

Master Thesis
May 16, 2013

First observation of the decay $B_c^+ \to B_s^0 \pi^+$

J.A. de Vries

Physics - track Particle & Astroparticle Physics

Nikhef National Institute for Subatomic Physics
Faculty of Science (FNWI)
University of Amsterdam

Supervisors
Dr. N. Tuning
Prof. Dr. S. Bentvelsen

Track coordinator
Prof. Dr. S. Bentvelsen

Abstract

A search is presented for the decay $B_c^+ \to B_s^0 \pi^+$ using the channel $B_s^0 \to D_s^- \pi^+$ with data collected with the LHCb detector in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$, corresponding to an integrated luminosity of 1.0 fb^{-1} and 2.0 fb^{-1} respectively. Knowledge on B_c^+ decays is limited and the decays of the type $B_c^+ \to B_s^0 X$ have not been observed thus far. In total 82277 B_s^0 candidates and 62 B_c^+ candidates are selected, resulting in $\frac{f_c}{f_s} \times \mathcal{B}(B_c^+ \to B_s^0 \pi^+) = (2.19 \pm 0.36\ (\text{stat})\ ^{+0.34}_{-0.31}\ (\text{syst})) \times 10^{-3}$. The ratio of branching ratios is estimated as $\mathcal{B}(B_c^+ \to B_s^0 \pi^+)/\mathcal{B}(B_c^+ \to J/\psi \pi^+) = 81.1 \pm 13.4\ (\text{stat})\ ^{+12.4}_{-11.3}\ (\text{syst}) \pm 15.9\ (\text{ext})$. This is the first observation of a B-meson decaying weakly to another B-meson, and is the largest known branching fraction of any weak B-decay.

keywords: particle, physics, b-physics, flavour, heavy, Bc

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Niels Tuning, for his continuous support. His positive attitude and constructive feedback kept me motivated throughout the year, and I am very grateful for all the time he spent on discussing various aspects of the analysis with me. I could have wished no more from any supervisor.

I would like to offer my special thanks to Rose Koopman and Francesco Dettori, for their contribution to the B_s^0 mass fit and the limit setting procedure, respectively. Without their expertise this analysis would be far from the state it is currently in. Their patience and effort are greatly appreciated.

My gratitude goes out to the B&Q working-group conveners, Giulia Manca and Vanya Belyaev, for their support and feedback during and outside of the meetings. Also, the possibility to present this analysis at CERN during the 51th LHCb Software and Analysis week was a fantastic opportunity. In addition, I am grateful for the help received by Vava Gligorov and Jibo Hé, and their preliminary study on this decay using 2011 data. Vava's supervision while being a summer student at CERN ensured a quick start-up of this project, while Jibo's continuing work on the $B_s^0 \to J/\psi\,\phi$ channel makes a constructive addition to this analysis.

I am very grateful for the quick pace at which the referees, Mariusz Witek and Zhenwei Yang, reviewed the analysis note, allowing for an unblinding approval just in time to be included in this thesis. I would like to thank them in advance for the contribution they will have in the approval of this study.

I would also like to thank the staff of the b-physics department at Nikhef, for all the assistance and answers I have received, relating to physics, statistics, hardware and software. In particular the physics discussions with my office-mate Jeroen van Leerdam were always very productive. I owe the possibility to study this decay, and the financing of my stay at CERN to Marcel Merk and Niels Tuning, to whom I am very grateful.

Any comments and suggestions received during working group meetings, after my presentation in the Software and Analysis week, and after unblinding were very welcome, and I would like to thank anyone who made the effort of contributing to this analysis in one way or another.

This analysis was only possible due to the outstanding performance of the LHC, and the LHCb detector. Therefore I would like to thank anyone involved in the building, operation and financing of the man-made marvels and technological masterpieces called LHC, LHCb, and the software required to do this analysis.

Contents

1 Preface

The universe as we know it is dominated by matter over antimatter. This is crucial for the existence of life, for if matter and antimatter would coexist in equal amounts, the universe would be dominated by light and energy from continuous annihilation processes and no massive structures such as suns and planets could ever form. If matter would be localised into regions, and large amounts of antimatter would exist in so-called 'antimatter regions' - regions of the universe where antimatter dominates - there would be photons emanating from the annihilation processes at the border of these regions. So far, these photons have not been observed, supporting the assumption that the universe is matter dominated.

On the other hand, when particles are created from energy - as what is generally accepted to have occurred shortly after the big bang - matter and antimatter are created in exactly equal amounts.

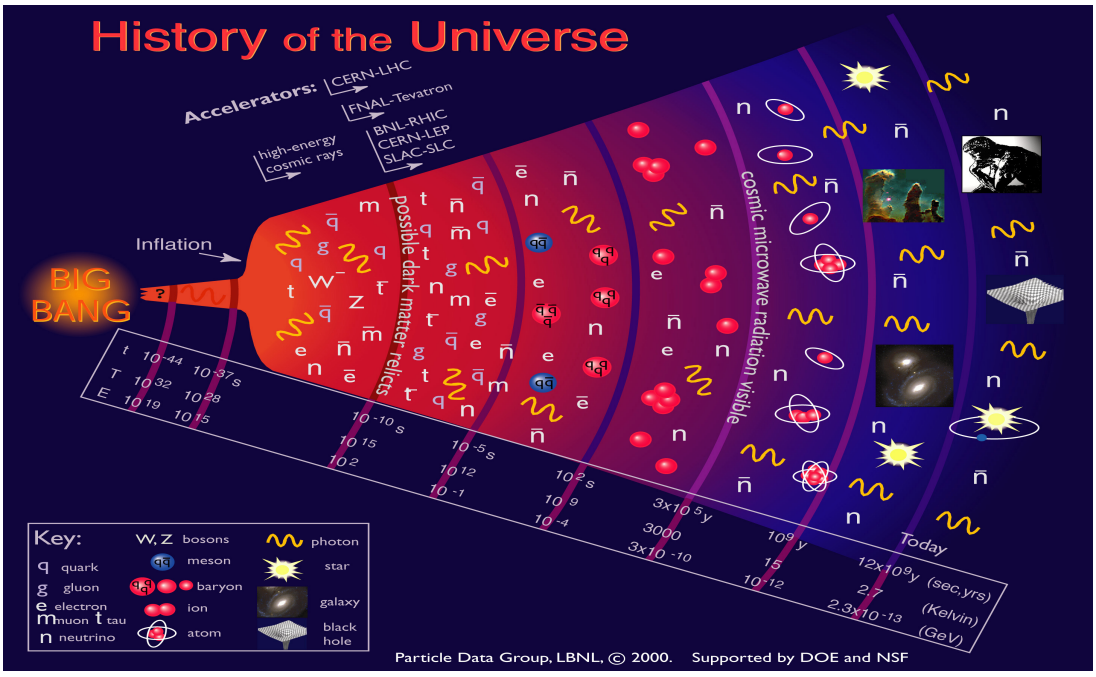


Figure 1: In the early universe, matter and antimatter is created in equal amounts. As the universe expands and cools down, most of the matter and antimatter annihilate, but a small fraction of matter remained.

The cause of this discrepancy, however fortunate, is one of the largest mysteries in fundamental physics. The Standard Model (SM) of particle physics does allow for a slight particle-antiparticle asymmetry - called Charge-Parity (CP) violation - visible in the decay of neutral mesons. CP violation is a necessary ingredient to obtain a matter-dominated universe. The search for CP violation, both within and outside the SM, is one of the main reasons for building the LHCb detector at CERN, which specialises in the detection of heavy meson decay products.

(Un)fortunately, the universe poses many other challenges that have yet to be explained by our understanding of physics. Even though the SM of particle physics works extremely well, standing up to many precise and stringent tests, it is known to have its limitations. For instance, insight into the properties of dark matter and dark energy is not described by our current knowledge, and some form of 'New Physics' is required to fully explain the physics in our universe. Furthermore, the continuing effort of combining theories of the very large - Gravity - with theories of the very small - the SM of particle physics - has not been very successful as of yet. Most models proposed by theorists that address these problems, e.g. supersymmetry (SUSY), predict new particles that could show up at LHC collision energies. Direct searches for new physics - where such a particle is created in interactions - are mostly done by the ATLAS and CMS experiments at CERN, while indirect searches - contributions of new particles in quantum loop corrections of heavy meson decays - are the specialty of LHCb. Possible effects of new particles are best studied in the decays of B-mesons. B-mesons consist of a b-quark and a d, u, s or c-quark. In this thesis the rare B_c^+ meson is studied. The decay of the B_c^+ meson to a B_s^0 meson is special, because it

1. allows to test QCD models in this interesting heavy-quarkonium system
2. might disturb the measurement of the B_s^0 decay time in analyses of B_s^0 decays
3. provides an interesting source of B_s^0 mesons.

No decay of B_c^+ to B_s^0 mesons have been observed to date.

2 Introduction

2.1 B_c motivation

The B_c^+ meson is unique as it is the only weakly decaying heavy-quarkonium system, consisting of the bottom and the charm quark, and so far it has only been studied limitedly. The possible decay of either quark compete with each other in the decay process of the B_c^+ meson. At present, only measurements exist in which the bottom quark decays weakly to a charm quark, while the decay of the B_c^+ meson to another b-meson (i.e. the case where the charm quark decays first) has never been observed, see Fig. 2. With the energy and luminosity of the LHC, and the sensitivity of the LHCb detector for heavy meson decays, these decays could possibly be observed for the first time.

A wide range of theoretical predictions on the branching ratio $\mathcal{B}(B_c^+ \to B_s^0\pi^+)$ exist [1], because e.g. corrections to Heavy Quark Effective Theory (HQET) are no longer of the order Λ_{QCD} but of the order of the charm quark mass, since there are now two heavy quark constituents. In addition, since the daughter B_s^0 is heavy the Q-value is small ("flies relatively slowly in the B_c^+ rest frame"), allowing for significant gluon exchange and making factorisation a worse approximation than in decays with light daughter particles. Study of the decay $B_c^+ \to B_s^0\pi^+$ allows to test these models. Also, precision studies of certain B_s^0 decays need an estimate of the amount of B_s^0 candidates originating from a B_c^+ decay. A first measurement in this area will thus be constructive to these analyses.

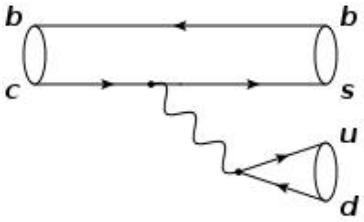


Figure 2: Diagram for the decay $B_c^+ \to B_s^0\pi^+$.

Furthermore, the study of time-dependent CP-violation in the Standard Model (SM) with the LHCb detector relies on measuring oscillations of neutral B_s^0 mesons. For these measurements it is required to identify the flavour of the B_s^0 meson at creation, which is done by a procedure called flavour tagging. Opposite-side tagging identifies the flavour of the other b-hadron in the system by finding a decay into a flavour-specific final state, while same-side kaon tagging uses the idea that the strange quark in the B_s^0 meson is pair-produced and its partner forms a kaon close to the B_s^0 in phase space. These methods are challenging in a hadronic environment, and their combined effective efficiency is small (for the $B_s^0 \to J/\psi\,\phi$ analysis it is about 4% [2]). However, when the B_s^0 originates from the decay $B_c^+ \to B_s^0 X^+$ the charge of the system X^+ indicates the flavour of the b-quark, resulting in a theoretical 100% efficient tag. These events can contribute to time-dependent CP violation measurements.

The B_c^+ meson was first measured at the Tevatron collider by CDF in 1998 in the semileptonic decay $B_c^+ \to J/\psi\,(\mu^+\mu^-)\ell^+X$ with $l = e, \mu$ [3]. Its mass was found to be $6.40 \pm 0.39 \pm 0.13$ GeV, its lifetime $0.46^{+0.18}_{-0.16} \pm 0.03$ ps, and the branching ratio $\mathcal{B}(B_c^+ \to (J/\psi \to \mu^+\mu^-)\ell^+\nu_\ell) \times \mathcal{B}(b \to B_c^+) = (5.2^{+2.4}_{-2.1}) \times 10^{-5}$. More stringent measurements with the fully reconstructed mode $B_c^+ \to J/\psi\,(\mu\mu)\pi^+$ did not occur until 2008 by CDF [4] and DØ [5], setting the current PDG average at $m_{B_c^+} = 6.277 \pm 0.006$ GeV and $\tau_{B_c^+} = 0.45 \pm 0.04$ ps.

Studies done by LHCb in 2012 resulted in the measurement $\mathcal{B}(B_c^+ \to J/\psi\pi^+\pi^-\pi^+)/\mathcal{B}(B_c^+ \to J/\psi\pi^+) = 2.41 \pm 0.30(\text{stat}) \pm 0.33(\text{syst})$ [6]. The ratio of production cross section times branching fraction of the decays $B_c^+ \to J/\psi\pi^+$ and $B^+ \to J/\psi K^+$ is measured to be $(0.68 \pm 0.10(\text{stat}) \pm 0.03(\text{syst.}) \pm 0.05(\text{lifetime}))\%$, for B_c^+ mesons with transverse momenta larger than 4 GeV/c and pseudorapidities between 2.5 and 4.5 [7]. In addition, the ratio of branching fractions of $\frac{\mathcal{B}(B_c^+ \to J/\psi D_s^+)}{\mathcal{B}(B_c^+ \to J/\psi\pi^+)} = 2.90 \pm 0.57(\text{stat}) \pm 0.24(\text{syst})$ and $\frac{\mathcal{B}(B_c^+ \to J/\psi D_s^{*+})}{\mathcal{B}(B_c^+ \to J/\psi D_s^+)} = 2.37 \pm 0.56(\text{stat}) \pm 0.10(\text{syst})$ were measured recently by LHCb in 2013 [8].

In this analysis we search for the exclusive decay $B_c^+ \to B_s^0\pi^+$, using the channel $B_s^0 \to D_s^-\pi^+$ [1].

2.2 Predicted signal yield

In order to estimate the signal yield of B_s^0 mesons originating from B_c^+ decays, the measurement of $\frac{\sigma(B_c^+)\times\mathcal{B}(B_c^+ \to J/\psi\pi^+)}{\sigma(B^+)\times\mathcal{B}(B^+ \to J/\psi K^+)} \approx (0.68 \pm 0.12)$ % is used [7]. In addition, the branching ratio $\mathcal{B}(B^+ \to J/\psi K^+)$ is averaged by the PDG [9] to be $(1.016 \pm 0.033) \times 10^{-3}$. A variety of theoretical models predict $\mathcal{B}(B_c^+ \to J/\psi\pi^+) = [0.06 - 0.18]\%$ [1]. Using a central value of 0.12%, this leads to the following estimate of the production ratio:

$$\frac{\sigma(B_c^+)}{\sigma(B_s^0)} = \frac{f_c}{f_s} = 0.68\% \times \frac{f_u}{f_s} \times \frac{\mathcal{B}(B^+ \to J/\psi K^+)}{\mathcal{B}(B_c^+ \to J/\psi\pi^+)} \approx 0.02,$$

where f_x denotes the fragmentation fraction of the b-quark to form a B_x meson, and where f_u/f_s is taken to be $\frac{1}{0.256}$ [10] assuming $f_u = f_d$. We use $\mathcal{B}(B_c^+ \to B_s^0\pi^+) = 10\%$ (in the predicted range $[2.5 - 16.4]\%$ again from Ref. [1]) resulting in a rough estimation of about 1 expected $B_c^+ \to B_s^0\pi^+$ candidate per 500 promptly produced B_s^0 mesons (ignoring the efficiencies of reconstructing and selecting $B_c^+ \to B_s^0\pi^+$ candidates). Throughout this thesis we use the notation $\frac{\sigma(b \to B_c^+)}{b} = f_c$, although strictly speaking, the production of B_c^+ mesons might not occur in the fragmentation process.

Recent NLO calculations [11] estimate $\mathcal{B}(B_c^+ \to J/\psi\pi^+) = (0.291^{+0.015+0.040}_{-0.042-0.027})\%$, which halves the above expectations.

[1] Charge conjugate states are implied throughout this note, unless explicitly stated otherwise.

3 Strategy

The goal of this analysis is to measure the branching ratio of the decay $B_c^+ \to B_s^0\pi^+$, or to set an upper limit, using the data collected by the LHCb detector from proton-proton (pp) collisions at the LHC in 2011 and 2012 corresponding to an integrated luminosity of about $3\,\text{fb}^{-1}$.

The signal region around the B_c^+ mass peak is blinded in the region $\pm 50\,\text{MeV}$ around the known mass of 6277 MeV [9], to prevent biasing the analysis and artificially creating a signal peak by adjusting parameter cuts.

3.1 Method

The efficiency corrected event yield equals the $b\bar{b}$ production cross-section times the hadronisation probability to form a B_c^+ meson, multiplied by the branching ratio,

$$\sigma(pp \to b\bar{b}) \times f_c \times \mathcal{B}(B_c^+ \to (B_s^0 \to X)\pi^+) = \frac{N(B_c^+ \to (B_s^0 \to X)\pi^+)}{\epsilon(B_c^+ \to (B_s^0 \to X)\pi^+)}.$$

If we normalise this quantity to the efficiency corrected yield of the $B_s^0 \to X$ decay,

$$\sigma(pp \to b\bar{b}) \times f_s \times \mathcal{B}(B_s^0 \to X) = \frac{N(B_s^0 \to X)}{\epsilon(B_s^0 \to X)}.$$

then the $b\bar{b}$ production cross-section, the efficiency to select $N(B_s^0 \to X)$ and the branching ratio $\mathcal{B}(B_s^0 \to X)$ cancel, to obtain the final result of this study:

$$\frac{f_c}{f_s} \times \mathcal{B}(B_c^+ \to B_s^0\pi^+) = \frac{N(B_c^+ \to (B_s^0 \to X)\pi^+)/\epsilon(B_c^+ \to B_s^0\pi^+)}{N(B_s^0 \to X)}. \tag{1}$$

Here $N(B_c^+ \to (B_s^0 \to X)\pi^+)$ is the observed yield of B_c^+ decays to the final state X and $N(B_s^0 \to X)$ the observed yield of B_s^0 decays. The factor $\epsilon(B_c^+ \to B_s^0\pi^+)$ accounts for the efficiency of reconstructing the bachelor pion in the decay $B_c^+ \to B_s^0\pi^+$, determined from simulation, as will be discussed in more detail in Sec. 6. The yield of $B_c^+ \to B_s^0\pi^+$ decays can then be compared to the yield of $B_c^+ \to J/\psi\phi$ decays, using external input, to obtain:

$$\frac{\mathcal{B}(B_c^+ \to B_s^0\pi^+)}{\mathcal{B}(B_c^+ \to J/\psi\pi^+)} = \frac{f_s}{f_d} \times \frac{(\frac{f_c}{f_s} \times \mathcal{B}(B_c^+ \to B_s^0\pi^+))}{\frac{\sigma(B_c^+)\times\mathcal{B}(B_c^+\to J/\psi\pi^+)}{\sigma(B^+)\times\mathcal{B}(B^+\to J/\psi K^+)}} \times \frac{1}{\mathcal{B}(B^+ \to J/\psi K^+)}. \tag{2}$$

With the NLO calculation of $\mathcal{B}(B_c^+ \to J/\psi\pi^+)$, an estimate of $\mathcal{B}(B_c^+ \to B_s^0\pi^+)$ can be given.

Reduction of combinatorial background is done in various steps. First the B_s^0 candidates pass the trigger selection and a loose offline preselection (called "stripping"). Then the B_s^0 candidates are selected using a D_s^+ mass window requirement, particle identification (PID) cuts and a boosted decision tree (BDT) [12] discriminating variable, 'BDT1'. The selection criteria are designed such, that the selection efficiency largely cancels in the ratio of Eq. 1. Finally the B_c^+ candidates are selected within a narrow B_s^0 mass window, using a PID cut on the bachelor particle, and the response of a second BDT variable, 'BDT2'. Both BDTs make use of the AdaBoost algorithm [13].

