Master Thesis
June 5, 2014

Measurement of the branching fractions of $B_s^0 \to D_s^{\mp} K^{\pm}$, $B_s^0 \to D_s^- \pi^+$ and $B^0 \to D_s^- K^+$

L.J. Bel

Abstract

The relative branching fraction of the decay $B_s^0 \to D_s^{\mp} K^{\pm}$ with respect to $B_s^0 \to D_s^- \pi^+$ is determined from pp collision data corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$. From the same data, the branching fractions of $B_s^0 \to D_s^- \pi^+$ and $B^0 \to D_s^- K^+$ are extracted with respect to $B^0 \to D^- \pi^+$. The obtained values for the $B_s^0 \to D_s^{\mp} K^{\pm}$ and $B^0 \to D_s^- K^+$ branching fractions are more precise than the existing world average values.

Contents

1 Preface

The Standard Model (SM) of particle physics (see Fig. 1) has achieved tremendous success in explaining the interactions between the various (anti-)particles. Many phenomena that occur on the smallest scales we can currently probe can be explained, and quantum field theory elegantly integrates quantum mechanics with special relativity.

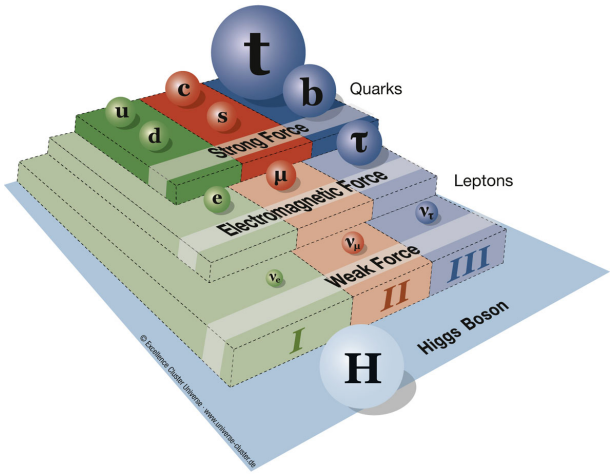

Figure 1: A visual representation of the Standard Model of particle physics. The size of the sphere corresponds to the particle's mass, and the colours and roman numerals indicate the generation of each particle. Each layer indicates the interaction (or force) that the particle above is sensitive to. For example, the quarks are the only fermions that are sensitive to all three forces.

Some long-standing problems, however, can not be solved by our current best understanding of the world of particles. Most notably, the difference between matter and anti-matter remains a mystery. Anti-matter is the exact opposite of matter in all regards (*i.e.*, all internal quantum numbers are inversed), and the two annihilate into photons (energy) when interacting (see Fig. 2). Because of this symmetry, one would expect matter and anti-matter to be created in equal amounts during the creation of the universe. This would mean that either all of it annihilated and nothing would exist, or that an amount of anti-matter equal to the amount of matter still exists somewhere in the universe. However, measurements indicate that more matter than anti-matter remains.

The inclusion of the Yukawa couplings in the SM – responsible for the interactions between matter and the Higgs boson – allows for a small violation in the matter–anti-matter symmetry. This symmetry is often called *CP* symmetry. This symmetry has been shown decades ago to be violated, but ever more precise measurements are still actively being pursued at the various particle detection experiments throughout the world. The best systems to study this *CP* symmetry involve particles that contain b quarks. The LHCb experiment at CERN is a dedicated experiment for the study of these b hadrons.

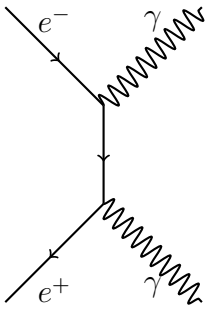

Figure 2: A particle and an anti-particle annihilate into two photons.

1.1 The LHC and the LHCb detector

The Large Hadron Collider (LHC) [1], located at CERN, Geneva, is the world's largest particle collider. Protons are injected as beams into a series of pre-accelerators, after which they enter the 27 km long LHC tunnel, up to 175 m deep beneath the surface. Inside this tunnel, the protons are accelerated to energies of up to 8 TeV and collided at one of the four interaction points in the LHC. At these points, huge particle detectors are located to detect the particles produced at the proton-proton collisions.

One of these experiments is the LHCb experiment [2] (see Fig. 3). This detector is very well suited for the study of b hadrons, particles that contain a b quark, in which CP-violating effects play an important role. LHCb is a forward-arm detector, justified by the fact that b hadrons are mainly produced in the forward region. It consists of a number of subdetectors, which together lead to a full reconstruction of the decay products of b hadrons [3].

Closest to the interaction point is the vertex locator (VELO) [4]. This silicon strip detector is located only 8 mm from the beam line, and together with the Tracker Turicencis (TT), Outer Tracker (OT) [5] and Inner Tracker (IT) [6], provides excellent particle tracking through the detector [7]. The LHCb magnet [8] bends the trajectories of charged particles, allowing for momentum measurements.

Particle identification (PID) is possible thanks to the two RICH subdetectors [9], together with the Electromagnetic and Hadronic Calorimeters (HCAL and ECAL) [10] and the muon system [11]. The RICH is capable of identifying any charged particle while the HCAL detects the energy deposit of both charged and neutral hadrons. The ECAL records the energy deposit of electrons and photons, and the muon system detects the muons, which leave tracks in the RICH but pass undetected through the calorimeters.

During the 2011 and 2012 data taking periods of the LHC, about 1400 bunches of approximately 10^{13} protons were circulating in both directions. Due to the strong focussing of the beams at the interaction points, about 10^{11} b hadrons were produced in LHCb per year, which leads to about 10^4 detected b hadron decays per year, depending on the decay probability to the final state of interest.

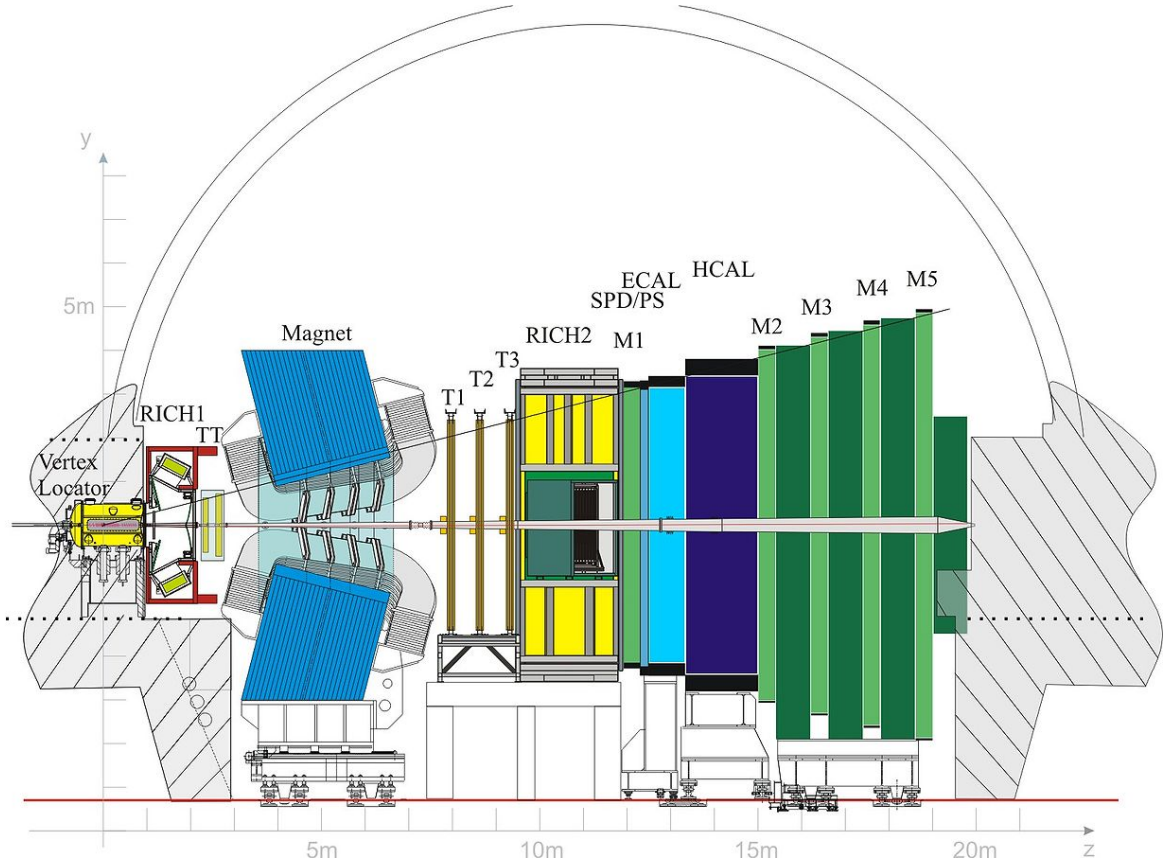

Figure 3: A schematic representation of the LHCb detector.

1.2 The B_s^0 meson

The B_s^0 meson[1] consists of an s (strange) quark and an anti-b (bottom) quark. In this particle, CP violation plays a role in certain decays: the rate with which the particle decays into a particular final state may be different to that of the same process with anti-particles. One such decay is the decay into a $D_s^{\mp}$ (containing $\overline{c}s$ or vice versa) and a $K^{\pm}$ meson (containing $\overline{s}u$ or vice versa).

The decay $B_s^0 \to D_s^{\mp} K^{\pm}$ is particularly sensitive to the so-called CP-violating observable γ. A measurement of γ using this decay will eventually be compared to other measurements of γ, that might hide the presence of new particles, potentially giving new insights into matter anti-matter differences. The first-order decay topologies of the decay $B_s^0 \to D_s^- K^+$ are displayed as Feynman diagrams in Fig. 4.

The physics parameter measured in this analysis is the $B_s^0 \to D_s^{\mp} K^{\pm}$ branching fraction, the probability for a B_s^0 meson to decay to the $D_s^{\mp} K^{\pm}$ final state. Because the number of produced $B_{(s)}^0$ mesons is not known precisely, the branching fraction is normalised to another branching fraction. In this case, the process $B_s^0 \to D_s^- \pi^+$ is used as normalisation.

[1]Throughout this document, inclusion of $\overline{B}_s^0$ in the notation B_s^0 is implied.

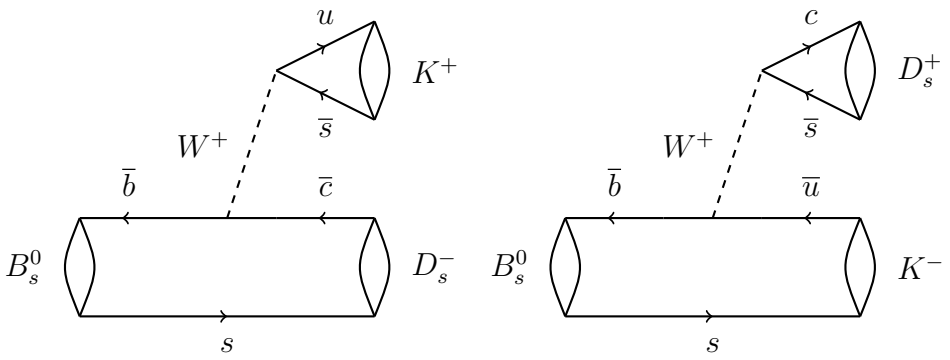

Figure 4: Feynman diagrams of the first-order contributions to the processes $B_s^0 \to D_s^- K^+$ (left) and $B_s^0 \to D_s^+ K^-$ (right).

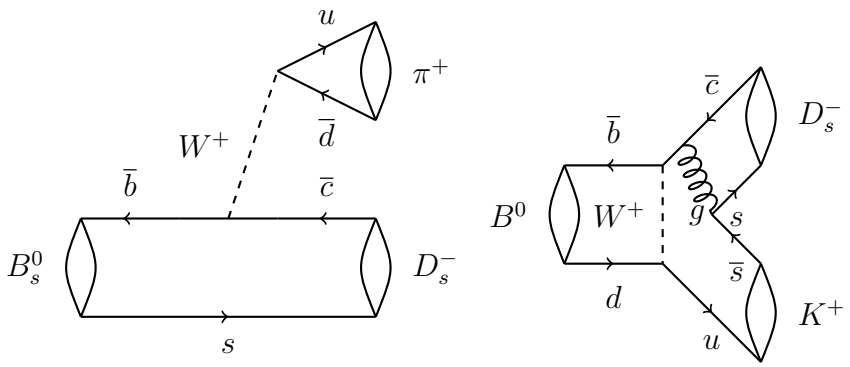

Figure 5: Feynman diagrams of the first-order contributions to the processes $B_s^0 \to D_s^- \pi^+$ (left) and $B^0 \to D_s^- K^+$ (right).

However, to know the branching fraction of $B_s^0 \to D_s^- \pi^+$, it is in turn normalised to the branching fraction of the decay $B^0 \to D^- \pi^+$. This value is obtained from the Particle Data Group [12], the international collaboration that aims to provide a comprehensive overview of all particle physics related parameters.

In the $B_s^0 \to D_s^\mp K^\pm$ branching fraction analysis, another process appears: the decay $B^0 \to D_s^- K^+$. The lowest-order Feynman diagram of this process is show in Fig. 5. The branching fraction of this decay is also measured, normalised to that of the decay $B^0 \to D^- \pi^+$. This means that in total three branching fractions are measured:

- The branching fraction $\mathcal{B}(B_s^0 \to D_s^- \pi^+)$, normalised to $\mathcal{B}(B^0 \to D^- \pi^+)$;
- The branching fraction $\mathcal{B}(B_s^0 \to D_s^\mp K^\pm)$, normalised to $\mathcal{B}(B_s^0 \to D_s^- \pi^+)$;
- The branching fraction $\mathcal{B}(B^0 \to D_s^- K^+)$, normalised to $\mathcal{B}(B^0 \to D^- \pi^+)$.

The measurement of the branching fraction of the decay $B_s^0 \to D_s^\mp K^\pm$ is a first step towards the precise determination of *CP* violation in this decay. In particular, an accurate determination is of interest, because:

- A previous measurement of $\mathcal{B}(B_s^0 \to D_s^\mp K^\pm)$ by LHCb [13], on only 10% of the currently available data, yielded a value incompatible with theoretical expectations [14]. By using the full data set, this analysis aims to confirm or refute this tension.

- To first order, only tree diagrams (such as the one in Fig. 4) play a role in the decay. Most theories for physics beyond the standard model (BSM) affect processes that involve loop diagrams. This branching fraction therefore gives a good gauge of the SM for other measurements to compare to.

- The difference between the branching fractions $\mathcal{B}(B_s^0 \to D_s^- K^+)$ and $\mathcal{B}(B_s^0 \to D_s^+ K^-)$ *does* provide a measurement of *CP* violation. Until now, statistics have been a limiting factor for this measurement, but it might be feasible using the full LHCb dataset.

2 Introduction

The decay $B_s^0 \to D_s^\mp K^\pm$ can occur through two different tree-level diagrams of similar magnitude. The branching fraction of this decay relative to that of $B_s^0 \to D_s^- \pi^+$ has been previously measured by the LHCb collaboration [13] using data of pp collisions recorded in 2011, corresponding to an integrated luminosity of $336\,\text{pb}^{-1}$, at a centre-of-mass energy of 7 TeV. This measurement reported a value of $0.0646 \pm 0.0043 \pm 0.0025$. This value has been found to be incompatible with Standard Model (SM) expectations [14]. This note presents an update of this measurement, using pp collision data corresponding to $3\,\text{fb}^{-1}$, taken in 2011 ($1\,\text{fb}^{-1}$) and 2012 ($2\,\text{fb}^{-1}$) at $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$, respectively.

The branching fraction of $B_s^0 \to D_s^\mp K^\pm$ is measured relative to that of the decay $B_s^0 \to D_s^- \pi^+$ because of the symmetry between the two processes [15]. The diagrams are related to one another by changing the final state d quark with an s quark (or the pion with a kaon), and vice versa. The difference is that $B_s^0 \to D_s^\mp K^\pm$ can decay into both charge conjugate final states, while the process $B_s^0 \to D_s^- \pi^+$ can only do so through B_s^0–$\overline{B}_s^0$ mixing (see also Table 1). In addition, the decay $B_s^0 \to D_s^\mp K^\pm$ can occur also through the exchange topology.

The branching fraction of $B_s^0 \to D_s^- \pi^+$ is subsequently determined relative to that of the process $B^0 \to D^- \pi^+$, using the relative production rates of the B^0 and B_s^0 mesons f_s/f_d, determined with semileptonic B decays [16].

The branching fraction of the decay $B^0 \to D_s^- K^+$ is also measured. It is normalised to that of $B^0 \to D^- \pi^+$. The process $B^0 \to D_s^- K^+$ is interesting because the lowest-order decay topology is an exchange topology (see Fig. 6), featuring the exchange of a W boson between the two quarks of the original B_s^0 meson. The measurement of $\mathcal{B}(B_s^0 \to D_s^\mp K^\pm)$ therefore provides a direct estimate of the magnitude of this type of diagrams.

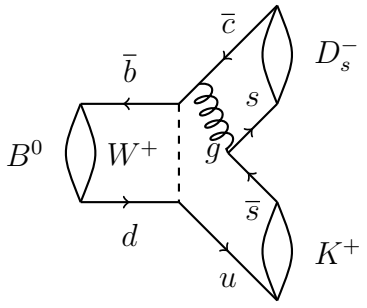

Figure 6: Feynman diagram of the process $B^0 \to D_s^- K^+$.

Table 1: The types of decay topologies of the decays relevant to this analysis.

Decay	$b \to c$ Tree	$b \to u$ Tree	Exchange
$B^0 \to D^- \pi^+$	✓		✓
$B_s^0 \to D_s^- \pi^+$	✓		
$B^0 \to D_s^- K^+$			✓
$B_s^0 \to D_s^\mp K^\pm$	✓	✓	✓

3 Dataset and trigger

When the LHC is operational, about 4×10^7 bunches of protons cross every second at so-called interaction points. Many of these events are of no particular interest to this analysis as they do not contain the relevant $B^0_{(s)}$ decays. The event selection consists of a number of steps. First, the events pass through a trigger system, described in section 3.1. Next, the particles undergo offline reconstruction and a first, loose, selection (see section 3.2). Finally, the events pass another round of selection requirements specific to the analysis, which is described in the next chapter.

3.1 Dataset and Trigger

This analysis is performed on pp collision data corresponding to $1\,\text{fb}^{-1}$ at $\sqrt{s} = 7\,\text{TeV}$ and $2\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$, for a total of $3\,\text{fb}^{-1}$. The data was taken using the triggers described in Refs. [17, 18].

The trigger consists of two stages: a hardware trigger, followed by a software trigger. The lowest trigger level is implemented in hardware, and selects events that contain a hadron with transverse energy greater than 3.6 GeV. The subsequent software trigger requires a two-, three- or four-track secondary vertex with a large sum of the transverse momentum, p_T, of the tracks and a significant displacement from the primary pp interaction vertices (PVs). At least one track should have $p_\text{T} > 1.7\,\text{GeV}/c$ and χ^2_IP with respect to any primary interaction greater than 16, where χ^2_IP is defined as the difference in χ^2 of a given PV reconstructed with and without the considered candidate.

3.2 Stripping selection

The first event selection ("stripping") is made based on kinematical and geometrical criteria. The stripping lines[2] proceed as follows. First, a $D^\pm$ or $D^\pm_s$ candidate is constructed by requiring a combination of three light hadrons (pions and kaons), each of which has a track χ^2/ndf no greater than 3.0, a transverse momentum (p_T) of at least $100\,\text{MeV}/c$, and a momentum p of at least $1000\,\text{MeV}/c$, typical for decay products of charmed mesons.

