi^{-}$ which are likely not coming from the decay of a b-hadron. Figure 19 shows the distribution of the $B^{0}_{(s)}$ visible and corrected mass in data after the stripping selection. The visible mass distributions are peaking in the region 3-5 GeV where we expect to observe the b-hadrons signals. Notice as well the discrepancy between same-sign and opposite-sign candidates in this region. The other important feature of these distributions is that in the side-band region i.e. above 5.5 GeV, the same-sign and opposite-sign distributions are similar in shape and decrease exponentially. Finally, narrow candidates\textsuperscript{7} excess is present in the opposite-sign distribution in the range 5200-5400 MeV for the $B^{0}$ candidates and in the range 5280-5480 MeV for the $B^{0}_{s}$ candidates. Those are fully reconstructed $B^{0}_{(s)} \rightarrow D_{s}^{-}\pi^{+}$ decays where the pion either is misidentified as a muon, or decays in-flight. These regions of the visible

\textsuperscript{3}In 2012 data with magnet polarity down for the $K^{+}K^{-}\pi^{-}$ mode, and full 2012 data for $K^{+}\pi^{-}\pi^{-}$ mode. The 2011 and magnet polarity up data present similar features.
### Table 3: Stripping selections for the $K^+K^−\pi^−$ (left) and $K^+\pi^−\pi^−$ (right) samples.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$K^+K^−\pi^−$ requirement</th>
<th>$K^+\pi^−\pi^−$ requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ProbNNghost}(\mu, \pi, K)$</td>
<td>$&lt; 0.5$</td>
<td></td>
</tr>
<tr>
<td>Minimum IP $\chi^2(\mu, \pi, K)$</td>
<td>$&gt; 4.0$</td>
<td>$&gt; 9.0$</td>
</tr>
<tr>
<td>$p_T(\mu)$</td>
<td>$&gt; 600$ MeV</td>
<td>$&gt; 800$ MeV</td>
</tr>
<tr>
<td>$p(\mu)$</td>
<td>$&gt; 3.0$ GeV</td>
<td>$&gt; 2.0$ GeV</td>
</tr>
<tr>
<td>$p_T(K), p_T(\pi)$</td>
<td>$&gt; 150$ MeV</td>
<td>$&gt; 300$ MeV</td>
</tr>
<tr>
<td>$p(K), p(\pi)$</td>
<td>$&gt; 1.5$ GeV</td>
<td>$&gt; 2.0$ GeV</td>
</tr>
<tr>
<td>$\text{PIDmu}(\mu)$</td>
<td>$&gt; 0.0$</td>
<td></td>
</tr>
<tr>
<td>$\text{PIDK}(K)$</td>
<td>$&gt; 4.0$</td>
<td>$&gt; 10.0$</td>
</tr>
<tr>
<td>$\text{PIDK}(\pi)$</td>
<td>$&gt; 4.0$</td>
<td>$&gt; 6.0$</td>
</tr>
<tr>
<td>$\text{Track } \chi^2/dof$</td>
<td>$&gt; 20.0$</td>
<td>$&gt; 100.0$</td>
</tr>
<tr>
<td>$\text{D vertex } \chi^2/dof$</td>
<td>$&gt; 8.0$</td>
<td>$&gt; 1.8$ GeV</td>
</tr>
<tr>
<td>$\text{D daughters’ } \Sigma p_T$</td>
<td>$&gt; 20.0$</td>
<td></td>
</tr>
<tr>
<td>$\text{D DIRA}$</td>
<td>$&gt; 0.99$</td>
<td></td>
</tr>
<tr>
<td>$m(D_s^-)$ (MeV)</td>
<td>$[1789.620; 2048.490]$</td>
<td>$[1789.620; 1949.620]$</td>
</tr>
<tr>
<td>$m(K^+K^-)$ (MeV)</td>
<td>$[979.455; 1059.455]$</td>
<td>$-$</td>
</tr>
<tr>
<td>$B$ vertex $\chi^2/dof$</td>
<td>$&lt; 20.0$</td>
<td>$&lt; 6.0$</td>
</tr>
<tr>
<td>$B$ DIRA</td>
<td>$&gt; 0.99$</td>
<td>$&gt; 0.999$</td>
</tr>
<tr>
<td>$m(D_{cs}\mu)$</td>
<td>$[0.0; 1000.0]$ GeV</td>
<td>$[2.5; 6.0]$ GeV</td>
</tr>
<tr>
<td>$v_z(D) - v_z(B)$</td>
<td>$&gt; -0.3$ mm</td>
<td>$&gt; 0.0$ mm</td>
</tr>
</tbody>
</table>

Figure 18: Distribution of the reconstructed $D_s^-$ mass in the $K^+K^-\pi^-$ (left) and $K^+\pi^-\pi^-$ (right) modes after the stripping selection. The vertical lines enclose the signal regions which are kept in the final selection. Only a subset of the full data sample is shown.

The candidates after stripping are then selected depending on their properties at trigger level. We require the candidates to be L0MuonTOS (Trigger On Signal i.e. we require that the muon candidate has fired the trigger) on the muon; then the muon is required to be TOS in one of the following lines: Hlt1TrackAllL0, Hlt1TrackMuon or Hlt1SingleMuonHighPT. At Hlt2 level, the $B_{cs}^0$ candidate is required to be TOS on one of the following topological lines: Hlt2TopoMu2BodyBDT, Hlt2TopoMu3BodyBDT or Hlt2TopoMu4BodyBDT. The offline requirements, presented in ta-
5.1 Selection

<table>
<thead>
<tr>
<th>Variable</th>
<th>$K^+K^-\pi^-$ requirement</th>
<th>$K^+\pi^-\pi^-$ requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T(K)$, $p_T(\pi)$</td>
<td>$&gt; 300$ MeV</td>
<td>$&gt; 500$ MeV</td>
</tr>
<tr>
<td>$p(K)$</td>
<td>$&gt; 2$ GeV</td>
<td>$&gt; 3$ GeV</td>
</tr>
<tr>
<td>$p(\pi)$</td>
<td>$&gt; 2$ GeV</td>
<td>$&gt; 5$ GeV</td>
</tr>
<tr>
<td>ProbNNk(K)</td>
<td></td>
<td>$&gt; 0.2$</td>
</tr>
<tr>
<td>ProbNNpi(\pi)</td>
<td></td>
<td>$&gt; 0.5$</td>
</tr>
<tr>
<td>ProbNNmu(\mu)</td>
<td></td>
<td>$&gt; 0.2$</td>
</tr>
<tr>
<td>$D$ vertex $\chi^2$/dof</td>
<td></td>
<td>$&lt; 6.0$</td>
</tr>
<tr>
<td>$m(D^-)$</td>
<td>$\in [1850; 1890]$ MeV</td>
<td></td>
</tr>
<tr>
<td>$m(D_s^-)$</td>
<td>$\in [1940; 2000]$ MeV</td>
<td>$-$</td>
</tr>
<tr>
<td>$m(K^+K^-)$</td>
<td>$\in [1008; 1032]$ MeV</td>
<td>$-$</td>
</tr>
<tr>
<td>$t_D$</td>
<td>$&gt; 0.1$ ps</td>
<td></td>
</tr>
<tr>
<td>$m_{corr}$</td>
<td>$\in [3000; 8500]$ MeV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$&gt; 3.1$ GeV</td>
<td></td>
</tr>
<tr>
<td>$m(D_s^-\mu)$</td>
<td>$\notin [5.200; 5.400]$ GeV (for $B^0$)</td>
<td>$\notin [5.200; 5.400]$ GeV</td>
</tr>
<tr>
<td></td>
<td>$\notin [5.280; 5.480]$ GeV (for $B^0_s$)</td>
<td></td>
</tr>
<tr>
<td>$m(\mu^+\mu^-)$</td>
<td>$\notin [3.040; 3.160]$ GeV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\notin [3.635; 3.735]$ GeV</td>
<td></td>
</tr>
<tr>
<td>$m(Kp\pi)$</td>
<td>$\notin [2.260; 2.310]$ GeV</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Offline selections for the $K^+K^-\pi^-$ (left) and $K^+\pi^-\pi^-$ (right) samples.

ble 4, are then imposed to the surviving candidates. The offline requirements are chosen such as to reduce the background under the signal peaks and to minimize the differences between same-sign and opposite-sign distributions in the side-band region. As explained previously, the idea is to use the same-sign data as a model for the combinatorial background, and as such we expect both distributions to be very similar in the side-bands, where only combinatorial background is expected. Additional tests have been done in the $B^0_s$ lifetime analysis in Ref. [22] to validate the SS data as a proxy of the combinatorial background.

There are also background sources coming from misreconstructed decays. These decays are the following:

\[
B^0_s \rightarrow \psi(\rightarrow \mu^+\mu^-)\phi(\rightarrow K^+K^-), \quad (19)
\]

\[
\Lambda^0_b \rightarrow \Lambda^+_c (\rightarrow p K^-\pi^+) \mu^-\bar{\nu}_\mu X, \quad (20)
\]

where $\psi$ from the decay in Eq. (19) denotes either $J/\psi$ or $\psi(2S)$ and one of the muons is misidentified as a pion. In the decay of Eq. (20) the proton is misidentified as a kaon for decays in the $K^+K^-\pi^-$ mode, while being misidentified as a pion for decays in the $K^+\pi^-\pi^-$ mode. These background decays are suppressed through mass vetoes. In order to suppress the decay from Eq. (19), we assign the muon’s mass to the pion candidate and veto the $J/\psi$ and $\psi(2S)$ masses in the invariant mass distribution $m(\mu\mu[\pi])$, see Fig. 20. Similarly, for the decay in Eq. (20) we assign the proton’s mass to the kaon candidate in the $K^+K^-\pi^-$ mode and to the pion candidate in the $K^+\pi^-\pi^-$ mode and then veto the $\Lambda^+_c$ mass in the $m(Kp[K/\pi]\pi)$...
Figure 19: Distribution of the $B_{s}^{0}$ visible (left column) and corrected (right column) mass after the stripping selection for $B_{s}^{0} \rightarrow D_{s}^{-} \rightarrow K^{+}K^{-}\pi^{-})\mu^{+}\nu_{\mu}X$ (top), $B^{0} \rightarrow D^{-} \rightarrow K^{+}K^{-}\pi^{-})\mu^{+}\nu_{\mu}X$ (middle) and $B^{0} \rightarrow D^{-} \rightarrow K^{+}K^{-}\pi^{-})\mu^{+}\nu_{\mu}X$ (bottom) candidates. Only a subset of the full data sample is shown.

We also require that the $K^{+}K^{-}$ pair comes from the decay of a $\phi(1020)$ by requiring a $m(KK)$ invariant mass comprised between 1008 and 1032 MeV for decays in the $K^{+}K^{-}\pi^{-}$ mode, see Fig. 21.

