Additional information to
LHCb-PAPER-2012-043
November 13, 2012

First evidence for the decay $B_s^0 \to \mu^+\mu^-$

The LHCb collaboration

Abstract

This note provides additional information and material to support LHCb-PAPER-2012-043.

1 Analysis strategy

The general structure of the analysis is based upon that published in Ref. [1]. First a loose selection removes most of the background while keeping very high efficiency for signal. The number of observed events is then compared to the number of expected signal and background events in two independent variables, the dimuon invariant mass and the output of a multi-variate discriminant, the Boosted Decision Tree (BDT) constructed using the TMVA package [2, 3].

In the present version of the analysis, only slight adjustments to the selection have been made to follow changes in the reconstruction software. Moreover, the treatment of the exclusive backgrounds is extended to include the $B^0 \to \pi^- \mu^+ \nu$ [4] and $B^{+(0)} \to \pi^{+(0)} \mu^+ \mu^-$ [5] channels. These contaminate significantly the mass sidebands, which are used to determine the expected background level in the signal search window.

The probability for a signal or background event to have a given value of the BDT output is extracted from data using $B^0_{(s)} \to h^+ h'^-$ candidates as signal (where $h^{(\prime)}$ can be a pion or a kaon) and sideband $B^0_{(s)} \to \mu^+ \mu^-$ candidates as background. The present analysis uses a PID-based selection to isolate the exclusive $B^0_{(s)} \to h^+ h'^-$ modes instead of an inclusive selection as was used in 2011 analysis [1].

The invariant mass lineshape of the signals is described by a Crystal Ball function [6]. The parameters that characterize the function are measured from data (the central value m and the mass resolution σ) or from simulated events (the transition point α and the exponent n). The parameters of the radiative tails have been extracted from a fit to the mass distribution of $B^0_s \to \mu^+ \mu^-$ simulated events where the resolution has been smeared in order to reproduce the measured value. The central values and the resolutions of the B^0 and B^0_s masses are used to define the search windows. The BDT output and invariant mass distributions for combinatorial background events in the signal regions are obtained using fits of the mass distribution of events in the mass sidebands in bins of the BDT output.

The two-dimensional space formed by the invariant mass and the BDT output is binned. For each bin we count the number of candidates observed in the data, and compute the expected number of signal events and the expected number of background events. The compatibility of the observed distribution of events in all bins with the distribution expected for a given branching fraction hypothesis is computed using the $\mathrm{CL_s}$ method [7]. In order to avoid unconscious bias the mass region $m_{\mu\mu} = [m(B^0) - 60\,\mathrm{MeV}/c^2, m(B^0_s) + 60\,\mathrm{MeV}/c^2]$ has been blinded until the finalization of all analysis choices.

The number of expected signal events, for a given branching fraction hypothesis, is obtained by normalizing to channels with well-known branching fractions: $B^+ \to J/\psi\, K^+$ and $B^0 \to K^+ \pi^-$. These channels are selected in ways as similar as possible to the signals in order to minimize systematic uncertainties. With respect to the 2011 measurement, the

less precise $B_s^0 \to J/\psi\,\phi$ normalization channel is not used in this analysis.[1]

2 Selection

The selection aims at reducing the data size to a manageable level while keeping the efficiency on the signal as high as possible. The selection criteria are mostly unchanged compared to the 2011 data analysis [1], apart from two adjustments necessary due to the different reconstruction in 2012: the maximum allowed track χ^2/ndf was decreased from 4 to 3, and a new selection criterion was introduced, aiming at removing fake track candidates produced by the pattern recognition. The latter has an efficiency above 99.5% for $B^0_{(s)} \to \mu^+\mu^-$ signal candidates and removes a significant part of the background tracks, especially in the $B^0_{(s)} \to h^+h^{'-}$ control sample.

The $B^0_{(s)} \to h^+h^{'-}$ inclusive sample is selected exactly in the same way as the $B^0_{(s)} \to \mu^+\mu^-$ signals (apart from the muon identification requirement) in order to be used as the main control sample in the calibration of the BDT and the invariant mass. The $J/\psi \to \mu\mu$ decay in the $B^+ \to J/\psi K^+$ normalization channel is selected in a way very similar to the $B^0_{(s)} \to \mu^+\mu^-$ signals. After an initial mass cut, the muons from the J/ψ are subject to a vertex fit to form the J/ψ mass. $K^\pm$ candidates are required to pass some track quality and impact parameter cuts.