Furthermore, each final state particle originating from a $B^0_{(s)}$ decay must be detached from the primary vertex. If the χ^2/ndf of the primary vertex fit, denoted χ^2_IP, increases by 4 or more when including the candidate track, the track is used.

In order to be combined, one of the $D^\pm_{(s)}$ daughters must have a track with $\chi^2/\text{ndf} < 2.5$, $p_\text{T} > 500\,\text{MeV}/c$, and $p > 5000\,\text{MeV}/c$. The invariant mass of the combination must lie between 1769.62 and 2068.49 MeV/c^2, representing a mass window extending

[2]The stripping lines used are the same as were used in the previous measurement, the `B02DKD2HHH` and `B02DPiD2HHH` `Beauty2CharmLine` stripping lines in the `B2DX` family. The stripping has been updated from version 17, used in the previous analysis, to version 20 and 20r1 in this analysis for the 2012 and 2011 data, respectively. The differences between versions 20 and 20r1 are limited to the ghost probabilty of the track of each individual particle being at most 0.4 and 0.3, respectively. The differences between these versions of stripping and version 17 are summarised in table 2.

Table 2: Differences between stripping versions 17, 20, and 20r1.

Variable	Stripping 17	Stripping 20	Stripping 20r1
Individual particles			
Track $\chi^2_{\rm IP}$	> 4	> 3	> 3
Track ghost probability		< 0.4	< 0.3
Bachelor particle			
Track χ^2/ndf	< 3	< 2.5	< 2.5
D, D_s			
Daughter track χ^2/ndf	< 3	< 2.5	< 2.5
Global cuts			
#Long tracks	< 500	< 250	< 250

from $100\,\text{MeV}/c^2$ below the $D^\pm$ mass to $100\,\text{MeV}/c^2$ above the $D_s^\pm$ mass. The $D^\pm_{(s)}$ candidate is also required to have a $p_{\rm T}$ of at least $1800\,\text{MeV}/c$, and it must have a distance of closest approach of at least 0.5 mm.

Finally, the $D^\pm$ or $D_s^\pm$ candidate must be combined with a bachelor pion or kaon. This bachelor particle is required to have a track χ^2/ndf smaller than 2.5 and $p_{\rm T} > 500\,\text{MeV}/c$, $p > 5000\,\text{MeV}/c$, $\chi^2_{\rm IP} > 4.0$, and track ghost probability < 0.3 (< 0.4 for the 2012 data). The $B^0_{(s)}$ candinate is a combination of the $D^\pm_{(s)}$ candidate with this bachelor particle with an invariant mass between 4750 and $6000\,\text{MeV}/c^2$ and a reconstructed lifetime of at least 2 ps. Some additional constraints are set on the final state particles before they are combined into a $B^0_{(s)}$ candidate: at least one of the tracks should have $p > 10\,\text{GeV}/c$ and $p_{\rm T} > 1700\,\text{MeV}/c$.

3.3 Simulated data

Simulated (Monte Carlo) data is used to estimate the invariant mass distributions of the background processes, as well as to obtain the shape of the invariant mass distributions of the signal decays, and to obtain the selection efficiency.

In the simulation, pp collisions are generated using Pythia [19] with a specific LHCb configuration [20]. Decays of hadronic particles are described by EvtGen [21], in which final state radiation is generated using Photos [22]. The interaction of the generated particles with the detector and its response are implemented using the Geant4 toolkit [23] as described in Ref. [24].

4 Offline selection and efficiencies

While the stripping provides an efficient selection of candidate events, it does not sufficiently reject background events. In order to obtain a cleaner data sample, another sequence of cuts is applied to the stripped sample. The corresponding efficiencies must be understood and estimated in order to eventually correct the measured event yields. Except where noted, the same offline selection is applied to each decay type.

4.1 Offline selection

Two different kinds of cuts are applied to the data sample: kinematic cuts, which are cuts on the kinematic properties (such as invariant mass and momentum) of the candidates; and particle identification (PID) cuts, which use information from the RICH subdetectors of LHCb. Each cut is outlined below.

First, the events are limited to a reconstructed $B^0_{(s)}$ meson mass between 5000 and 5800 MeV/c^2. Secondly, the reconstructed $D^\pm$ ($D^\pm_s$) candidate is required to fall in the mass window between 1844 and 1890 MeV/c^2 (1944 and 1990 MeV/c^2).

The next cuts are so-called fiducial cuts, which limit the data sample to the range detectable by the detector. This means limiting the p_{T} of the $B^0_{(s)}$ candidate to the range $[1500, 40000]$ MeV/c, and their pseudorapidity η to $[2, 5]$.

The final states can be produced through different decay channels as well as the ones relevant to this analysis, such as the charmless decay $B^0_s \to KKK\pi$ under the process $B^0_s \to D^\mp_s K^\pm$. To remove these background processes, the flight distance χ^2 of the $D^\pm_{(s)}$ candidate is required to be greater than 2.0.

After these cuts, a PID cut is applied to the bachelor particle. These are cuts on the difference of the log-likelihood of the two particle hypotheses, called the $\mathrm{DLL}_{K\pi}$ and $\mathrm{DLL}_{p\pi}$ variables. The cuts are outlined, for each different decay channel, in Table 3.

Table 3: PID cuts on the final state light hadron candidates. The PID cuts on the bachelor particle and on the kaon with opposite charge relative to the D^- meson are both tight, as these cuts distinguish the different decays. Note that the $\mathrm{DLL}_{p\pi} - \mathrm{DLL}_{K\pi}$ cuts in the $B^0_s \to D^-_s \pi^+$ and $B^0_s \to D^\mp_s K^\pm$ analyses are referring to the Λ^+_c veto procedure, as described in the text.

Cut	$B^0 \to D^-\pi^+$	$B^0_s \to D^-_s \pi^+$	$B^0_s \to D^\mp_s K^\pm$	Type
Bachelor	$\mathrm{DLL}_{K\pi} < 0$	$\mathrm{DLL}_{K\pi} < 0$	$\mathrm{DLL}_{K\pi} > 5$	Tight
$\pi^\pm$ from $D^\pm_{(s)}$	$\mathrm{DLL}_{K\pi} < 5$	$\mathrm{DLL}_{K\pi} < 5$	$\mathrm{DLL}_{K\pi} < 5$	Loose
	$\mathrm{DLL}_{p\pi} < 15$			
K^+ from $D^-_{(s)}$	$\mathrm{DLL}_{K\pi} > 0$	$\mathrm{DLL}_{K\pi} > 0$	$\mathrm{DLL}_{K\pi} > 0$	Loose
		$\mathrm{DLL}_{p\pi} - \mathrm{DLL}_{K\pi} < 0$	$\mathrm{DLL}_{p\pi} - \mathrm{DLL}_{K\pi} < 0$	
K^- from D^+_s	–	$\mathrm{DLL}_{K\pi} > 5$	$\mathrm{DLL}_{K\pi} > 5$	Tight

Next, a multivariate algorithm [25, 26] is used to select candidates consistent with the decay of a b hadron. This boosted decision tree (BDT) combines several kinematic and geometric variables of the event and outputs a single number in the interval $[-1, 1]$. The lower this number, the more likely the event is a background event; if the number is higher, the event is more signal-like. The BDT has been trained (calibrated) by supplying data in the relevant mass window, weighted with its similiarity to actual signal events; and a sample of background events in the upper mass sideband with an invariant $D^-_{(s)}\pi^+$ mass $> 5500\,\text{MeV}$. This BDT is the same as the one used in the time-dependent $B^0_s \to D^\mp_s K^\pm$ analysis [27]. The value resulting from this BDT is required to be greater than 0.3. This value rejects most of the background, without significantly suppressing the signal.

Finally, in the analyses containing a $D^\pm_s$ meson, a veto is applied to reduce the contribution of the $\Lambda^0_b \to \Lambda^+_c\pi^-$ ($\Lambda^+_c \to pK^-\pi^+$) background process, resulting from the proton-to-kaon misidentification by the RICH detector. This is done by reconstructing the final state K^+ meson from the D^-_s candidate under the p hypothesis and calculating the invariant D^-_s mass. If $\text{DLL}_{p\pi} - \text{DLL}_{K\pi}$ for a particular event is smaller than 0 and this mass is within $21\,\text{MeV}/c^2$ of the Λ^+_c mass $2286.46\,\text{MeV}/c^2$, the event is removed from the sample. This process, referred to from now on as the Λ^+_c veto procedure, ensures an almost fully efficient reduction of the background process $\Lambda^0_b \to \Lambda^+_c\pi^-$.

4.2 Selection efficiences

In order to correctly determine the ratio of branching fractions, the relative reconstruction and selection efficiencies of the channels must be taken into account. Each of the cuts outlined in the previous section has a specific efficiency for each process. All signal decays studied here have a very similar decay topology, but nevertheless these efficiencies may be different for each of the processes.

The efficiencies of the kinematic selection criteria are determined from simulation (see section 3.3), as the kinematic distributions are well modelled. These efficiencies are listed in Tables 4, 5, and 6. It should be noted that certain trigger requirements are applied within the stripping selection. The trigger cuts listed in the tables require in addition that the trigger decision was applied to the signal tracks.

In contrast, the PID performance to select kaons and pions is modelled less accurately, and needs to be determined from data. These efficiencies depend strongly on the kinematics of the particles, as well as properties of the events in which they occur. The kaon-pion efficiency is taken from a dedicated sample of $D^{*+} \to (D^0 \to K^-\pi^+)\pi^+$ decays, which provide a clean resonance peak at 2010 MeV in which the two resulting pions have the same charge. Because of the small mass difference between the $D^{*\pm}$ meson and the $\overset{(-)}{D}{}^0$ meson, the bachelor pion is slow and easily identifiable. Using the fact that the pion charges are identical, unambiguous identification of the final state particles is possible without requiring PID information from the RICH, yielding an accurate estimate of the pion-kaon PID efficiency. The proton-pion efficiency is determined in a similar way from $\Lambda \to p\pi^-$ decays.

The variables in which the PID performance is determined are track momentum, track

pseudorapidity η, and number of tracks in the event. PID performance histograms are generated by calculating, in bins of these three variables, the fraction of correctly identified particles, yielding a three-dimensional histogram with values between 0 and 1. This is a statistical method and does not yield exact values for specific decay channels, but it does provide a good data-driven estimate for the PID performance.

By weighting a Monte Carlo sample of simulated $B^0 \to D^-\pi^+$, $B_s^0 \to D_s^-\pi^+$ or $B_s^0 \to D_s^\mp K^\pm$ data to the PID performance histograms created this way, an estimate of the PID efficiency of those respective processes is obtained. The results of these performance estimations are summarised in Table 7. This reweighting is performed in the three variables mentioned above.

Table 4: Efficiencies of the kinematic cuts of $B^0 \to D^-\pi^+$.

	ε_{rel} (%)	ε_{cum} (%)
Generator level efficiency	16.10 ± 0.09	16.10 ± 0.09
Reconstruction and stripping	13.65 ± 0.03	2.20 ± 0.01
Trigger cuts	96.70 ± 0.10	2.12 ± 0.01
$B^0_{(s)}$ mass window cuts	99.32 ± 0.04	2.11 ± 0.01
$D^\pm_{(s)}$ mass window cuts	95.65 ± 0.11	2.02 ± 0.01
Fiducial cuts	99.53 ± 0.04	2.01 ± 0.01
Flight distance cuts	94.24 ± 0.12	1.89 ± 0.01
BDT cuts	95.62 ± 0.11	1.82 ± 0.01
Total		1.82 ± 0.01

Table 5: Efficiencies of the kinematic cuts of $B_s^0 \to D_s^- \pi^+$.

	ε_{rel} (%)	ε_{cum} (%)
Generator level efficiency	17.06 ± 0.38	17.06 ± 0.38
Reconstruction and stripping	14.37 ± 0.03	2.45 ± 0.05
Trigger cuts	93.53 ± 0.13	2.29 ± 0.05
$B_{(s)}^0$ mass window cuts	99.47 ± 0.04	2.28 ± 0.05
$D_{(s)}^{\pm}$ mass window cuts	97.27 ± 0.09	2.22 ± 0.05
Fiducial cuts	99.41 ± 0.04	2.21 ± 0.05
Flight distance cuts	87.78 ± 0.17	1.94 ± 0.04
BDT cuts	96.22 ± 0.10	1.88 ± 0.04
Λ_c^+ veto	99.08 ± 0.05	1.86 ± 0.04
Total		1.86 ± 0.04

Table 6: Efficiencies of the kinematic cuts of $B_s^0 \to D_s^{\mp} K^{\pm}$ and $B^0 \to D_s^- K^+$.

$B_s^0 \to D_s^{\mp} K^{\pm}$	ε_{rel} (%)	ε_{cum} (%)
Generator level efficiency	17.72 ± 0.07	17.72 ± 0.07
Reconstruction and stripping	15.08 ± 0.05	2.67 ± 0.01
Trigger cuts	93.64 ± 0.17	2.50 ± 0.01
$B_{(s)}^0$ mass window cuts	99.72 ± 0.04	2.49 ± 0.01
$D_{(s)}^{\pm}$ mass window cuts	97.54 ± 0.11	2.43 ± 0.01
Fiducial cuts	99.44 ± 0.05	2.42 ± 0.01
Flight distance cuts	87.91 ± 0.23	2.13 ± 0.01
BDT cuts	96.47 ± 0.13	2.06 ± 0.01
Λ_c^+ veto	99.13 ± 0.06	2.05 ± 0.01
Total		2.05 ± 0.01
$B^0 \to D_s^- K^+$	ε_{rel} (%)	ε_{cum} (%)
Generator level efficiency	15.27 ± 0.04	15.27 ± 0.04
Reconstruction and stripping	15.08 ± 0.05	2.30 ± 0.01
Trigger cuts	92.14 ± 0.80	2.12 ± 0.02
$B_{(s)}^0$ mass window cuts	99.47 ± 0.22	2.11 ± 0.02
$D_{(s)}^{\pm}$ mass window cuts	97.26 ± 0.49	2.05 ± 0.02
Fiducial cuts	99.20 ± 0.26	2.03 ± 0.02
Flight distance cuts	88.43 ± 0.95	1.80 ± 0.03
BDT cuts	96.02 ± 0.58	1.74 ± 0.03
Λ_c^+ veto	99.03 ± 0.29	1.72 ± 0.03
Total		1.72 ± 0.03

Table 7: Efficiencies of the PID cuts. Note that the PID efficiency for the selection of the bachelor particle is not independent of that of the $D^{\pm}_{(s)}$ daughters.

	PID efficiency (%)			
	$B^0 \to D^- \pi^+$	$B^0_s \to D^-_s \pi^+$	$B^0 \to D^-_s K^+$	$B^0_s \to D^{\mp}_s K^{\pm}$
Cut on bachelor	84.27 ± 0.06	84.24 ± 0.06	70.55 ± 0.85	69.70 ± 0.20
Cut on $D^{\pm}_{(s)}$ meson	87.91 ± 0.05	79.72 ± 0.10	80.42 ± 0.56	79.59 ± 0.14
Total	74.29 ± 0.07	67.76 ± 0.10	58.11 ± 0.82	56.64 ± 0.19

Table 8: $(\pi \leftrightarrow K)$-misidentification rates of the PID cuts. Note that again the bachelor and $D^{\pm}_{(s)}$ meson misidentification rates are not independent of one another.

	Misidentification rate (%)			
	$B^0 \to D^- \pi^+$	$B^0_s \to D^-_s \pi^+$	$B^0 \to D^-_s K^+$	$B^0_s \to D^{\mp}_s K^{\pm}$
Cut on bachelor	9.92 ± 0.05	10.14 ± 0.05	2.13 ± 0.05	2.12 ± 0.01
Cut on $D^{\pm}_{(s)}$ meson	15.42 ± 0.10	2.63 ± 0.02	2.79 ± 0.13	2.63 ± 0.02
Total	1.625 ± 0.018	0.261 ± 0.004	0.061 ± 0.003	0.057 ± 0.001

5 Fitting

5.1 Background components

Various background components play a role in the processes $B^0 \to D^-\pi^+$, $B_s^0 \to D_s^-\pi^+$, and $B_s^0 \to D_s^\mp K^\pm$. There are background contributions resulting from misidentification, such as $B_s^0 \to D_s^-\pi^+$ under $B_s^0 \to D_s^\mp K^\pm$, where one of the final state particles is wrongly identified. The excellent particle identification in LHCb reduces each of those by about $85-98\%$ (see Table 8), so only a small contribution remains. There are partially reconstructed background processes, such as $B_s^0 \to D_s^-(K^{*+} \to K^+\pi^0)$, where a neutral particle is not reconstructed. The invariant mass of these processes is shifted to lower values, and, because the momentum of the missed particle is unknown, smeared out. Finally, there is a contribution from combinatorial background – random tracks that accidentally form a $D_s^\mp K^\pm$ candidate in the relevant mass range.

The relevant background components are enumerated in Table 9. Each of these components is taken into account when fitting the invariant mass distribution, in order to obtain an accurate value for the signal yield. The shape of each of them, known as a "template", is taken from simulations, generated by Pythia versions 6 and 8 (see section 3.3). These events are simulated at a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$. To improve the accuracy of the fit, the yield of each of the background components is estimated using prior knowledge. This information includes the relative branching fractions, B^0 and B_s^0 meson production fractions (f_d and f_s, respectively), and reconstruction and selection efficiency. The reconstruction efficiency in turn depends on the (PID) performance and the efficiency of the stripping and track reconstruction applied to the raw data, as well as the efficiency of the cuts on the mass range of the $B_{(s)}^0$, the mass range of the $D_{(s)}^\pm$, and the BDT efficiency (see section 4).

The mass distribution for some background components are not distinct enough to allow for a precise yield estimate from the fit. Therefore, the yield is constrained in the fit, but not fixed. These components are marked in the "GC" (Gaussian constraint) column of Table 9. The corresponding Gaussian has its mean set to the calculated yield estimation, and its width to 10% of that number. This value is an approximation of the total error on the estimated yields, to correct for statistical and systematic deviations from that number.

The templates are obtained from fits to the simulated data, resulting in a non-parametric function. This fit takes for each data point a Gaussian distribution, then smears them out over the mass range to obtain a smooth shape [28]. In this process, the simulated data for magnet up and magnet down have been combined to increase the statistical precision. The resulting templates are shown in appendix A.

The PID performance is determined in bins of three different variables – hadron momentum, hadron pseudorapidity, and total number of tracks in the event – to give an estimate of the number of (in)correct identifications performed on a specific kind of particle in a sample. This is done separately for both magnet polarities. The relevant momentum and angular distributions of the final state kaons and pions are taken from simulation, which is assumed to have similar distributions in the three PID variables. The background

templates are then reweighted to the PID performance in bins of the three variables. The PID reweighting also contributes the only difference between the two magnet polarities in the background templates.

Table 9: Background and signal yield estimates of the different processes at hand: $B^0 \to D^-\pi^+$, $B_s^0 \to D_s^-\pi^+$, and $B_s^0 \to D_s^\mp K^\pm$. Type PR means partially reconstructed, M means misidentified, and S means signal. GC implies that the yield is Gaussian constrained. Prod. is the hadronisation fraction relative to that of the B^0 meson, f_d. PID is the probability of correctly identifying the final state particles of the given process. Rel. reco is the relative reconstruction efficiency, normalised to that of the parent process. This is a combination of mass range cuts, stripping efficiency, BDT cut, and Λ_c^+ veto, as appropriate per analysis. Rel. yield is the relative yield, again normalised to that of the parent process, and Exp. yield is the expected yield. The expected yield from $B^0 \to D^-\pi^+$ was taken from the fit, and the other yields are estimated relative to the $B^0 \to D^-\pi^+$ yield.