3.2 B_s^0 decay channels

There are several B_s^0 decay channels to be considered. The channel $B_s^0 \to J/\psi\,\phi$ (Fig. 3) is experimentally clean, due to the two muons in the final state and the narrow ϕ mass peak. The only expected background is of combinatorial nature, which is approximately constant in the invariant mass M($J/\psi\,\phi$) region close to M(B_s^0), leading to a straightforward estimation of the combinatorial background contribution. Also, the offline selection can be based purely on kinematic constraints. The total visible branching fraction to the final state particles is [14]

$$\begin{aligned}\mathcal{B}_{vis}(B_s^0 \to J/\psi\,\phi) &= \mathcal{B}(B_s^0 \to J/\psi\,\phi) \times \mathcal{B}(J/\psi \to \mu^+\mu^-) \times \mathcal{B}(\phi \to K^+K^-) \\ &= (1.05 \pm 0.10) \times 10^{-3} \times (5.93 \pm 0.06) \times 10^{-2} \times (48.9 \pm 0.5) \times 10^{-2} \\ &= (3.0 \pm 0.3) \times 10^{-5}. \end{aligned} \tag{3}$$

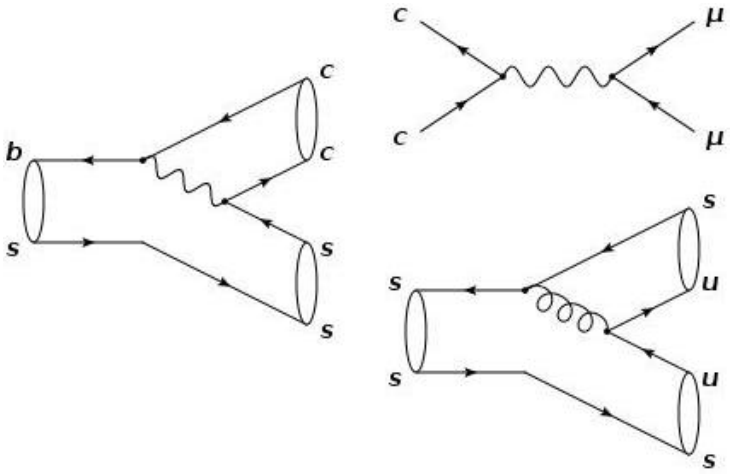


Figure 3: Diagrams for the full decay $B_s^0 \to (J/\psi \to \mu^+\mu^-)(\phi \to K^+K^-)$.

Another important channel, and the focus of this analysis, is the decay $B_s^0 \to D_s^-\pi^+$ (Fig. 4). Its visible branching ratio is [15]

$$\begin{aligned}\mathcal{B}_{vis}(B_s^0 \to D_s^-\pi^+) &= \mathcal{B}(B_s^0 \to D_s^-\pi^+) \times \mathcal{B}(D_s^- \to K^+K^-\pi^-) \\ &= (2.95 \pm 0.05 \pm 0.17 \pm^{0.18}_{0.22}) \times 10^{-3} \times (5.49 \pm 0.27) \times 10^{-2} \\ &= (1.62 \pm 0.25) \times 10^{-4}. \end{aligned} \tag{4}$$

This decay is a fully reconstructed B_s^0 decay with the highest branching ratio where the abundant final state $D_s^- \to K^+K^-\pi^-$ is considered. However, the decay $B_s^0 \to D_s^-\pi^+$ has contributions from various irreducible backgrounds, and their contribution to M($B_s^0\pi^+$) needs to be studied. This will be explained in Sec. 5. The possibility of adding the channel $B_s^0 \to K^+K^-$ is subject for future study.

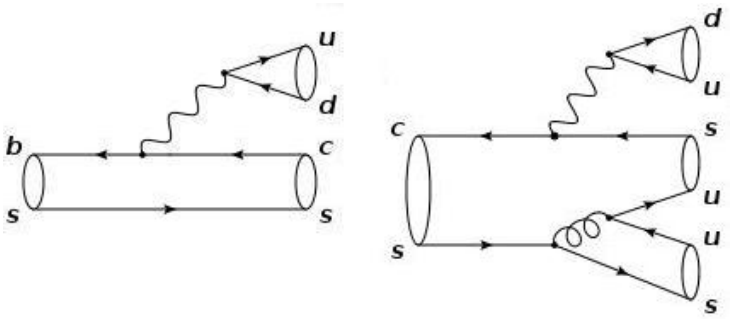


Figure 4: Diagrams for the full decay $B_s^0 \to (D_s^- \to K^+K^-\pi^-)\pi^+$.

4 The LHCb experiment

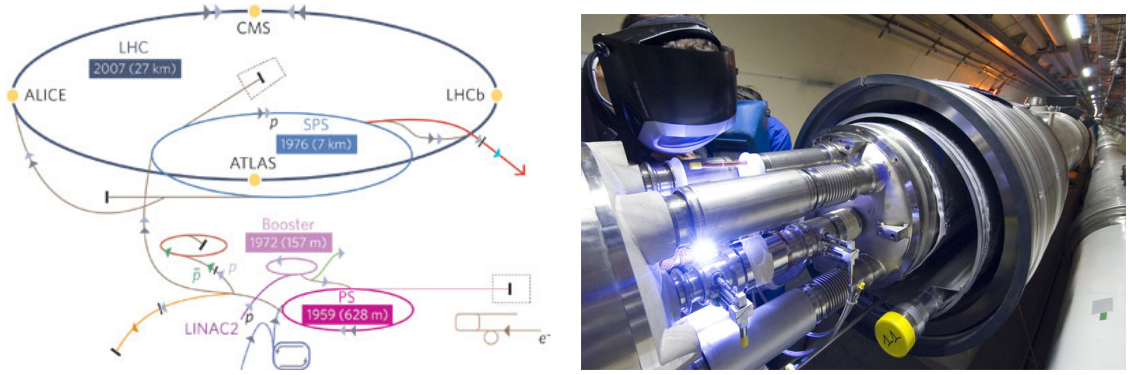
This section is dedicated to the description of the hardware and software setup of the experiment.

4.1 LHC

The Large Hadron Collider (LHC) [16] (Fig. 5) is a large synchrotron that accelerates protons (lead ions) up to energies of a few TeV per nucleon, and rests in the 27 km circumference tunnel of its predecessor, the Large Electron Positron Collider (LEP). The 3.8 m wide tunnel is located 100 m underground on average, near the CERN campus close to Geneva. A linear accelerator (LINAC2) accelerates protons from an ionized gas to about 50 MeV and injects them into a chain of pre-accelerators, including the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS), before the protons are injected into the LHC (Fig. 5).

The LHC consists of two 6.3 cm diameter rings with ultra high vacuum ($< 10^{-7}$ Pa, and $< 10^{-9}$ Pa near bunch crossing points) in which the proton beams move in opposite directions, surrounded by superconducting dipole magnets to bend the beam in a circle, and quadrupole and octopole magnets to focus the beam. The required magnetic field bending strength of 8.3 Tesla at design energy ($\sqrt{s} = 14\,\text{TeV}$) is achieved by cooling the superconducting electromagnets to 1.9 K with a liquid helium cryostat. The acceleration occurs in 8 straight Radio Frequency (RF) cavities, in which an oscillating field delivers about 2 MV to the protons during each pass, and squeezes the bunches tightly.

Four caverns are located around the tunnel in which the major LHC experiments are located (ATLAS, CMS, LHCb and ALICE). In the center of these experiments the bunches are crossed at a rate of 40 MHz, creating a wealth of high-energy proton-proton collisions. The design center of mass energy of the LHC is $\sqrt{s} = 14$ TeV but due to the 2008 quench accident, the LHC has run at $\sqrt{s} = 7$ TeV in 2011 and $\sqrt{s} = 8$ TeV in 2012. After the the long shutdown in 2013-2014 the LHC is expected to be able to safely reach the design energy.



(a) (b)

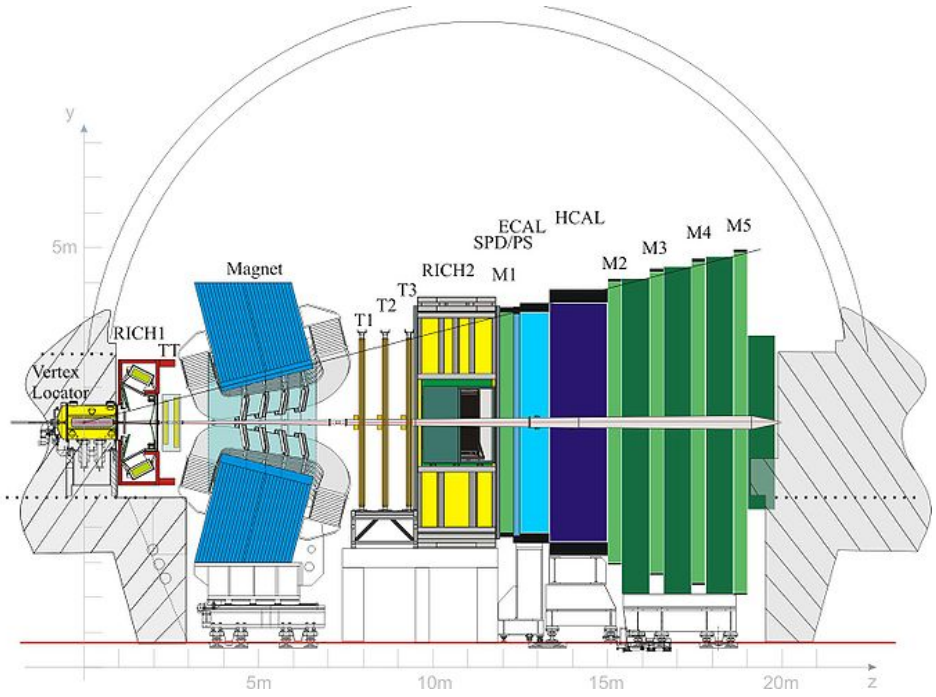
Figure 5: (a) An illustration of the different LHC accelerator rings, linacs and experiments. (b) An image of the LHC ring under construction. Visible is one of the two beam pipes with vacuum connections (middle), surrounded by four magnet bus bars (white caps), the helium vessel, vacuum vessel and radiation shield. *Source: [17]*

4.2 LHCb

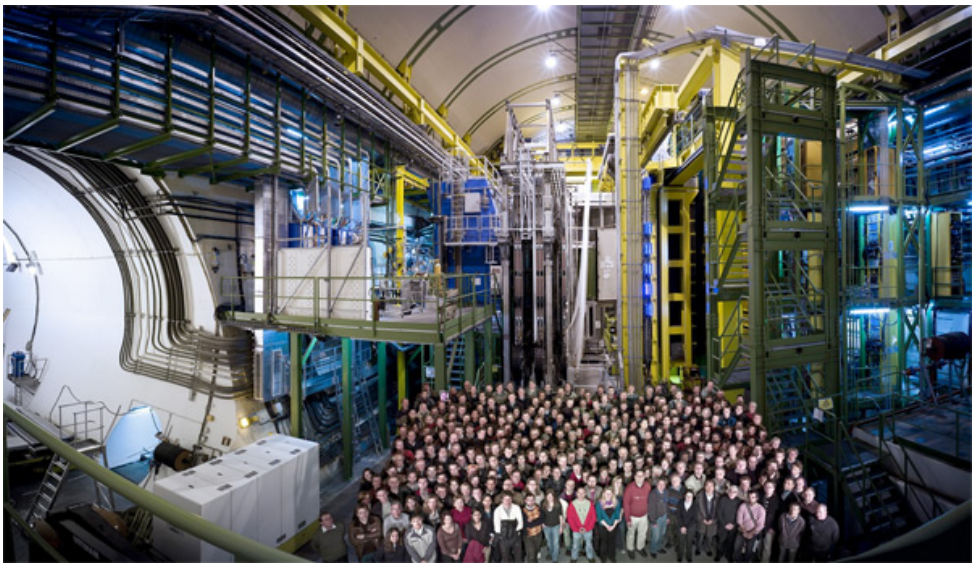
The Large Hadron Collider beauty (LHCb) experiment [18] (Fig. 6) is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, designed for the study of particles containing b or c quarks. The LHCb detector is a spectrometer in the sense that its main focus is to differentiate between final state particles by measuring the particle momentum, velocity and energy.

The motivation for covering only the forward region - as opposed to the more common hermetic cylindric detector shape - is that at the proton energies of the LHC, b-mesons are predominantly pair-produced in the same highly boosted direction[2]. The advantage of this setup is that LHCb can at relatively low cost cover a range of 10-300 mrad ($\eta \in [1.9, 5.3]$[3]) in the bending plane (x-z plane) of the magnet, and a range of 10-250 mrad ($\eta \in [2.1, 5.3]$) in the non-bending plane (y-z plane), stacking its detector layers one after another, instead of the full ϕ-covering multi-purpose detectors like ATLAS and CMS. Compared to these experiments, which have a tracking coverage of only up to $|\eta| < 2.5$, LHCb is ideally suited for detecting the decay products of particles that travel close to the beam line, such as the highly boosted b-quarks. In addition the vertex detector of LHCb is designed to have high vertex resolution to detect the displaced vertices of the relatively long-lived B-mesons, and determine their life-time accurately for oscillation measurements required for time-dependent CP violation analyses. Also the trigger system is specialized in heavy-meson final state detection, with a large computing farm performing online event reconstruction and multivariate analysis.

[2]At the collision energies of the LHC the most probable interaction creating a $b\bar{b}$ pair is by gluon fusion, and due to the PDFs of the constituents in the proton, one gluon is likely to be more boosted than



(a)



(b)

Figure 6: (a) A schematic overview of the LHCb detector and its subdetectors. (b) A picture of the LHCb detector in the cavern with a part of the collaboration in front. Visible from left to right are the magnet (blue), tracking stations, RICH2 (white), calorimetry (yellow) and muon chambers (green). *Source: [19].*

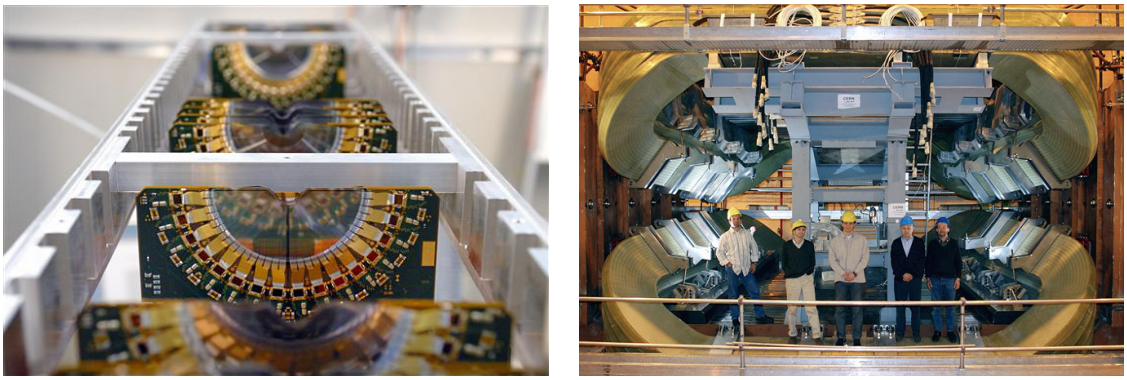
the other by far, resulting in two highly boosted b-mesons.

[3] $\eta = -\ln[\tan(\phi/2)]$

4.2.1 Tracking

The Vertex Locator (VELO) (Fig. 7a), together with the Tracker Turicensis (TT), Inner Tracker (IT) and Outer Tracker (OT) are responsible for the high precision tracking of charged decay products. The VELO consists of two arrays of semi-circular modules with silicon strip detectors. One module consists of 2 pieces that are glued back-to-back: one for a radial measurement, the other for an angular measurement. The two arrays are placed on both sides around the beam region to let the modules overlap into a cylindrical shape, separated from the beam by an RF box to preserve the ultra-high beam vacuum and prevent RF pickup from the beam. The VELO modules have a small inner radius of 8 mm, which allows for a lower limit of 15 mrad for the angle at which a track hits at least 3 VELO modules.

The TT and IT, also known as the Silicon Tracker (ST), are located upstream respectively downstream of the magnet. They consist of 1 resp. 3 stations of 4 layers of silicon microstrip detectors in the y direction, with an angle of $0°, -5°, 5°$ and $0°$ with respect to the first layer, in order to improve y-sensitivity. The TT is 160 cm wide and 130 cm high, and provides an intermediate detection layer between VELO hits and tracking hits after the magnet to reduce the amount of fake track combinations. The IT layers, with a total effective area of about 4 m^2, form the center of the three tracking stations T1, T2 and T3 downstream of the magnet and are completed by the surrounding OT modules, consisting of two layers of straw-tube drift cells with a total effective area of about 360 m^2. The purpose of these downstream tracking stations is measuring the deflection (thus momentum) of the tracks due to the warm magnet (Fig. 7b) with an integrated magnetic field strength of about 4 Tm.



(a) (b)

Figure 7: (a) An image of a VELO array before installation. (b) A picture of the LHCb's 1500 ton bending magnet. *Source: [19]*

The combined tracking system provides a momentum resolution that varies from 0.4% at 5 GeV/c to 0.6% at 100 GeV/c, and an impact parameter resolution of 20 µm for tracks with high transverse momentum.

4.2.2 Particle Identification

Particle identification (PID) is required to distinguish charged hadronic final states, and is mainly done with the Ring Imaging Cherenkov Detectors (RICH) 1 and 2, Electromagnetic and Hadronic Calorimeters (ECAL, HCAL) and the Muon system. The RICHes [20] use the principle of Cherenkov radiation: particles that travel faster than the speed of light in the medium emit photons in a cone with angle specific to the velocity. RICH1 is located upstream of the magnet in order to identify upstream tracks and uses aerogel as a radiator, covering the momentum range of 1-60 GeV/c, while RICH2 is located downstream of the magnet and uses CF_4 as radiator, covering the range up to approximately 100 GeV/c. The pattern of hits from the Hybrid Photon Detectors (HPDs) combined with the expectation from the reconstructed tracks under a certain particle hypothesis results in a delta log-likelihood (DLL) PID for each particle. For $\mathrm{DLL}(\pi - K) = 0$ the particle has equal probability to be either a pion or kaon.

Photon, electron and neutral hadron candidates are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic calorimeter and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers.

4.2.3 Trigger

The trigger [21] consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage which applies a full event reconstruction. The hardware stage reduces the event rate from 40 MHz to 1 MHz and is synchronous to the 40 MHz bunch crossings with fixed latencies.

The software trigger requires a two-, three- or four-track secondary vertex with a high sum of the transverse momentum, p_{T}, of the tracks and a significant displacement from the primary pp interaction vertices (PVs). The software trigger is split into two stages called HLT1 and HLT2. HLT1 reduces the rate from 1 MHz to 40 kHz that passes the hardware stage, with about 1 ms of processing time per event to make a decision. In this time window a partial reconstruction can take place using a Kalman filter. At least one track should have $p_{\mathrm{T}} > 1.7\,\mathrm{GeV}/c$ and impact parameter (IP) χ^2 with respect to the primary interaction greater than 16 to select the long-living b-decays. The IP χ^2 is defined as the difference between the χ^2 of the PV reconstructed with and without the considered track. HLT2 reduces the output rate to 2 kHz, after which the events are written to disk. Its 10 ms time window allows for forward tracking of all VELO tracks, the full off-line muon identification algorithms, and track association with ECAL clusters, corresponding to a full event reconstruction. A multivariate algorithm is used for the identification of secondary vertices consistent with the decay of a b-hadron.

4.3 Software

4.3.1 Reconstruction and stripping

Events that are written to disk are fully reconstructed offline. The difference with the online event reconstruction in HLT2 is the practically unlimited available time, which is used to employ many techniques using data from all subdetectors to improve on tracking and particle identification. After reconstruction all events are required to pass an offline selection called the "stripping". In this phase fiducial and inclusive cuts are made on the candidates, functioning as yet another data reduction step. The main trigger line used for this analysis is the so-called "topological" trigger[4], which is the widely used hadronic decay mode line, using a fast Bonsai boosted decision tree (BBDT) with topological variables [21]. Due to the contribution of $D_s^- \to \phi\pi^-$ resonances in the $D_s^- \to K^+K^-\pi^-$ decay an inclusive-ϕ[5] trigger line is added, which has tighter kaon identification cuts and allows relaxation of other constraints.

Since the above tasks require large amounts of computing power, LHCb is part of the Worldwide LHC Computing Grid (WLCG) [22]. It consists of a grid-based computer network infrastructure connecting over 170 computer centers (Tiers) in 36 countries, with over 200,000 processor cores. The reconstruction and stripping of the raw data is done on this Grid, after which the output is stored on the Grid in the over 150 PB large data storage, making the data available to all 60 institutes from 16 countries that participate in the LHCb experiment.