Figures 23 and 24 show the $m(D_{s}^{-})$, $m(D_{s}^{-}\mu)$ and $m_{corr}$ mass distributions in the $K^{+}K^{-}\pi^{-}$ and $K^{+}\pi^{-}\pi^{-}$ modes for candidates after full selection. The final number of signal candidates after the full selection is: $4.44 \cdot 10^{5}$ for $B_{s}^{0} \rightarrow D_{s}^{-} \rightarrow K^{+}K^{-}\pi^{-})\mu^{+}\nu_{\mu}X$ decays, $1.39 \cdot 10^{5}$ for $B^{0} \rightarrow D^{-} \rightarrow K^{+}K^{-}\pi^{-})\mu^{+}\nu_{\mu}X$ decays and $2.59 \cdot 10^{6}$ for $B^{0} \rightarrow D^{-} \rightarrow K^{+}\pi^{-}\pi^{-})\mu^{+}\nu_{\mu}X$ decays. Figure 25 shows the distributions of the $B_{s}^{0}$ visible and corrected mass for candidates restricted to the $D_{s}^{-}$ sidebands, i.e. in the range $m(KK\pi) \in [1.80; 1.85] \cup [1.89; 1.94] \cup [2.00; 2.03]$ GeV for the $K^{+}K^{-}\pi^{-}$ mode and $m(K\pi\pi) \in [1.80; 1.85] \cup [1.89; 1.95]$ GeV for the $K^{+}\pi^{-}\pi^{-}$ mode. We observe that for high visible and corrected masses, the opposite-sign and same-sign distributions are very similar both in shape and normalization. In the signal region (Fig. 24) however, there is still some discrepancy between OS and SS distributions at high masses. This is probably due to some unaccounted physics background. Nevertheless, the shapes are similar and we do not fix the normalization in the fit to the corrected mass, as explained in the section 5.4. Therefore, we consider such level of agreement between OS and SS distributions to be sufficient in
5.2 Suppressing low-mass physics backgrounds

Along with the background decays described in the previous section, which can be removed using mass vetoes, there are also other sources of physics background from $b$-hadrons which decay either semileptonically to a final state containing a $D_{s}^{−}(\rightarrow K^{+}K^{−}\pi^{−})\mu^{+}\nu_{\mu}X$ (top left), $B_{s}^{0} \rightarrow D^{−}(\rightarrow K^{+}K^{−}\pi^{−})\mu^{+}\nu_{\mu}X$ (top right) and $B_{s}^{0} \rightarrow D^{−}(\rightarrow K^{+}\pi^{−}\pi^{−})\mu^{+}\nu_{\mu}X$ (bottom). The vertical lines enclose the regions removed by the requirement. Only a subset of the full data sample is shown.

In order to take the same-sign distribution as a data-driven model for the combinatorial background. The effect of the residual differences can be included in the systematic uncertainties.

5.2 Suppressing low-mass physics backgrounds

Along with the background decays described in the previous section, which can be removed using mass vetoes, there are also other sources of physics background from $b$-hadrons which decay either semileptonically to a final state containing a $D_{s}^{−}$ meson and a muon, or to a final state containing a pair of charmed mesons, where one of them is fully reconstructed while the other decays semileptonically. Such background decays, described in detail in section 5.4, peak at low corrected masses. We would like to find a requirement which allows the suppression of a significant part of these backgrounds while still retaining a good proportion of signal. We should also take care not to sculpt too much the mass distributions, since we want to keep the discrimination power between the different components when doing the fit to the corrected mass. Such a requirement has been implemented in the $B_{s}^{0}$ lifetime analysis [22]. They reject all candidates such that:

$$p_{\perp}(D_{s}^{−}) > 1500 + 1.1 \cdot (m_{\text{corr}} - 4500)$$  \hspace{1cm} (21)$$

where $p_{\perp}(D_{s}^{−})$ and $m_{\text{corr}}$ are expressed in MeV. Using this requirement they keep $\mathcal{O}(90\%)$ signal while suppressing $\mathcal{O}(50\%)$ background, in MC simulation. We want to suppress even more background relatively to this.

In the two-dimensional histogram of $p_{\perp}(D_{s}^{−})$ and $m_{\text{corr}}$ we define an ellipse centred in $(x_0; y_0) = (4200 \text{ MeV}; 1500 + 1.1 \cdot (x_0 - 4500) \text{ MeV})$ with major and mi-
nor semi-axes \( a = 850 \) MeV and \( b = 150 \) MeV, respectively. The ellipse is rotated counter-clockwise by an angle \( \theta = \arctan(1.1) \). All the candidates falling inside this ellipse are rejected. In addition, the candidates where \( m_{\text{corr}} < 4800 \) MeV and the relation in Eq. (21) is fulfilled are also rejected. These two requirements are applied to the candidates where a \( D_s^- \) is reconstructed. In order to simplify the analysis, we want to keep similar efficiencies between numerator and denominator of equations (9) and (10). Since there is a mass difference of \( m(B^0_s) - m(B^0) = (87.35 \pm 0.20) \) MeV [3] between the two bottom mesons, we cannot treat in the same way the candidates where a \( D_s^- \) is reconstructed from those where a \( D^- \) is reconstructed. Therefore for the \( D^- \) candidates, we shift the requirements to the left in the \( (m_{\text{corr}}; p_\perp(D^-)) \) plane by a value corresponding to the mass difference between the two bottom mesons, \( i.e. \) the ellipse is centred in \( (x'_0; y'_0) = (4200 - 87.35 \text{ MeV}; 1500 + 1.1 \cdot (x'_0 - 4500) \text{ MeV}) \). We also remove candidates where \( m_{\text{corr}} < 4800 \) MeV and \( p_\perp(D^-) > 1500 + 1.1 \cdot (m_{\text{corr}} + 87.35 - 4500) \). Figure 26 shows the effect of the requirements for the background decays \( B^- \to D_s^{(*)} - K^- \mu^- \) and \( \Lambda^0_s \to \Lambda_s^+ D_s^{(*)+}(\pi^0) \) in MC. We see that the elliptical requirement encloses the peaking region of the \( (m_{\text{corr}}; p_\perp(D^-)) \) distribution. For the background decay \( B^- \to D_s^{(*)} - K^- \mu^- \), the efficiency in MC of the old requirement of Eq. (21) was \( \mathcal{O}(60\%) \) while for the two new requirements it is \( \mathcal{O}(46\%) \). Therefore the background rejection, relatively to the previous requirement, is significantly improved. The distributions for the other background components show similar features. Figure 27 shows the \( (m_{\text{corr}}; p_\perp(D^-)) \) distributions for the signal decays \( B^0_s \to D_s^{(*)} - \mu^+ \nu_\mu \) in the \( K^+ K^- \pi^- \) mode. The requirement does not spoil the difference in the signal
5.3 Cleaning the right tail of the signal

At high corrected mass values, only combinatorial background is expected. In this region, due to the poor resolution on the corrected mass, especially for candidates with smaller flight distance, the signal can exhibit a long tail on the right, which might be confused with part of the combinatorial background in the fit of the sample.
composition described in the next subsection. To remove candidates with poor mass resolution, the additional requirement is imposed:

$$m_{\text{corr}}^{\text{err}} < 0.06 \cdot m_{\text{corr}}$$  \hspace{1cm} (22)$$

where $m_{\text{corr}}^{\text{err}}$ is the error on the corrected mass for a given candidate. This requirement results in a $O(80\%)$ signal efficiency for all decay modes. This requirement removes combinatorial background, with an $O(70\%)$ efficiency for the $K^+ K^- \pi^-$ mode and $O(40\%)$ for the $K^+ \pi^- \pi^-$ mode. The effect of this requirement on the corrected mass distribution for the $B^0_{(s)} \rightarrow D^+_{(s)} \mu^+ \nu_{\mu} X$ and SS samples, in the 2011 dataset, can be observed on Fig. 28. The $B^0_{(s)} \rightarrow D^{*+}_{(s)} \mu^+ \nu_{\mu}$ mode as well as the 2012 dataset distributions are plotted in figures 40 and 41 in the appendix, section 9.4. We see that this requirement modifies the shape of the corrected mass distributions by reducing the yield at high $m_{\text{corr}}$ values. This effect is most pronounced in the $K^+ \pi^- \pi^-$ sample.
5.4 Fit of the corrected $B$ mass

In order to obtain the event yields of equations (9) and (10), we make a fit of the $B^0_s$ corrected mass. The purpose of the fit is to know the precise composition of the data sample. The signal and physics background components are described using MC simulation, while we use SS data to describe combinatorial background. Table 5 describes the different MC simulated decays used in the fit to the data, including the expected physics backgrounds. Table 25 in the appendix (section 9.3) describes in detail the different decays included in the MC simulation of the inclusive $B^0_s \rightarrow D_s^- (\rightarrow K^+K^-\pi^-)\mu^+\nu\mu X$ and $B^0 \rightarrow D^- (\rightarrow K^+K^-\pi^-)\mu^+\nu\mu X$ signals. We apply the full selection to the data as well as to the MC simulated data. We then construct histogram templates of the corrected mass distribution for the MC simulated and SS data components. The corrected mass distribution of data is fitted using a binned least-squares fit. The $\chi^2$ variable used to fit takes into account the limited statistics of the different templates:

$$\chi^2 = \sum_{i=1}^{N_{\text{bin}}} \frac{(n_i - p_i)\bar{f}^2}{\sigma^2_{n_i} + \sigma^2_{p_i}},$$

where $n_i$ and $p_i$ are the number of observed and predicted events, respectively, in the bin $i$; $\sigma_{n_i}$ and $\sigma_{p_i}$ are their associated statistical uncertainties and $N_{\text{bin}}$ is the total number of bins. The number of predicted events in bin $i$ is given by:

$$p_i = N_{\text{evt}} \sum_{j=1}^{N_{\text{comp}}} f_j h_{ji},$$

where $f_j$ is the $j$-th fraction associated to the component described by the histogram template $h_j$, which contributes with $h_{ji}$ candidates to the bin $i$; $N_{\text{comp}}$ is the total
number of components entering the fit. The histograms $h_j$ are normalized to unity in the full range of the corrected mass. The total number of events in data $N_{\text{evt}}$, is a normalization constant. The fit fractions are constrained to sum to one:

$$\sum_{i=1}^{N_{\text{comp}}} f_i = 1.$$  \hspace{1cm} (25)$$

In the case where the data is weighted, $n_i$ is the sum of weights and $\sigma^2_{n_i}$ is the sum of the squared weights, in the $i$-th bin. Eq. (23) is then minimised relatively to the set of free parameters $\vec{f} = \{f_1, \ldots, f_{N_{\text{comp}} - 1}\}$, which includes all fit fractions except one. The MC simulated decays composing the inclusive signal components $B^0_s \rightarrow D^-_s (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X, B^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X$ and $B^0 \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \mu^+ \nu_\mu X$, see Tab. 25, are regrouped in four distinct categories for the fit: the decays $B^0_{(s)} \rightarrow D^-_{(s)} \mu^+ \nu_\mu; B^0_{(s)} \rightarrow D^+_{(s)} \mu^+ \nu_\mu$, where the $D^{*-}_{(s)}$ decays to a $D^-_{(s)}$ and either a $\gamma$ or a $\pi^0; B^0_{(s)} \rightarrow D^{**+}_{(s)} \mu^+ \nu_\mu$, where $D^{**+}_{(s)}$ denotes the higher resonances than $D^{*-}_{(s)}$. This decay will be referred to as $B^0_{(s)} \rightarrow D^-_{(s)} \mu^+ \nu X$. Finally, we consider the inclusive semitauonic decays, where the tau lepton decays to a muon and a pair of neutrinos, which are referred to as $B^0_{(s)} \rightarrow D^-_{(s)} \tau^+ \nu X$. They are also referred to as category 1, 2, 3 and 4. Table 6 gives the proportion of each category relatively to the total signal in MC, after full selection.

As explained in the previous section, there are $b$-hadron decays which result in a similar final state as the signal and peak at low visible and corrected masses. Two requirements were specifically developed in order to reduce the contribution of these physics backgrounds. Nonetheless, a small fraction can survive in data and therefore we use MC simulated templates in order to include them in the fit. The contribution of these backgrounds relatively to the signal can be estimated in MC.
5.4 Fit of the corrected $B$ mass

<table>
<thead>
<tr>
<th>Decay type</th>
<th>Decay</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_0^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X$</td>
<td>Signals</td>
</tr>
<tr>
<td>$B_0^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X$</td>
<td></td>
</tr>
<tr>
<td>$B_0^0 \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \mu^+ \nu_\mu X$</td>
<td></td>
</tr>
</tbody>
</table>

$B_s^0 \rightarrow D_s^{(s)}_s D_s^{(s)}_s$  
$\Lambda^0_s \rightarrow \Lambda^{(s)}_s D_s^{(s)}_s (\rho^0)$  
$B^+ \rightarrow D_s^{(s)}_s D_s^{(s)}_s$  
$B^- \rightarrow D_s^{(s)}_s (\rightarrow D_s^{(s)}_s K^-) \mu^- \nu$  
$B^0 \rightarrow D_s^{(s)}_s (\rightarrow D_s^{(s)}_s K S) \mu^- \nu$  
$B_s^0 \rightarrow D_s^0 D_s^0 K$  

$B^+ \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \pi^+ \nu X$  
$B^+ \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \pi^+ \nu X$  

Table 5: Monte Carlo simulated decays used in the fit to the data. The top of the table contains the inclusive signal components. The exact composition of the simulated signal components is presented in table 25. The middle and bottom of the table present the physics backgrounds expected to contribute to the $B_0^0$ and $B^0$ samples, respectively.