Process	Background	Type	GC	$\mathcal{B}/\ 10^{-4}$	Prod.	PID	Rel. reco	Rel. yield	Exp. yield / 10^3
$B^0 \to D^-\pi^+$		S		27	1.0	0.74	1.00	1.0	454
	$B^0 \to D^-K^+$	M	✓	2.0	1.0	0.09	1.05	0.010	4.4
	$B^0 \to D^-\rho^+$	PR		78	1.0	0.74	0.16	0.516	234
	$B^0 \to D^{*-}\pi^+$	PR		28	1.0	0.73	0.74	0.238	108
	$B_s^0 \to D_s^-\pi^+$	PR, M	✓	30	0.26	0.18	0.41	0.017	7.3
	$\Lambda_b^0 \to \Lambda_c^+\pi^-$	M	✓	43	0.39	0.23	0.05	0.025	4.2
$B_s^0 \to D_s^-\pi^+$		S		30	0.26	0.67	1.00	1.0	69
	$B^0 \to D_s^-\pi^+$	S		0.2	1.0	0.67	1.02	0.049	3.3
	$B^0 \to D^-\pi^+$	M	✓	27	1.0	0.01	0.37	0.045	3.1
	$B_s^0 \to D_s^-\rho^+$	PR		74	0.26	0.68	0.21	0.559	38.4
	$B_s^0 \to D_s^{*-}\pi^+$	PR		20	0.26	0.67	0.93	0.658	45.1
	$\Lambda_b^0 \to \Lambda_c^+\pi^-$	M	✓	43	0.39	0.39	0.01	0.012	0.9
$B_s^0 \to D_s^\mp K^\pm$		S		2	0.26	0.56	1.00	1.0	4.1
	$B^0 \to D_s^-K^+$	S		0.2	1.0	0.56	1.02	0.419	1.7
	$B^0 \to D^-K^+$	M	✓	2	1.0	0.01	0.37	0.060	0.2
	$B_s^0 \to D_s^-\pi^+$	M	✓	30	0.26	0.02	0.93	0.398	1.6
	$B_s^0 \to D_s^\mp K^{*\pm}$	PR	✓	5	0.26	0.56	0.28	0.070	2.9
	$B_s^0 \to D_s^{*\mp} K^\pm$	PR	✓	2	0.26	0.56	0.86	0.858	3.5
	$B_s^0 \to D_s^{*-}\pi^+$	PR, M	✓	20	0.26	0.02	0.88	0.252	1.0
	$B_s^0 \to D_s^-\rho^+$	PR, M	✓	74	0.26	0.02	0.24	0.246	1.0
	$B_s^0 \to D_s^{*-}\rho^+$	PR, M	✓	97	0.26	0.02	0.03	0.049	0.2
	$\Lambda_b^0 \to \Lambda_c^+\pi^-$	M	✓	43	0.39	0.01	0.01	0.005	0.2
	$\Lambda_b^0 \to D_s^-p$	M	✓	0.1	0.39	0.36	0.78	0.034	0.1
	$\Lambda_b^0 \to D_s^{*-}p$	PR, M	✓	0.1	0.39	0.34	0.62	0.025	0.1

5.2 Signal shapes

The total signal yield is obtained by fitting the signal peak with a signal shape, of which different models have been tested. The tails on each side of the signal peak arise from detector resolution effects and from energy loss due to final state radiation. The various tested models share a Gaussian-like middle part, surrounded by two (possibly asymmetric) decreasing tails. The shape that was eventually chosen in the fits is the double Crystal Ball function [29]. This function consists of a central Gaussian part with on each side an exponential tail. The full definition of the double Crystal Ball function is given in Eq. (1).

$$
\begin{aligned}
&f(x;\overline{x},\sigma,\alpha_1,\alpha_2,n_1,n_2,\varepsilon) = \\
&\quad N \begin{cases}
\varepsilon \left(\dfrac{n_1}{|\alpha_1|}\right)^{n_1} \exp\left(-\dfrac{|\alpha_1|}{2}\right) \left(\dfrac{n_1}{|\alpha_1|} - |\alpha_1| - \dfrac{x-\overline{x}}{\sigma}\right)^{-n_1} & \text{for } \dfrac{x-\overline{x}}{\sigma} > \alpha_1 \\
(1-\varepsilon) \left(\dfrac{n_2}{|\alpha_2|}\right)^{n_2} \exp\left(-\dfrac{|\alpha_2|}{2}\right) \left(\dfrac{n_2}{|\alpha_2|} - |\alpha_2| - \dfrac{x-\overline{x}}{\sigma}\right)^{-n_2} & \text{for } \dfrac{x-\overline{x}}{\sigma} < \alpha_2 \\
\exp\left(-\dfrac{(x-\overline{x})^2}{2\sigma^2}\right) & \text{otherwise.}
\end{cases}
\end{aligned}
\tag{1}
$$

This function has seven parameters: the mean $\overline{x}$ and standard deviation σ of the Gaussians (shared between them), the exponential constants $\alpha_{1,2}$ of the two tails, and the distances $n_{1,2}$ from the mean at which the tails start (measured in units of σ). The parameter $\varepsilon \in [0,1]$ represents the relative fraction of the two Crystal Ball functions. This parameter has been fixed to 0.5 in all applications of this function throughout the full analysis, as the central Guassian is chosen to be the same for both Crystal Ball functions. N is not a parameter of the function, rather it is an overall normalization factor depending on the parameters.

Results of this fit to simulated data samples are shown in Fig. 7, and the parameter values in Table 10. Other functions that have been investigated for use in the mass fit, as well as results from fits to simulated signal samples, can be found in appendix B.

Table 10: Results of double Crystal Ball function fits to simulated signal samples.

Parameter	Fitted value
$B^0 \to D^- \pi^+$	
$\mathrm{Fract}_{\mathrm{Sig-CBs}}$	0.500000
α_1	1.5082 ± 0.0431
α_2	-1.9305 ± 0.0636
$\overline{\mathrm{x}}$	5284.0 ± 0.1
n_1	1.69 ± 0.09
n_2	3.04 ± 0.26
σ	16.9 ± 0.1
$B_s^0 \to D_s^- \pi^+$	
$\mathrm{Fract}_{\mathrm{Sig-CBs}}$	0.500000
α_1	1.5012 ± 0.0410
α_2	-1.7519 ± 0.0680
$\overline{\mathrm{x}}$	5370.4 ± 0.1
n_1	1.73 ± 0.08
n_2	3.47 ± 0.32
σ	16.4 ± 0.1
$B_s^0 \to D_s^{\mp} K^{\pm}$	
$\mathrm{Fract}_{\mathrm{Sig-CBs}}$	0.500000
α_1	1.6501 ± 0.0611
α_2	-1.6099 ± 0.0901
$\overline{\mathrm{x}}$	5370.4 ± 0.2
n_1	1.88 ± 0.13
n_2	6.27 ± 1.22
σ	15.8 ± 0.2

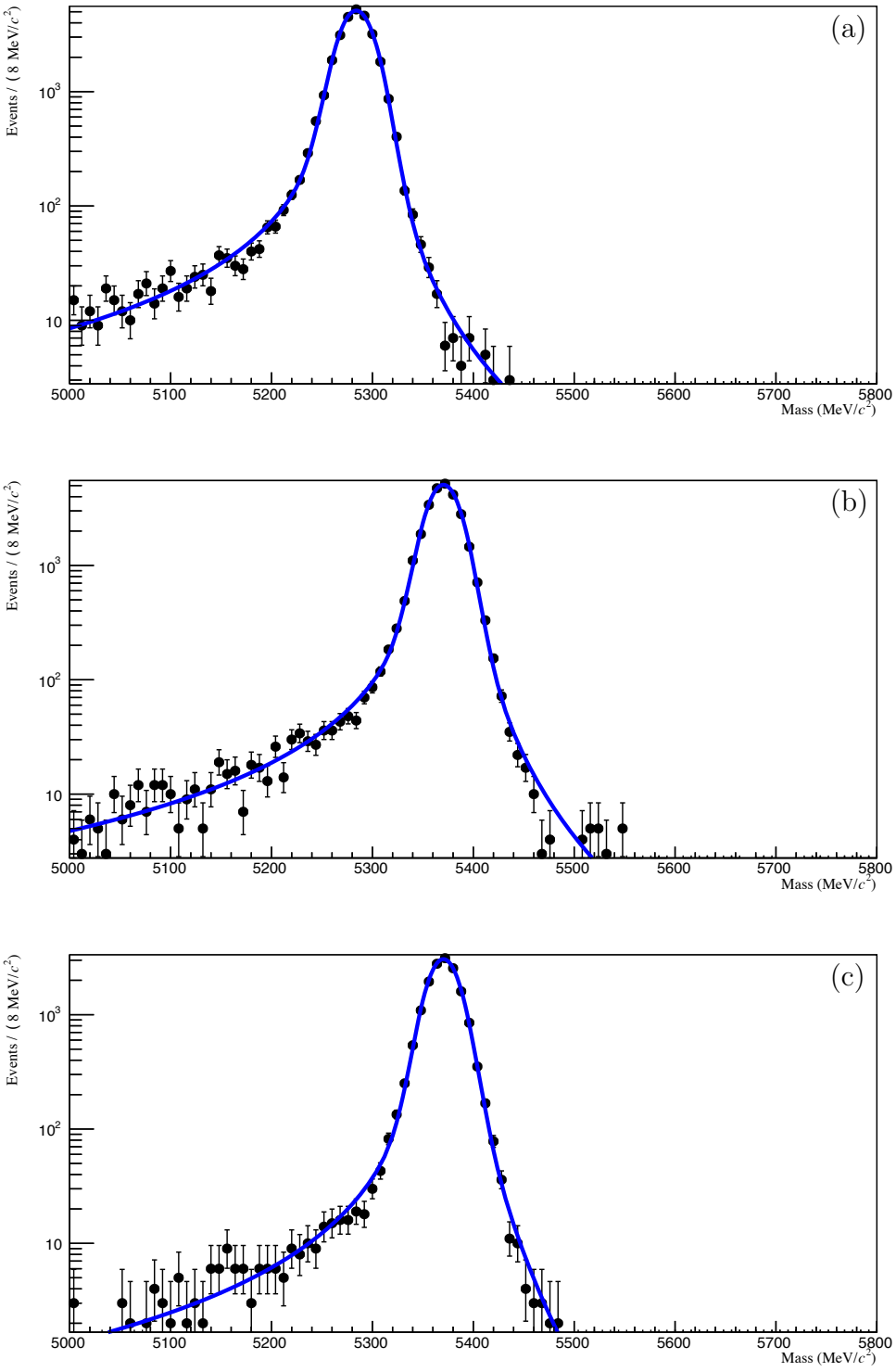

Figure 7: Double Crystal Ball function signal fits for (a) $B^0 \to D^- \pi^+$, (b) $B_s^0 \to D_s^- \pi^+$, and (c) $B_s^0 \to D_s^\mp K^\pm$.

5.3 Combinatorial

Combinatorial background events arise from the $D^{\pm}_{(s)}$ candidates, combined with a random track to form a fake $B^0_{(s)}$ candidate, or from fake $D^{\pm}_{(s)}$ candidates. Most of the contribution of the combinatorial background is rejected by the BDT selection (see section 4), but some still remains and must be accounted for in the fit.

The shape of the combinatorial background is estimated from wrong-sign samples. These are $D^{\pm}_{(s)}\pi^{\pm}$ and $D^{\pm}_{s}K^{\pm}$ candidates that can not have originated from $B^0_{(s)}$ decays. Like true combinatorial background, these events do not actually represent $B^0_{(s)}$ candidates; rather, they are combinations of random tracks. These samples are data samples aquired, reconstructed and stripped in the same way as the normal data samples. They are subsequently fitted with a function of the form

$$f(x; p_0, p_1) = p_0 + (1 - p_0)\exp(p_1 x).$$

The results of these fits are displayed in Table 11. The shape of the wrong-sign mass distributions for D^+_s candidates is consistent with a pure exponential, *i.e.* with a vanishing value for p_0. The plots can be found in App. C.

Table 11: Results from fits to the wrong-sign samples for each of the decays. The fits are performed in the $[5000, 5800]\,\mathrm{MeV}/c^2$ mass range for $B^0 \to D^-\pi^+$ and in the $[5100, 5800]\,\mathrm{MeV}/c^2$ mass range for $B^0_s \to D^-_s\pi^+$ and $B^0_s \to D^{\mp}_s K^{\pm}$, and are done separately for magnet up and down.

Wrong-sign sample	p_0	p_1
$B^0 \to D^{\mp}\pi^{\mp}$ (magnet up)	0.4512 ± 0.0448	-0.0031 ± 0.0003
$B^0 \to D^{\mp}\pi^{\mp}$ (magnet down)	0.1164 ± 0.1306	-0.0019 ± 0.0003
$B^0_s \to D^{\mp}_s\pi^{\mp}$ (magnet up)	0.0000 ± 0.0013	-0.00319 ± 0.00005
$B^0_s \to D^{\mp}_s\pi^{\mp}$ (magnet down)	0.0000 ± 0.0042	-0.00298 ± 0.00005
$B^0_s \to D^{\mp}_s K^{\mp}$ (magnet up)	0.0111 ± 0.7533	-0.0020 ± 0.0004
$B^0_s \to D^{\mp}_s K^{\mp}$ (magnet down)	0.0125 ± 0.7577	-0.0019 ± 0.0004

5.4 The fit to $B^0 \to D^-\pi^+$

The fit to $B^0 \to D^-\pi^+$ candidates is performed on the interval $[5000, 5800]\,\text{MeV}/c^2$. The background templates are obtained from simulated data with 2012 conditions at a centre-of-mass energy of 8 TeV (with the exception of the $B^0 \to D^{*-}\pi^+$ background, which was simulated with 2011 conditions, at a centre-of-mass energy of 7 TeV). All these shapes are shown in appendix A.1. The yields of the background components resulting from a misidentified final-state particle, such as the decays $B^0 \to D^-K^+$ or $B_s^0 \to D_s^-\pi^+$, are Gaussian constrained. Because the estimates of all background components (see Table 9) use the output of this fit (the $B^0 \to D^-\pi^+$ yield) as the input to calculate all the relative yields, an iterative approach has been applied, where the output of the fit was fed as input to the background estimates until the numbers stabilised.

The parameters of the double Crystal Ball function used to model the core of the signal shape, the mean $\overline{x}$ and standard deviation σ, were allowed to vary in the fit within the intervals $[5000, 5800]\,\text{MeV}/c^2$ and $[0, 20]\,\text{MeV}/c^2$, respectively. The parameters that describe the tails, $n_{1,2}$ and $\alpha_{1,2}$, were left free in the fit.

The fit is shown in Fig. 8, and the corresponding values are displayed in Table 12. The total $B^0 \to D^-\pi^+$ yield is about 460 000 events.

5.5 The fit to $B_s^0 \to D_s^-\pi^+$

The $B_s^0 \to D_s^-\pi^+$ fit is performed on the mass range $[5100, 5800]\,\text{MeV}/c^2$. The background templates are taken from 2012 simulations at a centre-of-mass energy of 8 TeV. These shapes can be found in appendix A.2. The tail parameters of the $B_s^0 \to D_s^-\pi^+$ signal shape are fixed to the values obtained in the signal shape fit to simulated events (see section 5.2). The mean of the $B^0 \to D_s^-\pi^+$ signal process is fixed to the B^0 mass, and its width and tail parameters are fixed to those of the $B_s^0 \to D_s^-\pi^+$ signal shape. As in the $B^0 \to D^-\pi^+$ fit, the yields of the misidentified background components, in this case only the decay $B^0 \to D^-\pi^+$, are Gaussian constrained around the values given in Table 9. The fit results are displayed in Table 13 and Fig. 9. The total $B_s^0 \to D_s^-\pi^+$ yield is about 76 000 events.

5.6 The fit to $B_s^0 \to D_s^\mp K^\pm$

The fit to $B_s^0 \to D_s^\mp K^\pm$, like the $B_s^0 \to D_s^-\pi^+$ fit, is performed on the mass range $[5100, 5800]\,\text{MeV}/c^2$, and its background templates are also taken from simulated data at a centre-of-mass energy of 8 TeV with 2012 conditions. These shapes can be found in appendix A.3. The yield of each of these background components is Gaussian constrained in the fit. The predicted $B_s^0 \to D_s^\mp K^\pm$ yield to which these constraints are normalised is calculated by multiplying the expected $B_s^0 \to D_s^-\pi^+$ yield with the estimated relative efficiencies.

The fit results are displayed in Table 14 and Fig. 10. The total $B_s^0 \to D_s^\mp K^\pm$ yield is about 5000 events, and the $B^0 \to D_s^-K^+$ yield about 2400.

Table 12: Fit results for $B^0 \to D^- \pi^+$.

	Magnet Up	Magnet Down
Parameters	Fit Results	Fit Results
$N_{B^0 \to D^- \pi^+}$	$217\,796 \pm 634$	$242\,643 \pm 679$
$N_{B^0 \to D^- K^+}$	$1\,678 \pm 198$	$2\,548 \pm 215$
$N_{B^0 \to D^- \rho^+}$	$121\,325 \pm 1\,470$	$136\,198 \pm 1\,580$
$N_{B^0 \to D^{*-} \pi^+}$	$42\,605 \pm 991$	$47\,133 \pm 1\,047$
$N_{B_s^0 \to D_s^- \pi^+}$	$4\,037 \pm 290$	$3\,270 \pm 317$
$N_{\Lambda_b^0 \to \Lambda_c^+ \pi^-}$	$3\,980 \pm 172$	$4\,417 \pm 180$
p_0	0.16597 ± 0.00902	0.15661 ± 0.00907
p_1	-0.00639 ± 0.00021	-0.00624 ± 0.00020
$N_{combinatorial}$	$46\,276 \pm 1\,120$	$49\,078 \pm 1\,222$
Common Parameters		
$\overline{x}$	$5\,284.18 \pm 0.04$	
σ	17.54 ± 0.05	
α_1	1.0530 ± 0.0215	
α_2	-1.5131 ± 0.0305	
n_1	6.43 ± 0.59	
n_2	118.25 ± 14.53	
Fixed Parameters		
$Fract_{Sig-CBs}$	0.50	

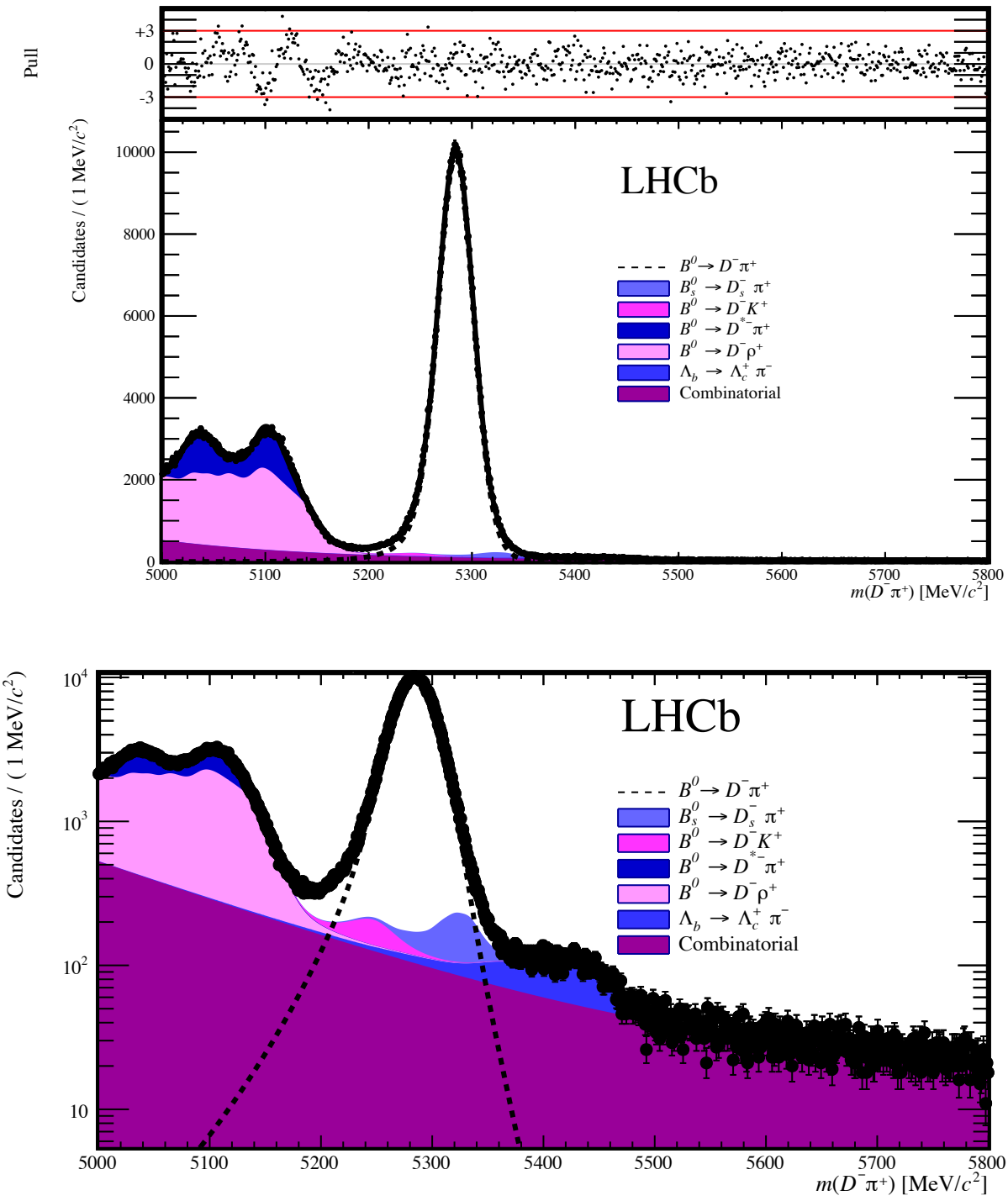

Figure 8: The mass fit to the process $B^0 \to D^-\pi^+$. The bottom figure shows the same plot in logarithmic scale.