4.3.2 Boosted Decision Trees

Further selections, as described in Sec. 5 and Sec. 6, will make use of so-called Boosted Decision Trees (BDTs) [12]. A BDT is a multivariate technique, trained using N user-specified variables in a separate signal and background sample. It can define a certain probability, ranging from -1 (background) to 1 (signal), of an event to be either signal or background depending on its position in the N-dimensional phase space, by dividing up the phase space into large sequences of yes/no decisions called 'trees' in the training. Boosting is a way to increase classification performance and stability with respect to statistical fluctuations in the training samples by training a series of decision trees (a forest, if you will), building an expansion out of them and minimizing the difference between the response of the expansion ($F(x)$) and the true training samples (y), measured by the so-called 'loss function' $L(F, y)$. The BDT used in this thesis uses a technique called 'gradient boosting' [13], which is an extension to the commonly used adaptive boosting and improves performance on low-statistics data samples. It does so by using a binomial log-likelihood ($L(F, y) = \ln(1 + e^{-2F(x)y})$) loss function instead of an exponential loss function ($L(F, y) = e^{-F(x)y}$).

A BDT is, in principle, sensitive to overtraining. If the signal (or background) sample is too small or the number of variables too large, the sample can get isolated in 'islands'

[4] 'Hlt2Topo(2 | 3 | 4)Body'
[5] 'Hlt2IncPhi'

of phase space during the training, instead of having a smooth distribution. This can make the BDT seem very effective, but its response will vary greatly depending on the subsample used and will not behave as expected on the real data. In order to check for this, both signal and background training samples are divided into two: the training will take place on one half (the training sample) and the trained BDT will be evaluated on the other half (the test sample). If the response distributions do not overlap, "overtraining" has occurred.

4.3.3 Simulation

In the simulation, *pp* collisions are generated using Pythia 6.4 [23] with a specific LHCb configuration [24]. Decays of hadronic particles are described by EvtGen [25] in which final state radiation is generated using Photos [26]. The interaction of the generated particles with the detector and its response are implemented using the Geant4 toolkit [27] as described in Ref. [28].

The production of B_c^+ mesons is simulated using a dedicated B_c^+ generator called BcVegPy [29] to model the dominant subprocess $gg \to B_c^+(B_c^{+*}) + \bar{c} + b$ interfaced with Pythia. Since $m_b \gg m_c \gg \Lambda_{QCD}$ the process can be calculated with perturbative QCD (pQCD), if the non-perturbative factor of combining $\bar{b}$ and c into B_c^+ is attributed to the decay constant $F_{B_c^+}$. A full pQCD calculation at the lowest order α_s^4 (36 contributing Feynman diagrams) is favoured over a (relatively simple) fragmentation approach, in order to retain information on the $\bar{c}$ and b quark jets. To make the calculation more efficient, the "particle helicity technique" is used: the diagrams with all helicities of massless external particles fixed are calculated first. All diagrams are then summed and the squared modulus is taken, averaging over the helicities of initial state particles and summing over the helicities of final state particles. Since the final states are not massless, this procedure has to be generalized, which is done by converting the problem into an equivalent massless one, and trying to extend the symmetries as much as possible as described in Ref. [29].

4.4 Data samples

The data samples used for this analysis are *pp* collisions collected by LHCb, corresponding to an integrated luminosity of $1\,\text{fb}^{-1}$ at $\sqrt{s} = 7\,\text{TeV}$ in 2011 and $2\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$ in 2012[6]. The polarity of the magnet is regularly changed to cancel possible systematic detector effects.

The MC samples used are 6M (Magnet Up) + 6M (Magnet Down) of prompt $B_s^0 \to (D_s^- \to K^+K^-\pi^-)\pi^+$ events[7], and 500K (Magnet Up) + 500K (Magnet Down) $B_c^+ \to (B_s^0 \to (D_s^- \to K^+K^-\pi^-)\pi^+)\pi^+$ events, both selected to have their decay products in the LHCb acceptance ("DecProdCut").

[6]The versions used are for 2011 data: reco12, stripping17, and for 2012 data: reco14, stripping20.
[7]MC11a: sim05, Reco12a, Stripping17

5 $\mathbf{B_s^0}$ selection

The events that pass the trigger are reconstructed and filtered through a preselection called the "stripping". After this preselection, inclusive B_s^0 candidates are formed, which are then selected more stringently. In this section the preselection, as well as the offline selection using PID and a BDT of the B_s^0 candidates are described. The selection of B_c^+ candidates is described in Sec. 6.

5.1 Stripping

The stripping selection for the $B_s^0 \to (D_s^- \to K^+K^-\pi^-)\pi^+$ decay[8] contains final state track quality cuts in order to assure a correct track reconstruction, (transverse) momentum requirements to cut particles that did not decay from a heavy meson, vertex χ^2 and maximum Distance of Closest Approach (DOCA) cuts for correct vertex reconstructions, a cut on the angle between the D_s^- momentum and vector from the PV to the D_s^- decay vertex (DIRA) to limit the available phase-space, PV distance and impact parameter (IP) significance cuts to assure the decay products do not originate from the PV (are 'detached'), and loose mass and PID requirements (for a full list, see Tab. 1). Pointing cuts, or cuts that constrain the B_s^0 candidate to have originated from the PV are avoided, in order to allow for the decay $B_c^+ \to B_s^0\pi^+$ and to avoid any differences in the selection of B_s^0 mesons originating from the PV compared to B_s^0 mesons from a B_c^+ decay.

The ghost probability of the final state tracks is calculated differently for 2011 and 2012 data, which is the reason for the difference in track χ^2 and ghost probability cuts. In addition, the cuts applied to the combination of $(B_s^0\pi^+)$ candidates into a B_c^+ decay vertex are shown in the bottom of Tab. 1. It consists of loose PID and mass requirements, a loose vertex fit χ^2, a p_T and track quality cut, and a PV IP χ^2 requirement to cut away most prompt pions. Tighter offline selection is required and is primarily done with a multivariate technique, described below.

[8]Stripping line: B02DPiNoIPD2HHHCFPIDBeauty2Charm

Table 1: The various cuts made in the stripping and first offline selection. The cuts in parenthesis are only applied to 2012 data.

	Parameter	Stripping	Offline Selection
	# Long Tracks	< 500	
	# PVs	≥ 1	
	Hlt2IncPhi TIS/TOS Trigger or		
	Hlt2Topo(2 | 3 | 4)Body TIS/TOS Trigger	True	
	All final state tracks GhostProb	(< 0.3)	< 0.5
D_s^-	D_s^- Daughters Track χ^2/ndf	< 4 (< 3)	
	D_s^- Daughters p_T	$> 100\,\text{MeV}$	
	D_s^- Daughters p	$> 1000\,\text{MeV}$	
	D_s^- Daughters Min. IP χ^2 (Best PV)	> 4	
	D_s^- One Daughter: Track χ^2/ndf	< 3 (< 2.5)	
	p_T	$> 500\,\text{MeV}$	
	p	$> 5000\,\text{MeV}$	
	D_s^- $\sum_{\text{daughters}} p_T$	$> 1800\,\text{MeV}$	
	D_s^- Daughters DOCA	$< 0.5\,\text{mm}$	
	D_s^- Vertex χ^2/ndf	< 10	
	D_s^- (Best PV - endVertex) Distance χ^2	> 36	
	D_s^- Best PV DIRA	> 0	
	$K^\pm_{D_s^-}$ PID($K-\pi$)	> -10	> 0
	$\pi^+_{D_s^-}$ PID($K-\pi$)	< 20	
	D_s^- Mass	$> 1868.49\,\text{MeV}$	$> 1944\,\text{MeV}$
		$< 2068.49\,\text{MeV}$	$< 1990\,\text{MeV}$
B_s^0	$\pi^+_{B_s^0}$ Track χ^2/ndf	< 2.5	
	$\pi^+_{B_s^0}$ p_T	$> 1700\,\text{MeV}$	
	$\pi^+_{B_s^0}$ p	$> 10000\,\text{MeV}$	
	$\pi^+_{B_s^0}$ Min. IP χ^2 (Best PV)	> 4	
	B_s^0 Mass	$> 4750\,\text{MeV}$	$> 5000\,\text{MeV}$
		$< 7000\,\text{MeV}$	$< 5800\,\text{MeV}$
	B_s^0 Vertex χ^2/ndf	< 10	
	B_s^0 Min. IP χ^2 (Best PV)	> 16	
	B_s^0 Min. IP (Best PV)	$> 0.1\,\text{mm}$	
	B_s^0 Best PV Lifetime	$> 0.2\,\text{ps}$	
B_c^+	$\pi^+_{B_c^+}$ p_T		$> 100\,\text{MeV}$
	$\pi^+_{B_c^+}$ Track χ^2/ndf		< 3
	$\pi^+_{B_c^+}$ Min. IP χ^2 (Best PV)		> 2
	$\pi^+_{B_c^+}$ PID($K-\pi$)		< 10
	B_c^+ Mass		$> 5000\,\text{MeV}$
	B_c^+ Vertex χ^2/ndf		< 100

5.2 Training BDT1 and offline selection

In order to reject combinatorial background from fake D_s^- candidates, a mass window of $(1944 < \mathrm{M}(D_s^-) < 1990)$ MeV is applied to the data before applying a BDT1 cut. The signal sample used for the training of BDT1 is the (prompt) B_s^0 MC sample, after passing the stripping selection described in Tab. 1. In order to ensure the candidates to be correctly reconstructed, the full decay chain of the reconstructed particles is required to match the true simulated particle, resulting in a sample size of 45.8k events. For the background sample the upper mass sideband is used, $M(D_s^-\pi^+) > (5367 + 100)\,\mathrm{MeV}$, of reconstructed $B_s^0 \to D_s^-\pi^+$ events in the 2011 data after stripping and applying a D_s^- mass window of $(1944 < \mathrm{M}(D_s^-) < 1990)$ MeV, while selecting one random candidate per event (in order to prevent biasing events with high multiplicity). The size of the background sample is 25.6k events. The lower mass sideband is not used since it will contain irreducible (partially reconstructed and misidentified) background. The samples are divided into two equally sized random subsets. One pair is used for the training, and the other for validation of the BDT output and as a check for overtraining.

The variables used in the training of BDT1 are motivated by previous analyses of the decays $B_s^0 \to D_s^-K^+$ and $B_s^0 \to D_s^-\pi^+$ [30] and consist of kinematic variables like the transverse momentum, transverse angle and flight distances with respect to their associated PV or their associated secondary vertex. In addition the cosine of the angle between the momentum of the particle and the direction vector of some vertex to the particle's decay vertex (DIRA) is used. The vertex χ^2 is a measure for the validity of the tracks for forming a vertex, opposed to the IP χ^2 which is used as a measure for a track not originating from a certain vertex. Again, no pointing variables like impact parameter of the B_s^0 candidate are used. The B_s^0 flight distance with respect to the PV is used since its discriminating power originates from the long lifetime of the B_s^0 meson and is thus inclusive for B_s^0 candidates originating from a B_c^+ meson. In addition no PID (RICH Delta Log-Likelihood (DLL) $(K-\pi)$) information is used in the training of BDT1 because these distributions are known to show a discrepancy between data and MC. The distributions of the irreducible (misidentified and partially reconstructed) backgrounds are known for a given, well-defined PID cut Ref. [10], and are needed in the B_s^0 mass fit. Therefore, a well-defined PID cut is used afterwards. A full list of the variables used in BDT1 and their relative discriminating importance is given in Tab. 2.

The result of the training is shown in Fig. 8. The discriminating power of the BDT1 is expressed in a so-called ROC curve, which shows the signal efficiency against the background rejection. In Fig. 8b, the BDT1 response of the training and evaluation samples is shown; the training sample is not different (i.e. unbiased) compared to the evaluation sample, as they overlap almost perfectly.

After the training of BDT1 a tight cut is made on the PID Delta Log-Likelihood (DLL) $(K-\pi)$ of the same-charged K^- from the D_s^- to be larger than 5 (since the same-charged K^- meson carries the charge of the s-quark of the D_s^- meson, thereby distinguishing the decay $D_s^- \to K^-K^+\pi^-$ from $D_s^- \to \pi^-K^+\pi^-$, see Fig. 4), and on the PID DLL$(K-\pi)$ of the pion bachelor from the B_s^0 to be smaller than 0. After these cuts, the efficiency of

Table 2: Variables used in the training of BDT1.

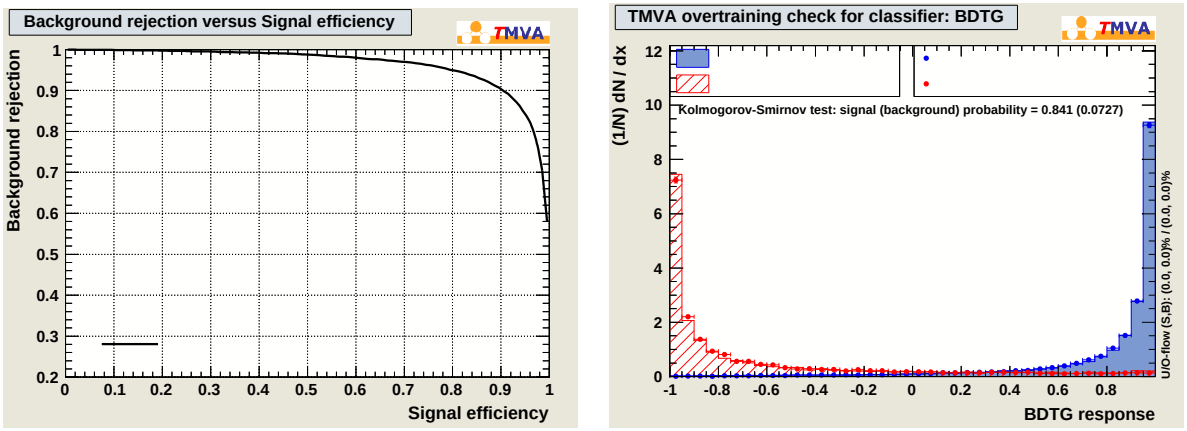
Rank	Variable	Relative importance
1	B_s^0 p_T	7.663e-02
2	$\pi_{B_s^0}$ $\cos(\theta)$	7.420e-02
3	$\pi_{B_s^0}$ p_T	6.917e-02
4	D_s^- p_T	6.742e-02
5	D_s^- Decay vertex χ^2/ndf	6.678e-02
6	B_s^0 Decay vertex χ^2/ndf	6.514e-02
7	D_s^- p	6.514e-02
8	B_s^0 p	6.325e-02
9	D_s^- Min. Daughter p_T	6.030e-02
10	$\pi_{B_s^0}$ p	5.998e-02
11	D_s^- $\cos(\theta)$	5.923e-02
12	D_s^- Flight Distance (oriVX)	4.966e-02
13	D_s^- Radial Flight Distance (ownPV)	4.646e-02
14	D_s^- DIRA (oriVX)	4.304e-02
15	B_s^0 Flight Distance (ownPV)	3.455e-02
16	B_s^0 Radial Flight Distance (ownPV)	2.198e-02
17	D_s^- IP χ^2 (ownPV)	2.067e-02
18	$\pi_{B_s^0}$ IP χ^2 (ownPV)	2.040e-02
19	D_s^- DIRA (ownPV)	1.503e-03
20	D_s^- Min. Daughter IP χ^2 (ownPV)	1.093e-02
21	D_s^- Flight Distance χ^2 (oriVX)	1.005e-03

different BDT1 working points is re-evaluated and shown in Fig. 9. This is done in order to choose the optimal working point after all additional cuts.

The green curve in Fig. 9 shows the $S/\sqrt{S+B}$ ratio and is used to choose the optimal working point. For the calculation of S and B the expected number of signal and background events in the signal region is multiplied by the efficiency at that BDT cut. The expected numbers of signal and background events in the signal region are estimated using the results of the fit of an earlier version of the analysis, and amount to 25k signal events and 70k background events. We choose a working point at BDT1 > 0.5, which is at the optimal working point with $S/\sqrt{S+B} = 133$. The signal efficiency for this point is 87% while the background retention is 7.4%, relative to the data samples after the preselection (stripping), D_s^- mass window and daughter PID cuts. A full list of signal efficiencies in the B_s^0 selection steps are listed in Tab. 13.

5.3 B_s^0 backgrounds and mass fit

As mentioned in Sec. 3.2, combinatorial background is not the only contributing background in this hadronic decay mode. Partially reconstructed background events like $B_s^0 \to D_s^{*-}\pi^+$



(a) ROC (b) Overtraining

Figure 8: Result of the BDT1 training. (a) The discriminating power of the BDT by plotting signal efficiency vs. background rejection. (b) The superposition of the BDT response of the training (dots) and test (bin) samples of signal (blue) and background (red). Both BDT distributions agree well, showing that there is no overtraining.

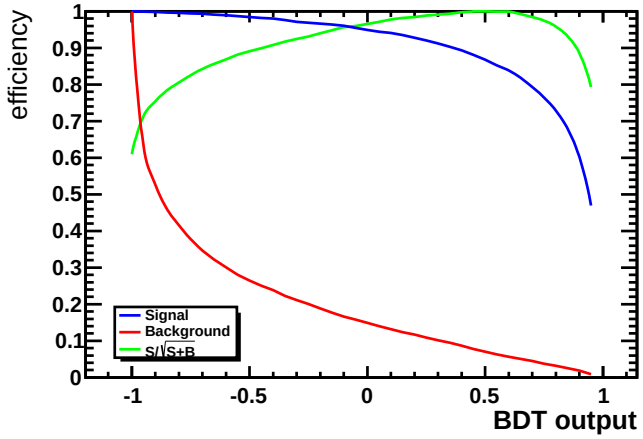
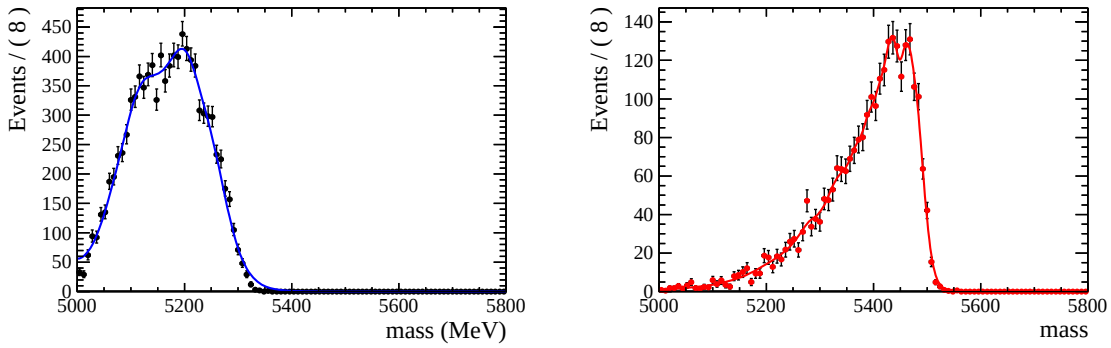


Figure 9: BDT1 efficiency curves after PID and D_s^- mass cuts.

and $B_s^0 \to D_s^- \rho^+$ decays in which a soft photon or π^0 is not reconstructed by the detector, will have a reconstructed mass below the nominal B_s^0 mass. In addition, misidentified background - mostly $B^0 \to D^- \pi^+$ decays in which a pion in the decay $D^- \to \pi^- \pi^- K^+$ has been misidentified as a kaon, and $\overline{\Lambda}_b^0 \to \overline{\Lambda}_c^- \pi^+$ decays, in which the proton of the decay $\overline{\Lambda}_c^- \to \pi^- \bar{p} K^+$ has been misidentified as a kaon - will have a contribution in the B_s^0 mass region. Also, the Cabbibo-suppressed decay $B^0 \to D_s^- \pi^+$ can have a small contribution. Since all these decay modes exhibit similar topology, it is unavoidable to include them in

the selection and are hence called 'irreducible'. In order to account for these contributions in the mass fit, MC samples of these decays are used to constrain the shapes of the mass distributions, see Fig. 10.