<table>
<thead>
<tr>
<th>Signal category</th>
<th>Relative fraction [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_0^0$</td>
<td>$B_0^0(K^+ K^- \pi^-)$</td>
</tr>
<tr>
<td>$B_0^0(s) \rightarrow D_s^{(s)}(\rightarrow \mu^+ \nu)$</td>
<td>27.13 ± 0.11</td>
</tr>
<tr>
<td>$B_0^0(s) \rightarrow D_s^{(s)}(\rightarrow \mu^+ \nu X)$</td>
<td>62.04 ± 0.17</td>
</tr>
<tr>
<td>$B_0^0(s) \rightarrow D_s^{(s)}(\rightarrow \mu^+ \nu \nu)$</td>
<td>10.20 ± 0.07</td>
</tr>
<tr>
<td>$B_0^0(s) \rightarrow D_s^{(s)}(\rightarrow \mu^+ \nu \nu X)$</td>
<td>0.63 ± 0.02</td>
</tr>
</tbody>
</table>

Table 6: Proportion of each of signal category relatively to the total signal in MC 2011, after full selection, for $B_0^0 \rightarrow D_s^{(s)}(\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X$, $B_0^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X$ and $B^0 \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \mu^+ \nu_\mu X$ decays.
for each component in the following way:

$$\frac{f_{\text{bkg}}}{f_{\text{sgn}}} = \frac{\epsilon_{\text{bkg}}}{\epsilon_{\text{sgn}}} \frac{f_{u,d,s,\Lambda}}{f_{d,s}} \times \frac{B(B \to D_{(s)}D_{(s)}K \mu \nu, \ldots)B(D_{(s)} \to \mu X)B(D_{(s)} \to \phi \pi, \phi \to KK)}{B(B_{(s)} \to D_{(s)} \mu X)B(D_{(s)} \to K K \pi)}, \quad (26)$$

where the relevant branching ratios $B$ are used, depending on the decay; $\frac{\epsilon_{\text{bkg}}}{\epsilon_{\text{sgn}}}$ is the ratio of the full efficiency between background and signal, and $f_{u,d,s,\Lambda}$ is the different fragmentation fractions. The decays of the $D_{(s)}$ mesons are forced to proceed through an intermediate $\phi$ state. In the case of $B \to D_{(s)}D_{(s)}$ decay types, one of the $D_{(s)}$ mesons is required to decay semi-leptonically. The branching ratios are taken from [3], while we take $f_u/f_s = f_d/f_s = 3.86 \pm 0.22$ and $f_{\Lambda}/f_s = 2.34 \pm 0.31$. These values are based on the LHCb determinations [14] and [15]. Table 7 shows $f_{\text{bkg}}/f_{\text{sgn}}$ for the background decays contributing to the $B^0_s \to D^-_s \to K^+ K^- \pi^- \mu^+ \nu_\mu X$ sample. In the case of the $B^0 \to D^- (\to K^+ K^- \pi^-) \mu^+ \nu_\mu X$ and $B^0 \to D^- (\to K^+ \pi^-) \mu^+ \nu_\mu X$ samples, the only significant physics background is $B^+ \to D^+ \mu X$ where $f_{\text{bkg}}/f_{\text{sgn}}$ is equal to $(11.43 \pm 2.13)\%$ and $(13.91 \pm 2.59)\%$, respectively. The other background decays’ contributions are negligible and can be treated as systematic uncertainty.

Notice that the requirements imposed to reduce the low mass backgrounds are effective, especially in the case of the $B^0_s \to D^-_s (\to K^+ K^- \pi^-) \mu^+ \nu_\mu X$ sample, where the contamination of the background relatively to the signal is less than one percent.

There is however a difficulty arising in the fit. Some components cannot be disentangled. Fig. 29 shows the mass templates for the different components in-

---

Figure 27: Two-dimensional histogram $p_{\perp}(D^-_s)$ as a function of $m_{\text{corr}}$ for the signal decays $B^0_s \to D^-_s \mu^+ \nu_\mu$, $B^0 \to D^-_s \mu^+ \nu_\mu$ (top row) and for $B^0 \to D^- \mu^+ \nu_\mu$, $B^0 \to D^+ \mu^+ \nu_\mu$ (bottom row). The elliptical and linear requirements are plotted in red and black, respectively. The right part of the figure shows the normalized distributions of $m_{\text{corr}}$ without requirement (blue) and with both new requirements (green).
5.4 Fit of the corrected $B$ mass

Figure 28: Normalized distribution of the corrected mass before (blue) and after (red) applying the $m^{corr}_{corr} < 0.06 \cdot m_{corr}$ requirement, for category 1 (left column) and SS (right column) 2011 samples for the $B^0 \rightarrow D^- (\rightarrow K_+ K^- \pi^-) \mu^+ \nu \mu X$ (top), $B^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu \mu X$ (middle) and $B^0 \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \mu^+ \nu \mu X$ modes.

cluded in the fit to the $B^0$ samples. We notice that the decays $B^0 \rightarrow D^- \mu^+ \nu X$ and $B^+ \rightarrow D^+ \mu^+ X$ have very similar distributions and therefore cannot be distinguished when doing the mass fit. For this reason, a single template describing both components is used in the fit. This template is obtained by summing both components with weights according to the expected fractions relatively to the signal. Similarly for the $B^0_d$ sample, some components have similar shapes, as shown in Fig. 30. The decays $B^0 \rightarrow D^{(*)+} D^{(*)-}$, $B^+ \rightarrow D^{(*)+} D^{(*)0}$, $\Lambda^0 \rightarrow D^{(*)-} \Lambda^+_c (\pi^0)$, $B^0_s \rightarrow D^{(*)-} D^{(*)+}$ and $B^0 \rightarrow D^{(*)+} (D^+) X$ cannot be separated and are therefore summed into a single template, with weights according to their expected fractions relatively to the signal. In the sum we only consider decays expected to contribute to more than 0.5% of the signal yield. Decays $\Lambda^0 \rightarrow D^{(*)-} \Lambda^+_c (\pi^0)$ and $B^0_s \rightarrow D^{(*)-} D^{(*)+}$ are thus neglected. The effect of this removal can be treated as a systematic uncertainty. Similarly, the decays $B^+ \rightarrow D^{(*)+} K^+ \mu^+ \nu$, $B^0 \rightarrow D^{(*)-} K^0_S \mu^+ X$ and $B^0_s \rightarrow D_s^- \tau^+ (\rightarrow \mu^+ \nu \nu) \nu X$ present comparable distributions and are therefore summed in a single histogram. Figure 31 shows the mass distributions of the components used in the fit: signal category 1 and 2, both physics background types and the SS distribution which models combinatorial background. Their mass distributions differ enough such that they can be reliably estimated by the fit.
### 5 MEASURING THE SIGNAL YIELDS IN DATA

<table>
<thead>
<tr>
<th>Category</th>
<th>Decay</th>
<th>( \mathcal{B} ) [10^{-4}]</th>
<th>( f_{q,A}/f_s )</th>
<th>( f_{bkg}/f_{sym} ) [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B \to DD )</td>
<td>( B^0 \to D^{(<em>)-}D^{(</em>)+} )</td>
<td>12.74 ± 1.60</td>
<td>3.86 ± 0.22</td>
<td>0.70 ± 0.23</td>
</tr>
<tr>
<td>( B \to DD )</td>
<td>( B^+ \to D^{(<em>)0}D_s^{(</em>)+} )</td>
<td>11.36 ± 1.29</td>
<td>3.86 ± 0.22</td>
<td>0.68 ± 0.23</td>
</tr>
<tr>
<td>( B \to DD )</td>
<td>( B_s^0 \to D_s^{(<em>)-}D_s^{(</em>)+} )</td>
<td>12.17 ± 3.93</td>
<td>1</td>
<td>0.37 ± 0.16</td>
</tr>
<tr>
<td>( B \to DK\mu\nu )</td>
<td>( B^- \to D_s^{(*)+}K^-\mu^-X )</td>
<td>6.10 ± 1.00</td>
<td>3.86 ± 0.22</td>
<td>0.80 ± 0.28</td>
</tr>
<tr>
<td>( B \to DK\mu\nu )</td>
<td>( B^0 \to D_s^{(*)-}K_{s0}^0\mu^+X )</td>
<td>6.10 ± 1.00</td>
<td>3.86 ± 0.22</td>
<td>0.75 ± 0.26</td>
</tr>
<tr>
<td>( B \to DDK )</td>
<td>( B_s^0 \to D^0D_s^-K^+ )</td>
<td>0.24 ± 0.09</td>
<td>1</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>( \Lambda_b )</td>
<td>( \Lambda_b^0 \to \Lambda_b^0D_s^{(*)-}(\pi^0) )</td>
<td>4.31 ± 1.69</td>
<td>2.34 ± 0.31</td>
<td>0.13 ± 0.07</td>
</tr>
</tbody>
</table>

Table 7: Background decays contributing to the \( B_s^0 \to D_s^- (\to K^+K^-\pi^-)\mu^+\nu_X \) 2011 sample split in four categories.

**Figure 29:** \( B^0 \) corrected mass distributions used in the fit to the signal decays \( B^0 \to D^- (\to K^+K^-\pi^-)\mu^+\nu_X \) (left) and \( B^0 \to D^- (\to K^+\pi^-\pi^-)\mu^+\nu_X \) (right).

### 5.5 Results

The samples have to be split and the fits made separately for the 2011 and 2012 datasets, since the conditions of data acquisition were different. Indeed, the centre-of-mass energy of the \( pp \) collision is \( \sqrt{s} = 7 \) TeV in 2011 and \( \sqrt{s} = 8 \) TeV in 2012. This leads to different production cross-sections of a \( b \) quark, fragmentation fractions ratios \( f_s/f_d \) and it might result in different efficiencies for the full set of selection requirements, which are adjusted (specially online) to cope with the different conditions.

The fit is first performed in the \( B^0 \) samples in order to validate the method and then applied to the \( B_s^0 \) sample. The fits for the \( B^0 \to D^- (\to K^+K^-\pi^-)\mu^+\nu_X \) and \( B^0 \to D^- (\to K^+\pi^-\pi^-)\mu^+\nu_X \) samples are presented in Fig. 32. The fit results are described in Tab. 8 where they are compared with the MC expectations. In the \( B^0 \) samples, the \( B^0 \to D^- \tau^+\nu X \) component has been neglected. Notice also that the expected fractions of Tab. 8 do not sum to one minus the combinatorial fraction, since the expected fraction of the neglected \( B^0 \to D^- \tau^+\nu X \) component (\( \sim 0.60\% \)) has not been included in the table. The fitted fractions are in agreement with these expected fractions. The \( \chi^2/dof \) for all samples are presented in Tab. 10.
5.5 Results

The values for both $B^0$ samples are mostly in agreement with the MC expectations and thus build confidence on the method. The fit projections for the $B^0 \rightarrow D_s^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X$ sample is presented in Fig. 33. Tables 8 and 9 report the fit results for the $B^0$ and the $B^0_s$ samples, respectively, for 2011 and 2012 datasets.