Table 13: Fit results for $B_s^0 \to D_s^- \pi^+$.

	Magnet Up	Magnet Down
Parameters	Fit Results	Fit Results
$N_{B_s^0 \to D_s^- \pi^+}$	$35\,827 \pm 233$	$39\,814 \pm 248$
$N_{B^0 \to D_s^- \pi^+}$	$1\,334 \pm 162$	$1\,449 \pm 170$
$N_{B_s^0 \to D_s^{*-} \rho^+}$	$15\,569 \pm 983$	$16\,481 \pm 1\,024$
$N_{B_s^0 \to D_s^{*-} \pi^+}$	$19\,403 \pm 945$	$22\,480 \pm 985$
$N_{B^0 \to D^- \pi^+}$	$1\,793 \pm 124$	$2\,079 \pm 135$
p_0	0.12968 ± 0.01523	0.10163 ± 0.01451
p_1	-0.00599 ± 0.00027	-0.00577 ± 0.00025
$N_{combinatorial}$	$17\,392 \pm 758$	$19\,081 \pm 801$
Common Parameters		
$\overline{x}$	$5\,371.28 \pm 0.08$	
σ	17.49 ± 0.08	
$mean_{B^0 \to D_s^- \pi^+}$	$5\,283.0 \pm 0.0$	
σ	17.49 ± 0.08	
Fixed Parameters		
α_1	1.5012	
α_2	-1.7519	
n_1	1.73	
n_2	3.47	
$Fract_{Sig-CBs}$	0.50	
$\alpha_1(B^0 \to D_s^- \pi^+)$	1.5012	
$\alpha_1(B^0 \to D_s^- \pi^+)$	-1.7519	
$n_1(B^0 \to D_s^- \pi^+)$	1.73	
$n_2(B^0 \to D_s^- \pi^+)$	3.47	
$Fract_{Sig-CBs}(B^0 \to D_s^- \pi^+)$	0.50	

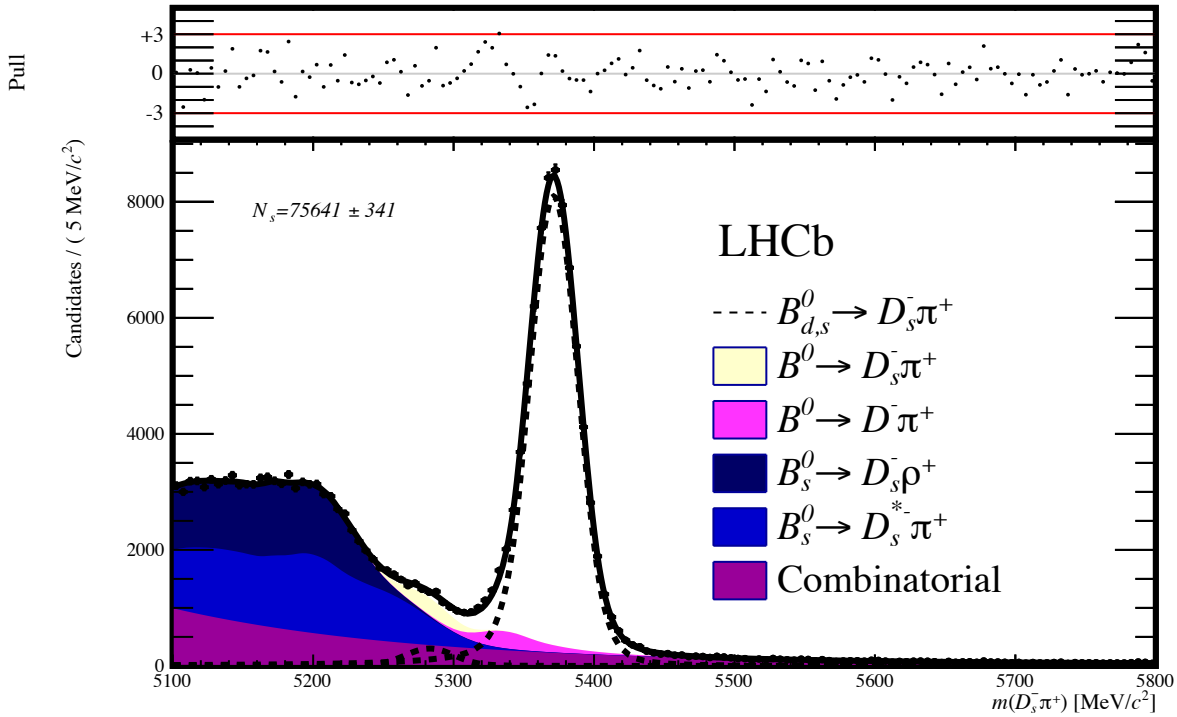

Figure 9: The mass fit to the process $B^0_s \to D^-_s\pi^+$.

Table 14: Fit results for $B_s^0 \to D_s^{\mp} K^{\pm}$.

	Magnet Up	Magnet Down
Parameters	Fit Results	Fit Results
$N_{B_s^0 \to D_s^{\mp} K^{\pm}}$	$2\,384 \pm 69$	$2\,646 \pm 72$
$N_{B^0 \to D_s^- K^+}$	$1\,190 \pm 65$	$1\,182 \pm 68$
$N_{B^0 \to D^- K^+}$	125 ± 12	131 ± 13
$N_{B_s^0 \to D_s^- \pi^+}$	953 ± 55	$1\,026 \pm 60$
$N_{B_s^0 \to D_s^{\mp} K^{*\pm}}$	819 ± 101	803 ± 107
$N_{B_s^0 \to D_s^{*\mp} K^{\pm}}$	$1\,005 \pm 131$	$1\,261 \pm 139$
$N_{B_s^0 \to D_s^{*-} \pi^+}$	488 ± 50	542 ± 54
$N_{B_s^0 \to D_s^{*-} \rho^+}$	464 ± 48	517 ± 53
$N_{B_s^0 \to D_s^{*-} \rho^+}$	96 ± 10	104 ± 11
$N_{\Lambda_b^0 \to D_s^- p}$	67 ± 7	75 ± 7
$N_{\Lambda_b^0 \to D_s^{*-} p}$	50 ± 5	55 ± 5
p_0	0.00000 ± 0.00567	0.00000 ± 0.00716
p_1	-0.00238 ± 0.00028	-0.00222 ± 0.00027
$N_{\text{combinatorial}}$	$2\,971 \pm 224$	$3\,135 \pm 231$
Common Parameters		
$\overline{x}$	$5\,371.33 \pm 0.39$	
σ	17.75 ± 0.36	
$\text{Mean}_{B^0 \to D_s^- K^+}$	$5\,283.5 \pm 0.0$	
σ	17.75 ± 0.36	
Fixed Parameters		
α_1	1.6501	
α_2	-1.6099	
n_1	1.88	
n_2	6.27	
$\text{Fract}_{\text{Sig}-\text{CBs}}$	0.50	
$\alpha_1(B^0 \to D_s^- K^+)$	1.6501	
$\alpha_2(B^0 \to D_s^- K^+)$	-1.6099	
$n_1(B^0 \to D_s^- K^+)$	1.88	
$n_2(B^0 \to D_s^- K^+)$	6.27	
$\text{Fract}_{\text{Sig}-\text{CBs}}(B^0 \to D_s^- K^+)$	0.50	

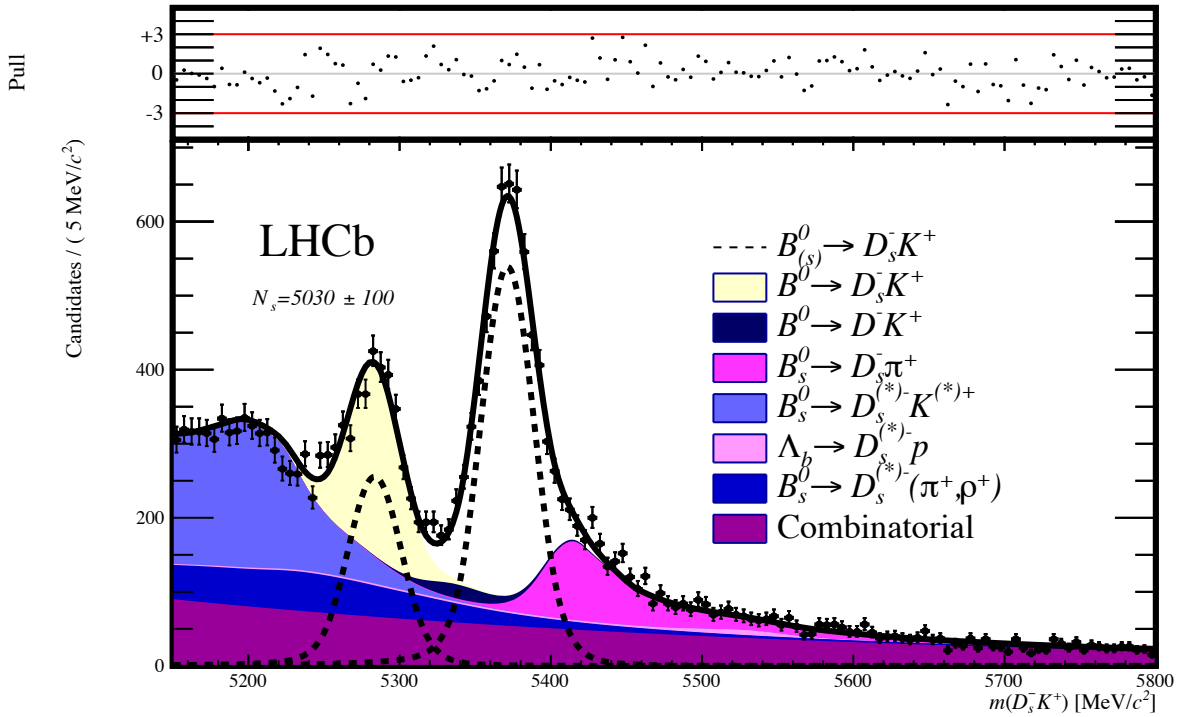

Figure 10: The mass fit to the process $B_s^0 \to D_s^{\mp} K^{\pm}$.

6 Systematic uncertainties and consistency checks

The systematic uncertainties in this analysis arise from uncertainties on the efficiency due to the kinematic and PID cuts (see section 4) and from assumptions in the fit model (described in section 5). Section 6.5 describes checks performed to assert the internal consistency of the analysis.

6.1 Uncertainty on selection efficiency

The efficiency of the kinematic selection carries a systematic uncertainty. This uncertainty is dominated by the uncertainty on the BDT selection efficiency, which relies on the agreement between data and simulation. This agreement has been studied in detail in the analysis of $B^0 \to D^-\pi^+$ and $\Lambda_b^0 \to \Lambda_c^+\pi^-$ [30, 31], and the same method has been used in this analysis. The method gauges the systematic uncertainty on the selection efficiency by applying the BDT cut to two samples of simulated data: one which is reweighed to more closely resemble actual data, and one which is not. The difference in efficiency then gives a measure of the systematic uncertainty on the BDT selection. The uncertainty on the BDT efficiency is 2.8%.

6.2 Uncertainties from PID selection

The uncertainties on the PID efficiency have been calculated for bachelor kaons and pions as well as $D^\pm_{(s)}$ daughters in Ref. [32]. These values have been obtained by using a simulated reference sample to determine the PID efficiency, rather than D^* data. The difference in PID efficiency then gives a measure of the systematic on the PID. In the systematic uncertainties of each signal decay (see Table 15) the PID uncertainties of both particles are listed. In the systematic uncertainties on the ratios (see Table 17), only the PID uncertainties have been taken into account for the particles that are different between the two decays (*e.g.* only the bachelor particle in the ratio of the decays $B_s^0 \to D_s^-\pi^+$ and $B_s^0 \to D_s^\mp K^\pm$).

6.3 Uncertainties from the fit model

To test the robustness of the fit model, and to obtain a numerical systematic uncertainty, several variations of the fit model are applied. Each time, the change in signal yield is observed, and taken as a systematic uncertainty. Table 15 lists all these uncertainties. Below, each of them is motivated and described in more detail.

Different combinatorial shape
: A variation of the combinatorial shape is tested: rather than fitting to a constant plus an exponential, the combinatorial is modelled using only an exponential.

Vary background estimates by $\pm 10\%$
: The means of the Gaussian constraints are increased or decreased by 10%. The respective widths are recalculated as 10% of the new value.

Different background templates

The background templates are made using data simulated with 2011 conditions at $\sqrt{s} = 7\,\text{TeV}$, rather than 2012 ($\sqrt{s} = 8\,\text{TeV}$). A small shift in the mass values and small differences in the shape can be caused by the different centre-of-mass energy (see Fig. 12). In the $B^0 \to D^-\pi^+$ fit, the tail parameters are fixed to the values obtained in the original fit, to prevent this variation from being taken over by changes in the tail parameters.

Reduced fit range ($B^0 \to D^-\pi^+$ only)

The effect of uncertainties in the lower mass sideband of the fit model on the signal yield is estimated by a separate fit in the mass range $[5100, 5800]\,\text{MeV}/c^2$, ignoring the events in the range $[5000, 5100]\,\text{MeV}/c^2$. The results of this fit are displayed in Fig. 13.

Fix tail parameters ($B^0 \to D^-\pi^+$ only)

In contrast to the fits to the processes $B_s^0 \to D_s^-\pi^+$ and $B_s^0 \to D_s^\mp K^\pm$, the tail parameters of the decay $B^0 \to D^-\pi^+$ are left free in the fit, as fixing them reduces the quality of the fit. In this variation, they are fixed to the values resulting from the signal shape fit to simulated data (see section 5.2). For convenience, the parameter values are reproduced in Table 16. The double Crystal Ball shapes resulting from those tail parameter values are shown in Fig. 11. The other two parameters are left free in the fit, and to provide a proper comparison, they have been set to the values from the data fit in the figure.

Because in the fits to the other decay modes the tail parameters are fixed to the values obtained from simulation, this variation can not be applied. However, if the assumption is made that the data-simulation differences are similar across the different simulated decays, the value obtained for the $B^0 \to D^-\pi^+$ analysis can be used in the others as well.

Omit $\Lambda_b^0 \to D_s^{-(*)}p$ ($B_s^0 \to D_s^\mp K^\pm$ only)

The decay $\Lambda_b^0 \to D_s^{-(*)}p$ has not been measured yet, and the branching fraction used to estimate its yield in this analysis is the current best upper limit. If the actual branching fraction is lower by any significant amount, the decay would not be measurable anymore. This is checked here by omitting it from the fit.

Table 15: Systematic uncertainties on each of the fits. The fit model variations are described in more detail in the text. The totals are calculated by adding the relevant values in quadrature.

	Relative change in signal yield (%)			
	$B^0 \to D^-\pi^+$	$B_s^0 \to D_s^-\pi^+$	$B^0 \to D_s^- K^+$	$B_s^0 \to D_s^\mp K^\pm$
Different combinatorial shape	0.3	0.03	0.04	0.02
Background estimates +10%	0.2	0.3	0.2	0.3
Background estimates −10%	0.03	0.4	0.5	0.1
Different background templates	0.6	0.1	1.5	0.1
Reduced fit range	0.3	–	–	–
Fix tail parameters	2.1	2.1	2.1	2.1
Omit $\Lambda_b^0 \to D_s^{(*)-} p$	–	–	2.2	1.1
Particle identification	1.3	1.4	1.4	1.4
BDT	2.8	2.8	2.8	2.8
Total	4.1	4.3	5.1	4.4

Table 16: Fit results for signal shape parameters, from the fit to simulated signal events and the fit to data, where the tail parameters are floated.