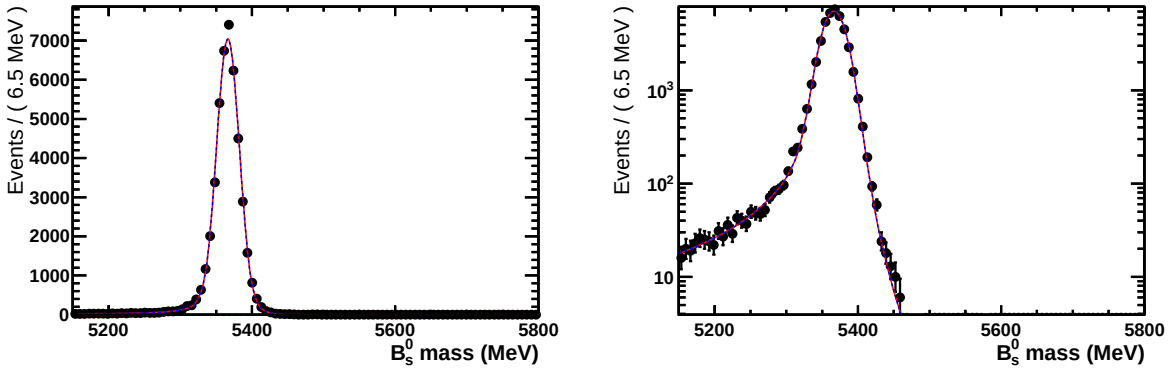
(a) $B_s^0 \to D_s^{*-}\pi^+$ (simulation) (b) $\Lambda_b^0 \to \Lambda_c^+\pi^+$ (simulation)

Figure 10: Invariant mass shapes of various backgrounds that contribute to the distribution of $B_s^0 \to D_s^-\pi^+$ candidates.

The simultaneous fitting of the partially reconstructed and misidentified fully reconstructed shapes on MC, and the combination with combinatorial background and signal shape on data is done in the same way as in Ref. [31]. The MC and data samples are split in Magnet Up (MU) and Magnet Down (MD) because PID performance was slightly different between the two samples in the 2010 data, although the effect was found to be negligible in the 2011 data. One random candidate per event is taken for the fit, since multiple candidates are mainly due to the correct B_s^0 candidate combined with a wrong π^+ from the PV. The signal shape is taken from a fit to simulated events after full trigger, reconstruction, stripping and offline selection chain, see Fig. 11. After considering various models [31] it was decided to use a double Crystal Ball function (DCB) with the same values for the mean and resolution, in order to take into account the radiative tail (lower side) and detector resolution effects (upper side). A Crystal Ball (CB) function is a Gaussian with a power-law tail, according to the following parameterisation:

$$f_{\rm CB}(m,\alpha,n,\mu,\sigma) = N \cdot \begin{cases} e^{-\frac{(m-\mu)^2}{2\sigma^2}}, & \text{for} \frac{m-\mu}{\sigma} > \alpha \\ A \cdot (B - \frac{m-\mu}{\sigma})^{-n}, & \text{for} \frac{m-\mu}{\sigma} \leq \alpha \end{cases} \tag{5}$$

where $A = (\frac{n}{|\alpha|})^n \cdot e^{-\frac{|\alpha|^2}{2}}$ and $B = \frac{n}{\alpha} - |\alpha|$. Here N is the normalisation factor, m is the invariant mass, μ is the mean, σ the mass resolution, α is the point where the shape changes from Gaussian to a power-law tail and n describes the size of the tail. In the fit the common mean and resolution are allowed to vary, while the description of the tails is fixed from simulation, see Fig. 11 and Tab. 3. The ratio of the two CB functions is fixed to 0.5.



(a) Linear scale (b) Log scale

Figure 11: The signal shape is obtained from a DCB fit to the mass distribution of simulated B_s^0 events. The reconstructed B_s^0 candidates are required to match the true generated B_s^0 decays.

Table 3: Fit parameter values for the B_s^0 simulation signal fits in Fig. 11.

Parameter	Fit Results
mean	5366.46 ± 0.08 MeV
width	16.01 ± 0.07 MeV
α_L	1.71 ± 0.01
α_R	-1.87 ± 0.02
n_L	1.38 ± 0.03
n_R	7.10 ± 0.49
N_{sig}	45507 ± 213

The shape of the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ misidentified background is extracted from simulated events, reweighed according to the $(p \to K)$ misidentification probability in data[9], following Ref. [10] and its yield is allowed to vary in the fit. The shape of the misidentified background $B^0 \to D^-\pi^+$ is determined from a fit to data by selecting a clean sample, then assuming a wrong mass hypothesis and reweighing according to the $(\pi \to K)$ misidentification probability. Since its contribution is peaking in the signal region, the yield is constrained using the misidentification probability with a Gaussian constraint to account for the uncertainty. Since the number of events is expected to differ between 2011 Magnet Up and Magnet Down, and 2012, the $B^0 \to D^-\pi^+$ constrained yield is scaled with the number of events that pass the selection, and amounts to 794 ± 63 and 1109 ± 87 events for 2011, and 2505 ± 200 and 2445 ± 196 events for 2012, magnet up and down respectively.

[9]The misidentification probability is determined as a function of momentum and pseudorapidity using clean channels with identified final states, without using the RICH. See section 6.4.

The shapes of the partially reconstructed background $B_s^0 \to D_s^{*-}\pi^+$ and $B_s^0 \to D_s^-\rho^+$ are obtained using simulated events, corrected for the momentum spectra and PID efficiencies in the data, using non-parametric functions. The shapes are fixed in the fit, while the yields are allowed to vary. The yield of $B^0 \to D_s^-\pi^+$ decays is fixed to 1/30 of the $B_s^0 \to D_s^-\pi^+$ yield, using the ratio of branching ratios (1/137) and different production rates of B^0 and B_s^0 mesons, $f_s/f_d = 0.256$. The mean reconstructed mass of the B_s^0 meson is shifted with about 3 MeV compared to the PDG [9] value of 5366.77 ± 0.24 MeV, due to misalignment and incomplete magnetic field mapping and is allowed to vary in the fit. The result of the fit to the mass distribution of B_s^0 candidates is presented in Sec.7.

6 $\mathbf{B_c^+}$ selection

The main task of the B_c^+ selection is to separate true B_c^+ decays from combinations of B_s^0 candidates with a prompt, random, charged pion. Since the mass difference between the B_s^0 and the B_c^+ mesons is small, the transverse momentum of the π^+ is not expected to be particularly large. In addition, the B_c^+ meson has a shorter lifetime compared to the B_s^0 meson, and therefore it is difficult to distinguish the bachelor pion from the B_c^+ decay, from prompt pions. The large number of prompt pions, results in multiple pions to form topologically 'correct' combinations within the same event: an average of about 10 B_c^+ candidates per event is found in the 2011 data after the full B_s^0 selection.

6.1 Training BDT2

To remove most of the combinatorial background consisting of a random pion from the PV as bachelor, a second BDT - from now on referred to as BDT2 - is used in which pointing variables can now be utilised. The B_s^0 and B_c^+ IP χ^2, DIRA's and lifetime significances, as well as kinematic variables like the (transverse) momentum are used in the training of BDT2. In addition, an algorithm called the decay tree fitter [32] (DTF) uses the constraints that the B_c^+ has to originate from the PV and the masses of B_s^0 and D_s^- mesons are equal to the PDG mass when the whole decay tree is refitted, creating more accurate estimates for the vertices and lifetimes. The DTF χ^2 is the χ^2 of the whole decay tree refit and acts as a sum of all independent track and vertex χ^2 values. A full list of variables used in the training of BDT2 and their relative importance can be found in Tab. 4.

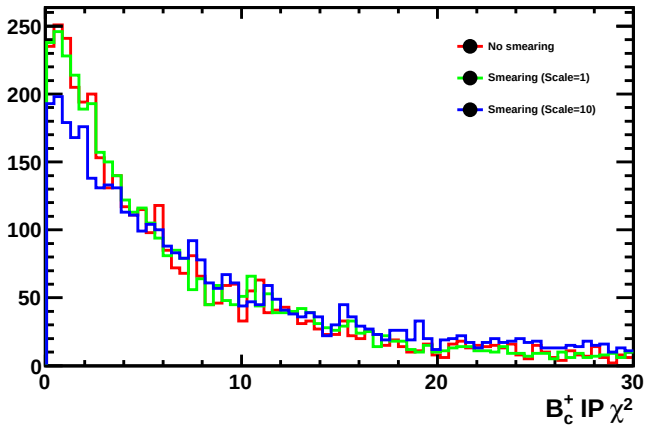


Figure 12: The IP χ^2 distribution of B_c^+ MC with no, intermediate, and large smearing of the IP χ^2, to match the distribution in data.

For the signal sample simulated $B_c^+ \to B_s^0\pi^+$ events, described in Sec. 4, are used. In order to improve the data to MC comparison the MC tracks are smeared with a smearing scale of 1 (for a comparison of the effect of smearing on the IP χ^2 see Fig. 12). Since some

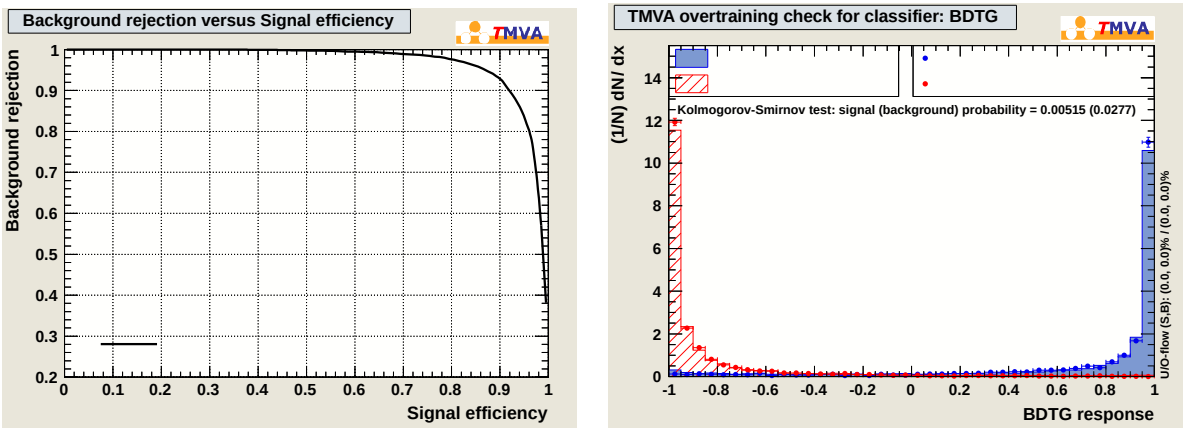
candidates pass the stripping line only after smearing, and vice versa, the preselection was done after this smearing. After the events pass the stripping line, all final state particles are required to coincide with the true generated final state particles, resulting in 9543 events used for the training.

For the background sample the B_c^+ upper sideband in the data ($M(B_s^0\pi^+) > 6277 + 50\,\text{MeV}$) is used, after the full B_s^0 selection described in Sec. 5. In addition a narrow B_s^0 mass window of $M(B_s^0) = 5371 \pm 36\,\text{MeV}$ is applied. In order to prevent a bias towards events with high multiplicity, one random candidate per event is selected after which 20.6k events remain. Similar to the training of BDT1, the samples are randomly divided into two subsamples: one for training and one for testing, in order to check for overtraining.

Table 4: Variables used in the training of BDT2.

	Variable	Relative importance
1	$\pi^+_{B_c^+}$ $\cos(\theta)$	1.004e-01
2	$\pi^+_{B_c^+}$ p_T	9.156e-02
3	$\pi^+_{B_c^+}$ p	7.862e-02
4	B_c^+ p_T	6.907e-02
5	B_s^0 DTF $c\tau$	6.785e-02
6	B_c^+ DTF p	6.511e-02
7	DTF χ^2	6.360e-02
8	B_s^0 DTF Decay Length (oriVX)	6.193e-02
9	B_s^0 DTF p	5.924e-02
10	B_c^+ DTF cτ	5.462e-02
11	B_s^0 $\cos(\theta)$	5.285e-02
12	B_c^+ Endvertex χ^2/NDF	5.044e-02
13	B_c^+ IP χ^2 (ownPV)	4.780e-02
14	B_c^+ DTF Decay Length (oriVX)	4.717e-02
15	B_s^0 IP χ^2 (ownPV)	3.398e-02
16	B_c^+ DIRA (ownPV)	2.159e-02
17	B_s^0 DIRA (oriVX)	1.828e-02
18	$\pi^+_{B_c^+}$ IP χ^2 (ownPV)	1.592e-02

The result of the training is shown in Fig. 13. The ROC-curve shows high background suppression while still retaining large signal yield. As is visible from the distributions in Fig. 13(b), overtraining of BDT2 does not occur.



(a) ROC (b) Overtraining

Figure 13: Result of the BDT2 training. (a) The discriminating power of the BDT by plotting signal efficiency vs. background rejection. (b) The superposition of the BDT response of the training (dots) and test (bin) samples of signal (blue) and background (red). Both BDT distributions agree well, showing that there is no overtraining.

The efficiency of the BDT2 response cut is re-evaluated using the smeared B_c^+ simulated sample as signal after passing the B_s^0 selection and mass window, and requiring a match to the generated decay. The background sample used is again the upper side band in the data after all selection steps, but including all candidates per event. The resulting efficiency of events passing a BDT2 cut is shown in Fig. 14. In order to select an optimal working point assuming a low estimated signal to background ratio, we use the Punzi function [33],

$$f_{\text{Punzi}}(t) = \frac{\epsilon_{Sig}(t)}{a/2 + \sqrt{B(t)}}, \tag{6}$$

where $\epsilon(t)$ is the signal efficiency, $B(t)$ is the number of expected background events, a is the desired significance, and t is the BDT2 response. We use an estimate of $B(t = -1) \approx 20000$. The Punzi function with $a = 3$ optimizes the cut for a 3-sigma significant result, independent of the signal yield. Using the estimated 1 in 500 B_c^+ originating from a B_c^+ decay from Sec. 3 we estimate about 40 (100) signal events in the 2011 (2012) data after B_s^0 selection.

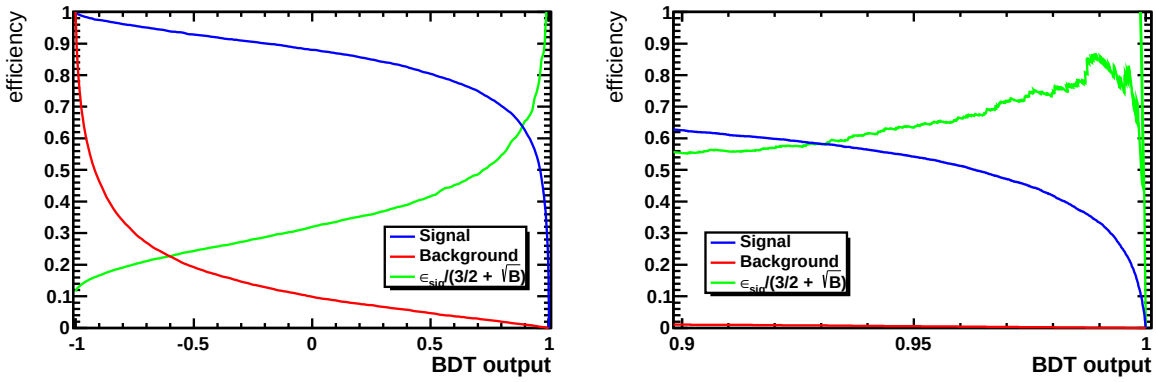


Figure 14: BDT2 efficiency curves after applying the B_s^0 selection and mass cut.

The efficiency as a function of the BDT2 cut is shown in Fig. 14, and is listed in Tab. 5, where the efficiency uncertainties are estimated using Poisson statistics and the differential fractional-square addition error propagation technique[10], which are slightly larger than the binomial errors.

The working point is chosen at BDT2 > 0.97, such that the figure of merit in Eq. 6 is large while still including a significant number of events. (For comparison, note that 90% of the $B_s^0 \to \mu^+\mu^-$ [34] signal sensitivity originates from 30% of the events in the purest bins.)

Table 5: Selection efficiencies of BDT2 working points after applying the B_s^0 selection and B_s^0 mass window cut.

BDT2 Cut	ϵ_{Sig}	ϵ_{Bkg}
$>$ -1.0	100%	100%
> 0.5	$(80.4 \pm 1.1)\%$	$(4.71 \pm 0.15)\%$
> 0.9	$(62.5 \pm 1.0)\%$	$(1.04 \pm 0.07)\%$
> 0.95	$(54.2 \pm 0.9)\%$	$(0.55 \pm 0.05)\%$
$>$ **0.97**	$(\mathbf{47.2 \pm 0.9})\%$	$(\mathbf{0.30 \pm 0.04})\%$
> 0.99	$(33.3 \pm 0.7)\%$	$(0.08 \pm 0.02)\%$

In order to compare different methods for the optimisation of the BDT2 working point, Fig. 15 shows the normalized curves for Eq. 6 with $a = 3$ and $a = 5$, as well as for the figure-of-merit $S/\sqrt{S+B}$, which maximizes for the highest signal significance. All three figures-of-merit are large in the region $0.95 <$ BDT2 < 0.99.

[10] $s_f = \sqrt{(\frac{\partial f}{\partial x})^2 s_x^2 + (\frac{\partial f}{\partial y})^2 s_y^2 + ...}$

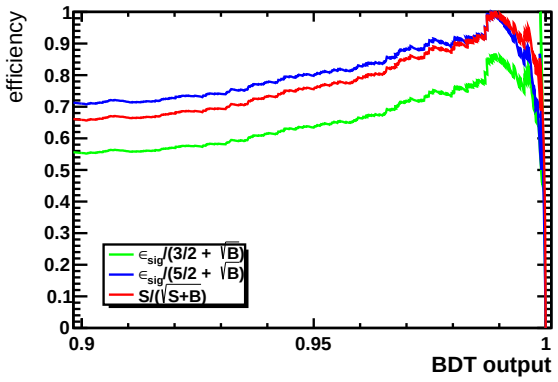


Figure 15: Various methods of estimating the optimal working point. Shown are Punzi curves optimized for a significance of 3 and 5 sigma, as well as the curve that maximises $S/\sqrt{S+B}$.

6.2 Candidates in data

After the B_s^0 selection, a mass window on the $B_s^0\pi^+$ candidates of $6000 < \mathrm{M}(B_s^0\pi^+) < 7000$ MeV, as well as a loose PID cut on the bachelor pion of $\mathrm{DLL}(K-\pi) < 5$ is applied. A list of the yields of all cuts made in the data is shown in Tab. 6.

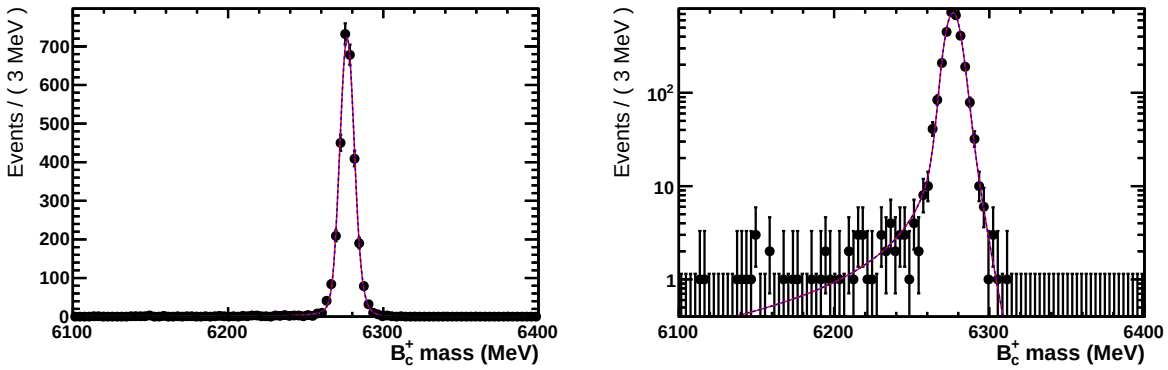
Table 6: Number of selected $B_s^0\pi^+$ candidates in the data after subsequent selection criteria.