The yields ratios defined in equations (9) and (10) are given in Tab. 11. The yields are obtained by multiplying the fractions obtained in the fit of the corrected mass, presented in tables 8 and 9, with the total number of events in the dataset. In order to compare 2011 and 2012 values for a given yields ratio, we do the following:

$$\frac{|r_{2011} - r_{2012}|}{\sqrt{\sigma^2_{r_{2011}} + \sigma^2_{r_{2012}}}},$$  \hfill (27)

where $r_X$ is the yields ratio for the year $X$ and $\sigma_{r_X}$ is its associated statistical uncertainty. Finally, $N_{B_s}/N_B$ and $N_{B_s}^*/N_B^*$ display 1.66$\sigma$ and 0.95$\sigma$ agreements between 2011 and 2012 datasets, respectively.
5 MEASURING THE SIGNAL YIELDS IN DATA

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Component</th>
<th>$K^+K^-\pi^-$ Fit fraction [%]</th>
<th>Expectation [%]</th>
<th>$K^+\pi^-\pi^-$ Fit fraction [%]</th>
<th>Expectation [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>$B^0 \rightarrow D^-\mu^+\nu$</td>
<td>46.75 ± 1.34</td>
<td>45.46 ± 2.80</td>
<td>53.34 ± 0.84</td>
<td>48.74 ± 3.07</td>
</tr>
<tr>
<td></td>
<td>$B^0 \rightarrow D^+\mu^-\nu$</td>
<td>31.30 ± 1.76</td>
<td>30.74 ± 1.11</td>
<td>28.01 ± 1.20</td>
<td>32.27 ± 1.25</td>
</tr>
<tr>
<td></td>
<td>$B^0/B^+ \rightarrow D^-\mu^+\nu X$</td>
<td>12.94 ± 0.95</td>
<td>14.19 ± 2.34</td>
<td>17.64 ± 0.58</td>
<td>17.49 ± 2.90</td>
</tr>
<tr>
<td></td>
<td>Combinatorial</td>
<td>9.01 ± 0.88</td>
<td>–</td>
<td>0.93 ± 0.06</td>
<td>–</td>
</tr>
<tr>
<td>2012</td>
<td>$B^0 \rightarrow D^-\mu^+\nu$</td>
<td>48.80 ± 0.82</td>
<td>48.59 ± 2.75</td>
<td>52.24 ± 0.64</td>
<td>51.74 ± 2.98</td>
</tr>
<tr>
<td></td>
<td>$B^0 \rightarrow D^-\mu^-\nu$</td>
<td>31.41 ± 1.12</td>
<td>33.65 ± 0.90</td>
<td>29.86 ± 0.91</td>
<td>34.40 ± 0.98</td>
</tr>
<tr>
<td></td>
<td>$B^0/B^+ \rightarrow D^-\mu^+\nu X$</td>
<td>13.17 ± 0.57</td>
<td>10.49 ± 1.68</td>
<td>17.31 ± 0.45</td>
<td>12.63 ± 2.02</td>
</tr>
<tr>
<td></td>
<td>Combinatorial</td>
<td>6.62 ± 0.39</td>
<td>–</td>
<td>0.60 ± 0.06</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 8: Result of the fit of the $B^0$ corrected mass for the $K^+K^-\pi^-$ and $K^+\pi^-\pi^-$ samples for 2011 (top) and 2012 (bottom) datasets. The expected fractions are indicated as well. The $B^0 \rightarrow D^-\tau^+\nu X$ component has been neglected in the fits of both samples. The total number of events are: $N_{tot}^{K^+K^-\pi^-} = 31706\pm 178$ in the $K^+K^-\pi^-$ sample and $N_{tot}^{K^+\pi^-\pi^-} = 553922\pm 744$ in the $K^+\pi^-\pi^-$ sample in 2011; $N_{tot}^{K^+K^-\pi^-} = 80499\pm 284$ in the $K^+K^-\pi^-$ sample and $N_{tot}^{K^+\pi^-\pi^-} = 1450760\pm 1204$ in the $K^+\pi^-\pi^-$ in 2012.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Component</th>
<th>Fit fraction [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>$B^0_s \rightarrow D^- \mu^+\nu$</td>
<td>32.88 ± 0.92</td>
</tr>
<tr>
<td></td>
<td>$B^0_s \rightarrow D^- \mu^-\nu$</td>
<td>55.40 ± 1.68</td>
</tr>
<tr>
<td></td>
<td>$B^0_s \rightarrow D^-\mu^+(D_s)X$</td>
<td>4.91 ± 1.26</td>
</tr>
<tr>
<td></td>
<td>$B^0_s \rightarrow D^-\mu^+(K\mu\nu)(\tau\nu)$</td>
<td>2.43 ± 0.38</td>
</tr>
<tr>
<td></td>
<td>Combinatorial</td>
<td>4.39 ± 0.29</td>
</tr>
<tr>
<td>2012</td>
<td>$B^0_s \rightarrow D^- \mu^+\nu$</td>
<td>31.70 ± 0.71</td>
</tr>
<tr>
<td></td>
<td>$B^0_s \rightarrow D^- \mu^-\nu$</td>
<td>59.98 ± 1.34</td>
</tr>
<tr>
<td></td>
<td>$B^0_s \rightarrow D^-\mu^+(D_s)X$</td>
<td>1.56 ± 1.03</td>
</tr>
<tr>
<td></td>
<td>$B^0_s \rightarrow D^-\mu^+(K\mu\nu)(\tau\nu)$</td>
<td>3.27 ± 0.31</td>
</tr>
<tr>
<td></td>
<td>Combinatorial</td>
<td>3.49 ± 0.15</td>
</tr>
</tbody>
</table>

Table 9: Result of the fit of the $B^0_s$ corrected mass for the $B^0_s \rightarrow D^-\rightarrow K^+K^-\pi^-\mu^+\nu X$ sample in the 2011 (top) and 2012 (bottom) datasets. The total number of events are: $N_{tot} = 102916\pm 321$ in 2011 and $N_{tot} = 260228\pm 510$ in 2012.

<table>
<thead>
<tr>
<th>Sample</th>
<th>2011 $\chi^2/dof$</th>
<th>p-value [%]</th>
<th>2012 $\chi^2/dof$</th>
<th>p-value [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^0 \rightarrow D^-\rightarrow K^+K^-\pi^-\mu^+\nu X$</td>
<td>95.44/88</td>
<td>27.57</td>
<td>98.92/89</td>
<td>22.16</td>
</tr>
<tr>
<td>$B^0 \rightarrow D^-\rightarrow K^+\pi^-\pi^-\mu^+\nu X$</td>
<td>100.93/88</td>
<td>16.36</td>
<td>99.72/89</td>
<td>20.54</td>
</tr>
<tr>
<td>$B^0_s \rightarrow D^-\rightarrow K^+K^-\pi^-\mu^+\nu X$</td>
<td>92.90/88</td>
<td>33.98</td>
<td>110.11/88</td>
<td>5.55</td>
</tr>
</tbody>
</table>

Table 10: Results of the fits of all samples: $\chi^2/dof$ and its associated p-value.
6 Computation of the efficiencies

In the previous section, the event yields ratios of equations (9) and (10) have been obtained through the fits to the corrected mass distributions, see Tab. 11. The following step is to compute the efficiencies ratios \( \epsilon_s/\epsilon_d \) and \( \epsilon_s^*/\epsilon_d^* \). The different requirements leading to the full selection have been tuned such as to minimize the differences between numerator and denominator in order to simplify the analysis, in particular to suppress systematic uncertainties in the ratio of efficiencies. The requirements leading to the final selection can be separated into multiple steps:

\[
\epsilon_{\text{tot}} = \epsilon_{\text{geom}} \cdot \epsilon_{\text{stripping}} \cdot \epsilon_{\text{trigger}} \cdot \epsilon_{\text{offline}} \cdot \epsilon_{\text{PID}},
\]

where \( \epsilon_{\text{tot}} \) is the full efficiency, \( \epsilon_{\text{geom}}, \epsilon_{\text{stripping}}, \epsilon_{\text{trigger}}, \epsilon_{\text{offline}} \) and \( \epsilon_{\text{PID}} \) are the efficiencies associated to the geometrical acceptance of the LHCb detector, the stripping selection, the trigger requirements, the offline selection (not taking into account the particle identification, PID, requirements) and the PID requirements, respectively.
6 Computer of the Efficiencies

Table 11: Events yields ratio for \( B_s^0 \rightarrow D_s^- \mu^+ \nu \mu \) and \( B_s^0 \rightarrow D_s^*^- \mu^+ \nu \mu \).

<table>
<thead>
<tr>
<th>Variable</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_{B_s}/N_B )</td>
<td>2.28 ± 0.09</td>
<td>2.10 ± 0.06</td>
</tr>
<tr>
<td>( N_{B_s^<em>}/N_B^</em> )</td>
<td>5.75 ± 0.37</td>
<td>6.17 ± 0.26</td>
</tr>
</tbody>
</table>

The splitting of the full efficiency into multiple contributions allows us to better determine the sources of differences between numerator and denominator and to make use of control samples of data, when possible, for their estimation.

6.1 Geometric efficiency

The first term appearing in the decomposition is the geometric efficiency \( \epsilon_{geom} \). It is the efficiency associated to the geometrical acceptance of the LHCb detector. To compute this term, we produce MC samples where only the geometrical acceptance of the detector is simulated and not the rest of the reconstruction process. We produce samples separately for the \( B_s^0 \rightarrow D_s^- \mu^+ \nu \), \( B_s^0 \rightarrow D_s^*^- \mu^+ \nu \), \( B_s^0 \rightarrow D^- \mu^+ \nu \), and \( B^0 \rightarrow D^*^- \mu^+ \nu \) decays. The results are given in tables 12 and 13. This efficiency term is the most compatible with efficiency ratios of one.

6.2 Stripping efficiency

This is the efficiency associated to the requirements at stripping level. They are computed in MC. The results are given in tables 14 and 15. This efficiency term is the least compatible with efficiency ratios of one with values of \( \sim 1.5 \). The reason for this large discrepancy between numerator and denominator is that we require that the kaon pair comes from the decay of a \( \phi \) meson. However, the ratio of resonant to non-resonant branching fractions is greater for the \( B_s^0 \) sample than for the \( B^0 \)
This particular requirement is the main cause of the discrepancy.

Table 12: Geometric efficiencies split into 2011 and 2012 datasets for the $B_{s}^{0} \rightarrow D_{s}^{-}(\rightarrow K^{+}K^{-}\pi^{-})\mu^{+}\nu_{\mu}X$ and $B^{0} \rightarrow D^{-}(\rightarrow K^{+}K^{-}\pi^{-})\mu^{+}\nu_{\mu}X$ decays.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Component</th>
<th>$B_{s}^{0}$ $\epsilon_{\text{geom}}$ [%]</th>
<th>$B^{0}$ $(K^{+}K^{-}\pi^{-})$ $\epsilon_{\text{geom}}$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>$B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu$</td>
<td>16.70 ± 0.06</td>
<td>16.88 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>$B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu$</td>
<td>16.81 ± 0.06</td>
<td>16.70 ± 0.06</td>
</tr>
<tr>
<td>2012</td>
<td>$B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu$</td>
<td>17.10 ± 0.06</td>
<td>17.02 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>$B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu$</td>
<td>17.06 ± 0.06</td>
<td>17.10 ± 0.06</td>
</tr>
</tbody>
</table>

Table 13: Geometric efficiencies ratio for $B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu_{\mu}$ and $B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu_{\mu}$ decays.

<table>
<thead>
<tr>
<th>Geometric</th>
<th>Component</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_{s}/\epsilon_{d}$</td>
<td>0.989 ± 0.005</td>
<td>1.005 ± 0.005</td>
<td></td>
</tr>
<tr>
<td>$\epsilon_{s}^{<em>}/\epsilon_{d}^{</em>}$</td>
<td>1.006 ± 0.005</td>
<td>0.997 ± 0.005</td>
<td></td>
</tr>
</tbody>
</table>

Indeed, the corresponding branching fractions are the following [3]:

$$
\mathcal{B}(D_{s}^{-} \rightarrow K^{+}K^{-}\pi^{-}) = (5.45 \pm 0.17)\%,
$$

$$
\mathcal{B}(D_{s}^{-} \rightarrow \phi(\rightarrow K^{+}K^{-}\pi^{-}) = (2.27 \pm 0.08)\%,
$$

$$
\mathcal{B}(D^{-} \rightarrow K^{+}K^{-}\pi^{-}) = (9.96 \pm 0.26) \times 10^{-3},
$$

$$
\mathcal{B}(D^{-} \rightarrow \phi(\rightarrow K^{+}K^{-}\pi^{-}) = (2.77^{+0.09}_{-0.10}) \times 10^{-3}.
$$

The stripping requirement on the $\phi$ resonance is shown in Fig. 34. We notice that the efficiency of the requirement is indeed greater for the $B_{s}^{0}$ than for the $B^{0}$ samples. This particular requirement is the main cause of the discrepancy.