Parameter	Value from simulation	Value from data fit
$\overline{\mathrm{x}}$	5284.0 ± 0.1	$5\,284.18 \pm 0.04$
σ	16.9 ± 0.1	17.54 ± 0.05
α_1	1.5082 ± 0.0431	1.0530 ± 0.0215
α_2	-1.9305 ± 0.0636	-1.5131 ± 0.0305
n_1	1.69 ± 0.09	6.43 ± 0.59
n_2	3.04 ± 0.26	118.25 ± 14.53

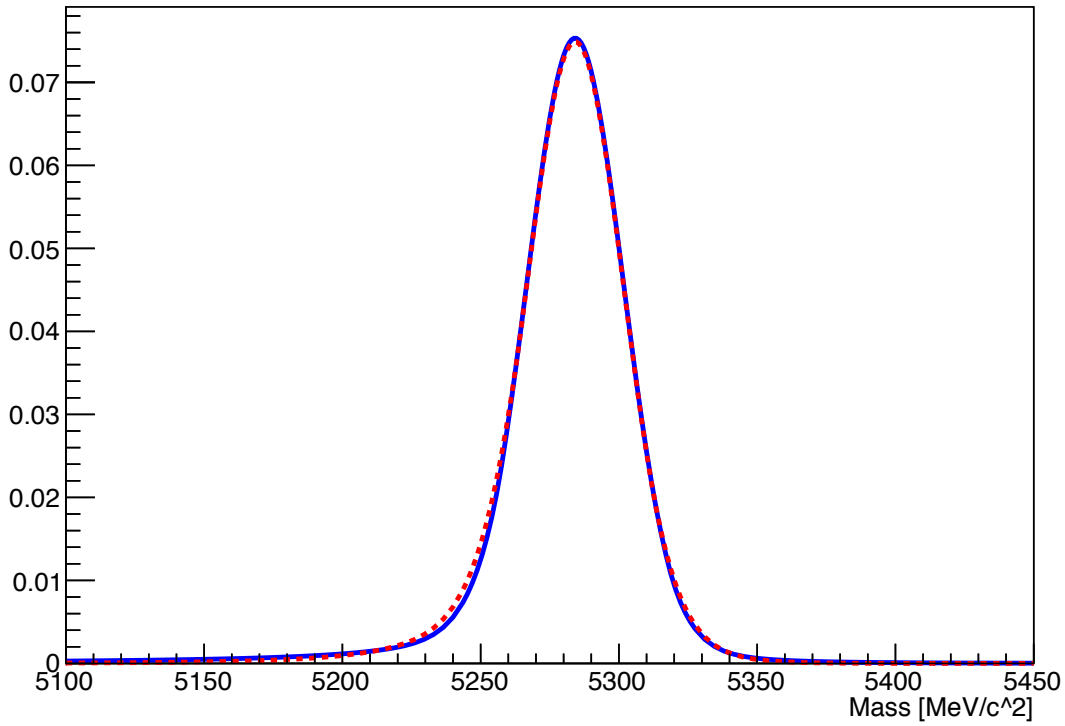

Figure 11: Comparison of double Crystal Ball functions with tail parameters obtained from simulated $B^0 \to D^- \pi^+$ events (blue curve) and data (red dashed curve), in arbitrary units. The mean and σ are set to the values obtained in the data fit.

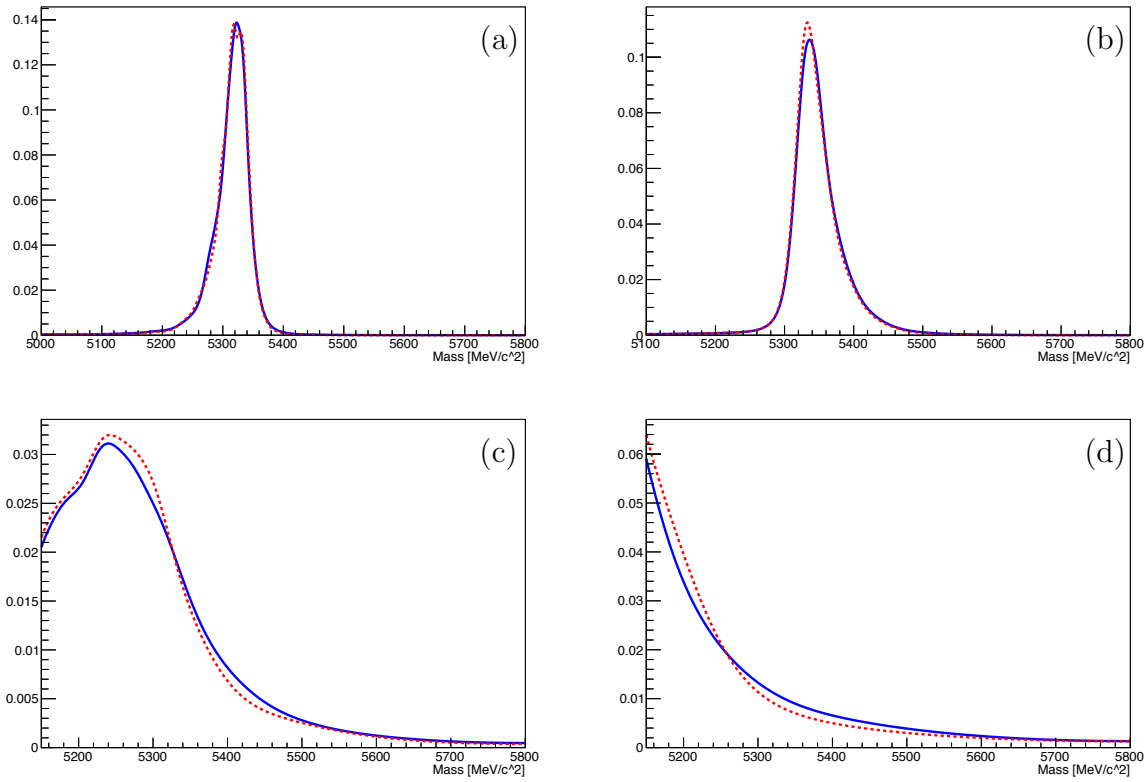

Figure 12: Comparisons of partially reconstructed and misidentified data simulated under 2011 (red dashed curve) and 2012 (blue curve) conditions, in arbitrary units, for (a) $B_s^0 \to D_s^- \pi^+$ under $B^0 \to D^- \pi^+$, (b) $B^0 \to D^- \pi^+$ under $B_s^0 \to D_s^- \pi^+$, (c) $B_s^0 \to D_s^{*-} \pi^+$ under $B_s^0 \to D_s^{\mp} K^{\pm}$, and (d) $B_s^0 \to D_s^{*-} \rho^+$ under $B_s^0 \to D_s^{\mp} K^{\pm}$. Note that each shape used in the fit is normalised with a parameter corresponding to the yield of that particular background component.

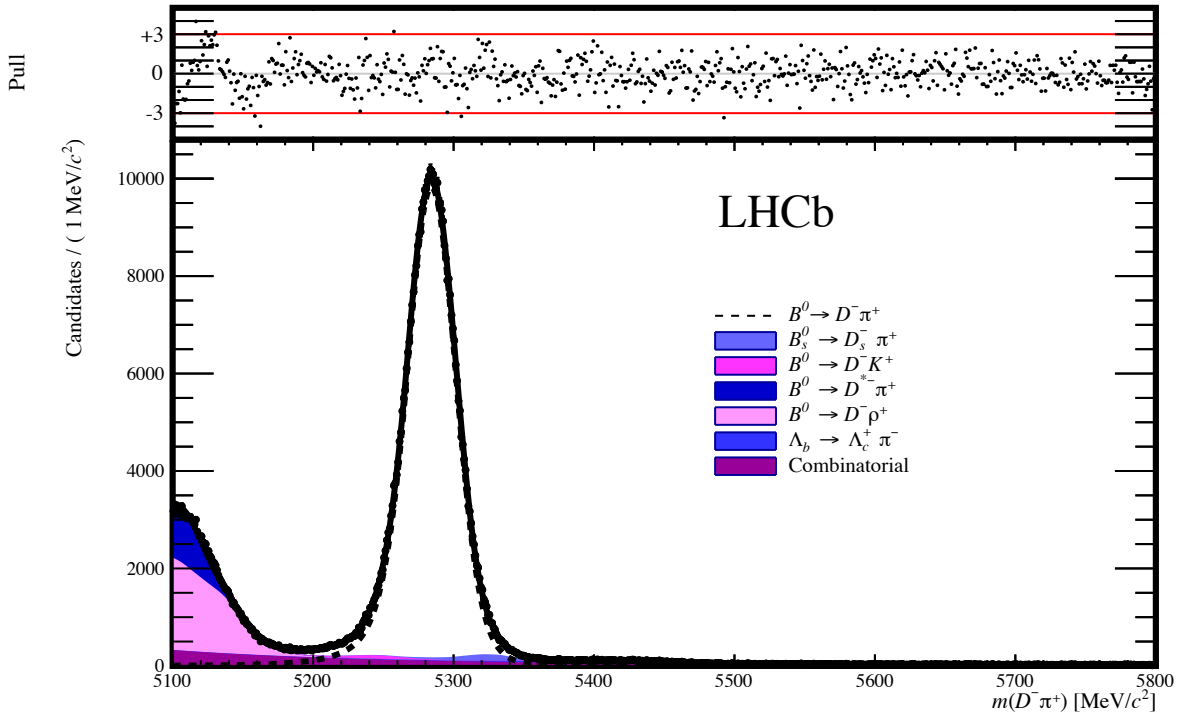

Figure 13: Mass fit to the process $B_s^0 \to D_s^-\pi^+$, in the reduced mass range $[5100, 5800]\,\mathrm{MeV}/c^2$.

6.4 Total systematic uncertainties per ratio

When calculating ratios of branching fractions, some systematic uncertainties appear in both the numerator and the denominator. To avoid double counting these errors, only a selection of uncertainties is used for the final systematic error on each ratio of branching fractions. The resulting systematic uncertainties are listed in Table 17.

$B_s^0 \to D_s^- \pi^+$ relative to $B^0 \to D^- \pi^+$

Of the systematic uncertainties on the PID, only those related to differently identified particles in the final state need to be taken into account. In this case, that amounts to the $D_{(s)}^{\pm}$ meson final state pions and kaons. The resulting uncertainty is 1.1%.

Of the fit variations, the different combinatorial shape is fully taken into account, as the combinatorial shapes may be different for the two processes. For the constraint mean shifts and different background templates, only the largest error between the two analyses is used, as these represent fluctuations in the simulated data which are assumed to be similar. The constrained fit range and fixed signal tail parameters are unique to the $B^0 \to D^- \pi^+$ analysis, and as such are also taken fully into account. In particular, any uncertainty on the signal shape tail parameters of the $B_s^0 \to D_s^- \pi^+$ peak is assumed to be covered this way. Adding all these numbers in quadrature, an uncertainty of 2.8% is obtained.

The uncertainty on the offline selection of 2.8% is applied once, as the two selection procedures are very similar, and the BDT cut is identical.

$B_s^0 \to D_s^{\mp} K^{\pm}$ relative to $B_s^0 \to D_s^- \pi^+$

To determine the systematic uncertainty on the ratio of $B_s^0 \to D_s^{\mp} K^{\pm}$ and $B_s^0 \to D_s^- \pi^+$, similar arguments are used as above. This leads to a systematic uncertainty from PID of 1.0%.

The different combinatorial shape uncertainties are again taken fully into account. The omission of $\Lambda_b^0 \to D_s^{-(*)} p$ is only present in the $B_s^0 \to D_s^{\mp} K^{\pm}$ analysis, and is used. The other uncertainties are all expected to be similar in the two analyses, and again for each only the largest value is used. The total uncertainty obtained this way is 1.2%.

The uncertainty on the offline selection is again set to 2.8%.

$B^0 \to D_s^- K^+$ relative to $B^0 \to D^- \pi^+$

In this ratio, the PID uncertainty is a combination of the previous two ratios. The value is 1.5%.

The fit variations are processed similarly to above: the combinatorial is taken fully into account, and of the constraint means and background templates only the largest of the two values. The constrained fit range, fixed tail parameters and omitted $\Lambda_b^0 \to D_s^{-(*)} p$ are all unique to one of the analyses, and are all taken into account. The resulting total systematic uncertainty is 3.5%.

The uncertainty on the offline selection is, once again, 2.8%.

Table 17: The systematic uncertainties on the ratios of branching fractions, obtained as described in the text. The totals are obtained by adding the other values in quadrature.

Systematic uncertainty	Value (%)		
	$\frac{B_s^0 \to D_s^- \pi^+}{B^0 \to D^- \pi^+}$	$\frac{B_s^0 \to D_s^\mp K^\pm}{B_s^0 \to D_s^- \pi^+}$	$\frac{B^0 \to D_s^- K^+}{B^0 \to D^- \pi^+}$
Particle identification	1.1	1.0	1.5
Fit model	2.8	1.2	3.5
BDT	2.8	2.8	2.8
Total	4.1	3.1	4.7

6.5 Consistency checks

In order to check the consistency of each fit as well as between the fits, a number of cross-checks is performed.

6.5.1 Misidientified background yields

For some background components resulting from a misidentified final state particle, the yield of that background can be compared to the expected yield, in parti. The expected yield is calculated by taking the number of signal decay events, multiplied with the misidentification probability, and corrected for the relative efficiencies (see Tables 4, 5, and 6), branching fractions, and hadronisation probabilities of the two processes. The branching fractions used for this fit are the current world average values [12], except for the branching fraction of the decay $B_s^0 \to D_s^- \pi^+$, which is obtained from Ref. [33]. The hadronisation fraction $f_s/f_d = 0.259 \pm 0.015$ is also obtained from Ref. [33]. The yields of the misidentified background components, as listed in Table 18, are constrained in the nominal fit, and hence no discrepancy is expected by construction. Indeed, no tension is observed.

Table 18: Consistency checks comparing the expected yield of misidentified background components to the fitted yield. For convencience, the fitted misidentified yields include the errors from both the fit and the expected yield. The deviation is the difference between the expected and the fitted yield, in units of the corresponding error. Other processes are possible, but they are not observed due to the relative branching fractions.

	Expected yield	Fitted yield	Deviation
$B^0 \to D^-\pi^+$ fit			
$B^0 \to D^- K^+$	$4\,651 \pm 395$	$4\,257 \pm 468$	-0.73
$B_s^0 \to D_s^- \pi^+$	$7\,981 \pm 654$	$7\,463 \pm 750$	-0.60
$B_s^0 \to D_s^- \pi^+$ fit			
$B^0 \to D^- \pi^+$	$3\,339 \pm 271$	$3\,872 \pm 330$	1.39
$B_s^0 \to D_s^\mp K^\pm$ fit			
$B^0 \to D^- K^+$	254 ± 34	256 ± 39	0.05
$B_s^0 \to D_s^- \pi^+$	$2\,086 \pm 238$	$1\,979 \pm 252$	-0.35

6.5.2 Magnet polarity

The fit of the background components is done separately for two data samples: data corresponding to the LHCb magnet creating an upwards pointing magnetic field, and data corresponding to a downwards pointing magnetic field. The signal yields of the two samples, and the deviation from being equal to one another (after correcting for the difference in luminosity), are show in Table 19.

Table 19: Consistency checks comparing the expected relative yield of each magnet polarity. The luminosity values are the integrated luminosity in that particular sample, as a fraction of the total integrated luminosity used in the analysis. The yield values are results from the fit, and the errors are the statistical errors and systematic errors on the yield (arising from uncertainties in the fit model), added in quadrature. The deviation in σ is the number of standard deviations the expected yield differs from the observed yield. The deviation in % is the deviation as a fraction of the total yield (for both magnet polarities).

	Magnet up yield	Magnet down yield	Deviation (σ)	Deviation (%)
Luminosity	47.7%	52.3%		
$B^0 \to D^-\pi^+$	$217\,796 \pm 634$	$242\,643 \pm 679$	2.55	0.43
$B_s^0 \to D_s^-\pi^+$	$35\,827 \pm 233$	$39\,814 \pm 248$	0.97	0.37
$B^0 \to D_s^-K^+$	$1\,190 \pm 65$	$1\,182 \pm 68$	0.73	2.44
$B_s^0 \to D_s^\mp K^\pm$	$2\,384 \pm 69$	$2\,646 \pm 72$	0.20	0.33

7 Branching fractions

This section discusses the signal yields obtained from the fits and the resulting extraction of branching fractions. Section 7.1 contains the $B_s^0 \to D_s^- \pi^+$ branching fraction, section 7.2 the branching fraction of $B_s^0 \to D_s^\mp K^\pm$, and section 7.3 discusses the relative and absolute branching fraction of $B^0 \to D_s^- K^+$.

7.1 Branching fraction of $B_s^0 \to D_s^- \pi^+$

The branching fraction of the decay $B_s^0 \to D_s^- \pi^+$ is extracted using the following formula:

$$\mathcal{B}(B_s^0 \to D_s^- \pi^+) = \mathcal{B}(B^0 \to D^- \pi^+) \frac{\varepsilon_{B^0 \to D^- \pi^+}}{\varepsilon_{B_s^0 \to D_s^- \pi^+}} \frac{f_d}{f_s} \frac{N_{B_s^0 \to D_s^- \pi^+}}{N_{B^0 \to D^- \pi^+}} \frac{\mathcal{B}(D^- \to K^+ \pi^- \pi^-)}{\mathcal{B}(D_s^- \to K^- K^+ \pi^-)}. \quad (2)$$

In this formula, f_d/f_s represents the ratio of hadronisation fractions of B^0 mesons and B_s^0 mesons; N_X is the measured yield of decay X, and ε_X the efficiency of process X. For f_s/f_d, the hadronic determination cannot be used, as that *relies* on the relative branching ratios that we determine. Instead, the independent semi-leptonic determination of f_s/f_d is used $f_s/f_d = 0.268^{+0.0234}_{-0.0215}$ [16]. Updated values for the ratio of D and D_s branching fractions increases the value of f_s/f_d by 1.3%, whereas updated values of the ratio of lifetimes, $(\tau_{B0} + \tau_{B-})/2\tau_{Bs} = 1.056 \pm 0.012$, reduces it by 1.3%. The uncertainty due to the new inputs vary accordingly, resulting in $f_s/f_d = 0.268^{+0.020}_{-0.018}$. Finally, we use the semileptonic value, without any uncertainty on the D and D_s branching fractions, as they cancel in the determination of $\mathcal{B}(B_s^0 \to D_s^- \pi^+)$,

$$\frac{f_s}{f_d} = 0.268^{+0.018}_{-0.016}.$$

For the branching fractions appearing in the right-hand side of Eq.(2), the following values are used [12, 33]:

$$\begin{aligned} \mathcal{B}(B^0 \to D^- \pi^+) &= (2.68 \pm 0.13) \times 10^{-3}, \\ \mathcal{B}(D^- \to K^+ \pi^- \pi^-) &= (9.13 \pm 0.19) \times 10^{-2}, \\ \mathcal{B}(D_s^- \to K^- K^+ \pi^-) &= (5.42 \pm 0.14) \times 10^{-2}. \end{aligned} \quad (3)$$

The relative yields and efficiencies can be found in Table 20. The systematic error on the branching fraction is a combination of the errors on the numbers in Eqs. (3) and the systematic uncertainties described in section 6. The yields used to extract the branching fraction are the combined yields for magnet up and down data.

Combining all the relevant numbers gives the following result:

$$\mathcal{B}(B_s^0 \to D_s^- \pi^+) = (2.98 \pm 0.02 \pm 0.16^{+0.18}_{-0.20}) \times 10^{-3},$$

where the errors are statistical, systematic, and due to f_s/f_d, respectively.

Table 20: Numbers entering the calculation of the branching fraction of the process $B_s^0 \to D_s^- \pi^+$ relative to $B^0 \to D^- \pi^+$. The error on the ratio of yields is statistical, and that on the ratio of efficiencies is systematic, from the BDT and the PID.

$N_{B_s^0 \to D_s^- \pi^+} / N_{B^0 \to D^- \pi^+}$	0.1643 ± 0.0008
$\varepsilon_{B^0 \to D^- \pi^+} / \varepsilon_{B_s^0 \to D_s^- \pi^+}$	$1.078 \quad \pm 0.023$

7.2 Branching fraction of $B_s^0 \to D_s^\mp K^\pm$

The $B_s^0 \to D_s^\mp K^\pm$ branching fraction is calculated in much the same way as that of $B_s^0 \to D_s^- \pi^+$ in the previous section. The relevant numbers can be found in Table 21. The

Table 21: Numbers entering the calculation of the branching fraction of the process $B_s^0 \to D_s^\mp K^\pm$ relative to $B_s^0 \to D_s^- \pi^+$. The errors are statistical and systematic, respectively.