	2011		2012	
Cut	Events	(candidates/event)	Events	(candidates/event)
Trig/reco/stripping	984k	(15.8)	2888k	(19.4)
D_s^- mass window	742k	(15.4)	2155k	(18.9)
D_s^- PID cuts	352k	(14.5)	1017k	(17.7)
BDT1 > 0.50	109k	(9.9)	271k	(10.3)
B_s^0 mass window	26.3k	(9.7)	63.2k	(10.0)
$\pi^+_{B_c^+}$ PID cut	26.1k	(8.6)	63.1k	(8.9)
B_c^+ mass range	23.8k	(3.3)	57.5k	(3.7)
BDT2 > 0.9	552	(1.09)	1788	(1.08)
BDT2 > 0.95	330	(1.06)	948	(1.06)
BDT2 > 0.97	215	(1.04)	621	(1.05)
BDT2 > 0.99	85	(1.00)	216	(1.02)

6.3 Mass Fit

According to Ref. [35] it is preferred to treat multiple candidates by including all candidates, because of a low expected signal yield and the lack of a correct definition of the 'best'

candidate. After unblinding, the B_c^+ invariant mass will be fitted using an unbinned maximum likelihood fit. The shape used for the signal fit is a double Crystal Ball (DCB) with transition points and slopes determined from simulation (Fig. 16, Tab. 7), and fixed in the final fit. The B_s^0 mass is fixed to 5366.8 MeV, resulting in a small width due to the small Q-value in the decay $B_c^+ \to B_s^0\pi^+$, and the mean and width are left free in the final fit, in order not to be affected by data-simulation discrepancies. The simulation-data correspondence is further explored in Sec. 8.2. The signal yield will be left free.



(a) Linear scale (b) Log scale

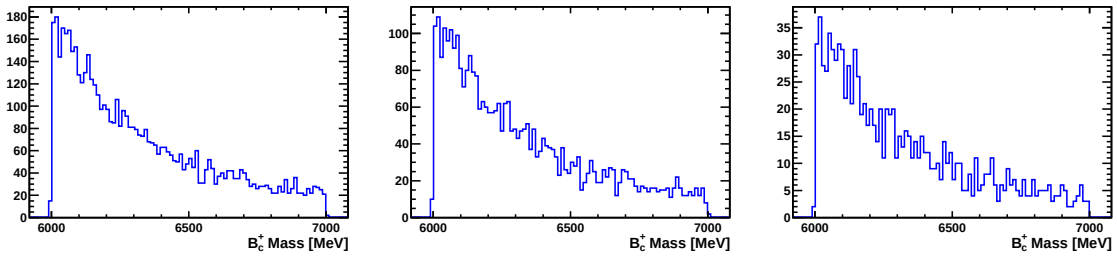
Figure 16: The signal shape is obtained from a DCB fit to the mass distribution of simulated B_c^+ events. The reconstructed B_c^+ candidates are required to match the true generated B_c^+ decays and pass the cuts BDT1 > 0.50 and BDT2 > 0.97.

Table 7: Fit parameter values for the signal fits to simulated B_c^+ events in Fig. 16.

Parameter	Fit Results
mean	6276.7 ± 0.1
width	4.70 ± 0.07
α_L	1.95 ± 0.03
α_R	-1.26 ± 0.06
n_L	1.28 ± 0.12
n_R	19.7 ± 0.9
N_{sig}	2999 ± 55

The shape for the combinatorial background is chosen as an exponential function, motivated by the shape of the prompt B_s^0 MC after full B_c^+ selection (Fig. 17). The exponent is loosely constrained from a fit to the sidebands, as is done in Sec. 7.3. Partially reconstructed backgrounds are expected to be largely rejected due to the lower limit on the

mass, $M(B_s^0\pi^+) > 6000\,\text{MeV}$, and possible remaining partially reconstructed backgrounds in the mass region below the B_c^+ mass are modelled by a gaussian.



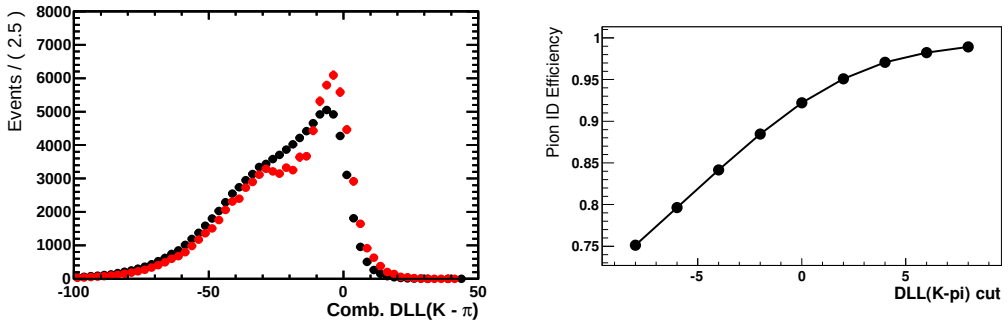
(a) BDT2 > 0.95 (b) BDT2 > 0.97 (c) BDT2 > 0.99

Figure 17: Invariant mass distribution of $(B_s^0\pi^+)$ candidates in the simulated B_s^0 sample, representing the combinatorial background, for various values of the BDT2 cut.

6.4 PID Efficiency

In measuring the ratio of efficiency corrected event yields (see Eq. 1), the efficiency of selecting $B_c^+ \to B_s^0\pi^+$ from B_s^0 candidates needs to be evaluated. In the B_c^+ selection a loose requirement on the particle identification of $\text{PID}(K-\pi) < 5$ is made on the bachelor pion, whose efficiency cannot be determined from MC due to the known discrepancy between the PID distributions in MC and data. In order to estimate this efficiency a calibration sample of D^* decays is used, which are purely selected using kinematic variables, to find the PID performance from data. This calibration sample is then reweighed to match the p_T and η distributions of the bachelor pion in the B_c^+ MC sample in order to obtain the efficiency of a certain PID cut.

The RICH performance is in principle a function of all topological parameters in the event. Commonly used parameters used to bin the performance are the (transverse) momentum, pseudorapidity and occupancy (number of tracks in the event). We choose to use 3 bins in the number of tracks from 0 to 400, 7 bins in momentum divided between 50 and 10000 MeV and 4 bins in pseudorapidity between 1.5 and 5.5. The result of the reweighing is shown in Fig. 18(a).



(a) (b)

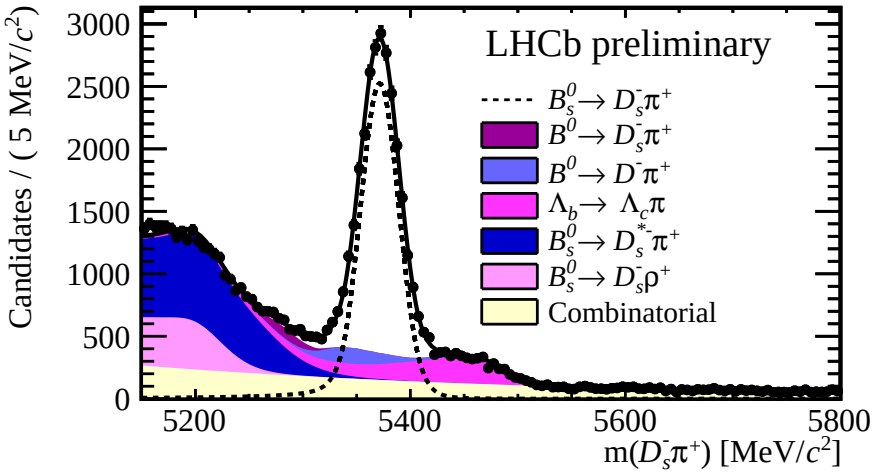
Figure 18: (a) The PID DLL($K-\pi$) distribution of pions in the D^* calibration sample before (black) and after (red) reweighing to the kinematic distributions of the bachelor pion in the simulated B_c^+ sample. (b) The pion selection efficiency as a function of DLL($K-\pi$) value, for pions with momentum and rapidity distributions corresponding to bachelor pions from $B_c^+ \to B_s^0\pi^+$ decays. A cut of DLL($K-\pi$) < 5 is applied to the bachelor pion.

After the reweighing of the PID distribution, it can be integrated over all bins up to a certain PID DLL($K-\pi$) value, resulting in the efficiency curve shown in Fig. 18(b). The total efficiency of the B_c^+ selection is discussed in Sec. 7.

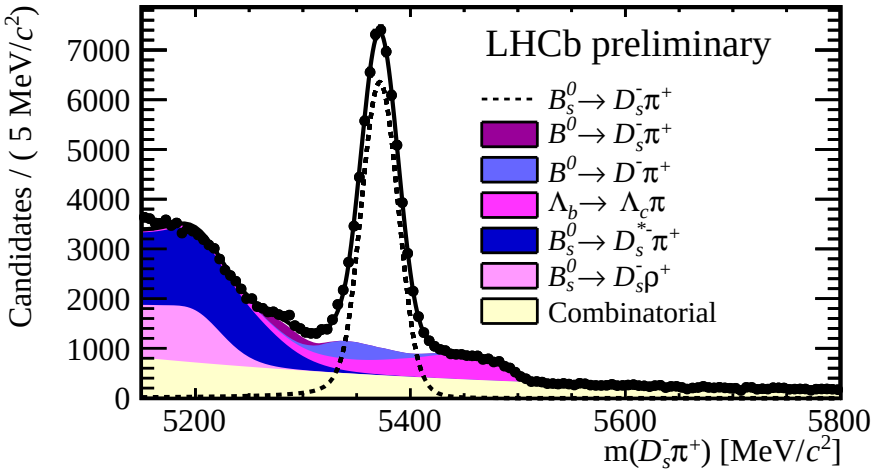
7 Event yields

7.1 B_s^0 Mass Fit

The result of the maximum likelihood fit to the $D_s^-\pi^+$ candidates, as described in Sec. 5.3, is shown in Fig. 19. The resulting fit parameters are listed in Tab. 8.



(a) 2011



(b) 2012

Figure 19: B_s^0 mass fit after PID and D_s^- mass cuts, for BDT1 > 0.50, of (a) the 2011 data and (b) the 2012 data, corresponding to 1 fb^{-1} and 2 fb^{-1}, respectively.

The result of the B_s^0 mass fit yields 23644 ± 298 $B_s^0 \to D_s^-\pi^+$ events in the 2011 data, and 57497 ± 494 $B_s^0 \to D_s^-\pi^+$ events in the 2012 data. The yield in 2012 is slightly higher than a factor 2 compared to the yield in 2011, as expected from the relative

integrated luminosities and the higher $b\bar{b}$ production cross section at $\sqrt{s} = 8\,\text{TeV}$. When performing the B_c^+ selection a B_s^0 mass window of (5371 ± 36) MeV is applied. Sources of background resulting from the B_s^0 selection are the residual combinatorial background and the misidentified backgrounds under the B_s^0 mass peak, whose yields are shown in Tab. 9.

Table 8: Fit results for $B_s^0 \to D_s^- \pi^+$, at BDT1 > 0.50.

	2011		2012	
	Magnet Up	Magnet Down	Magnet Up	Magnet Down
N_{Sig}	9701 ± 132	13943 ± 166	28582 ± 245	28915 ± 249
$N_{combinatorial}$	6455 ± 413	10404 ± 590	26186 ± 875	25141 ± 878
a_1	-0.0023 ± 0.0003	-0.0027 ± 0.0002	-0.0026 ± 0.0001	-0.0025 ± 0.0001
$N_{B^0 \to D\pi}$	788 ± 59	1198 ± 82	2634 ± 172	2949 ± 171
$N_{part-reco}$	9416 ± 277	12723 ± 397	26376 ± 583	27728 ± 582
Frac $\frac{B_s^0 \to D\rho}{B_s^0 \to D^* \pi}$	0.285 ± 0.028	0.297 ± 0.026	0.342 ± 0.019	0.315 ± 0.018
$N_{\Lambda_b^0 \to \Lambda_c^+ \pi}$	2852 ± 152	3889 ± 190	8110 ± 285	8245 ± 286
Common				
$mean_{Sig}$(MeV)	5371.5 ± 0.2		5371.5 ± 0.1	
σ_{Sig}(MeV)	17.9 ± 0.2		17.3 ± 0.1	
Fixed				
$Fract_{Sig-CBs}$	0.50		0.50	
α_1	1.71		1.71	
α_2	−1.87		−1.87	
n_1	1.38		1.38	
n_2	7.10		7.10	
$mean_{B^0 \to D_s^+ \pi}$	5283.0		5283.0	

Table 9: Yields of various contributions in Fig. 19 in the B_s^0 mass window 5371±36 MeV, and the percentage of the total yield in this mass window.

	2011		2012	
Source	Yield	(% of window)	Yield	(% of window)
Signal	21602	77.4%	52912	75.5%
Comb	2179	7.8%	6640	9.5%
Part.Reco	148	0.5%	349	0.5%
$B^0 \to D^- \pi^+$	1212	4.3%	3405	4.9%
$\overline{\Lambda}_b^0 \to \overline{\Lambda}_c^- \pi^+$	1779	6.4%	4322	6.2%

During 2011 approximately 0.4 (0.6) fb^{-1} was recorded with magnet polarity up (down), which is reflected in the relative sample size of 'Magnet Up' and 'Magnet Down' data. The event yields for both magnet polarities are in good agreement.

7.2 $\epsilon(B_c^+ \to B_s^0\pi^+)$

The efficiency of the B_c^+ selection with respect to the B_s^0 selection is obtained as follows. The B_s^0 mass window efficiency is obtained from simulated B_c^+ events that are matched to the generated event tree. The PID cut efficiency is obtained with data driven techniques described in Sec. 6.4 with the error assigned later as a systematic uncertainty (Sec. 8). The BDT2 efficiency is obtained from Fig. 14 and Tab. 5. The efficiencies are summarized in Tab. 10.

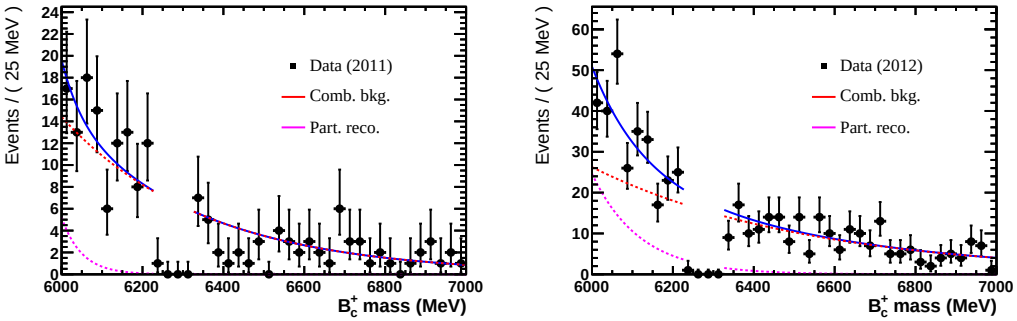
Due to the extra track in the final state of the $B_c^+ \to B_s^0\pi^+$ decay with respect to the normalisation channel, a detector acceptance effect has to be taken into account. For this, the fraction of generator efficiencies is used, of the prompt B_s^0 and $B_c^+ \to B_s^0\pi^+$ simulated samples, resulting in $13.18/16.23 = 0.812$ with negligible statistical error.

Table 10: Signal event efficiencies of the B_c^+ selections. The same efficiency is assumed in 2012, as obtained from 2011 simulation.

Cut	Rel. Efficiency
B_s^0 mass window	$(91.4 \pm 0.6)\%$
$\pi_{B_c^+}$ PID(K-π) < 5	97.8%
BDT2 cut	$(47.2 \pm 0.9)\%$
Generator	81.2%
Total	$(34.3 \pm 0.7)\%$

7.3 Procedure of limit setting

In order to define the limit setting (in case of no signal) before unblinding, we explore the sidebands in the data after all B_s^0 and B_c^+ selection steps in Fig. 20 for various BDT2 working points. Due to the blinding, the gaussian fit to the partially reconstructed contribution is expected to behave incorrectly, and instead a second exponent represents the partially reconstructed background in the sideband fit. The results of the fit to the sidebands are listed in Tab. 11.



(a) 2011 (b) 2012

Figure 20: Sidebands in the data, for BDT2 > 0.97.

Table 11: Fit parameters of the exponential fit to the B_c^+ mass sidebands using BDT2 > 0.97, shown in Fig. 20. The quoted values for the yield correspond to the total fitted number of combinatorial background events in the range [6000:7000] MeV/c, excluding the blinded region.

	2011	2012
Expo(comb. bkg)	-0.0028 ± 0.0004	-0.0019 ± 0.0005
N(comb. bkg)	195.2 ± 20.7	475.0 ± 95.2
Expo(part. reco.)	-0.02 ± 0.05	-0.0085 ± 0.0043
N(part. reco.)	10.5 ± 15.1	116.5 ± 92.0

The determination of the sensitivity to a certain exclusion limit is done by using CLs as a significance estimator [36] (i.e. p-value with respect to background-only hypothesis). The B_c^+ mass sideband fits are used to estimate the number of background events in the signal region. In the case that no signal events are observed, the exclusion limits correspond to the blue line in Fig. 21 using 2011 data and BDT2 > 0.95. If the estimated number of events - using $\mathcal{B}(B_c^+ \to B_s^0\pi^+) = 10\%$ as in Sec. 2.2 and the efficiencies from Tab. 10 - are observed, the red line in Fig. 21 is obtained.

The number of expected signal events is determined as follows. Using the measured values in Sec. 2.2 including the latest calculation on $\mathcal{B}(B_c^+ \to J/\psi\pi^+)$ we obtain

$$\frac{f_c}{f_s} = (0.93 \pm 0.24)\%. \tag{7}$$

Interpreting the range of theoretical predictions for the $B_c^+ \to B_s^0\pi^+$ branching ratio in Sec. 2.2 as $\mathcal{B}(B_c^+ \to B_s^0\pi^+) = (9.5 \pm 7.0)\%$ we obtain

$$\frac{f_c}{f_s} \times \mathcal{B}(B_c^+ \to B_s^0\pi^+) = (0.088 \pm 0.069)\%.$$

Using the observed B_s^0 yields and the efficiencies from the previous subsections, we obtain an expectation of the B_c^+ yield,

$$\begin{aligned} N_{\text{exp}}(B_c^+) &= 7.1 \pm 5.6 \quad (2011), \\ N_{\text{exp}}(B_c^+) &= 17.4 \pm 13.6 \quad (2012), \end{aligned} \tag{8}$$

for BDT2 > 0.97 .

Changing the observed B_c^+ yield into a branching ratio is done by the following relation,

$$\mathcal{B}(B_c^+ \to B_s^0\pi^+) = \frac{N(B_c^+)}{\frac{f_c}{f_s} \times N(B_s^0) \times \epsilon(B_c^+ \to B_s^0\pi^+)} = N(B_c^+)/X,$$

where $X = 92.8 \pm 24.1$ for 2011, and $X = 240.6 \pm 62.3$ for 2012 data.