Figure 34: Stripping requirement on the $\phi$ resonance for $B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu_{\mu}$ (left) and $B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu_{\mu}$ (right) samples in the 2012 dataset. The distribution of $m(K^{+}K^{-})$ is plotted in blue for $B^{0} \rightarrow D_{s}^{-}\mu^{+}\nu_{\mu}$ decays and in red for $B_{s}^{0} \rightarrow D_{s}^{-}\mu^{+}\nu_{\mu}$. We notice that the efficiency of the requirement is higher for the $B_{s}^{0}$ samples.
6 COMPUTATION OF THE EFFICIENCIES

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Component</th>
<th>$B^0_s (K^+K^-\pi^-)$</th>
<th>$B^0 (K^+K^-\pi^-)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>$B^0_s \rightarrow D^0_{(s)}\mu^+\nu$</td>
<td>1.552 $\pm$ 0.006</td>
<td>1.017 $\pm$ 0.003</td>
</tr>
<tr>
<td></td>
<td>$B^0_s \rightarrow D^*_0\mu^+\nu$</td>
<td>1.508 $\pm$ 0.003</td>
<td>1.044 $\pm$ 0.003</td>
</tr>
<tr>
<td>2012</td>
<td>$B^0_s \rightarrow D^0_{(s)}\mu^+\nu$</td>
<td>1.526 $\pm$ 0.006</td>
<td>1.025 $\pm$ 0.003</td>
</tr>
<tr>
<td></td>
<td>$B^0_s \rightarrow D^*_0\mu^+\nu$</td>
<td>1.497 $\pm$ 0.003</td>
<td>1.031 $\pm$ 0.003</td>
</tr>
</tbody>
</table>

Table 14: Stripping efficiencies split into 2011 and 2012 datasets for the $B^0_s \rightarrow D^0_{(s)}\rightarrow K^+K^-\pi^-\mu^+\nu, X$ and $B^0 \rightarrow D^-\rightarrow (K^+K^-\pi^-)\mu^+\nu, X$ decays.

<table>
<thead>
<tr>
<th>Stripping</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_s/\epsilon_d$</td>
<td>1.525 $\pm$ 0.007</td>
<td>1.488 $\pm$ 0.007</td>
</tr>
<tr>
<td>$\epsilon^<em>_{s}/\epsilon^</em>_{d}$</td>
<td>1.445 $\pm$ 0.006</td>
<td>1.452 $\pm$ 0.006</td>
</tr>
</tbody>
</table>

Table 15: Stripping efficiencies ratio for $B^0_s \rightarrow D^0_{(s)}\mu^+\nu, X$ and $B^0_s \rightarrow D^*_0\mu^+\nu, X$ decays. We notice a large departure from one.

6.3 Trigger efficiency

The trigger system is designed to reduce the LHC beam crossing rate of 40 MHz to a few kHz [25]. This allows to reduce the amount of stored data by only keeping events which are potentially useful for physics analysis and rejecting poorly reconstructed events. The trigger is split in two stages. The first stage, L0, is based on hardware, i.e., uses logic electronic, while the second stage is the High Level Trigger (HLT) which is implemented in software using C++ applications. The High Level Trigger is decomposed into HLT1 and HLT2.

At the L0 stage, the event rate goes from 40 MHz to 1 MHz. Information from four detectors is combined in order to distinguish between hadron, electron or photon showers. These detectors are the Pre-Shower Detector (PS), Scintillating Pad Detector (SPD), the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL). Together they reconstruct the hadrons, electrons and photons clusters with the highest transverse energy $E_T$, defined as:

$$E_T = \sum_{i=1}^{4} E_i \sin \theta_i, \quad (33)$$

where the energy is deposed in $2 \times 2$ cells, $E_i$ is the energy deposited in cell $i$ and $\theta_i$ is the angle between the $z$-axis and a neutral particle assumed to be coming from the mean position of the interaction envelope hitting the centre of the cell [26]. The muon system composed of the five muon stations (M1-M5) reconstructs the two muon tracks with the largest $p_T$. By requiring L0MuonTOS we impose a minimal threshold on the highest muon $p_T$ of 1480 and 1760 MeV for 2011 and 2012 datasets, respectively [26].

The HLT1 stage processes the full data from the L0 stage and reduces the event rate to approximately 30 kHz. Using the VELO, muon chambers and tracker sta-
6.3 Trigger efficiency

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Component</th>
<th>$\epsilon_{L0}$ [%]</th>
<th>$\epsilon_{HLT1}$ [%]</th>
<th>$\epsilon_{HLT2}$ [%]</th>
<th>$\epsilon_{L0}$ [%]</th>
<th>$\epsilon_{HLT1}$ [%]</th>
<th>$\epsilon_{HLT2}$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>$B_s^0 \rightarrow D_{s0}^{*+} \mu^- \nu$</td>
<td>95.74 ± 0.04</td>
<td>91.26 ± 0.06</td>
<td>79.22 ± 0.09</td>
<td>95.51 ± 0.05</td>
<td>90.79 ± 0.07</td>
<td>74.76 ± 0.10</td>
</tr>
<tr>
<td></td>
<td>$B_{s0}^0 \rightarrow D_{s0}^{*+} \mu^- \nu$</td>
<td>96.56 ± 0.02</td>
<td>93.10 ± 0.03</td>
<td>74.95 ± 0.06</td>
<td>96.58 ± 0.05</td>
<td>93.09 ± 0.07</td>
<td>68.72 ± 0.13</td>
</tr>
<tr>
<td>2012</td>
<td>$B_s^0 \rightarrow D_{s0}^{*+} \mu^- \nu$</td>
<td>89.43 ± 0.06</td>
<td>89.46 ± 0.06</td>
<td>82.36 ± 0.08</td>
<td>89.37 ± 0.05</td>
<td>88.84 ± 0.05</td>
<td>78.71 ± 0.07</td>
</tr>
<tr>
<td></td>
<td>$B_{s0}^0 \rightarrow D_{s0}^{*+} \mu^- \nu$</td>
<td>91.70 ± 0.03</td>
<td>90.88 ± 0.04</td>
<td>78.43 ± 0.06</td>
<td>91.97 ± 0.05</td>
<td>90.57 ± 0.05</td>
<td>73.37 ± 0.09</td>
</tr>
</tbody>
</table>

Table 16: Trigger efficiencies split into L0, HLT1 and HLT2 requirements, 2011 and 2012 datasets for the $B_s^0 \rightarrow D_s^- (\rightarrow K^+K^-\pi^-)\mu^+\nu_\mu X$ and $B^0 \rightarrow D^- (\rightarrow K^+K^-\pi^-)\mu^+\nu_\mu X$ decays. Notice the significant difference of $\epsilon_{HLT2}$ between the $B_s^0$ and $B^0$ decay modes.

At the HLT2 stage, the tracks are used to form composite particles. The event rate is reduced to $\sim 2$ kHz mainly by imposing topological requirements. These requirements implemented with a boosted decision tree (BDT) [26], combine $N$ tracks to a common displaced vertex. By imposing Hlt2TopoMu2,3,4BodyBDT, we are requiring 2,3 or 4 tracks to come from the same displaced vertex. More precisely, we require that the $K^+K^-\pi^-\mu^+$ tracks come from the same vertex. We expect a difference in efficiency between the $B^0$ and the $B_s^0$ samples, since the lifetime of the $D^-$ and $D_s^-$ mesons are quite different, $\tau(D^-) = (500 \pm 7)$ fs and $\tau(D_s^-) = (1040 \pm 7)$ fs [3]. Since the lifetime of the $D_s^-$ meson is greater, its decay vertex will be more displaced on average relatively to the $B_s^0$ secondary vertex. Therefore, the efficiency of the topological triggers, which require the muon to be attached to the $K^+K^-\pi^-$ vertex, is expected to be smaller for the $B^0 \rightarrow D^- (\rightarrow K^+K^-\pi^-)\mu^+\nu_\mu X$ decay mode compared to that for $B_s^0 \rightarrow D_s^- (\rightarrow K^+K^-\pi^-)\mu^+\nu_\mu X$ decays.

In order to control more accurately the source of efficiency difference between $B^0$ and $B_s^0$ decay modes, the full trigger efficiency is split into the L0, HLT1 and HLT2 efficiencies associated to the requirements imposed at these respective levels:

$$\epsilon_{\text{trigger}} = \epsilon_{L0} \cdot \epsilon_{HLT1} \cdot \epsilon_{HLT2}. \quad (34)$$

The trigger efficiency is computed in MC by imposing the different requirements on the sample after stripping selection. The results are reported in Tab. 16. Notice that $\epsilon_{HLT2}$ is significantly different between the $B_s^0$ and $B^0$ decay modes, as expected. Note also that $\epsilon_{L0}$ differs between 2011 and 2012 because the $p_T(\mu)$ threshold requirement has been tightened to cope with the increase of tracks multiplicity. Tab. 17 shows the full trigger efficiencies ratios for category 1 and 2 split into 2011 and 2012 samples.
6 COMPUTATION OF THE EFFICIENCIES

<table>
<thead>
<tr>
<th>Trigger</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_s/\epsilon_d$</td>
<td>1.068 ± 0.002</td>
<td>1.054 ± 0.002</td>
</tr>
<tr>
<td>$\epsilon_s^<em>/\epsilon_d^</em>$</td>
<td>1.091 ± 0.002</td>
<td>1.070 ± 0.002</td>
</tr>
</tbody>
</table>

Table 17: Trigger efficiencies ratio for $B^0_{(s)} \rightarrow D^+_{(s)}\mu^+\nu_\mu$ and $B^0_{(s)} \rightarrow D^{*-}_{(s)}\mu^+\nu_\mu$ decays.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Component</th>
<th>$B^0_s$</th>
<th>$B^0 (K^+K^-\pi^-)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B^0_{(s)} \rightarrow D^-<em>{(s)}\mu^+\nu</em>\mu$</td>
<td>36.57 ± 0.12</td>
<td>35.52 ± 0.13</td>
</tr>
<tr>
<td></td>
<td>$B^0_{(s)} \rightarrow D^+_{(s)}\mu^+\nu$</td>
<td>34.83 ± 0.08</td>
<td>34.94 ± 0.16</td>
</tr>
<tr>
<td>2012</td>
<td>$B^0_{(s)} \rightarrow D^-_{(s)}\mu^+\nu$</td>
<td>35.97 ± 0.11</td>
<td>34.59 ± 0.09</td>
</tr>
<tr>
<td></td>
<td>$B^0_{(s)} \rightarrow D^+_{(s)}\mu^+\nu$</td>
<td>33.81 ± 0.07</td>
<td>33.83 ± 0.11</td>
</tr>
</tbody>
</table>

Table 18: Offline efficiencies split into 2011 and 2012 datasets for the $B^0_s \rightarrow D^- (\rightarrow K^+K^-\pi^-)\mu^+\nu_\mu X$ and $B^0 \rightarrow D^- (\rightarrow K^+K^-\pi^-)\mu^+\nu_\mu X$ decays.

6.4 Offline efficiency

The efficiency, $\epsilon_{\text{offline}}$, corresponds to the set of offline requirements presented in Tab. 4 except those using the PID variables. It is computed in MC relatively to the samples after the trigger requirements. The individual results are given in table 18 while the efficiencies ratios are given in Tab. 19.