$N_{B_s^0 \to D_s^\mp K^\pm} / N_{B_s^0 \to D_s^- \pi^+}$	0.0665 ± 0.0080
$\varepsilon_{B_s^0 \to D_s^- \pi^+} / \varepsilon_{B_s^0 \to D_s^\mp K^\pm}$	$1.098 \quad \pm 0.023$

resulting ratio of branching fractions is

$$\frac{\mathcal{B}(B_s^0 \to D_s^\mp K^\pm)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = 0.0730 \pm 0.0015 \pm 0.0021,$$

where the first error is statistical and the second systematic. Combining this with the result for $\mathcal{B}\ (B_s^0 \to D_s^- \pi^+)$, the following value is obtained:

$$\mathcal{B}(B_s^0 \to D_s^\mp K^\pm) = (2.22 \pm 0.05 \pm 0.13^{+0.13}_{-0.15}) \times 10^{-4},$$

where the errors are statistical, systematic, and resulting from f_s/f_d, respectively. The error from f_s/f_d enters through the branching fraction of $B_s^0 \to D_s^- \pi^+$.

7.3 Branching fraction of $B^0 \to D_s^- K^+$

Finally, the $B^0 \to D_s^- K^+$ branching fraction, measured relative to that of $B^0 \to D^- \pi^+$, is determined using the numbers in Table 22. The resulting relative and absolute branching fractions are

$$\frac{\mathcal{B}(B^0 \to D_s^- K^+)}{\mathcal{B}(B^0 \to D^- \pi^+)} = 0.0117 \pm 0.0005 \pm 0.0007,$$

where the errors are statistical and systematic, respectively, and

$$\mathcal{B}(B^0 \to D_s^- K^+) = (3.14 \pm 0.12 \pm 0.14 \pm 0.18) \times 10^{-5}, \tag{4}$$

Table 22: Numbers entering the calculation of the branching fraction of the process $B^0 \to D_s^- K^+$ relative to $B^0 \to D^- \pi^+$. The errors are statistical and systematic, respectively.

$N_{B^0 \to D_s^- K^+} / N_{B^0 \to D^- \pi^+}$	$(5.15 \pm 0.20) \times 10^{-3}$
$\varepsilon_{B^0 \to D^- \pi^+} / \varepsilon_{B^0 \to D_s^- K^+}$	1.351 ± 0.030

where the first error is statistical, the second systematic, and the last one resulting from the branching fractions of $B^0 \to D^- \pi^+$, $D^- \to K^+ \pi^- \pi^-$, and $D_s^- \to K^+ K^- \pi^-$, respectively.

8 Conclusion and discussion

The reported branching fraction measurements of $B_s^0 \to D_s^\mp K^\pm$ and $B^0 \to D_s^- K^+$ are the current most precise values. The branching fraction of $B_s^0 \to D_s^- \pi^+$ is compatible with the current world average value, which is dominated by the systematic uncertainty of the previous LHCb analysis.

8.1 Ratio of branching fractions $\mathcal{B}(B_s^0 \to D_s^\mp K^\pm)/\mathcal{B}(B_s^0 \to D_s^- \pi^+)$

The ratio of branching fractions of $B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^- \pi^+$ has a theoretical lower limit of [14]

$$\frac{\mathcal{B}(B_s^0 \to D_s^\mp K^\pm)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} \geq \frac{\varepsilon}{\mathcal{C}} \left[1 - y_s^2 \cos^2 \delta_s \cos^2(\phi_s + \gamma)\right] = 0.080 \pm 0.007. \tag{5}$$

In this formula, ε depends directly on CKM elements, $y_s = \Delta\Gamma_s/2\Gamma_s$ is the relative decay width difference of the B_s^0 meson system, δ_s is the relative strong phase between the two ($b \to c$ and $b \to u$) tree topologies of the $B_s^0 \to D_s^\mp K^\pm$ decay, and $\phi_s + \gamma$ the relative weak phase. The factor $\mathcal{C}$ depends on the differences in kinematics and decay topologies of the decays $B_s^0 \to D_s^\mp K^\pm$ and $B_s^0 \to D_s^- \pi^+$:

$$\mathcal{C} = \frac{\Phi_{D_s\pi}}{\Phi_{D_sK}} \mathcal{N}_F \mathcal{N}_a \mathcal{N}_E.$$

Here, the Φ are phase-space factors, $\mathcal{N}_F$ contains the form factor and decay constant ratios, $\mathcal{N}_a$ describes the ratio of the colour-allowed tree amplitudes of the decays, and $\mathcal{N}_E$ contains corrections for the exchange topology, which only plays a role in $B_s^0 \to D_s^\mp K^\pm$ (see Table 1).

In the previous analysis of this ratio by the LHCb collaboration [13], the reported value was $0.0646 \pm 0.0043 \pm 0.0025$, lower than both the theoretical lower bound and the measurement reported in this analysis. The value of the ratio of branching fractions presented here is compatible both with the previous measurement, and with the theoretical bound.

The theoretical lower bound could be reduced if one of the parameters has a different value than was assumed in the calculation. Most parameters in Eq. (5) are either relatively well known (such as the CKM elements) or do not affect the ratio much (such as the cosine term, which is suppressed by $y_s^2 = 0.01$). Of all the parameters playing a role, $\mathcal{N}_E$ is of particular interest, as it can be probed via decay processes that only occur through exchange topologies. By definition it equals

$$\mathcal{N}_E = \left| \frac{T}{T+E} \right|^2,$$

where T and E are directly related to the amplitudes of the corresponding tree and exchange diagrams, respectively.

Since branching fractions are related to the square of the amplitude,

$$\mathcal{B}(X \to Y) = \tau_X |A(X \to Y)|^2 \Phi_{X,Y}, \tag{6}$$

where τ_X is the lifetime of particle X and $\Phi_{X,Y}$ is a phase-space factor, a measurement of the branching fraction does not contain information on the (complex) value of the amplitude itself. Therefore, in order to get a handle on T and E, other decays are used, which only occur through either tree or exchange diagrams [34]. The values of $|T|$ and $|E|$ can then be extracted using SU(3) symmetries between the three lightest quarks. In particular, the following decay amplitudes are used:

$$\begin{aligned}
\mathcal{B}(B^0 \to D^- K^+) &= (2.0 \pm 0.2) \times 10^{-4} \sim |A(B^0 \to D^- K^+)|^2 = |T|^2, \\
\mathcal{B}(B^0 \to D_s^- K^+) &= (3.1 \pm 0.3) \times 10^{-5} \sim |A(B^0 \to D_s^- K^+)|^2 = |E|^2, \\
\mathcal{B}(B^0 \to D^- \pi^+) &= (2.7 \pm 0.1) \times 10^{-3} \sim |A(B^0 \to D^- \pi^+)|^2 = |T+E|^2,
\end{aligned} \tag{7}$$

of which the second is given in Eq. (4) and the other two are taken from Ref. [12].

To obtain information on the complex phases of these values, the values of $|T|/|T+E|$ and $|E|/|T+E|$ are compared. In the complex plane, these form, together with $|T+E|/|T+E| = 1$, a triangle, whose apex is determined by the imaginary parts of the amplitudes. Recalculating these values using the numbers in Eq. (7) yields

$$\begin{aligned}
\frac{|T|}{|T+E|} &= 0.99 \pm 0.11, \\
\frac{|E|}{|T+E|} &= 0.074 \pm 0.005.
\end{aligned} \tag{8}$$

Because of the stability of the central values of the numbers in Eq.(8) compared to those used to calculate the same quantity in Ref. [34], the resulting numbers are also stable. In fact, since the central value of $|T|/|T+E|$ remains the same, the lower bound on the ratio of branching fractions is also unchanged. That value was calculated using precisely measured numbers from decays into D^* mesons, $\mathcal{N}_E^* = 0.966 \pm 0.056$. An extra systematic uncertainty of 5% was assigned for using the result for D mesons. Even when taking into account the reduction in uncertainty obtained by the value of $|T|/|T+E|$ in Eq. (8), the uncertainty is not competitive with respect to $\mathcal{N}_E^*$: $\mathcal{N}_E = 0.98 \pm 0.22$. Therefore, the lower bound is unchanged. The value for the ratio of branching fractions obtained in this analysis is compatible with the lower bound within one standard deviation.

8.2 Further research

The logical next step is to repeat the analysis on $B_s^0 \to D_s^{\mp} K^{\pm}$ in a time-dependent fashion. Doing so should reveal the B_s^0-$\overline{B}_s^0$ oscillations which lead to a measurement of the CKM angle γ. This is a complicated fit which has to take into account effects on the measured decay time distribution, amongst which detector acceptance and cross-feed of partially reconstructed, misidentified and combinatorial backgrounds. One way to reduce the combinatorial background is to further optimise and tighten the BDT cut, a process that was not applied in this analysis because the remaining combinatorial in fact allows its shape to be determined. Since the process $B_s^0 \to D_s^{\mp} K^{\pm}$ only occurs through tree diagrams, very few New Physics models predict an enhancement of γ. On the other hand, this allows such a measurement that serves as a benchmark for other analyses with greater sensitivity to New Physics, like the charmless B_s^0 decays to two hadrons which involve (penguin) loop diagrams.

Another interesting analysis that could be feasible with the currently available data is that of a separate analysis of $D_s^- K^+$ and $D_s^+ K^-$ final states. The relative contributions of CKM elements in these two processes enable a direct detection of *CP* violation in the yield difference between the two final states. The available statistics may limit the success of such an analysis.

Appendices

A Fit templates

This section contains the templates for the backgrounds used in the mass fits: $B^0 \to D^-\pi^+$ in section A.1, $B_s^0 \to D_s^-\pi^+$ in A.2, and $B_s^0 \to D_s^\mp K^\pm$ in A.3. The total yield of each background depends on the number of events in the sample of simulated data and is not actually used in the fit; each template is instead scaled to a predicted yield, as described in section 5.1.

A.1 $B^0 \to D^-\pi^+$ fit templates

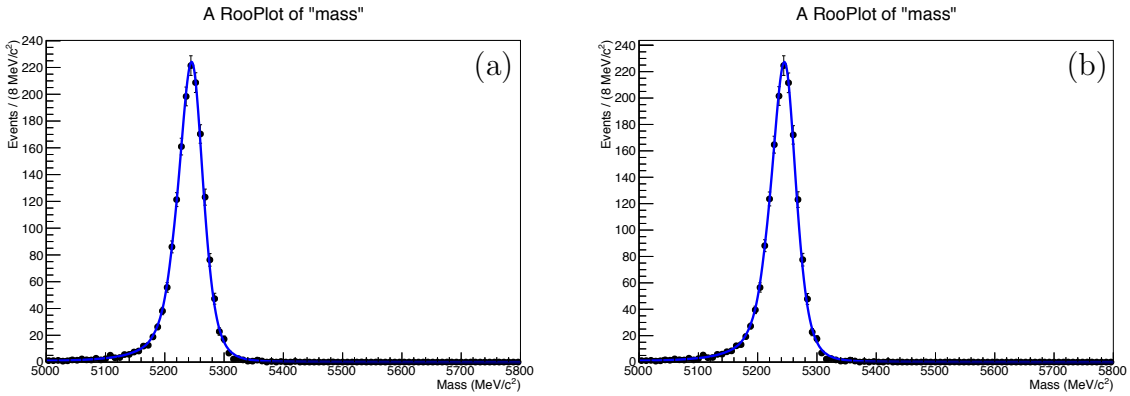

Figure 14: $B^0 \to D^-K^+$ template for (a) magnet down and (b) magnet up.

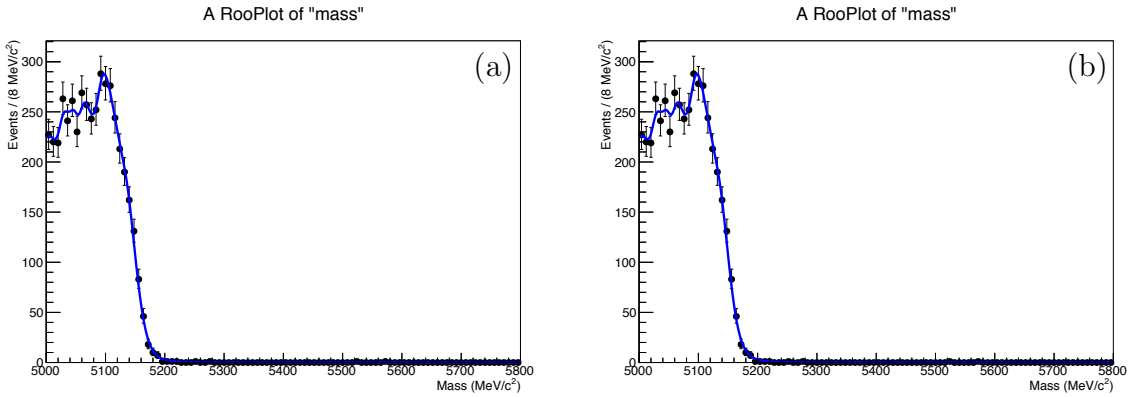

Figure 15: $B^0 \to D^-\rho^+$ template for (a) magnet down and (b) magnet up.

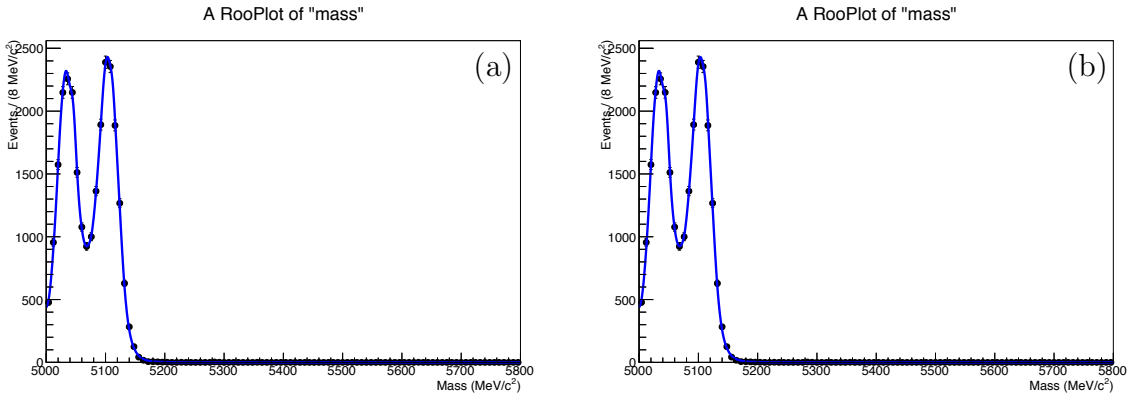

Figure 16: $B^0 \to D^{*-}\pi^+$ template for (a) magnet down and (b) magnet up.

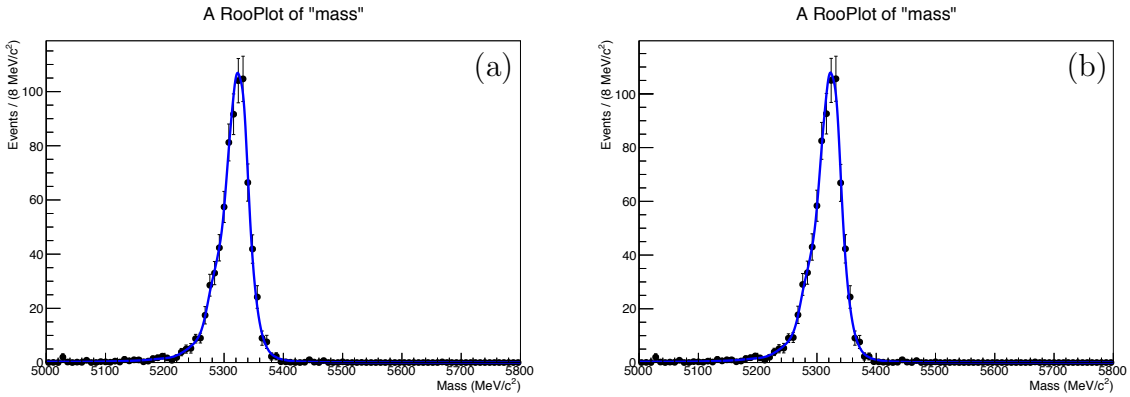

Figure 17: $B_s^0 \to D_s^-\pi^+$ template for (a) magnet down and (b) magnet up.

A.2 $B_s^0 \to D_s^- \pi^+$ fit templates

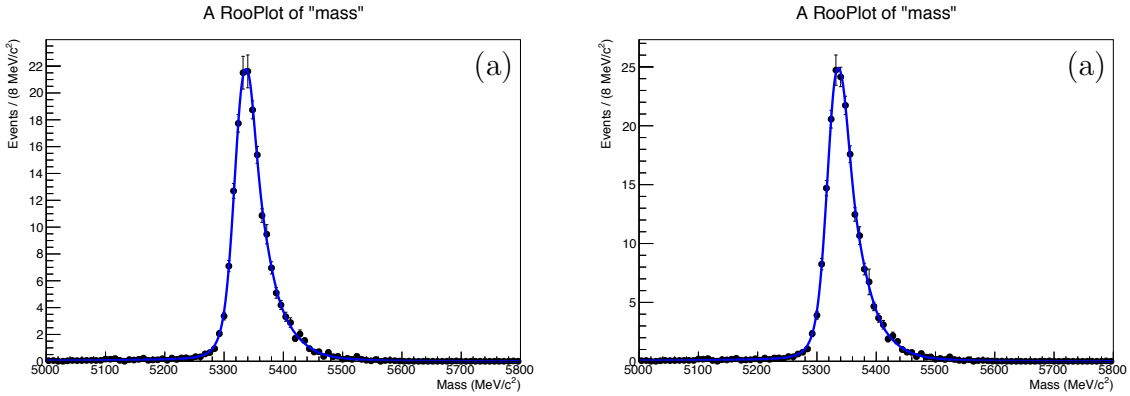

Figure 18: $B^0 \to D^- \pi^+$ template for (a) magnet down and (b) magnet up.

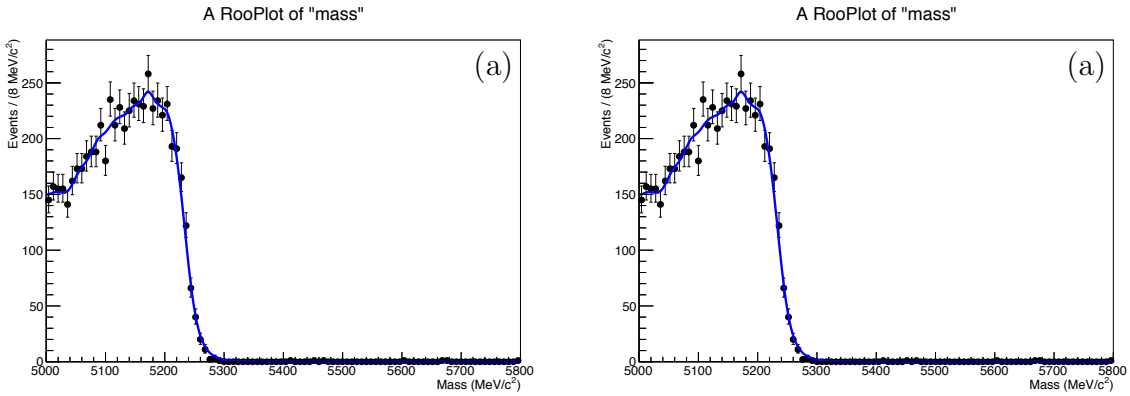

Figure 19: $B_s^0 \to D_s^- \rho^+$ template for (a) magnet down and (b) magnet up.