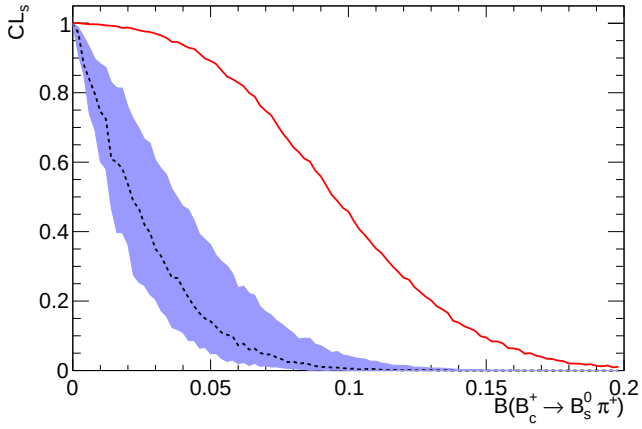
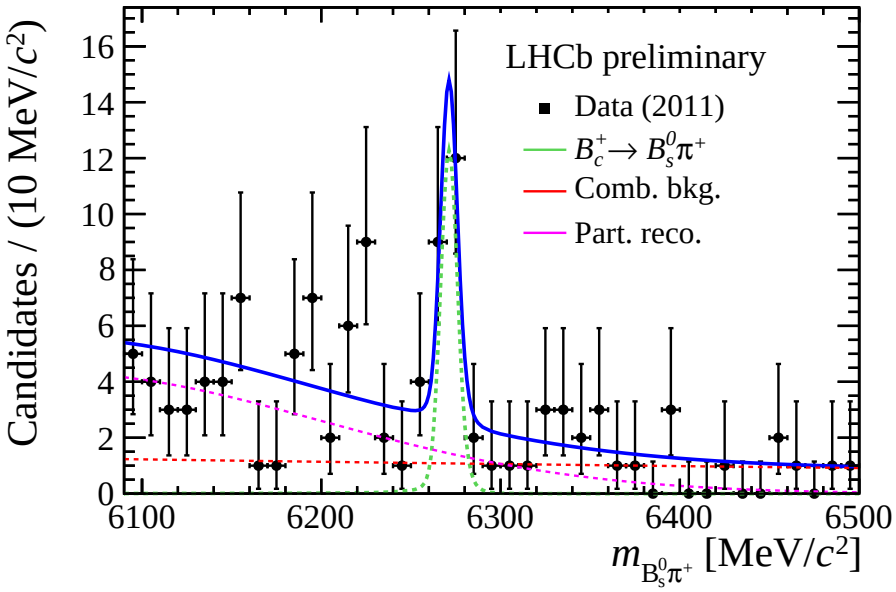


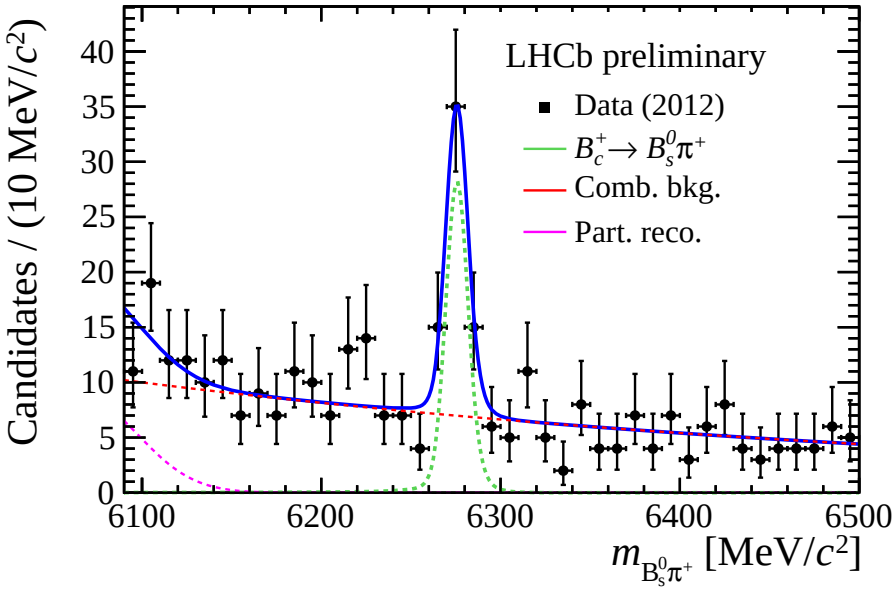
Figure 21: CLs for 2011+2012 data with BDT2 > 0.97. The blue dashed line are the CLs expected in case of no signal observation, with blue zone as $\pm 1\sigma$. The red line would be the observed exclusion curve, assuming $\mathcal{B}(B_c^+ \to B_s^0\pi^+) = 9.5\%$.

7.4 B_c^+ Mass Fit

The result of the unblinded fit to the B_c^+ mass is shown in Fig. 33. The resulting fit parameters are listed in Tab. 12.



(a) 2011



(b) 2012

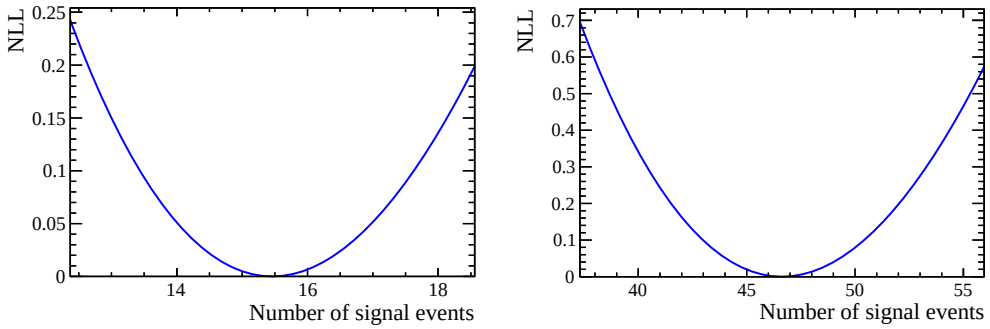
Figure 22: B_c^+ mass fit to the 2011 data (a) and 2012 data (b), using BDT2 > 0.97. A plot of the full fit range is available in Appendix B.

The resulting yields fall well within the wide range of predictions of Eq. 8. In order to estimate the significance of the result, the p-value with respect to the null-hypothesis (i.e. no signal) is determined using log-likelihood scans, resulting in a 2.1σ significance

for 2011 data and 3.7σ significance for 2012 data. The log-likelihood scans as a function of the signal yield are shown in Fig. 23. For illustration purposes, the fit using 2011 and 2012 data combined is shown in Fig. 34.

Table 12: Results of the B_c^+ mass fits, using BDT2 > 0.97, in Fig. 33. In addition, the fit results of the 2011+2012 data in Fig. 34 are shown.

	2011	2012	2011+2012
Significance (σ)	0.70	6.07	6.58
N($B_c^+ \to B_s^0\pi^+$)	15.5 ± 5.0	46.3 ± 8.9	62.7 ± 10.5
Mean (MeV)	6271.3 ± 1.8	6275.7 ± 1.4	6274.3 ± 1.1
Width (MeV)	4.7 ± 1.5	6.2 ± 1.2	6.1 ± 1.0
Expo(Comb. bkg)	-0.0007 ± 0.0009	-0.0021 ± 0.0002	-0.0017 ± 0.0004
N(Comb. bkg)	93.6 ± 26.5	526 ± 35	600 ± 68
Mean(Part. reco.)	6049 ± 64	6067 ± 9	6055 ± 38
Width(Part. reco.)	150 ± 41	29 ± 11	107 ± 40
N(Part. reco.)	102 ± 27	61 ± 27	182 ± 66



(a) 2011 (b) 2012

Figure 23: Log-likelihood of the B_c^+ mass fit as a function of the signal yield, for BDT2 > 0.97.

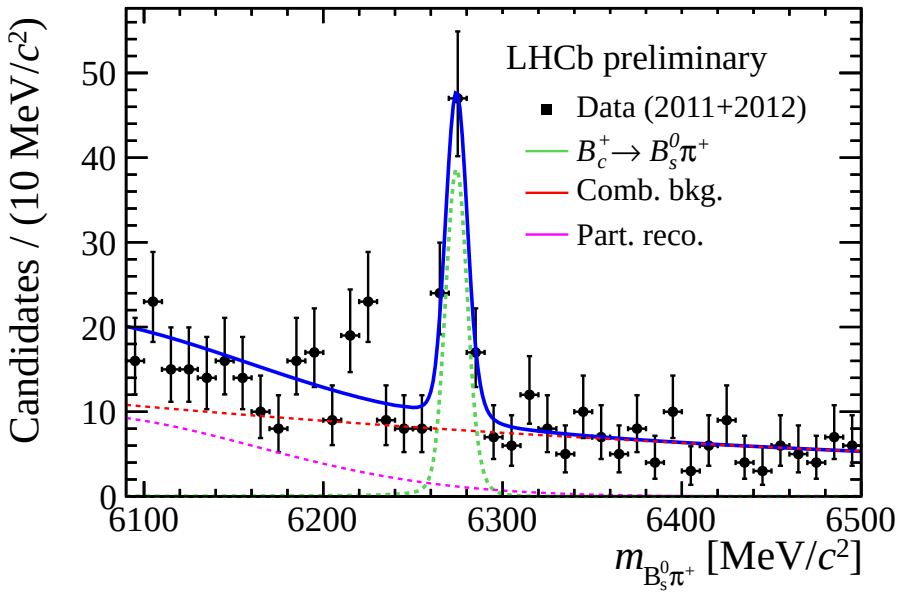


Figure 24: B_c^+ mass fit to the 2011 + 2012 data, using BDT2 > 0.97. A plot of the full fit range is available in Appendix B.

8 Systematic Uncertainties

This section presents the estimation of various sources of systematic uncertainties.

8.1 Cancellation of $\epsilon(B_s^0)$

The systematic uncertainty assigned to the difference in efficiency of selecting prompt B_s^0 or B_s^0 from B_c^+ decays, which cancels in Eq. 1, is estimated using simulated B_c^+ and B_s^0 samples, after passing the stripping and offline selection cuts listed in Tab. 1. The PID cut efficiency in the B_s^0 selection is not expected to differ significantly between the two samples as the momentum and pseudorapidity distributions of the final state particles is very similar between prompt B_s^0 decays and B_s^0-from-B_c^+ decays. The efficiencies of the different selection steps in the B_s^0 selection are summarized in Tab. 13.

Table 13: Comparison of the subsequent B_s^0 selection efficiencies between B_s^0 decays from B_c^+ decays and prompt B_s^0 decays, estimated using simulated events.

Cut	B_s^0 from B_c^+ eff.	Prompt B_s^0 eff.
Trigger/Reco/Stripping	$(3.43 \pm 0.02)\%$	$(3.339 \pm 0.005)\%$
$(1944 < M(D_s^-) < 1990)\,\mathrm{MeV}$	$(98.5 \pm 0.5)\%$	$(98.6 \pm 0.4)\%$
BDT1 > 0.50	$(84.8 \pm 0.5)\%$	$(86.5 \pm 0.4)\%$
Total	$(2.87 \pm 0.03)\%$	$(2.85 \pm 0.02)\%$

The difference in selection efficiency is within the statistical uncertainty. The statistical uncertainty on the simulated samples of 1.4% is assigned as a systematic uncertainty.

8.1.1 B_s^0 fit

In order to study the effect of the B_s^0 fit shape on the yield, the parameters that are fixed from simulation are varied by 10%, which is the difference between the tail parameters from simulation and the tail parameters obtained when fitting for $B^0 \to D^-\pi^+$ in data. In addition, a fit is performed with the exponential constant in the combinatorial background fixed to half its fitted value, in order to estimate the error caused by potentially missed sources of partially reconstructed backgrounds. The results are listed in Tab. 14

Table 14: The yields of fit pdf adjustments on the B_s^0 yield in the 2011 and 2012 data, and gain percentage with respect to the regular fit pdf yield.

	2011		2012	
Adjustment	MU yield (%)	MD yield (%)	MU yield (%)	MD yield (%)
$\alpha_1 = 1.53$	9874 (1.78)	14213 (1.94)	29200 (2.16)	29535 (2.14)
$\alpha_1 = 1.88$	9567 (-1.38)	13734 (-1.50)	28123 (-1.61)	28452 (-1.60)
$\alpha_2 = -1.68$	9743 (0.43)	14001 (0.42)	28688 (0.37)	29026 (0.38)
$\alpha_2 = -2.06$	9673 (-0.29)	13903 (-0.29)	28507 (-0.26)	28837 (-0.27)
$n_1 = 1.24$	9752 (0.53)	14019 (0.55)	28746 (0.57)	29080 (0.57)
$n_1 = 1.52$	9660 (-0.42)	13881 (-0.44)	28448 (-0.47)	28780 (-0.47)
$n_2 = 4.0$	9748 (0.48)	14009 (0.47)	28707 (0.44)	29044 (0.45)
$n_2 = 11.0$	9685 (-0.16)	13919 (-0.17)	28536 (-0.16)	28869 (-0.16)
$a_1 = -0.0012$	9855 (1.59)	14239 (2.12)	29089 (1.77)	29425 (1.76)

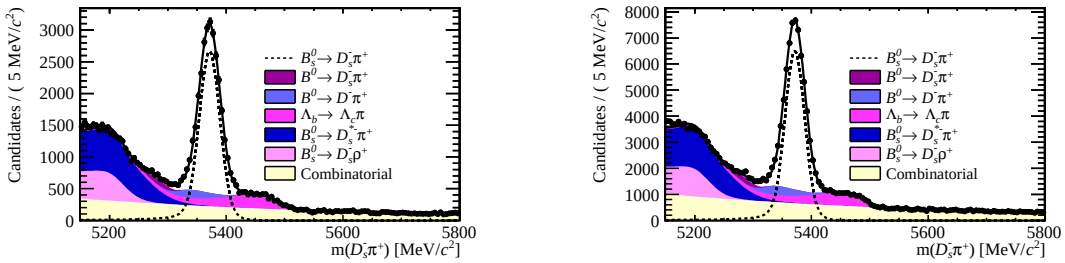
Since this error does not cancel in the ratio, a conservative systematic uncertainty of 3% is assigned to the choice of B_s^0 signal shape.

8.2 Event selection

8.2.1 BDT1

In order to estimate the error on the BDT1 efficiency, the B_s^0 mass fit is redone for a choice of BDT1 > 0.0 (see Fig. 25), for which the signal efficiency should be 95% instead of 87% (see Fig. 9). An estimate of the signal yield can be done by correcting the observed yield for BDT1 > 0.5 with the expected efficiencies, which results in 25745 ± 285 (62700 ± 475) expected number of events for 2011 (2012) data. A discrepancy between the calculated and observed number of events is a measure for the uncertainty on the value of the BDT1 efficiency, as estimated from simulation. Since simulated events are used as the signal sample, and data is used as the background sample in training BDT1, it is an estimate for the error on the correspondence between simulated events and data. The result of the fit is shown in Tab. 15.

The yield of B_s^0 candidates in the fit with BDT1 > 0.00 cut is 24863 ± 309 (57973 ± 501) for 2011 (2012) data, which is a discrepancy of 3.4% (7.5%) compared to the yield expected using the BDT1 efficiency curve. Conservatively assuming that the error does not fully cancel in the ratio, half the difference is assigned as a systematic uncertainty. The large error for 2012, however, indicates a significant difference between 2011 simulated events and 2012 data, and 2012 simulated events are required to train BDT1 and BDT2 for the 2012 data, to reduce this uncertainty.



(a) 2011 (b) 2012

Figure 25: The B_s^0 mass fit of the 2011 and 2012 data, after PID and D_s^- mass cuts, for BDT1 > 0.00.

Table 15: Fit results for $B_s^0 \to D_s^- \pi^+$, at BDT1 > 0.00.

	2011		2012	
	Magnet Up	Magnet Down	Magnet Up	Magnet Down
N_{Sig}	10205 ± 137	14658 ± 172	28672 ± 249	29301 ± 252
$N_{combinatorial}$	9834 ± 448	15292 ± 589	38241 ± 905	36346 ± 880
a_1	-0.0019 ± 0.0002	-0.0020 ± 0.0002	-0.0021 ± 0.0001	-0.0018 ± 0.0001
$N_{B^0 \to D\pi}$	780 ± 60	1199 ± 82	2580 ± 174	2947 ± 172
$N_{part-reco}$	9608 ± 288	13310 ± 378	26006 ± 576	27939 ± 553
Frac $\frac{B_s^0 \to D\rho}{B_s^0 \to D^* \pi}$	0.316 ± 0.029	0.323 ± 0.026	0.347 ± 0.021	0.316 ± 0.019
$N_{\Lambda_b^0 \to \Lambda_c^+ \pi}$	2823 ± 167	4209 ± 209	7686 ± 312	8220 ± 313
Common				
$mean_{Sig}$(MeV)	5371.5 ± 0.2		5371.6 ± 0.1	
σ_{Sig}(MeV)	17.8 ± 0.2		17.1 ± 0.1	
Fixed				
$Fract_{Sig-CBs}$	0.50		0.50	
α_1	1.71		1.71	
α_2	-1.87		-1.87	
n_1	1.38		1.38	
n_2	7.10		7.10	
$mean_{B^0 \to D_s^+ \pi}$	5283.0		5283.0	

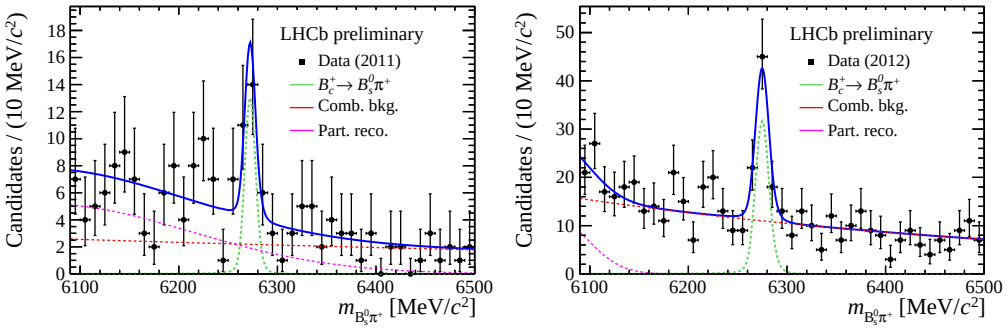
8.2.2 BDT2

In order to estimate the systematic uncertainty on the BDT2 efficiency, the same exercise as in Sec. 8.2.1 is repeated, using the observed yield and the BDT2 efficiency curve (Fig. 14) to find the expected B_c^+ yield using a different BDT2 working point. The results are shown

in Fig. 26 and Tab. 16.

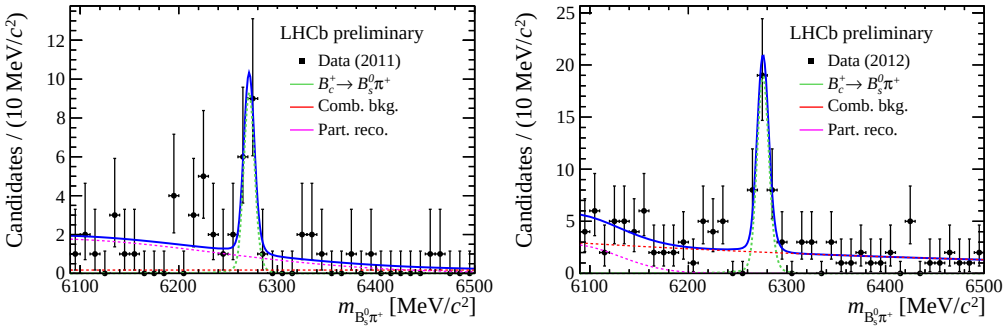
The discrepancies between the expected and observed yields are not assigned as a systematic uncertainty. Instead, the simulation-data discrepancy is studied in more detail in the next subsection, and the values obtained here are used as a consistency check.

The discrepancy between 2012 data and 2011 simulation has been postulated in Sec. 8.2.1. In order to study this effect for BDT2 we compute the 2012/2011 ratio of the sideband yields for BDT2 > 0.97, resulting in $601/205 = 2.93$. The ratio of number of events in the B_s^0 mass window is $70082/27910 = 2.51$ The difference of 14% is assigned as a systematic uncertainty on the 2012 selection efficiency.



(a) 2011, BDT2 > 0.95

(b) 2012, BDT2 > 0.95



(c) 2011, BDT2 > 0.99

(d) 2012, BDT2 > 0.99

Figure 26: B_c^+ mass fit to the data, using various values for the BDT2 working point.

Table 16: Results of the B_c^+ mass fits in Fig. 26.

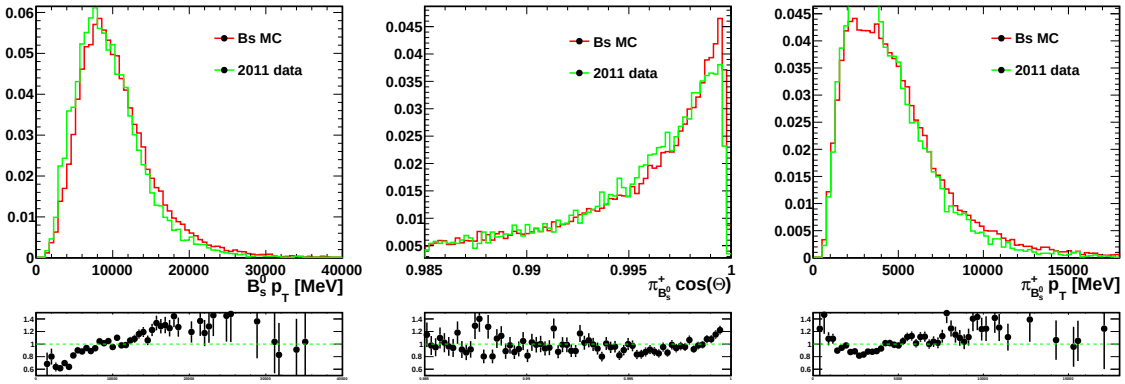
	2011		2012	
BDT2 cut	> 0.95	> 0.99	> 0.95	> 0.99
Significance (σ)	0	1.16	5.82	5.93
N($B_c^+ \to B_s^0\pi^+$)	19.0 ± 6.0	13.4 ± 4.3	57.5 ± 10.7	29.8 ± 6.3
Mean (MeV)	6272.4 ± 2.0	6271.4 ± 2.1	6274.9 ± 1.5	6275.7 ± 1.4
Width (MeV)	5.5 ± 1.7	5.4 ± 1.7	6.9 ± 1.3	5.9 ± 1.1
Expo(Comb. bkg)	-0.0009 ± 0.0007	-0.000 ± 0.001	-0.0019 ± 0.0002	-0.0020 ± 0.0005
N(Comb. bkg)	185.2 ± 42	17 ± 7	829 ± 43	153.9 ± 20.1
Mean(Part. reco.)	6065 ± 57	6064 ± 91	6064 ± 9	6082 ± 18
Width(Part. reco.)	147 ± 42	172 ± 56	34 ± 11	45 ± 20
N(Part. reco.)	127 ± 41	49 ± 9	97 ± 31	30 ± 15
Nexpected	17.8 ± 5.8	8.2 ± 4.5	53.2 ± 10.3	32.7 ± 6.3
Rel. diff.	6.7%	10.9%	8.2%	-8.8%

8.3 Data and simulation discrepancy in BDTs

In order to further investigate the source of the systematic error on the BDT efficiencies (i.e. the difference between simulation and data), the following is done.