6.5 PID efficiency

The particle identification (PID) is the process of assigning a particle hypothesis to a given track e.g. pion, kaon, electron, muon, . . . PID is provided by the RICH detectors, the calorimeter system (SPD, PS, ECAL and HCAL) and the muon chambers.

Each of the three sub-systems assigns a mass hypothesis to a given track by maximising a likelihood variable $L$ [18]. For instance, the RICH system considers all tracks and all radiators and constructs a likelihood variable $L^{\text{RICH}}$ based on the propagation of one cone of photons per mass hypothesis and its comparison of the predicted and measured pixel hits. The identities assigned to the tracks are varied and the likelihood is recomputed. The procedure is iterated such as to maximise $L$. The calorimeters and muon systems have a similar procedure. A global likelihood variable is then constructed by taking into account the information of each system,

<table>
<thead>
<tr>
<th>Offline</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_s/\epsilon_d$</td>
<td>1.029 ± 0.005</td>
<td>1.040 ± 0.004</td>
</tr>
<tr>
<td>$\epsilon_s^<em>/\epsilon_d^</em>$</td>
<td>0.997 ± 0.005</td>
<td>0.999 ± 0.004</td>
</tr>
</tbody>
</table>

Table 19: Offline efficiencies ratio for $B^0_{(s)} \rightarrow D^-_{(s)}\mu^+\nu_\mu$ and $B^0_{(s)} \rightarrow D^{*-}_{(s)}\mu^+\nu_\mu$ decays.
6.5 PID efficiency

for example for a kaon:

\[ L(K) = L^{RICH}(K) \cdot L^{CALO}(\text{non } e) \cdot L^{MUON}(\text{non } \mu). \]  

(35)

In practice, instead of using the likelihood of a mass hypothesis for a given track, we use the difference between the logarithm of the likelihood of the track under the mass hypothesis of interest and the pion mass hypothesis:

\[ PID_x = \Delta \ln L_{X\pi} = \ln L(X) - \ln L(\pi), \]  

(36)

for the mass hypothesis of particle X.

Another type of PID variables is the ProbNN variables. The ProbNN variables combine information from the three PID sub-systems, i.e. the RICH detectors, calorimeters and muon system, and from the tracking system. This information is used as input to train a neural network. The result is a Bayesian probability value for a given mass hypothesis.

In the selection some PID requirements are imposed in order to assign an identity to a given track. For example, in the \( D^- \rightarrow K^+K^−\pi^- \) selection, we require the kaon candidate to pass \( \text{ProbNN}_k > 0.2 \) in order to accept it as a kaon, otherwise the event is rejected. However, it is possible that some other particle, e.g. a pion, initially identified as a kaon passes the requirement. That would be the case of a pion mis-identified as a kaon. Notice that for the decays in the \( K^+K^−\pi^- \) mode, the two kaons \( K_1 \) and \( K_2 \) can be distinguished by their electric charge. We define \( K_1 \) as the kaon whose charge is opposite to the charge of the pion. Therefore \( K_1 \) can be identified by its electric charge alone and no ProbNN requirement is applied on it.\(^4\) In order to estimate the identification efficiency and the mis-identification rate of a given PID requirement, large control samples are extracted from data. These samples consist of decay modes which can be reconstructed with kinematics alone, without the use of PID information, e.g. \( K_S^0 \rightarrow \pi^+\pi^- \), \( \Lambda^0 \rightarrow \text{p}\pi^- \), \( D^{*+} \rightarrow D^0(\rightarrow K^−\pi^+)\pi^+ \) and \( J/\psi \rightarrow \mu^+\mu^- \). An example of the efficiency and mis-identification rate of the kaon as a function of the track momentum, for kaons and pions from a \( D^{*+} \rightarrow D^0\pi^+ \) calibration sample, is presented in Fig. 35.

The MC simulations do not reproduce well the PID variables and therefore are unreliable to compute the efficiencies associated to such requirements. The efficiency of a given PID requirement depends on the kinematics of the tracks and the occupancy of the detectors. It can be parametrized as a function of the momentum, \( p \), the pseudo-rapidity, \( \eta \), and the number of tracks in the events, \( n\text{Tracks} \). The average efficiency associated to a given PID requirement can be expressed as:

\[ \epsilon = \int \epsilon(\vec{x})f(\vec{x})d\vec{x}, \]  

(37)

where \( \vec{x} = (p, \eta, n\text{Tracks}) \), \( \epsilon(\vec{x}) \) is the universal efficiency function and \( f(\vec{x}) \) is the sample’s distribution of \( \vec{x} \). A bad modelling of \( \epsilon(\vec{x}) \) in MC is not a concern in the ratio of efficiencies as long as the distributions of these kinematical variables \( f(\vec{x}) \) are the same for numerator and denominator, i.e. \( f_s^{(*)}(\vec{x})/f_d^{(*)}(\vec{x}) \approx 1 \). While we expect

\(^4\) On the contrary, in the \( K^+\pi^-\pi^- \) mode both pions have the same electric charge and therefore the \( \text{ProbNN}_\pi > 0.5 \) requirement is imposed on both of them.
these distributions to be similar for $B^0_s \to D^{*-} \mu^+ \nu_\mu$ decays, some discrepancy is expected for $B^0_s \to D^{*-} \mu^+ \nu_\mu$ decays. Indeed, since in most cases $D_s^{*-} \to D^- \gamma$ while $D^{*-} \to D^- \pi^0$ the distributions are expected to be slightly different because of the different final state, hence kinematics. Fig. 36 shows the momentum distribution ratio of the muon, pion and the second kaon ($K_2$) in the 2011 sample. The distributions for the 2012 sample are in the appendix, section 9.5. These ratios are fitted by a constant in order to assess the compatibility with one. The $\chi^2/dof$ of these fits are presented in Tab. 20. The muon distributions are the most compatible with a ratio of one while the pion distributions show the greatest discrepancy. The values for $K_1$ are also present for illustration, while being irrelevant for the computation of the efficiency, since no ProbNN cut is applied on this track in the $K^+ K^- \pi^-$ mode. The same study is done for the ratio of pseudo-rapidity distribution. Unfortunately, the global number of tracks per event, $nTracks$, was not included in the tuples produced after the stripping selection. Instead, the study was performed on the variable $PVNTRACKS$ which is the number of tracks coming from the primary vertex. This variable is naturally correlated with $nTracks$ and therefore still provides information. The distributions of $PVNTRACKS$ are mainly in agreement with the hypothesis of $f^{(*)}_s/f^{(*)}_d = 1$. For this reason, we do not include it in the correction of $\epsilon(\vec{x})$ and will consider the efficiency as a function of the momentum and pseudo-rapidity only: $\epsilon(p, \eta)$. The distributions of $f^{(*)}_s/f^{(*)}_d$ for the pseudo-rapidity and the number of tracks from the primary vertex are presented in the appendix, section 9.5.

Since the distributions of momentum and pseudo-rapidity are different between numerator and denominator, the efficiency $\epsilon(p, \eta)$ has to be corrected. In order to correct it we use the PIDCalib package [27]. The idea is to use calibration samples with kinematically clean decays, as explained previously, to compute the correct efficiency of the ProbNN requirements. A two-dimensional binning of the $(p, \eta)$ distribution is constructed. The binning has to be chosen such that the bins are sufficiently small in order for the efficiency to be constant across the bin, while
6.5 PID efficiency

Figure 36: Ratios of the momentum distribution in the $B^0 \rightarrow D_s^- \mu^+ \nu_\mu$ mode relatively to the $B^0 \rightarrow D^- \mu^+ \nu_\mu$ mode (left column) and in the $B^0_s \rightarrow D_s^+ \mu^+ \nu_\mu$ mode relatively to the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ mode (right column) as a function of the momentum for the 2011 samples, in MC. The distributions are for the muon (top row), the pion (middle row) and the second kaon (bottom row). The red line is a constant fit.

being sufficiently large in order to minimise statistical fluctuation. For a given requirement, the efficiency $\epsilon(p, \eta)$ is computed using the calibration samples and the corresponding histogram is saved. We construct these histograms for every ProbNN requirement. The average corrected efficiency is then computed by taking the $(p, \eta)$ distribution from our signal MC and for every event falling into a particular $(p, \eta)$ bin we assign the combined calibration efficiency of every ProbNN requirement. This procedure takes into account kinematic correlations between the tracks. The efficiency is then averaged over the full MC sample distributions. The computation of the efficiency is made separately for 2011 and 2012 samples, $B^0_{(s)} \rightarrow D^-_{(s)} \mu^+ \nu_\mu$ and $B^0_{(s)} \rightarrow D^{*+}_{(s)} \mu^+ \nu_\mu$ decays, and for magnet polarity up and down. The values of these efficiencies are presented in table 26 in the appendix, section 9.5. The weighted average between the efficiencies for magnet polarity Up and Down is made using the number of events in data after full selection separated by polarity. The final ProbNN efficiency ratios are presented in Tab. 21.

In the above computation of the efficiencies we have used the MC description of the $(p, \eta)$ distributions. However it is possible that the MC does not reproduce
7 PROSPECTS ON SYSTEMATIC UNCERTAINTIES

<table>
<thead>
<tr>
<th>Particle type</th>
<th>$B_{(s)}^0 \rightarrow D_{(s)}^{-}\mu^+\nu_\mu$</th>
<th>$B_{(s)}^0 \rightarrow D_{(s)}^{*^{-}}\mu^+\nu_\mu$</th>
<th>$B_{(s)}^0 \rightarrow D_{(s)}^{-}\mu^+\nu_\mu$</th>
<th>$B_{(s)}^0 \rightarrow D_{(s)}^{*^{-}}\mu^+\nu_\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pion</td>
<td>2.62</td>
<td>1.78</td>
<td>1.98</td>
<td>2.88</td>
</tr>
<tr>
<td>Muon</td>
<td>0.85</td>
<td>1.84</td>
<td>1.14</td>
<td>2.39</td>
</tr>
<tr>
<td>$K_1$</td>
<td>1.64</td>
<td>1.74</td>
<td>1.80</td>
<td>1.50</td>
</tr>
<tr>
<td>$K_2$</td>
<td>1.49</td>
<td>1.75</td>
<td>1.35</td>
<td>2.12</td>
</tr>
</tbody>
</table>

Table 20: Value of the $\chi^2$/dof for the constant fits of $f_s^{(*)}/f_d^{(*)}$ for all samples.

We observe that the muon displays the best agreement with a flat distribution and that $B_{(s)}^0 \rightarrow D_{(s)}^{*^{-}}\mu^+\nu_\mu$ decays are more in tension with the flat hypothesis than $B_{(s)}^0 \rightarrow D_{(s)}^{-}\mu^+\nu_\mu$ decays.

<table>
<thead>
<tr>
<th>Variable</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_s/\epsilon_d$</td>
<td>$0.997 \pm 0.005$</td>
<td>$0.997 \pm 0.004$</td>
</tr>
<tr>
<td>$\epsilon_s^{<em>}/\epsilon_d^{</em>}$</td>
<td>$0.998 \pm 0.005$</td>
<td>$0.998 \pm 0.003$</td>
</tr>
</tbody>
</table>

Table 21: ProbNN efficiencies ratio for $B_{(s)}^0 \rightarrow D_{(s)}^{-}\mu^+\nu_\mu$ and $B_{(s)}^0 \rightarrow D_{(s)}^{*^{-}}\mu^+\nu_\mu$ decays.