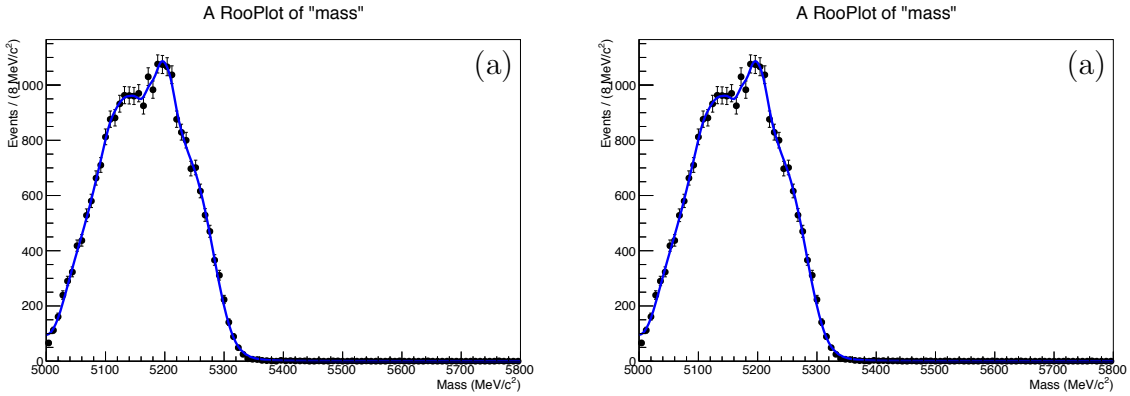

Figure 20: $B_s^0 \to D_s^{*-} \pi^+$ template for (a) magnet down and (b) magnet up.

A.3 $B_s^0 \to D_s^{\mp} K^{\pm}$ fit templates

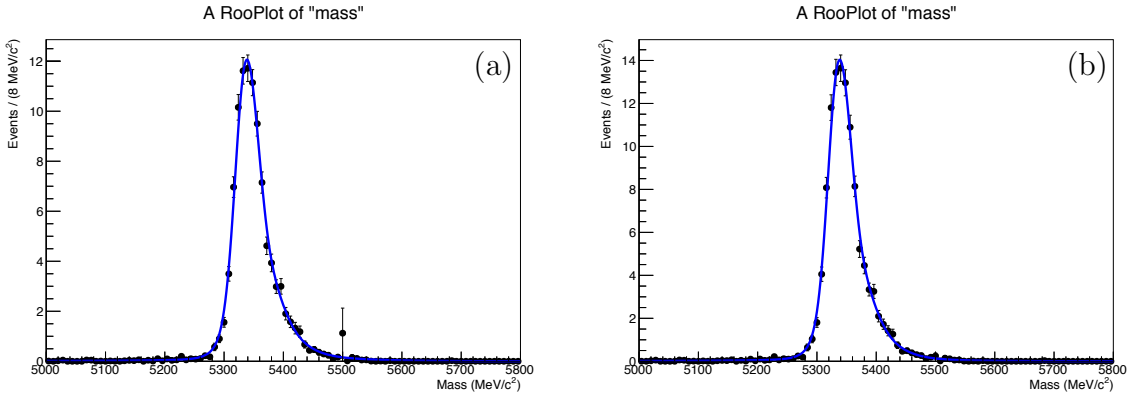

Figure 21: $B^0 \to D^- K^+$ template for (a) magnet down and (b) magnet up.

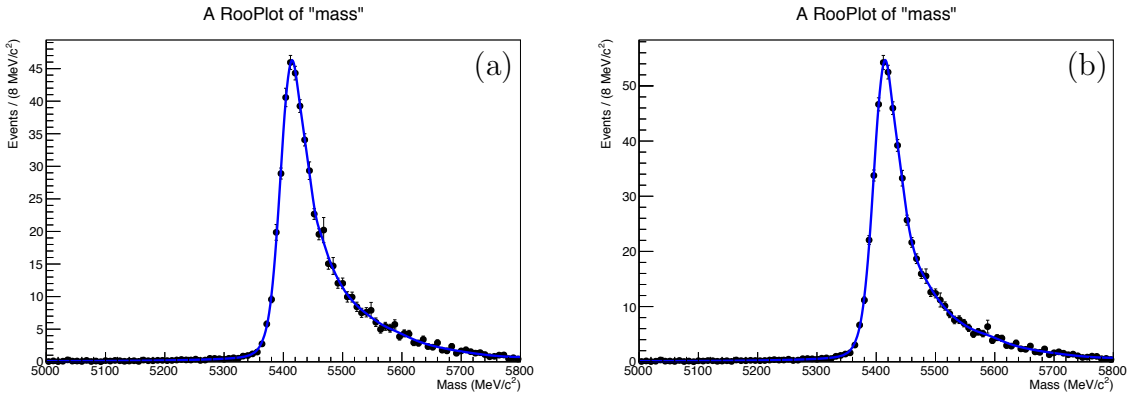

Figure 22: $B_s^0 \to D_s^- \pi^+$ template for (a) magnet down and (b) magnet up.

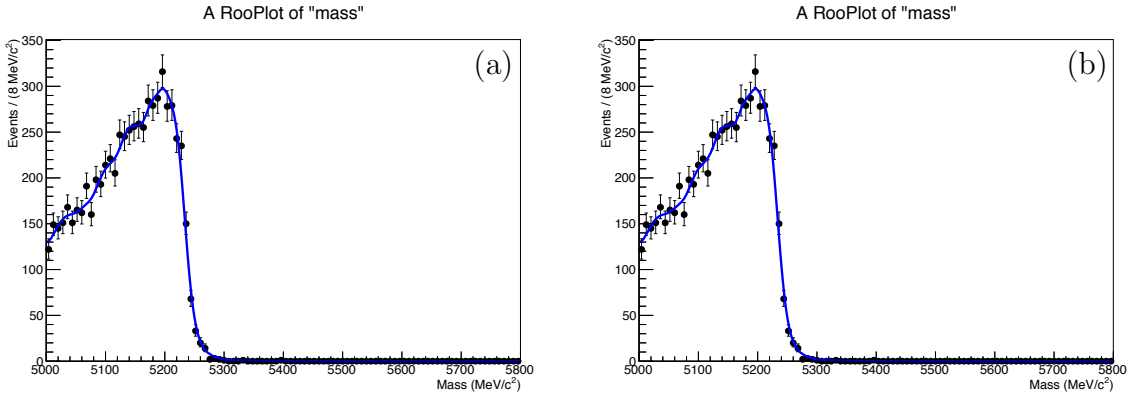

Figure 23: $B_s^0 \to D_s^{\mp} K^{*\pm}$ template for (a) magnet down and (b) magnet up.

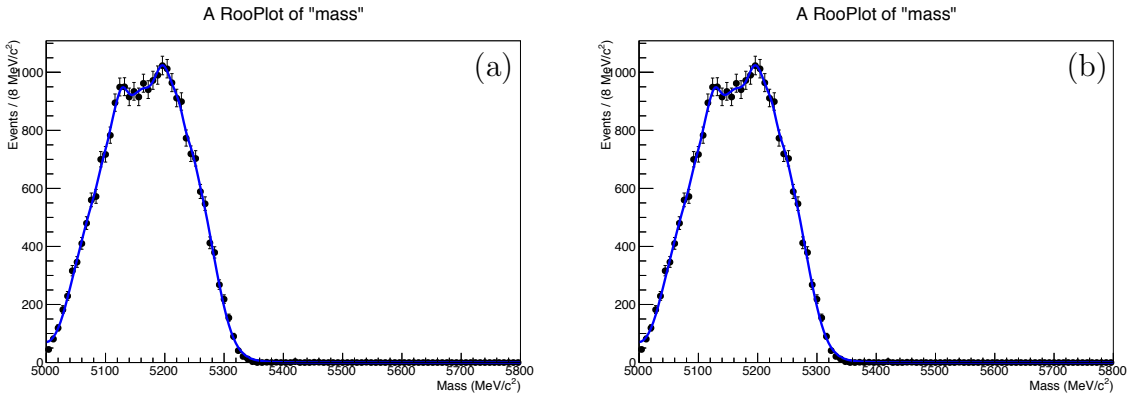

Figure 24: $B_s^0 \to D_s^{*\mp} K^{\pm}$ template for (a) magnet down and (b) magnet up.

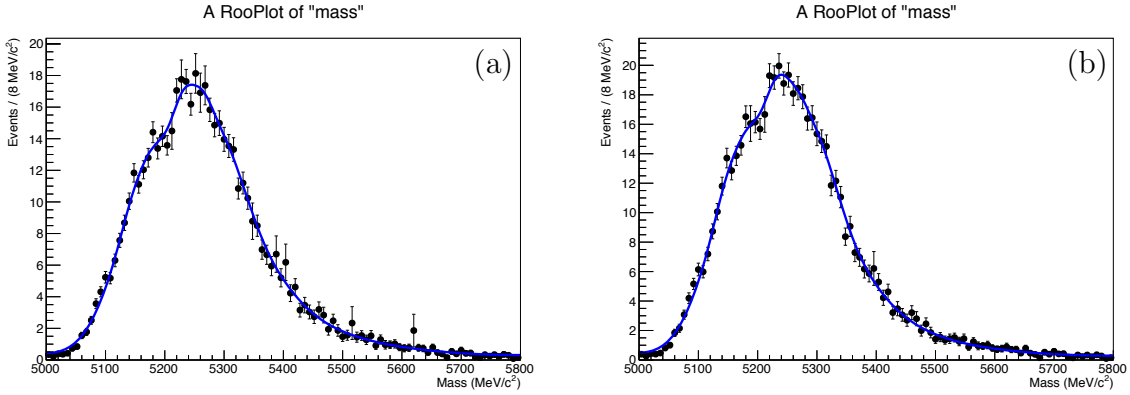

Figure 25: $B_s^0 \to D_s^{*-} \pi^+$ template for (a) magnet down and (b) magnet up.

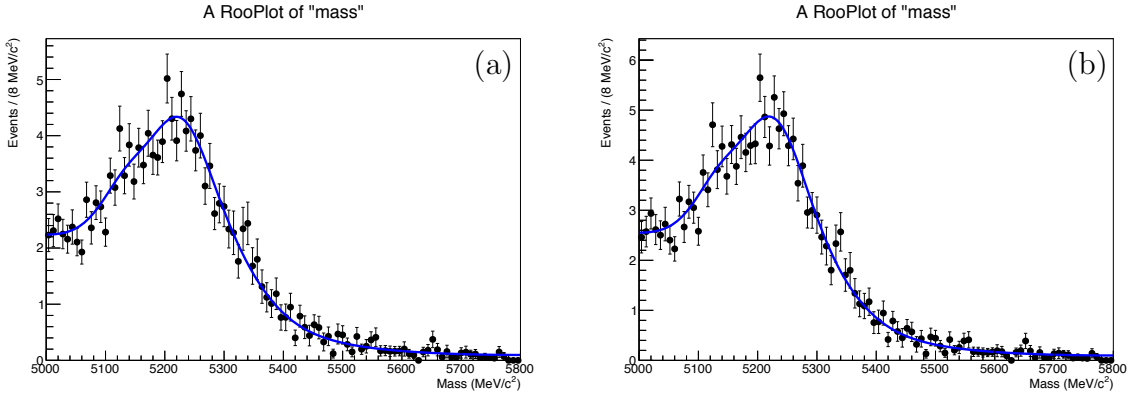

Figure 26: $B_s^0 \to D_s^- \rho^+$ template for (a) magnet down and (b) magnet up.

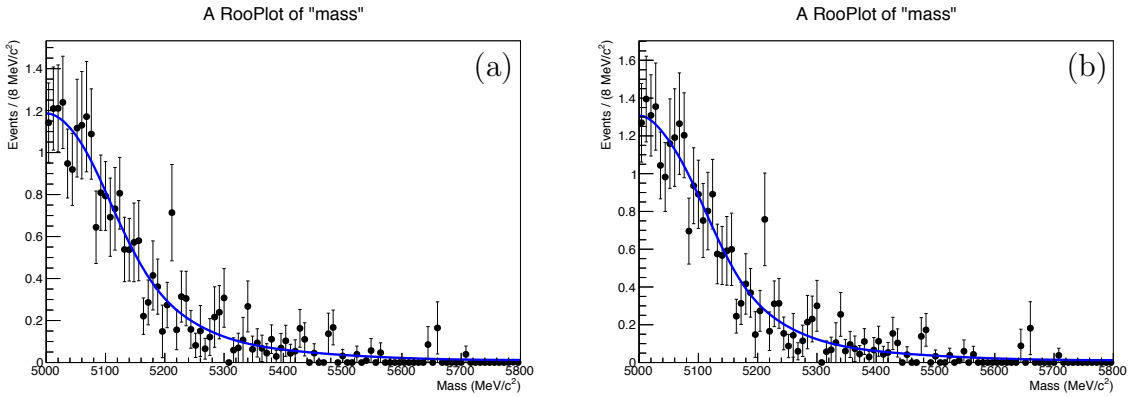

Figure 27: $B_s^0 \to D_s^{*-}\rho^+$ template for (a) magnet down and (b) magnet up.

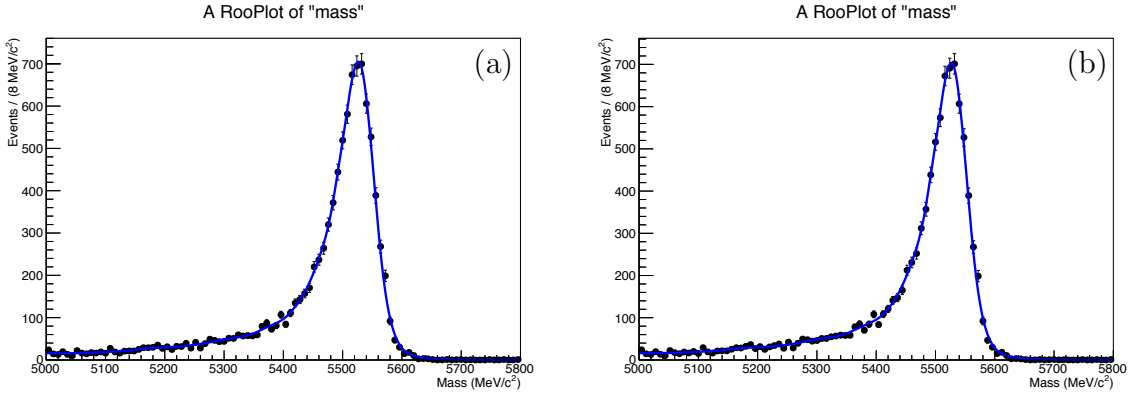

Figure 28: $\Lambda_b^0 \to D_s^- p$ template for (a) magnet down and (b) magnet up.

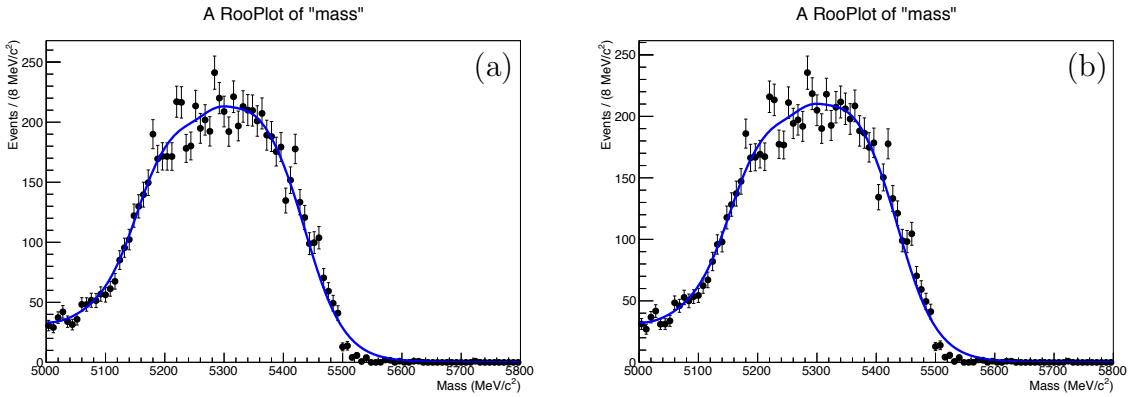

Figure 29: $\Lambda_b^0 \to D_s^{*-} p$ template for (a) magnet down and (b) magnet up.

B Signal shape fits

In order to determine the most accurate representation of the shape of the signal decay processes, several shapes have been tried by fitting to a simulated sample of each of those processes. In each of the following functions, N represents an overall, parameter-dependent normalisation such that the integral of the function over $\mathbb{R}$ is identically 1.

B.1 Cruijff function fit

Another probability distribution that has been tried is the so-called Cruijff function:

$$f(x;\overline{x},\sigma_L,\sigma_R,\alpha_L,\alpha_R) = N \begin{cases} \exp\left(\dfrac{-(x-\overline{x})^2}{2\sigma_L+\alpha_L(x-\overline{x})^2}\right) & \text{for } x \leq \mu \\ \exp\left(\dfrac{-(x-\overline{x})^2}{2\sigma_R+\alpha_R(x-\overline{x})^2}\right) & \text{for } x > \mu. \end{cases} \tag{9}$$

The fits to this function can be found in Fig. 30, and the corresponding results for the parameters in Table 23.

Table 23: Results of Cuijff function fits to simulated signal samples.

Parameter	Fitted value
$B^0 \to D^-\pi^+$	
α_{L}	0.16172 ± 0.00008
α_{R}	0.10173 ± 0.00003
$\overline{\mathrm{x}}$	5284.6 ± 78.8
σ_{L}	14.96367 ± 0.00578
σ_{R}	14.70005 ± 0.00932
$B_s^0 \to D_s^-\pi^+$	
α_{L}	0.15731 ± 0.00134
α_{R}	0.10532 ± 0.00202
$\overline{\mathrm{x}}$	5371.5 ± 0.3
σ_{L}	14.45136 ± 0.18365
σ_{R}	14.67946 ± 0.18644
$B_s^0 \to D_s^{\mp}K^{\pm}$	
α_{L}	0.14220 ± 0.00188
α_{R}	0.09154 ± 0.00335
$\overline{\mathrm{x}}$	5371.2 ± 0.3
σ_{L}	13.69576 ± 0.23122
σ_{R}	14.62148 ± 0.25504

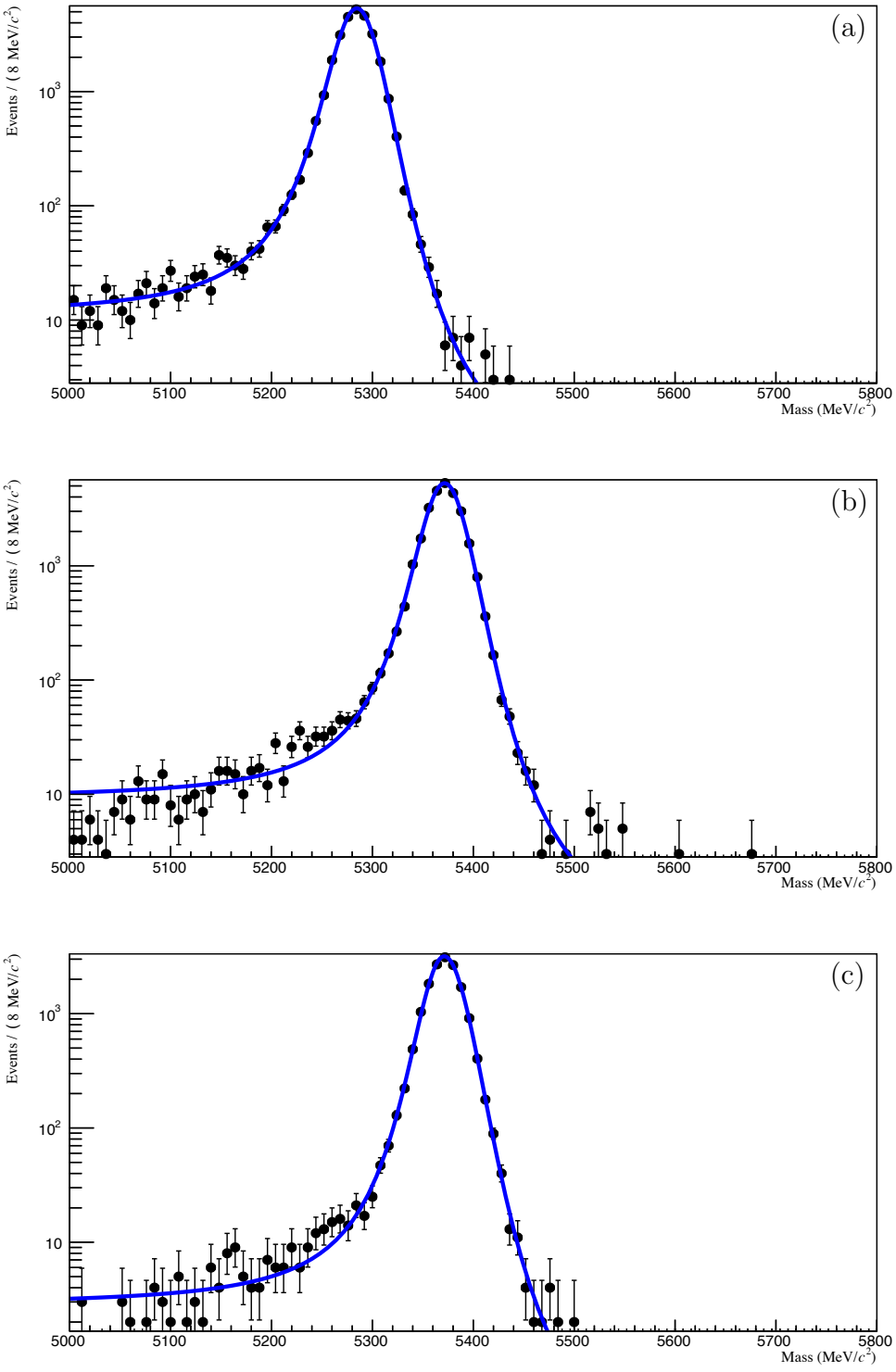

Figure 30: Cruijff function signal fits for (a) $B^0 \to D^- \pi^+$, (b) $B_s^0 \to D_s^- \pi^+$, and (c) $B_s^0 \to D_s^\mp K^\pm$.