8.3.1 Simulation and data plots

A relatively clean B_s^0 data sample can be obtained by applying the B_s^0 selection and choosing one random candidate per event, using a higher BDT1 cut of BDT1 > 0.7, a tight mass window of 5731 ± 30 MeV and applying a veto on the misidentified background with a width of 15 MeV. Using the 2011 data and selecting one random candidate per event, 13.2k events remain. The resulting distributions for the B_s^0 selection can then be compared to B_s^0 simulated events that match the generated events and have also passed the BDT1 > 0.7 cut (31.9k events). The distributions of the three most discriminating variables of the BDT1 training are shown in Fig. 27, while Appendix A shows an extended selection of variables. Data and simulation agree well.



(a) B_s^0 p_T (b) $\pi^+_{B_s^0}$ cos(Θ) (c) $\pi^+_{B_s^0}$ p_T

Figure 27: Data - simulation comparison for the input variables to the BDT1. Below each figure the ratio of the distributions from data and simulation are shown.

8.3.2 Track smearing

To estimate the uncertainty on the BDT2 efficiency, the variable with largest data-simulation discrepancy is varied, applying no track smearing in the B_c^+ simulation, and comparing the efficiency of the BDT2 cut. The track smearing influences the IP χ^2, the lifetimes and decay lengths, and the decay tree fitter χ^2 distributions of the simulation, which are the most discriminating variables of the BDT2 training. The results are shown in Tab 17. For comparison, the results from Tab. 5 are listed as well.

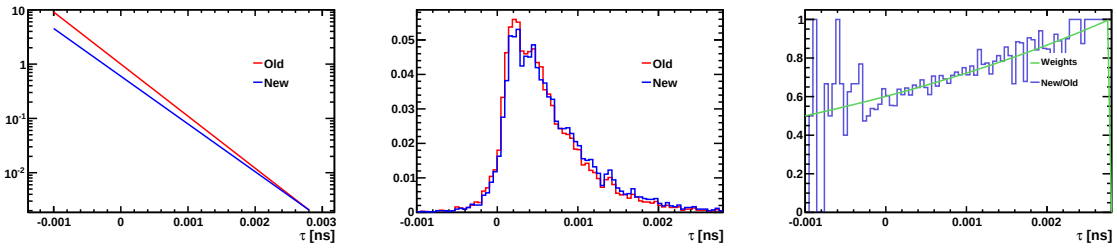
Table 17: The effect of a different smearing scale in the simulation on the BDT2 signal efficiency.

Cut	Scale = 0.0 ϵ_{Sig}	Scale = 1.0 ϵ_{Sig}	Relative diff.
BDT2 > -1.0	100%	100%	
BDT2 > 0.5	$(78.2 \pm 1.1)\%$	$(80.4 \pm 1.1)\%$	$(2.7 \pm 1.9)\%$
BDT2 > 0.9	$(60.1 \pm 1.0)\%$	$(62.5 \pm 1.0)\%$	$(3.8 \pm 2.3)\%$
BDT2 > 0.95	$(51.1 \pm 0.9)\%$	$(54.2 \pm 0.9)\%$	$(5.7 \pm 2.3)\%$
BDT2 > 0.97	$(44.5 \pm 0.9)\%$	$(47.2 \pm 0.9)\%$	$(5.7 \pm 2.7)\%$
BDT2 > 0.99	$(29.7 \pm 0.7)\%$	$(33.3 \pm 0.7)\%$	$(10.8 \pm 3.0\%$

Not smearing the track momenta in simulation, needed to better describe the data, changes the selection efficiency by a relative 16%. A variation of the smearing scale to 0.9 and 1.1 is deemed more realistic, and hence 20% of the effect is taken as a systematic uncertainty, 1.1%.

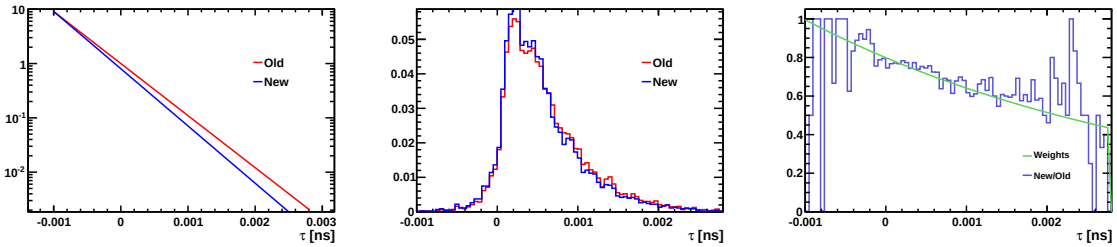
8.3.3 B_c^+ lifetime

The value for the B_c^+ lifetime is measured to be 0.453 ± 0.041 ps [9]. Since the uncertainty is 10%, which is large compared to the B_s^0 lifetime, it could be modelled incorrectly in the simulated samples used. To estimate the systematic uncertainty on the B_c^+ selection efficiency, due to the uncertainty on the B_c^+ lifetime, a different lifetime in the B_c^+ simulation is obtained by rejecting events as a function of the B_c^+ decay time. The simulated lifetime is changed in order to correspond with $0.453 + 0.041$ ps and $0.453 - 0.041$ ps, and the effect on the BDT2 efficiency - where lifetime-related variables are used in the selection - is shown in Tab. 18. The theoretical variation is normalized to always be below the old distribution in the used window (Fig. 28(a) and Fig. 29(a)), and the fraction of the two in 80 bins in τ is assigned as weights. The events in the data are then assigned a random value in the range $]0, 1]$, and discarded if the value exceeds the weight in the particular bin. The samples with the new lifetime distributions (see Fig. 28(b) and Fig. 29(b)) are then used to reproduce the efficiency curve as in Fig. 14.



(a) Theoretical curves (b) 2011 simulation (normalized) (c) Fraction

Figure 28: Lifetime distributions of the originally generated sample (red) and the lifetime up-shifted sample using $\tau = 0.453 + 0.041$ ps (blue) for (a) the theoretical curves, (b) the normalized data before and after discarding events, (c) the fraction of the new and old data and the weights.



(a) Theoretical curves (b) 2011 simulation (normalized) (c) Fraction

Figure 29: Lifetime distributions of the originally generated sample (red) and the lifetime down-shifted sample using $\tau = 0.453 - 0.041$ ps (blue) for (a) the theoretical curves, (b) the normalized data before and after discarding events, (c) the fraction of the new and old data and the weights.

Table 18: The result of the BDT2 efficiencies of the lifetime adjusted samples from Fig. 28 and Fig. 29.

BDT2 cut	$\epsilon(\tau = 0.453 - 0.041\,\text{ps})$	$\epsilon(\tau = 0.453\,\text{ps})$	$\epsilon(\tau = 0.453 + 0.041\,\text{ps})$
> 0.95	$(52.8 \pm 1.1)\%$	$(54.2 \pm 0.9)\%$	$(57.6 \pm 1.2)\%$
> 0.97	$(45.6 \pm 1.0)\%$	$(47.2 \pm 0.9)\%$	$(50.4 \pm 1.1)\%$
> 0.99	$(31.9 \pm 0.8)\%$	$(33.3 \pm 0.7)\%$	$(36.4 \pm 0.9)\%$

A systematic uncertainty of $+6.8\%(-3.5\%)$ due to the B_c^+ lifetime uncertainty is assigned to the efficiency.

8.4 Bachelor PID efficiency error

In order to get an estimate for the systematic error on the bachelor PID DLL$(K - \pi)$ cut efficiency, the procedure of reweighing the PID distribution in Sec. 6.4 is repeated, using simulated events of the D^* calibration decay as the calibration sample. The difference of the two reweighed PID distributions is shown in Fig. 30. At a cut value of $\text{PID}_{\text{bachelor}}(K - \pi) < 5$, the difference between the reweighed ID efficiencies is about 0.4% of the efficiency, which is assigned as a systematic uncertainty.

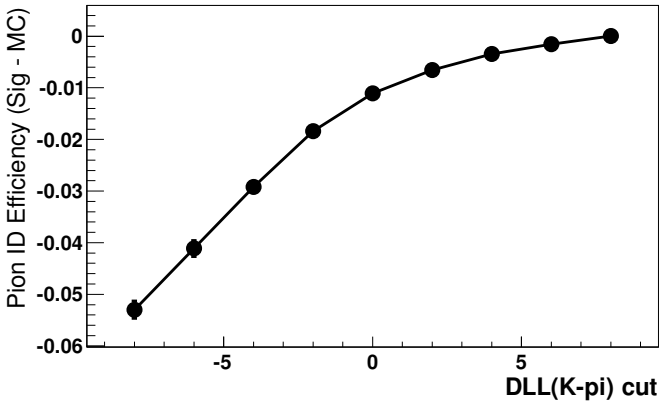


Figure 30: The pion ID efficiency difference between using real and simulated calibration samples.

8.5 Bachelor misidentification

The relative branching fraction of $B_c^+ \to B_s^0 K^+$ to $B_c^+ \to B_s^0 \pi^+$ is about $\frac{\Phi_{K^+}}{\Phi_{\pi^+}} \times |\frac{V_{us}}{V_{ud}}|^2 \approx 84.9\% \times 5.4\% = 4.6\%$[11], neglecting QCD effects. The efficiency of selecting $B_c^+ \to B_s^0 K^+$ after the PID DLL$(K-\pi) < 5$ cut is unknown due to a lack of $B_c^+ \to B_s^0 K^+$ simulated events, but is estimated to be about 5%. This results in a contribution from misidentified $B_c^+ \to B_s^0 K^+$ events of 0.2%, which is assigned as a systematic uncertainty.

8.6 B_c^+ selection effect on B_s^0 yield

In the first B_s^0 selection step (Sec. 5.1) we require $B_s^0\pi^+$ candidates with loose cuts, resulting in about 16 candidates per event in the 2011 data. In order to study the potential bias it has on the B_s^0 yield when comparing to a pure B_s^0 selection, this inclusive B_c^+ selection is compared to a B_s^0-only selection, applied to the simulated B_s^0 sample. The resulting yields are 35691 events with 1.0 candidates per event for the B_s^0 selection, and 35202 with 13.7 candidates per event for the inclusive B_c^+ selection. A correction of the 1.4% missed B_s^0 candidates is applied to the B_s^0 yield, and a systematic uncertainty of 1.0% is assigned to the B_s^0 yield.

8.7 Bachelor tracking efficiency

Since the normalisation channel has one final state track less, we need to include the tracking efficiency of the bachelor π^+ originating from the B_c^+ decay in the final result. Using the results from Ref. [37], we assign a 0.6% systematic uncertainty to this effect.

[11] $\Phi(A \to B + C)_{\text{(pseudo)scalar}} = \frac{1}{16\pi} \frac{\lambda^{1/2}(m_A^2, m_B^2, m_C^2)}{m_A^3}$, where $\lambda(x,y,z) = x^2 + y^2 + z^2 - 2xy - 2xz - 2yz$

8.8 B_c^+ fit

8.8.1 Varying the mean

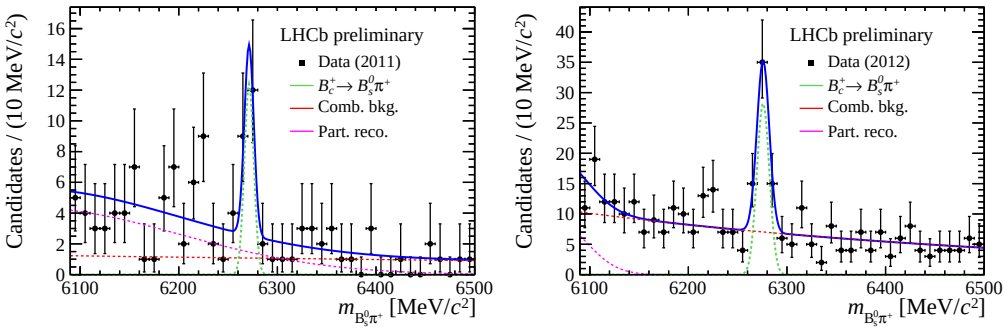
The effect of a variation of the mean of the B_c^+ signal shape, is studied by gaussian constraining the mean to the fitted value $\pm 1\sigma$. The results are shown in Tab. 19. A 1.5% (0.6%) systematic uncertainty due to the uncertainty of the mean is assigned to the B_c^+ yield for 2011 (2012) data.

Table 19: Results of the B_c^+ mass fits for BDT2 > 0.97, using a shifted value for the central value of the gaussian constraint on the mean.

	2011		2012	
$\text{Mean}_{\text{constr}}$ (MeV)	$+ 1\sigma$	$- 1\sigma$	$+ 1\sigma$	$- 1\sigma$
N($B_c^+ \to B_s^0\pi^+$)	15.2 ± 4.8	15.7 ± 4.9	46.6 ± 8.7	46.6 ± 8.6
Mean (MeV)	6272.2 ± 1.2	6270 ± 1.3	6276.4 ± 1.0	6275.0 ± 1.0
Width (MeV)	4.7 ± 1.0	4.9 ± 1.0	6.3 ± 0.8	6.2 ± 0.8
Expo(Comb. bkg)	-0.0007 ± 0.0009	-0.0007 ± 0.0009	-0.0020 ± 0.0002	-0.0020 ± 0.0002
N(Comb. bkg)	93 ± 27	94 ± 26	522 ± 32	522 ± 32
Mean(Part. reco.)	6050 ± 63	6048 ± 65	6067 ± 9	6067 ± 9
Width(Part. reco.)	150 ± 41	150 ± 41	31 ± 9	31 ± 9
N(Part. reco.)	102 ± 27	102 ± 27	65 ± 22	65 ± 22
Rel. yield incr.	-1.9%	1.3%	0.6%	0.6%

8.8.2 Gaussian signal shape

The signal fit is redone using a gaussian as signal model, in order to estimate an error on the signal shape. The results are shown in Fig. 31 and Tab. 20. The relative difference in signal yield is assigned as a systematic uncertainty.



(a) 2011, BDT2 > 0.97 (b) 2012, BDT2 > 0.97

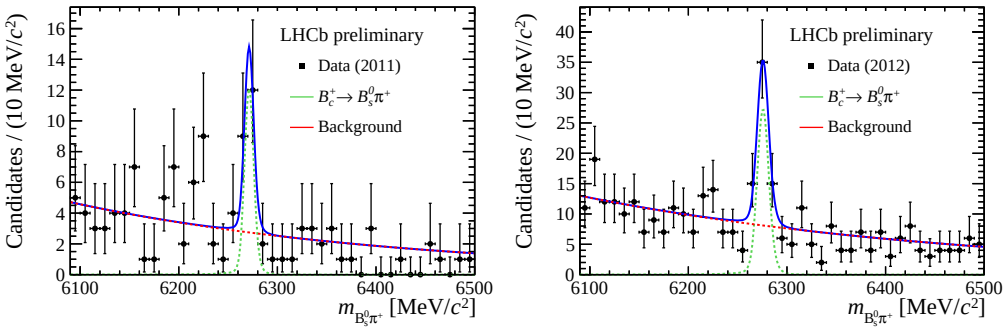
Figure 31: B_c^+ mass fit to the data, using a gaussian as signal model.

Table 20: Results of the B_c^+ mass fits in Fig. 31, using a gaussian as signal model, for BDT2 > 0.97.

	2011	2012
N($B_c^+ \to B_s^0\pi^+$)	14.7 ± 4.7	44.6 ± 8.4
Mean (MeV)	6271.3 ± 1.7	6275.8 ± 1.3
Width (MeV)	4.7 ± 1.2	6.3 ± 1.0
Expo(Comb. bkg)	-0.0007 ± 0.0009	-0.0021 ± 0.0002
N(Comb. bkg)	93 ± 27	526 ± 31
Mean(Part. reco.)	6051 ± 63	$6067pm9$
Width(Part. reco.)	150 ± 40	30 ± 9
N(Part. reco.)	103 ± 27	64 ± 21
Rel. yield incr.	-5.2%	-3.7%

8.8.3 Single exponential background

The background model is varied to a single exponential as a systematic test, in order to study the effect of not considering any partially reconstructed background. The exponential constants are gaussian constrained from a fit to the sidebands, and their yields are left free. The result is shown in Fig. 32 and Tab. 21. The relative difference in signal yield is assigned as a systematic uncertainty.



(a) 2011, BDT2 > 0.97 (b) 2012, BDT2 > 0.97

Figure 32: B_c^+ mass fit to the data, using a single exponential as background model.

Table 21: Results of the B_c^+ mass fits in Fig. 32, using a single exponential as background model, for BDT2 > 0.97.

	2011	2012
N($B_c^+ \to B_s^0\pi^+$)	14.9 ± 4.9	42.4 ± 8.6
Mean (MeV)	6271.3 ± 1.8	6275.7 ± 1.4
Width (MeV)	4.6 ± 1.4	5.9 ± 1.2
N(bkg)	196 ± 14	591 ± 25
Exponent	-0.0030 ± 0.0002	-0.0025 ± 0.0001
Rel. yield incr.	-3.9%	-8.4%

8.9 Summary

The contributions of various systematic uncertainties are shown in Tab. 22

Table 22: The various contributions of systematic uncertainties. The total systematic uncertainty is the quadratic sum of the individual contributions.

Source	2011	2012
$\epsilon(B_s^0)$	1.4%	1.4%
B_s^0 fit model	3.0%	3.0%
BDT1 cut (Data-Sim)	1.7%	3.8%
$B_s^0\pi^+$ preselection	1.0%	1.0%
Total N(B_s^0)	3.9%	5.1%
Track smearing	1.1%	1.1%
B_c^+ lifetime	$^{+6.8}_{-3.5}$	$^{+6.8}_{-3.5}$
PID efficiency	0.4%	0.4%
Tracking eff.	0.6%	0.6%
BDT2 sideband eff.	0.0%	14.0%
Total $\epsilon(B_c^+ \to B_s^0\pi^+)$	$^{+6.9}_{-3.7}$%	$^{+15.6}_{-14.5}$%
Bachelor misID	0.2%	0.2%
B_c^+ fit mean shift	1.5%	0.6%
B_c^+ fit gaussian	5.2%	3.7%
B_c^+ fit single exp.	3.9%	8.4%
Total N(B_c^+)	6.7%	9.2%

9 Results

The values and uncertainties for the B_s^0 yield, $B_c^+\pi^+$ selection efficiency and B_c^+ yield obtained in the previous sections are summarised in Tab. 23. The B_s^0 yield is corrected by +1.4% for the $B_s^0\pi^+$ selection effect discussed in Sec. 8.6. The systematic error on the efficiency is the quadratic sum of the statistical error from simulation in Tab. 10 and the systematic error from Tab. 22.

Table 23: Summary of the obtained values and relative uncertainties.