In order to investigate this possibility, we can compare the distributions of background-subtracted data with MC. To minimise the background impact, we consider only candidates with $m_{corr} > 5.6$ GeV, where mostly combinatorial background is present. We subtract this background by using SS data, with the expected fraction from the fit of the sample composition. We neglect small residual contaminations from physics backgrounds. In the MC distributions, we sum the $B_{(s)}^0 \rightarrow D_{(s)}^{-}\mu^+\nu_\mu$ and $B_{(s)}^0 \rightarrow D_{(s)}^{*^{-}}\mu^+\nu_\mu$ histograms according to the fractions derived in the fit of the sample composition. The figures 37 and 38 show this ratio for the momentum and pseudo-rapidity distributions for the $B_{s}^0$ sample. While it is mainly flat and consistent with one for $\eta$ there is a greater discrepancy for the momentum distribution at low values. Although the method can be refined and must be validated e.g. on the $K^+\pi^-\pi^-$ sample, this test provides an indication that the momentum distribution is not perfectly described in the MC and that some systematic uncertainty might be considered. Notice that, since the measurement uses the ratio of efficiencies, if the discrepancy in the momentum distribution is similar for the $B_{s}^0$ and the $B^0$ samples, then it should be suppressed in the ratio and only appears as a second-order effect.

7 Prospects on systematic uncertainties

The analysis is based on some assumptions which can introduce systematic uncertainties. There are two main sources of uncertainties: those related to the fit of the sample composition, therefore on the yields, and those related to the efficiencies estimations.

In order to assess the systematics on the yields, the fit of the sample composition
We observe some discrepancy for low momenta values.

can be made using different assumptions and measuring the change of the estimated yields. The signal models of the $B_s^0$ decays used in the simulations assume some values of the form factors which are inferred from $B^0$ measurements considering SU(3) symmetry (exchange of the $s$ and $d$ quarks), as no experimental information are present. Difference of the true form factors with respect to the ones used in simulation can give a systematic bias, as the mass templates can depend on them. Therefore, one could see how large is the variation of the estimated $B_s^0$ yields when using different mass templates according to different reasonable variations of the form factors models. From studies in the lifetime analysis, this is expected to be the largest source of systematics, contributing with variations of the estimated signal fraction of about 2% [22].

We observe some discrepancy for low $\eta$ values, however there are few events in these regions. The core of the distribution is approximately flat at one for all particles.
Another systematic bias on the yields can be due to the assumptions done for the backgrounds. Poor knowledge exists on the composition of the $B_s^0 \to D_s^{**-} \mu^+ \nu_\mu X$ decays. In a similar manner as before (for the form factors models), one must consider reasonable variations of the fraction of each excited state resonance and address the impact on the estimated $B_s^0$ yields.

Some of the backgrounds were neglected in the fit, when merging together mass templates that had similar shapes. Also, the fraction used to lump together different backgrounds is affected by uncertainties from their measured branching ratios and efficiencies. Variations of these must be considered, for instance we could add the following components: the $B_s^0 \to D^- \tau^+ \nu_\tau X$ decays, the $\Lambda^0_b$ decays, the $B^0(s) \to DD$ decays and the $B \to DDK$ decays.

Concerning the systematics on the efficiencies, we should quantify the discrepancy between the distributions in MC and data and address the question to which level these discrepancies cancel in the ratio of efficiency (although such a cancellation is expected to be a good approximation). Similarly to the first systematic listed above for the signal yields, the efficiencies can also depend on the model used for the generation of the signal decays. This source of systematic will not cancel in the ratio and need a careful estimation. The $K^+ \pi^- \pi^-$ sample, which posses a lot of statistics, can be used for instance to check the difference between data and MC for the trigger efficiencies.

8 Conclusion and future prospects

The $B^0$ mesons have been abundantly studied, notably by the $B$-factories, and provided a tool to make precision tests of the SM. In particular, semileptonic $B^0$ decays yield results which are in tension with the property of lepton universality. In order to obtain additional experimental information, similar studies should be made using $B_s^0$ mesons. However, these mesons are only currently being produced in large numbers by hadron colliders, and in particular the knowledge of semileptonic branching ratios are fairly poor. In particular, the exclusive branching ratios $\mathcal{B}(B_s^0 \to D^- \mu^+ \nu_\mu)$ and $\mathcal{B}(B_s^0 \to D_s^{*-} \mu^+ \nu_\mu)$ have not been measured yet. The objective of this project was to obtain a first estimation of these two branching ratios.

In these semileptonic decays, one or more particles cannot be reconstructed. In order to cope with this difficulty, a study was performed to identify a variable which allows separation between the different decay modes included in the data sample. The corrected mass, $m_{corr}$, was determined to be the best variable for this purpose. A study of a new approximation method for the $q^2$ and $\theta_l$ variables was also realised. These variables are needed to describe the differential decay rates of $B^0 \to D^{(*)-} \mu^+ \nu_\mu$ decays, which could give access to the form factors measurement. The method gives promising results for $q^2$ while it is not conclusive for $\theta_l$.

Using Run 1 data, the composition of the $B_s^0 \to D^- (\to K^+ K^- \pi^-) \mu^+ \nu_\mu X$ sample was determined by a fit to the corrected mass. The selection was tuned to suppress physics backgrounds which otherwise might have introduced large source of systematic uncertainties in the estimated signal yields. All the efficiencies were computed in MC except for those associated to the PID requirements which were computed
measurements, respectively. Notice that we use Table 24: Measured branching ratios for $B^0 \to D^\pm_s \mu^\pm \nu_\mu$ and $B^0_s \to D^*_s \mu^+ \nu_\mu$ decays.

<table>
<thead>
<tr>
<th>Ratio</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_s/\epsilon_d$</td>
<td>1.654 ± 0.016</td>
<td>1.634 ± 0.014</td>
</tr>
<tr>
<td>$\epsilon'_s/\epsilon'_d$</td>
<td>1.577 ± 0.015</td>
<td>1.545 ± 0.013</td>
</tr>
</tbody>
</table>

Table 22: Full efficiencies ratio for $B^0 \to D^\pm_s \mu^\pm \nu_\mu$ and $B^0_s \to D^*_s \mu^+ \nu_\mu$ decays. The first uncertainty is statistical, the second one is due to the uncertainties of the branching ratios taken as input.

Table 23: Ratios of signal branching fractions for $B^0_s \to D^\pm_s \mu^\pm \nu_\mu$ and $B^0_s \to D^*_s \mu^+ \nu_\mu$ decays. The first uncertainty is statistical, the second one is due to the uncertainties of the branching ratios taken as input.

\begin{align*}
R &= \frac{N_B \mathcal{B}(D^- \to K^+ K^- \pi^-) \epsilon_d f_d}{N_B \mathcal{B}(D_s^- \to K^+ K^- \pi^-) \epsilon_s f_s}, \\
R' &= \frac{N_{B^0_s} \mathcal{B}(D^+_s \to D^- \pi^+) \mathcal{B}(D^- \to K^+ K^- \pi^-) \epsilon'_d f_d}{N_{B^0_s} \mathcal{B}(D^- \to D^- \pi^-) \mathcal{B}(D^- \to K^+ K^- \pi^-) \epsilon'_s f_s}.
\end{align*}

Notice that we use $\mathcal{B}(D^\pm_s \to K^+ K^- \pi^-)$ instead of $\mathcal{B}(D^\pm_s \to \phi \pi^-)$ since the requirement on the resonant decay is already taken into account by the stripping efficiency. These variables are presented in Tab. 23, where the uncertainty is split in two terms: the first is statistical while the second is due to the uncertainties of the branching ratios taken as input from external measurements (see equations (9) and (10)). In figure 39 $R$ and $R'$ as a function of $f_s/f_d$ are shown. We can observe the possibility of obtaining a value of the fragmentation fractions ratio if the values of $R$ and $R'$ are provided (e.g. from lattice QCD calculations).

Without taking into account systematic uncertainties, we provide a first estimation of the exclusive branching ratios. We can define the following variables:

\begin{align*}
R &= \frac{N_B \mathcal{B}(D^- \to K^+ K^- \pi^-) \epsilon_d f_d}{N_B \mathcal{B}(D_s^- \to K^+ K^- \pi^-) \epsilon_s f_s}, \\
R' &= \frac{N_{B^0_s} \mathcal{B}(D^+_s \to D^- \pi^+) \mathcal{B}(D^- \to K^+ K^- \pi^-) \epsilon'_d f_d}{N_{B^0_s} \mathcal{B}(D^- \to D^- \pi^-) \mathcal{B}(D^- \to K^+ K^- \pi^-) \epsilon'_s f_s}.
\end{align*}

Considering the current value of $f_s/f_d$ measured by LHCb at $\sqrt{s} = 7$ TeV, and taking the known values of the $B^0_s$ branching fractions, we obtain the results of the exclusive $B^0_s$ branching ratios presented in Tab. 24. Their weighted average is:
CONCLUSION AND FUTURE PROSPECTS

Figure 39: Ratios of signal branching fractions for 2011 (left column) and 2012 (right column), for $R$ (top row) and $R'$ (bottom row). The central value is given by the black curve while the blue curves are the edges of the 1σ bands. The red lines indicate the value of $f_s/f_d$, with its uncertainty, used in the analysis and the associated $R$ and $R'$. 

\[ B(B^0_s \to D^- \mu^+\nu_\mu) = (2.03 \pm 0.05{\text{(stat)}} \pm 0.14{\text{(BR)}} \pm 0.12(f_s/f_d))\% \quad (40) \]
\[ B(B^0_s \to D^{*-} \mu^+\nu_\mu) = (4.35 \pm 0.16{\text{(stat)}} \pm 0.22{\text{(BR)}} \pm 0.25(f_s/f_d))\% \quad (41) \]

where the first uncertainty is statistical, the second and the third are due to the uncertainties of the branching ratios and $f_s/f_d$ taken as input from external measurements, respectively. This value of $f_s/f_d$ that we take as input is partly based on a semileptonic measurement, as explained previously. While the method of this semileptonic measurement is mainly different from ours, they also use $B(D_s^- \to K^+K^−\pi^−)$ which results in some correlated uncertainties which need further investigation. These exclusive branching ratios are measured for the first time. The measurement is still incomplete but provides a first estimation and is consistent with expected results so far. Notice that the precision of the measurement is mainly limited by external measurements.

The next step is to start the study of the systematic uncertainties. As future prospects, LHCb has already on tape a data sample corresponding to almost 2 fb$^{-1}$ of integrated luminosity collected in Run 2 at $\sqrt{s} = 13$ TeV. Due to the larger production cross section of $B$ mesons at this centre-of-mass energy, such a sample
provides a larger number of $B^0_s$ mesons with respect to that of the full Run 1 data set. This data can be analysed to make the first measurement of $f_s/f_d$ at 13 TeV, or to increase the statistical precision of the branching ratio measurements reported here.
References


[6] Belle collaboration, M. Huschle et al., Measurement of the branching ratio of $\bar{B}^0 \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \rightarrow D^{(*)}l^-\bar{\nu}_l$ decays with hadronic tagging at Belle, Phys. Rev. D 92, 072014 (2015), arXiv:1507.03233 [hep-ex].

[7] Belle collaboration, A. Abdesselam et al., Measurement of the branching ratio of $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \rightarrow D^{(*)}l^-\bar{\nu}_l$ decays with a semileptonic tagging method, arXiv:1603.06711 [hep-ex].


[20] BABAR collaboration, B. Aubert et al., *Measurements of the Semileptonic Decays $\bar{B} \rightarrow Dl\bar{\nu}$ and $\bar{B} \rightarrow D^*l\bar{\nu}$ Using a Global Fit to $DXl\bar{\nu}$ Final States*, Phys. Rev. D 79, 012002 (2009), arXiv:0809.0828 [hep-ex].