B.2 Double Apollonios fit

The final function shape that has been tried is the double-sided Apollonios function [35]:

$$
\begin{aligned}
&f(x;\overline{x},b,\delta,a,n_L,n_R) = \\
&N\begin{cases}
\exp\left(-b\sqrt{1+a^2}\right)\left(\dfrac{(n_R\sqrt{1+a^2}-a)/(ba)-a}{(n_R\sqrt{1+a^2}-a)/(ba)-(x-\overline{x})/\delta}\right)^{n_R} & \text{for } \dfrac{x-\overline{x}}{\delta} > a \\
\exp\left(-b\sqrt{1+(x-\overline{x})^2/\delta^2}\right) & \text{for } \left|\dfrac{x-\overline{x}}{\delta}\right| \leq a \\
\exp\left(-b\sqrt{1+a^2}\right)\left(\dfrac{(n_L\sqrt{1+a^2}-a)/(ba)+a}{(n_L\sqrt{1+a^2}-a)/(ba)-(x-\overline{x})/\delta}\right)^{n_L} & \text{for } \dfrac{x-\overline{x}}{\delta} < -a
\end{cases}
\end{aligned}
\tag{10}
$$

These fits are shown in Fig. 31 and the fitted parameter values in Table 24.

Table 24: Results of double Apollonios function fits to simulated signal samples.

Parameter	Fitted value
$B^0 \to D^-\pi^+$	
a	2.47722 ± 0.00013
b	2.34961 ± 0.01529
δ	22.33186 ± 0.00080
$\overline{\text{x}}$	5283.5 ± 0.3
n_L	1.48933 ± 0.01279
n_R	7.40865 ± 0.28704
$B_s^0 \to D_s^-\pi^+$	
a	2.80994 ± 0.02180
b	1.98925 ± 0.01376
δ	19.53861 ± 0.05426
$\overline{\text{x}}$	5371.0 ± 0.2
n_L	1.60698 ± 0.04267
n_R	7.08174 ± 0.37706
$B_s^0 \to D_s^\mp K^\pm$	
a	2.90476 ± 0.01279
b	2.14334 ± 0.02289
δ	19.68621 ± 0.08198
$\overline{\text{x}}$	5371.2 ± 0.1
n_L	1.77328 ± 0.09643
n_R	8.86130 ± 1.33386

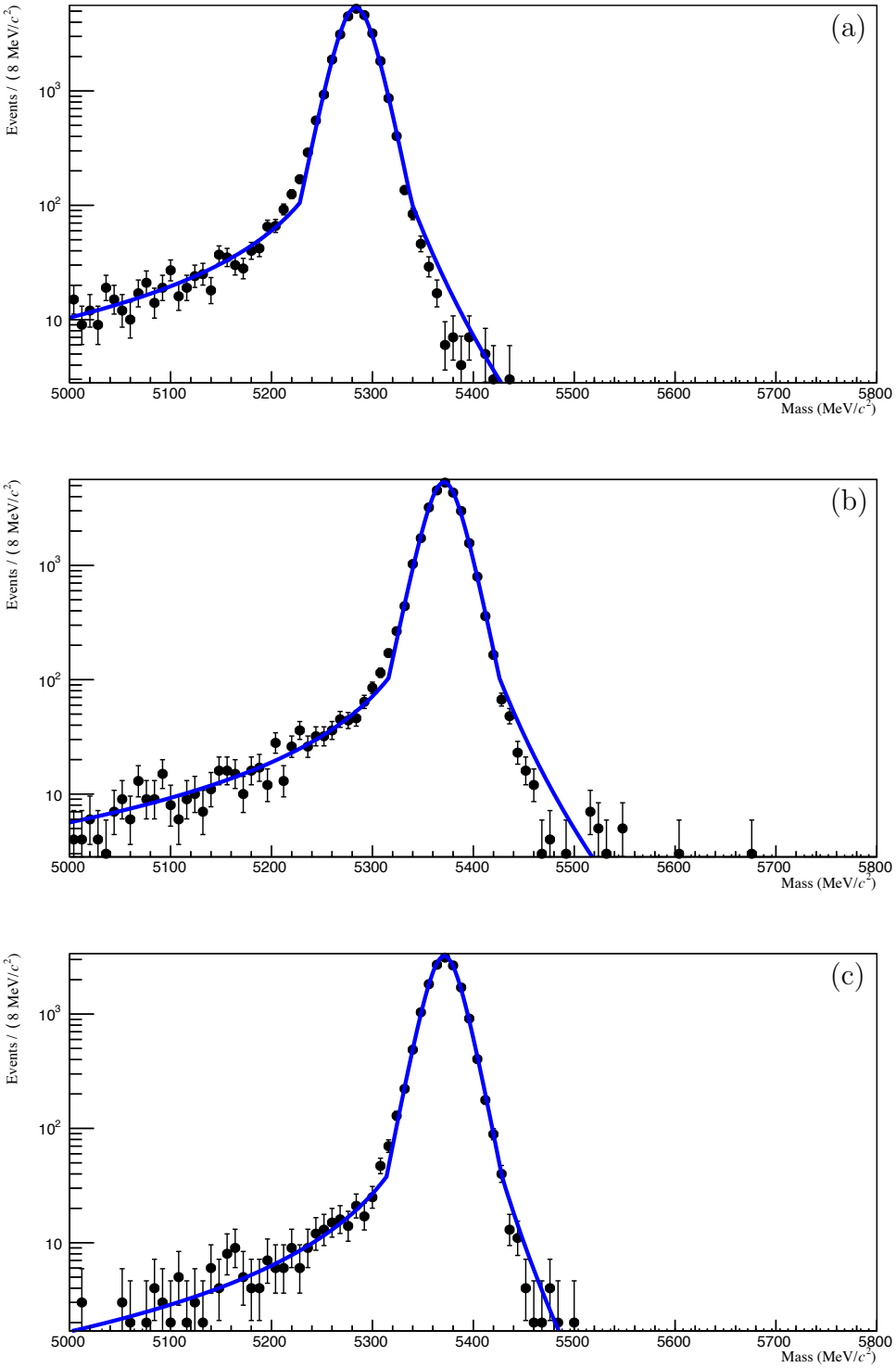

Figure 31: Double Apollonios function signal fits for (a) $B^0 \to D^- \pi^+$, (b) $B_s^0 \to D_s^- \pi^+$, and (c) $B_s^0 \to D_s^\mp K^\pm$.

C Wrong-Sign fits

Fig. 32, 33, and 34 show the wrong-sign samples for the three different decay processes, fitted with a function of the form

$$f(x; p_0, p_1) = p_0 + (1 - p_0)\exp(p_1 x).$$

The results of these fits are displayed in Table 11.

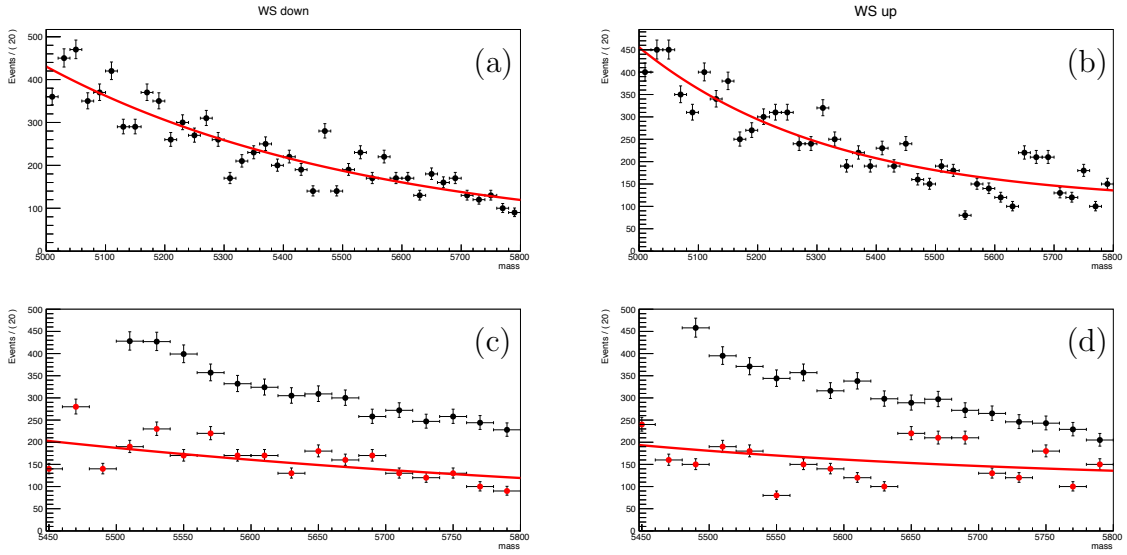

Figure 32: $D^{\pm}\pi^{\pm}$ (wrong-sign) data and a fit to that data, for (a) magnet down and (b) magnet up. Also plotted are the same data and fit (in red) together with the normal (right-sign) data (in black), in close-up on the mass range $[5450, 5800]$ MeV, for (c) magnet down and (d) magnet up.

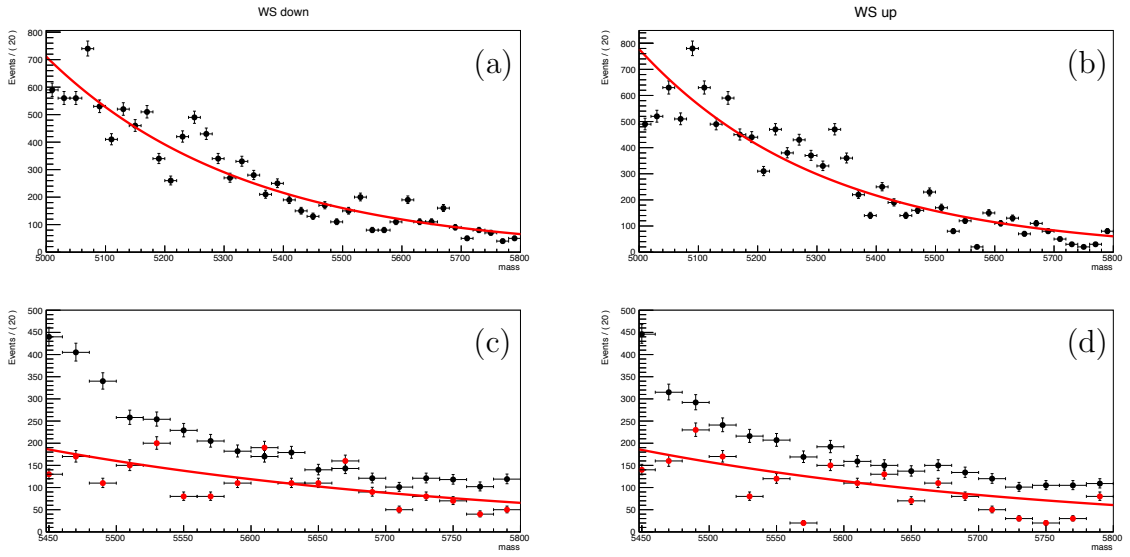

Figure 33: $D_s^{\pm}\pi^{\pm}$ (wrong-sign) data and a fit to that data, for (a) magnet down and (b) magnet up. Also plotted are the same data and fit (in red) together with the normal (right-sign) data (in black), in close-up on the mass range $[5450, 5800]$ MeV, for (c) magnet down and (d) magnet up.

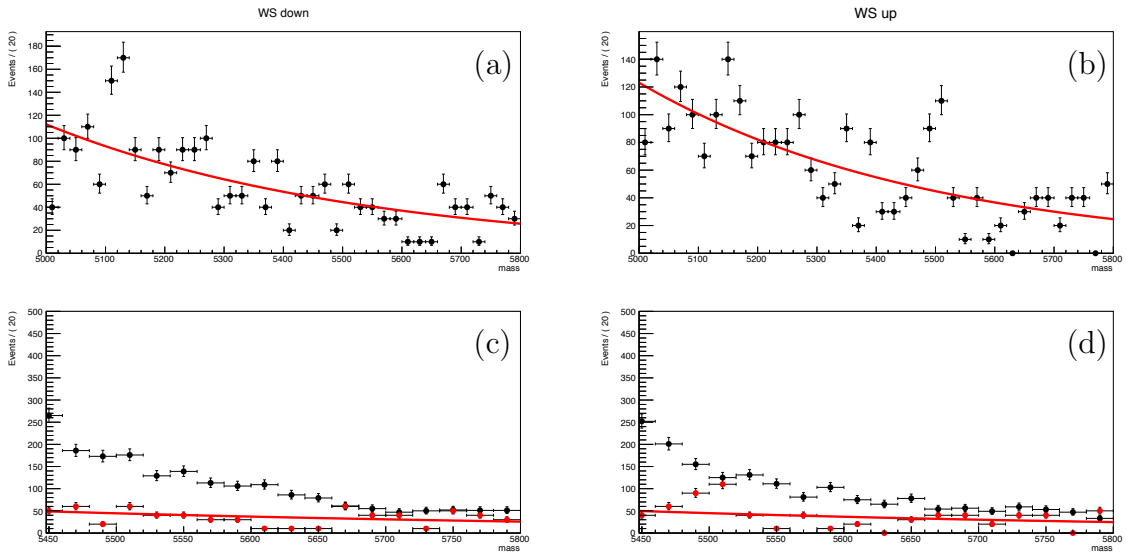

Figure 34: $D_s^{\pm}K^{\pm}$ (wrong-sign) data and a fit to that data, for (a) magnet down and (b) magnet up. Also plotted are the same data and fit (in red) together with the normal (right-sign) data (in black), in close-up on the mass range $[5450, 5800]$ MeV, for (c) magnet down and (d) magnet up.

References

[1] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) 8001.

[2] LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3** (2008) 8005.

[3] LHCb Collaboration, *LHCb online system, data acquisition and experiment control: Technical Design Report*, CERN-LHCC-2001-040. LHCB-TDR-007.

[4] LHCb Collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011. LHCB-TDR-005.

[5] LHCb Collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024. LHCB-TDR-006.

[6] LHCb Collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029. LHCB-TDR-008.

[7] R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, LHCb-DP-2013-002, in preparation.

[8] LHCb Collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007. LHCB-TDR-001.

[9] LHCb Collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037. LHCB-TDR-003.

[10] LHCb Collaboration, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036. LHCB-TDR-002.

[11] LHCb Collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010. LHCB-TDR-004.

[12] Particle Data Group, J. Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86** (2012) 010001, and 2013 partial update for the 2014 edition.

[13] LHCb collaboration, R. Aaij *et al.*, *Measurements of the branching fractions of the decays* $B^0_s \to D^\mp_s K^\pm$ *and* $B^0_s \to D^-_s \pi^+$, JHEP **06** (2012) 115, arXiv:1204.1237.

[14] K. De Bruyn *et al.*, *Exploring* $B_s \to D_s^{(*)\pm} K^\mp$ *Decays in the Presence of a Sizable Width Difference* $\Delta\Gamma_s$, Nuclear Physics B **868** (2012) 351, arXiv:1208.6463.

[15] R. Fleischer, *New Strategies to Obtain Insights into CP Violation Through* $B_s \to D_s^\pm K^\mp, D_s^{*\pm} K^\mp, ...$ *and* $B_d \to D^{\pm\mp}, D^{*\pm\mp}, ...$ *Decays*, Nuclear Physics B **671** (2003) 459, arXiv:hep-ph/0304027.

[16] LHCb collaboration, R. Aaij *et al.*, *Measurement of b hadron production fractions in 7 TeV pp collisions*, Phys. Rev. **D85** (2012) 032008, arXiv:1111.2357.

[17] LHCb Collaboration, *LHCb trigger system: Technical Design Report*, CERN-LHCC-2003-031. LHCB-TDR-010.

[18] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, arXiv:1210.6861.

[19] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175; T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[20] I. Belyaev *et al.*, *Handling of the generation of primary events in* Gauss*, the LHCb simulation framework*, Nuclear Science Symposium Conference Record (NSS/MIC) **IEEE** (2010) 1155.

[21] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[22] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, arXiv:hep-ph/0506026.

[23] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270; Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[24] M. Clemencic *et al.*, *The LHCb simulation application,* Gauss*: design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[25] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[26] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577, arXiv:physics/0408124.

[27] LHCb collaboration, *Measurement of the time-dependent CP-violation parameters in* $B_s^0 \to D_s^{\mp} K^{\pm}$, LHCb-CONF-2012-029.

[28] K. Cranmer, *Kernel Estimation in High-Energy Physics*, Comput. Phys. Commun. **136** (2001) 198, arXiv:hep-ex/0011057.

[29] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[30] LHCb collaboration, R. Aaij *et al.*, *Measurement of the p_T and η dependences of Λ_b^0 production and of the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ branching fraction*, LHCb-PAPER-2014-004, in preparation.

[31] L. Carson, R. Koopman, I. Sepp, and N. Tuning, *Measurement of the $p_{\rm T}$ and η dependences of Λ_b^0 production, and $\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+\pi^-)$, using hadronic decays*, LHCb-ANA-2013-023.

[32] LHCb collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP **04** (2013) 001, arXiv:1301.5286.

[33] LHCb collaboration, *Updated average f_s/f_d b-hadron production fraction ratio for 7 TeV pp collisions*, LHCb-CONF-2013-011.

[34] R. Fleischer, N. Serra, and N. Tuning, *Tests of factorization and $su(3)$ relations in b decays into heavy-light final states*, Phys. Rev. D **83** (2011) , arXiv:1012.2784.

[35] D. Santos and F. Duperthuis, *Mass distributions marginalized over per-event errors*, 2013. arXiv:1312.5000.