	2011			2012		
	Value	Stat. error	Syst. error	Value	Stat. error	Syst. error
$N(B_s^0)$	23975	1.3%	3.9%	58302	0.9%	5.1%
$\epsilon(B_c^+ \to B_s^0\pi^+)$	0.343	-	$^{+6.9}_{-3.8}$%	0.343	-	$^{+15.6}_{-14.5}$%
$N(B_c^+)$	15.5	32.3%	6.7%	46.3	19.2%	9.2%

Adding 2011 and 2012 data results in $N(B_s^0) = 82277 \pm 610$ (stat) $\pm$ 3117 (syst), $N(B_c^+) = 61.8 \pm 10.2$ (stat) $\pm$ 4.4 (syst), and $\epsilon(B_c^+ \to B_s^0\pi^+) = (34.3\ ^{+4.5}_{-3.9}$ (syst))%. The resulting value for the branching ratio using Eq. 1 is

$$\frac{f_c}{f_s} \times \mathcal{B}(B_c^+ \to B_s^0\pi^+) = (2.19\ \pm 0.36\ (\text{stat})\ ^{+0.34}_{-0.31}\ (\text{syst})) \times 10^{-3} \tag{9}$$

In addition, we can quote the result relative to $\mathcal{B}(B_c^+ \to J/\psi\pi^+)$ using Eq. 2:

$$\frac{\mathcal{B}(B_c^+ \to B_s^0\pi^+)}{\mathcal{B}(B_c^+ \to J/\psi\pi^+)} = 81.1\ \pm\ 13.4\ (\text{stat})\ ^{+12.4}_{-11.3}\ (\text{syst})\ \pm\ 15.9\ (\text{ext}), \tag{10}$$

where we used $f_u = f_d$, $\frac{\sigma(B_c^+)\times\mathcal{B}(B_c^+\to J/\psi\pi^+)}{\sigma(B^+)\times\mathcal{B}(B^+\to J/\psi K^+)} = (0.68 \pm 0.12)$ % [7], $\mathcal{B}(B^+ \to J/\psi K^+) = (1.016 \pm 0.033) \times 10^{-3}$ [9], and $f_s/f_d = 0.256 \pm 0.020$ [10], and where the third uncertainty is due to the measurements external to this analysis.

Using the recent theoretical value $\mathcal{B}(B_c^+ \to Jpsi\pi^+) = (0.291^{+0.015+0.040}_{-0.042-0.027})$% [11], we obtain an estimate for the branching ratio,

$$\mathcal{B}(B_c^+ \to B_s^0\pi^+) = (\ 23.6\ \pm\ 3.9\ (\text{stat})\ ^{+3.6}_{-3.3}\ (\text{syst})\ \pm\ 4.6\ (\text{ext})\ ^{+3.5}_{-4.1}\ (\text{theo})\ )\%,$$

where the fourth uncertainty is due to the theoretical uncertainty on the $B_c^+ \to J/\psi\pi^+$ branching ratio.

10 Conclusion

The $B_c^+ \to B_s^0\pi^+$ and $B_s^0 \to D_s^-\pi^+$ yields are determined from the data collected by LHCb in 2011 and 2012, corresponding to an integrated luminosity of $1\,\text{fb}^{-1}$ and $2\,\text{fb}^{-1}$, respectively. The observed signal significance of the decay $B_c^+ \to B_s^0\pi^+$ over the background-only hypothesis of 2011 and 2012 data combined, is 6.7σ, resulting in the first observation of the decay $B_c^+ \to B_s^0\pi^+$, which is the first observation of a weak B-to-B decay. This decay is the b-decay with the highest branching ratio observed thus far.

The efficiency of the B_c^+ selection steps are estimated using data-driven techniques for the mass window and PID performance, and use the efficiency curve of BDT2. Combined with the systematic studies done in Sec. 8, a value for the branching ratio of $\frac{f_c}{f_s} \times \mathcal{B}(B_c^+ \to B_s^0\pi^+) = (2.19 \pm 0.36\ (\text{stat})\ ^{+0.34}_{-0.31}\ (\text{syst})) \times 10^{-3}$ is obtained. This result is within the large range of theoretical estimates, and provides a measure for the amount of B_s^0 mesons originating from a B_c^+ decay, used in precision studies. The ratio of branching ratios is estimated as $\mathcal{B}(B_c^+ \to B_s^0\pi^+)/\mathcal{B}(B_c^+ \to J/\psi\pi^+) = 81.1 \pm 13.4\ (\text{stat})\ ^{+12.4}_{-11.3}\ (\text{syst}) \pm 15.9\ (\text{ext})$.
The observed B_c^+ yield is considered too small to extract a significant tagging performance, and the uncertainty is still too large to discriminate between various QCD models.
We have estimated $f_c/f_s \approx 1\%$, and the number of B_s^0 mesons originating from a B_c^+ decay to be smaller than 0.5%, thus limiting the effect on B_s^0 analyses.

10.1 Discussion

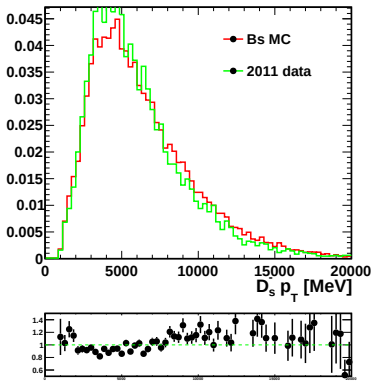
The accuracy of the final result is limited by the small number of B_c^+ candidates, which will improve by including the $B_s^0 \to J/\psi\,\phi$ or $B_s^0 \to K^+K^-$ final states.
The three largest sources of systematic uncertainty are due to the uncertainty of the B_c^+ lifetime, the lack of simulated events under 2012 conditions, and the systematics of the signal and background model in the B_c^+ mass fit. Improvements on this can be made by more accurate measurements of the B_c^+ lifetime, and by using 2012 simulated events for the training of BDT1 and BDT2 for 2012 data. The partially reconstructed background shapes in the B_c^+ mass fit could be obtained from simulation.
In future studies, the B_s^0 selection should be done independantly from the selection of $B_s^0\pi^+$ candidates to reduce this systematic uncertainty. Further optimization of the B_c^+ selection could be done by including more discriminating variables in the training of BDT2, e.g. the track multiplicity in a cone around the signal candidate, resulting in a more signal efficient BDT2 and increasing the B_c^+ yield.

By including more data collected by the LHCb detector after long shutdown 1, starting in 2014, the statistical error on the branching ratio will drop, and the observed B_c^+ yield could be large enough to attempt a study on the tagging efficiency of B_s^0 mesons originating from a B_c^+ decay.

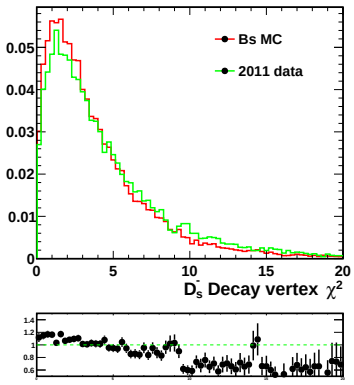
The B_c^+ yield in 2012 is larger then one would expect from simple luminosity scaling. Besides a different selection efficiency for 2012, a possible explanation for this effect could be that the production mechanism of a B_c^+ meson is mostly due to hard scatters in the primary interaction, as opposed to fragmentation, which could have a different center-of-mass energy dependence.

Appendix

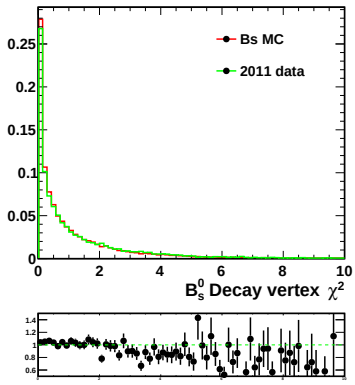
A MC Data comparison



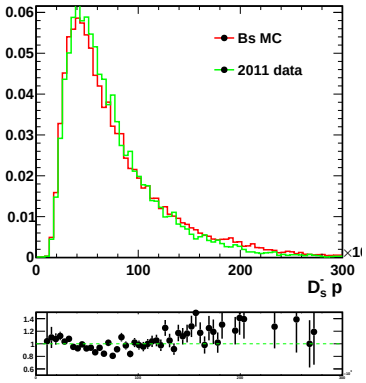
(a) D_s^- p_T



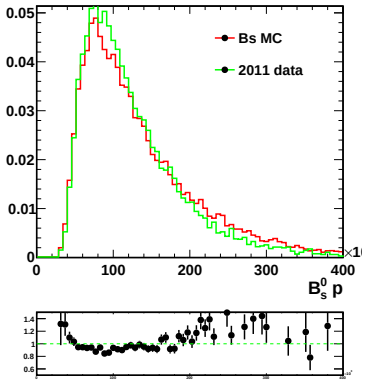
(b) D_s^- Decay Vertex χ^2



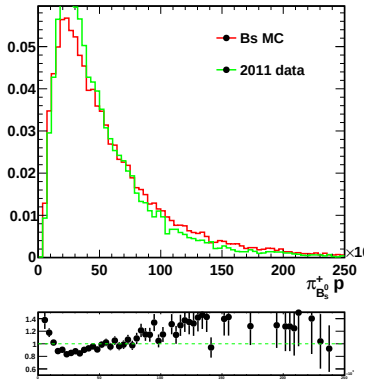
(c) B_s^0 Decay Vertex χ^2



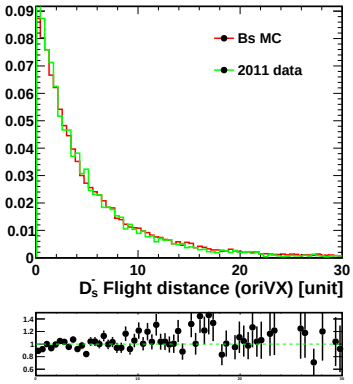
(d) D_s^- p



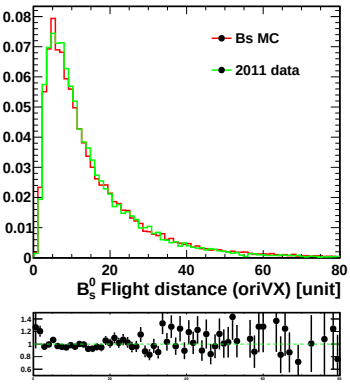
(e) B_s^0 p



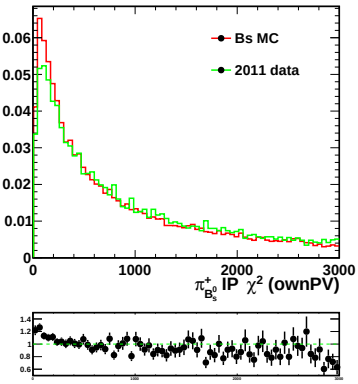
(f) $\pi^+_{B_s^0}$ p



(g) D_s^- Flight Distance

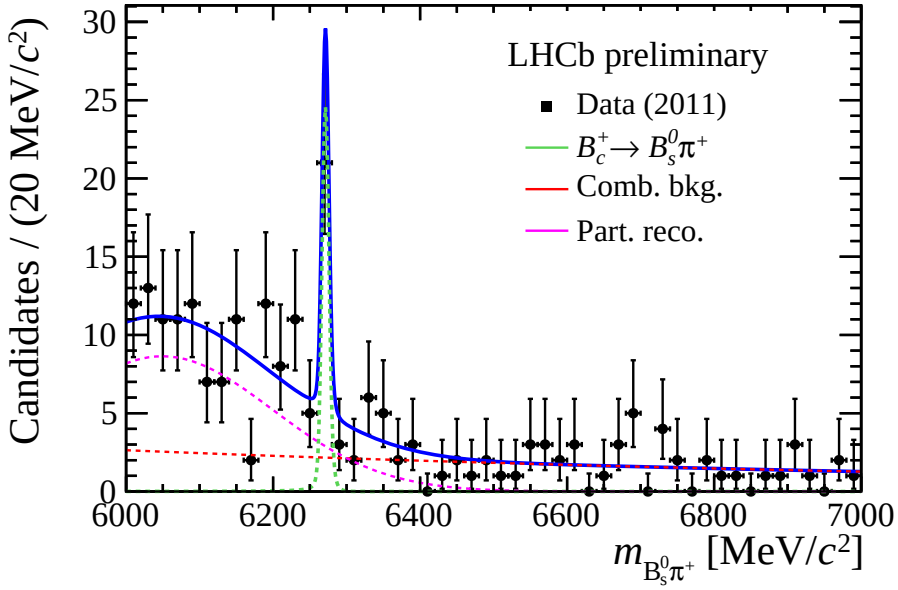


(h) B_s^0 Flight Distance

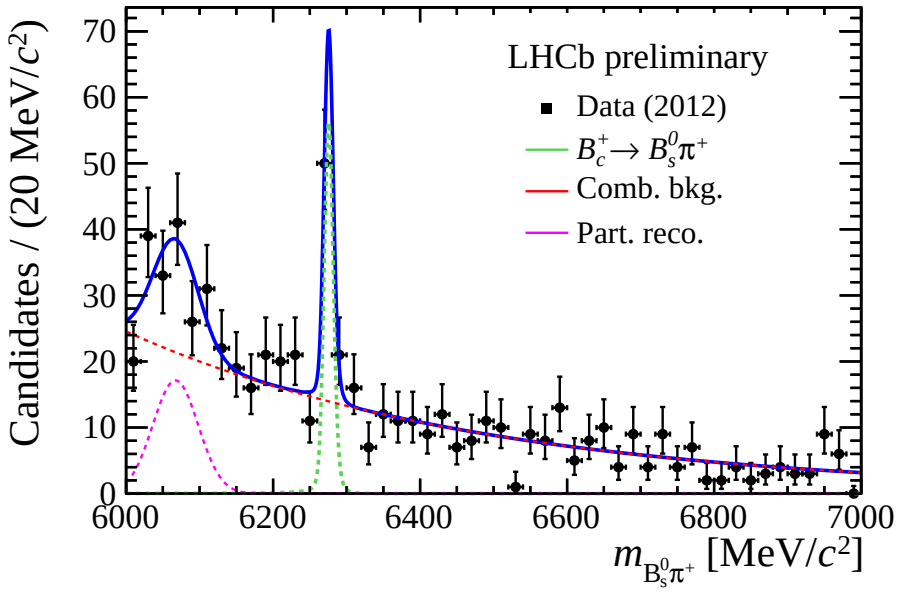


(i) $\pi^+_{B_s^0}$ IP χ^2

B B_c^+ Mass fits



(j) 2011



(k) 2012

Figure 33: B_c^+ mass fit to the 2011 data (a) and 2012 data (b), using BDT2 > 0.97.

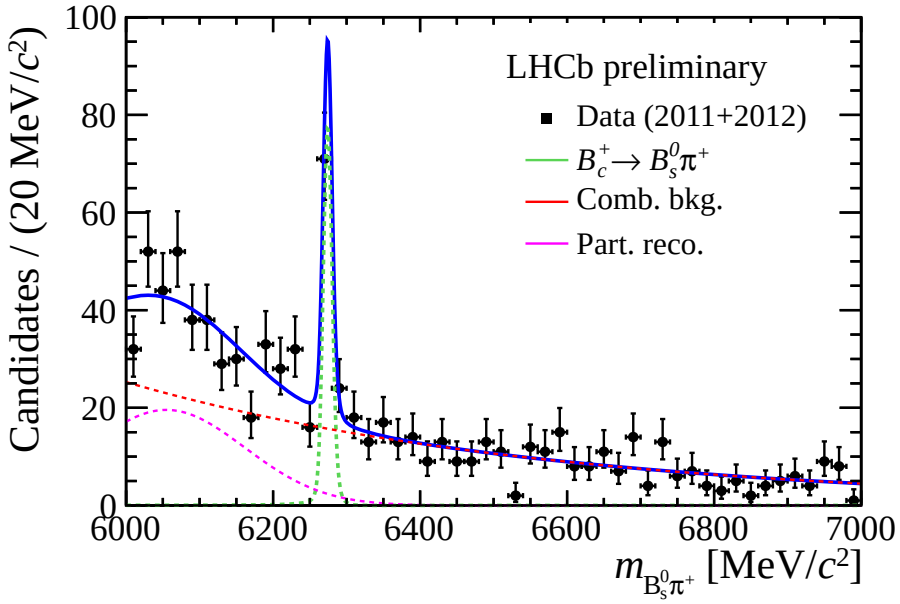


Figure 34: B_c^+ mass fit to the 2011 + 2012 data, using BDT2 > 0.97.

References

[1] M. A. Ivanov, J. G. Korner, and P. Santorelli, *Exclusive semileptonic and nonleptonic decays of the B_c meson*, Phys. Rev. **D73** (2006) 054024, arXiv:hep-ph/0602050.

[2] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation and the B_s^0 meson decay width difference with $B_s^0 \to J/\psi K^+K^-$ and $B_s^0 \to J/\psi\pi^+\pi^-$ decays*, arXiv:1304.2600.

[3] CDF collaboration, F. Abe *et al.*, *Observation of the B_c^+ meson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. **81** (1998) 2432, arXiv:9804014.

[4] CDF collaboration, T. Aaltonen *et al.*, *Observation of the decay $B_c^\pm \to J/\psi\pi^\pm$ and measurement of the $B_c^\pm$ mass*, Phys. Rev. Lett. **100** (2008) 182002, arXiv:0712.1506.

[5] D0 collaboration, V. Abozov *et al.*, *Measurement of the B_c^+ meson mass in the exclusive decay $B_c^+ \to J/\psi\pi^+$*, Phys. Rev. Lett. **101** (2008) 012001, arXiv:0802.4258.

[6] LHCb collaboration, R. Aaij *et al.*, *First observation of the decay $B_c^+ \to J/\psi\pi^+\pi^-\pi^+$*, Phys. Rev. Lett. **108** (2012) 251802, arXiv:1204.0079.

[7] LHCb collaboration, R. Aaij *et al.*, *Measurements of B_c^+ production and mass with the $B_c^+ \to J/\psi\pi^+$ decay*, Phys. Rev. Lett. **109** (2012) 232001, arXiv:1209.5634.

[8] LHCb collaboration, R. Aaij *et al.*, *Observation of $B_c^+ \to J/\psi D_s^+$ and $B_c^+ \to J/\psi D_s^{*+}$ decays*, arXiv:1304.4530.

[9] Particle Data Group, J. Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86** (2012) 010001.

[10] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of fragmentation functions f_s/f_d and the dependence on B meson kinematics* , JHEP **1304** (2013) 001, arXiv:1301.5286.

[11] C. Qiao *et al.*, *B_c^+ exclusive decays to charmonia and light mesons in QCD factorization at next-to-leading order accuracy*, arXiv:1209.5859.

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

[13] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, Jour. Comp. and Syst. Sc. **55** (1997) 119.

[14] LHCb collaboration, R.Aaij *et al.*, *Amplitude analysis and branching fraction measurement of $\bar{B}_s^0 \to J/\psi K^+K^-$*, Phys. Rev. D **87** (2013) 072004, arXiv:1302.1213.

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

[16] *LHC design report*, http://lhc.web.cern.ch/lhc/lhc-designreport.html.

[17] *LHC public website*, http://home.web.cern.ch/about/accelerators/large-hadron-collider.

[18] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[19] *LHCb public website*, http://lhcb-public.web.cern.ch/lhcb-public/.

[20] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, CERN-LHCb-DP-2012-00.

[21] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** P04022, arXiv:1211.3055.

[22] T. Board, *LCG computing grid, technical design report*, CERN-LHCC-2005-024.

[23] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[24] 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.

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

[26] 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.

[27] 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.

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

[29] C. Chang *et al.*, *BCVEGPY: An event generator for hadronic production of the* B_c^+ *meson*, arXiv:0309120.

[30] S. Ali *et al.*, *Time-dependent CP analysis of* $B_s^0 \to D_s^{\mp} K^{\pm}$ *and* $B_s^0 \to D_s^- \pi^+$, LHCb-ANA-2012-068.

[31] C. L. *et al.*, *Measurement of the ratio of b-hadron production fractions* f_s/f_d *with hadronic decays in 2011*, LHCb-ANA-2011-070.

[32] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.

[33] G. Punzi, *Sensitivity of searches for new signals and its optimization*, arXiv:0308063.

[34] LHCb collaboration, R. Aaij *et al.*, *First evidence for the decay* $B_s^0 \to \mu^+\mu^-$, Phys. Rev. Lett. **110** (2013) 021801, arXiv:1211.2674.

[35] P. Koppenburg, *Dealing with multiple candidates*, LHCb-INT-2011-009.

[36] A. Read, *Presentation of search results: the CLs technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002) 10 2693.

[37] A. Jaeger *et al.*, *Measurement of the track finding efficiency*, LHCb-PUB-2011-025.