9 Appendix

9.1 Computation of \( m_{\text{corr}} \)

In section 4.1 we stated that the corrected mass is the \( B_0^{(s)} \) invariant mass where the longitudinal momentum of the neutrino relatively to the \( B_0^{(s)} \) flight direction has been neglected. Let us prove this assertion. We work in the \( B_0^{(s)} \) center-of-mass frame and neglect the neutrino’s mass. The \( B_0^{(s)} \) mass is then given by:

\[
m_B^2 = (E_D + E_\mu + |\vec{p}_\nu|; \vec{0})^2 \tag{42}
\]

\[
= E_D^2 + E_\mu^2 + 2E_D E_\mu + |\vec{p}_\nu|^2 + 2|\vec{p}_\nu|(E_D + E_\mu) \tag{43}
\]

where \( E_D \) and \( E_\mu \) are the energy of the \( D^{(s)} \) and the muon respectively. But we have

\[
m(D^{(s)}\mu)^2 = (E_D + E_\mu; -\vec{p}_\nu)^2 \tag{44}
\]

\[
= E_D^2 + E_\mu^2 + 2E_D E_\mu - |\vec{p}_\nu|^2 \tag{45}
\]

therefore

\[
m_B^2 = m(D^{(s)}\mu)^2 + 2(p_T^2 + p_L^2) + 2\sqrt{p_T^4 + p_L^4} \sqrt{m(D^{(s)}\mu)^2 + p_T^2 + p_L^2} \tag{46}
\]

where \( p_T \) and \( p_L \) are respectively the transverse and longitudinal parts of the neutrino’s momentum, in the center-of-mass frame, relatively to the \( B_0^{(s)} \) momentum in the laboratory frame. Then, neglecting the \( p_L \) terms since they are not measurable, we obtain the formula for the corrected mass (16). Notice that \( p_T \) is Lorentz invariant for a boost along the \( B_0^{(s)} \) flight direction and thus has the same value in the lab frame.

9.2 Computation of \( p_\parallel(\nu) \)

As explained in section 4.3, it is possible to obtain the longitudinal component of the neutrino’s momentum, relatively to the \( B_0^{(s)} \) flight direction, \( p_\parallel(\nu) \) using kinematical constraints. However, this leads to an ambiguity since there are two possible solutions. Let us do the calculation. In the laboratory frame, using energy-momentum conservation, we have the following relations (see Fig. 6):

\[
E_B = E_{D\mu} + p_\nu \tag{47}
\]

\[
\vec{p}_B = \vec{p}_\parallel(D\mu) + \vec{p}_\parallel(\nu) \tag{48}
\]

\[
\vec{0} = \vec{p}_\perp(D\mu) + \vec{p}_\perp(\nu) \tag{49}
\]

where \( E_X \) is the energy of the \( X \) system, \( \vec{p}_{\parallel(\perp)}(X) \) is the momentum parallel (perpendicular) of the \( X \) system, relatively to the \( B_0^{(s)} \) momentum and \( p_\nu \) is the norm of the neutrino’s momentum. Using \( E^2 = p^2 + m^2 \) we obtain:

\[
m_B^2 = E_{D\mu}^2 + 2E_{D\mu} p_\nu + p_\nu^2 - \left( p_\parallel^2(D\mu) + p_\parallel^2(\nu) + 2\vec{p}_\parallel(D\mu) \cdot \vec{p}_\parallel(\nu) \right) \tag{50}
\]

\[
= m_{D\mu}^2 + 2p_\perp^2(D\mu) + 2\left( E_{D\mu} p_\nu - \vec{p}_\parallel(D\mu) \cdot \vec{p}_\parallel(\nu) \right) \tag{51}
\]

\[\text{5The neutrino’s mass is neglected and the convention } c = 1 \text{ is used.}\]
rearranging the terms and squaring them, we obtain:

\[
E_{D\mu}^2 \left( p_{\parallel}^2(\nu) + p_{\perp}^2(D\mu) \right) = \frac{1}{4} \left( m_B^2 - m_{D\mu}^2 - 2p_{\perp}^2(D\mu) \right)^2 + \left( m_B^2 - m_{D\mu}^2 - 2p_{\perp}^2(D\mu) \right) \vec{p}_{\parallel}(D\mu) \cdot \vec{p}_{\parallel}(\nu) + p_{\parallel}^2(D\mu)p_{\parallel}^2(\nu)
\] (52)

therefore we obtain a second order equation for \( |\vec{p}_{\parallel}(\nu)| = p_{\parallel}(\nu) \):

\[
a \cdot p_{\parallel}^2(\nu) + b \cdot p_{\parallel}(\nu) + c = 0
\] (53)

with the coefficients:

\[
a = \left( m_{D\mu}^2 + p_{\perp}^2(D\mu) \right)
\] (54)

\[
b = \pm \left( m_B^2 - m_{D\mu}^2 - 2p_{\perp}^2(D\mu) \right) p_{\parallel}(D\mu)
\] (55)

\[
c = E_{D\mu}^2 p_{\perp}^2(D\mu) - \frac{1}{4} \left( m_B^2 - m_{D\mu}^2 - 2p_{\perp}^2(D\mu) \right)^2
\] (56)

where the ambiguity on the sign of \( b \) comes from the fact that we do not know whether \( \vec{p}_{\parallel}(\nu) \) is parallel or anti-parallel to \( \vec{p}_{\parallel}(D\mu) \). Therefore we obtain:

\[
p_{\parallel}(\nu)^\pm = \left| \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \right|
\] (57)

knowing that \( p_{\parallel}(\nu) \) is a positive quantity, we are able to determine the sign of \( b \). However, even if we know the correct \( b \), there is still an ambiguity due to double solution from the second order equation. Therefore, when we construct the neutrino’s momentum we do not know which value is correct: \( \vec{p}(\nu) = \vec{p}_{\perp}(\nu) + \vec{p}_{\parallel}(\nu)^\pm \). Thus, in the computation of \( q^2 = (p_{\mu} + p_{\nu})^2 \), we are left with two solutions.

### 9.3 Composition of Monte Carlo samples

Table 25 describing in detail the composition of the MC samples used for the fits is reported here.

### 9.4 Error on \( m_{\text{corr}} \) requirement for 2011 and 2012 datasets

Here are reported, in figures 40 and 41, the corrected mass distributions before and after applying the requirement \( m_{\text{corr}}^c < 0.06 \cdot m_{\text{corr}} \), for 2011 and 2012 datasets.

### 9.5 ProbNN results

Figures 42, 43, 44 and 45 show the ratio \( f_s^c / f_d^c \) for the kinematical distributions \( p, \eta \) and \( PVNTRACKS \). Table 26 reports the ProbNN efficiencies for the \( B^0_s \to D^- (\to K^+K^-\pi^-)\mu^+\nu_\mu X \) and \( B^0 \to D^- (\to K^+K^-\pi^-)\mu^+\nu_\mu X \) decays split into magnet polarities Up and Down; for 2011 and 2012 datasets.
\begin{table}[h]
\centering
\begin{tabular}{lll}
\hline
Decay & $B$ [%]
\hline
$B_s^0 \rightarrow D_s^- \mu^+ \nu$ & 2.1000 \\
$B_s^0 \rightarrow D_s^+ (\rightarrow D_s^- X) \mu^+ \nu$ & 5.1000 \\
$B_s^0 \rightarrow D_s^0 (\rightarrow D_s^- X) \mu^+ \nu$ & 0.7000 \\
$B_s^0 \rightarrow D_s^0 (\rightarrow D_s^- X) \mu^+ \nu$ & 0.4000 \\
$B_s^0 \rightarrow D_s^0 (\rightarrow D_s^- X) \mu^+ \nu$ & 0.4000 \\
$B_s^0 \rightarrow D_s^0 (\rightarrow D_s^- X) \mu^+ \nu$ & 0.3800 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D_s^- X) \tau^+ (\rightarrow \mu^+ \nu \nu) \nu$ & 0.2770 \\
$B_s^0 \rightarrow D_s^0 (\rightarrow D_s^- X) \tau^+ (\rightarrow \mu^+ \nu \nu) \nu$ & 0.0310 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D_s^- X) \tau^+ (\rightarrow \mu^+ \nu \nu) \nu$ & 0.0310 \\
$B_s^0 \rightarrow D_s^- \mu^+ \nu$ & 2.1700 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 1.6218 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.1848 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.1652 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.1436 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.1197 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0902 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0616 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0294 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0237 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0198 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0194 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0190 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0841 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0110 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0087 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0069 \\
$B_s^0 \rightarrow D_s^- (\rightarrow D^- X) \mu^+ \nu$ & 0.0053 \\
\hline
\end{tabular}
\caption{Decays included in the simulation of the inclusive $B_s^0 \rightarrow D_s^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu \nu X$ (top) and $B^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu \nu X$ (bottom) decays. The right column indicates the branching fractions used for each decay.}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{llllllll}
\hline
Dataset & Component & $\epsilon_{\uparrow}$ [%] & $\epsilon_{\downarrow}$ [%] & $\epsilon_{\uparrow}$ [%] & $\epsilon_{\downarrow}$ [%]
\hline
2011 & $B_s^0 \rightarrow D_s^- \mu^+ \nu$ & 86.99 ± 0.29 & 87.07 ± 0.29 & 87.21 ± 0.35 & 87.29 ± 0.35 \\
& $B_s^0 \rightarrow D_s^- \mu^+ \nu$ & 86.78 ± 0.19 & 86.83 ± 0.19 & 86.99 ± 0.41 & 87.05 ± 0.41 \\
2012 & $B_s^0 \rightarrow D_s^- \mu^+ \nu$ & 87.24 ± 0.28 & 87.66 ± 0.28 & 87.54 ± 0.25 & 87.97 ± 0.25 \\
& $B_s^0 \rightarrow D_s^- \mu^+ \nu$ & 87.09 ± 0.19 & 87.49 ± 0.19 & 87.26 ± 0.29 & 87.67 ± 0.29 \\
\hline
\end{tabular}
\caption{ProbNN efficiencies split into magnet polarities Up and Down. 2011 and 2012 datasets for the $B_s^0 \rightarrow D_s^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu \nu X$ and $B^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu \nu X$ decays.}
\end{table}
Figure 40: Normalized distribution of the corrected mass before (blue) and after (red) applying the $m_{\text{corr}}^\text{corr} < 0.06 \cdot m_{\text{corr}}$ requirement, for the $B^0_{(s)} \rightarrow D^{-}_s \mu^+ \nu_\mu$ (left column) mode and SS data (right column) in the 2012 dataset for the $B^0 \rightarrow D^{-} (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X$ (top), $B^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_\mu X$ (middle) and $B^0 \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \mu^+ \nu_\mu X$ modes.
Figure 41: Normalized distribution of the corrected mass before (blue) and after (red) applying the $m_{\text{corr}}^{\text{corr}} < 0.06 \cdot m_{\text{corr}}$ requirement, for $B^0 \rightarrow D^{*+} \mu^+ \nu_{\mu}$ decays for 2011 (left) and 2012 (right) samples for the $B^0_s \rightarrow D^-_s (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_{\mu} X$ (top), $B^0 \rightarrow D^- (\rightarrow K^+ K^- \pi^-) \mu^+ \nu_{\mu} X$ (middle) and $B^0 \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \mu^+ \nu_{\mu} X$ modes.
Figure 42: Ratios of the momentum distribution $f_s(p)/f_d(p)$ (left column) and $f_s^*(p)/f_d^*(p)$ (right column) for the 2012 samples, in MC. The distributions are for the muon (top row), the pion (middle row) and the second kaon (bottom row). The red line is a constant fit.
Figure 43: Ratios of the pseudo-rapidity distribution $f_s(\eta)/f_d(\eta)$ (left column) and $f_s^*(\eta)/f_d^*(\eta)$ (right column) for the 2011 samples, in MC. The distributions are for the muon (top row), the pion (middle row) and the second kaon (bottom row). The red line is a constant fit. The departure from 1 is mainly at the extremal $\eta$ values.
Figure 44: Ratios of the pseudo-rapidity distribution $f_s(\eta)/f_d(\eta)$ (left column) and $f_s^{*}(\eta)/f_d^{*}(\eta)$ (right column) for the 2012 samples, in MC. The distributions are for the muon (top row), the pion (middle row) and the second kaon (bottom row). The red line is a constant fit. The departure from 1 is mainly at the extremal $\eta$ values.

Figure 45: Ratios of the PVTRACKS distribution $f_s/f_d$ (left column) and $f_s^{*}/f_d^{*}$ (right column) for the 2011 and 2012 samples, in MC. The red line is a constant fit. All samples agree with $f_s^{*}/f_d^{*} = 1$ except for $f_s/f_d$ in the 2012 sample.