The fiducial large momenta, transversa momenta and B lifetime cuts applied to reject the unphysical signal candidates are unchanged from Ref. [1] and have an efficiency of $\epsilon = 99.9\%$ on Monte Carlo $B_s^0 \to \mu^+\mu^-$ signal events.

3 BDT and invariant mass calibration

After the selection each event is given a probability to be signal or background in a two-dimensional space defined by two independent variables: invariant mass and BDT.

Geometrical and kinematic information not fully exploited in the selection is combined via the BDT for which nine variables are employed [8]. Ordered by their background rejection power, they are: the B-candidate impact parameter (IP), the minimum IP significance of the muons, the sum of the degrees of isolation of the muons (the number of good two-track vertices a muon can make with other tracks in the event), the decay time, the transverse momentum ($p_{\rm T}$), and isolation [9] of the B candidate, the distance of closest approach between the two muons, the minimum $p_{\rm T}$ of the muons, and the cosine of the angle between the muon momentum in the dimuon rest frame and the vector perpendicular to the B candidate momentum and to the beam axis. No data events were used for the choice of the variables and the subsequent training of the BDT, to avoid biasing the results.

The BDT output is independent of the invariant mass for signal inside the search window and is defined such that for the signal it is approximately uniformly distributed

[1] The branching fraction $\mathcal{B}(B_s^0 \to J/\psi\,\phi)$ is known to $\sim 30\%$, to be compared with $\sim$3% relative accuracy of the other two.

between zero and one, while for the background it peaks at zero.

The binning of the BDT and invariant mass distributions has been optimized during the previous analysis [1] using the simulation to maximize the separation between the median of the test statistic distribution expected for background and SM $B_s^0 \to \mu^+\mu^-$ signal and that expected for background only. The BDT range is divided into eight bins with boundaries $[0.0, 0.25, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]$. For the 2012 dataset, only one bin is considered in the BDT range 0.8–1.0 due to the lack of events in the mass sidebands for BDT > 0.9. Over 97% of test statistic separation comes from the bins with BDT>0.5. The BDT is trained using simulated samples: $B^0_{(s)} \to \mu^+\mu^-$ for signal and $b\bar{b} \to \mu\mu X$ for background.

3.1 Calibration of likelihoods for signal: BDT

For this analysis we use the same BDT as used for the 2011 result. The distribution of four out of the nine BDT input variables, is shown in figure 1 for signal $B_s^0 \to \mu^+\mu^-$ and $b\bar{b} \to \mu\mu X$ Monte Carlo events and for dimuon background events from 2011 data sidebands. The nine input variables have the same distributions in 2012 data as in 2011, and no retraining of the BDT discriminant was deemed necessary. This has been verified for the signal by looking at the variables in the normalisation channel $B^+ \to J/\psi\, K^+$ where $J/\psi \to \mu^+\mu^-$ is used as a signal proxy. The only exception is the isolation which has a shift to smaller values (worse isolation) in 2012 as we have a higher average number of visible pp interactions per bunch crossing, leading to a higher average number of tracks. This results in a slight degradation in the performance of the BDT operator.

To calibrate the BDT for signal we rely on the exclusive channels $B^0_{(s)} \to h^+h^{'-}$ as a proxy $B^0_{(s)} \to \mu^+\mu^-$. To distinguish between the different channels and therefore to identify the correct mass hypothesis of the final state particles, we apply cuts on the $\Delta LL_{K-\pi}$ variable[2] of the final state particles. These cuts can induce biases as the separation power of the $\Delta LL_{K-\pi}$ variable is highly dependent on the kinematics of the final state particle and the multiplicity of the event. The possible bias is corrected by weighting each event by the efficiency of the $\Delta LL_{K-\pi}$ cut, which is measured from $D^{*\pm} \to (D^0 \to K^\pm\pi^\mp)\pi^\pm$ data as a function of a suitable set of kinematical and multiplicity variables.

Figure 2 shows the signal BDT PDF. The results are summarized in Table 1. The BDT calibration based on $B^0_{(s)} \to h^+h^{'-}$ 2012 data, agrees within the uncertainty with that for 2011 data, even if a systematic shift of the central values towards smaller BDT values, due to the higher multiplicity in 2012 with respect to the 2011 data, is present.

3.2 Calibration of likelihoods for signal: invariant mass

The invariant mass distribution for the signal is described by a Crystal Ball function. The peak positions are obtained from the invariant mass distribution of $B^0_{(s)} \to h^+h'^-$ events separated into different decay channels following the $m_{\pi^+\pi^-}$, $m_{K\pi}$, and $m_{K^+K^-}$

[2]This is a discriminating variable defined for each track, build from the difference of single K and π log-likelihood PID variables.

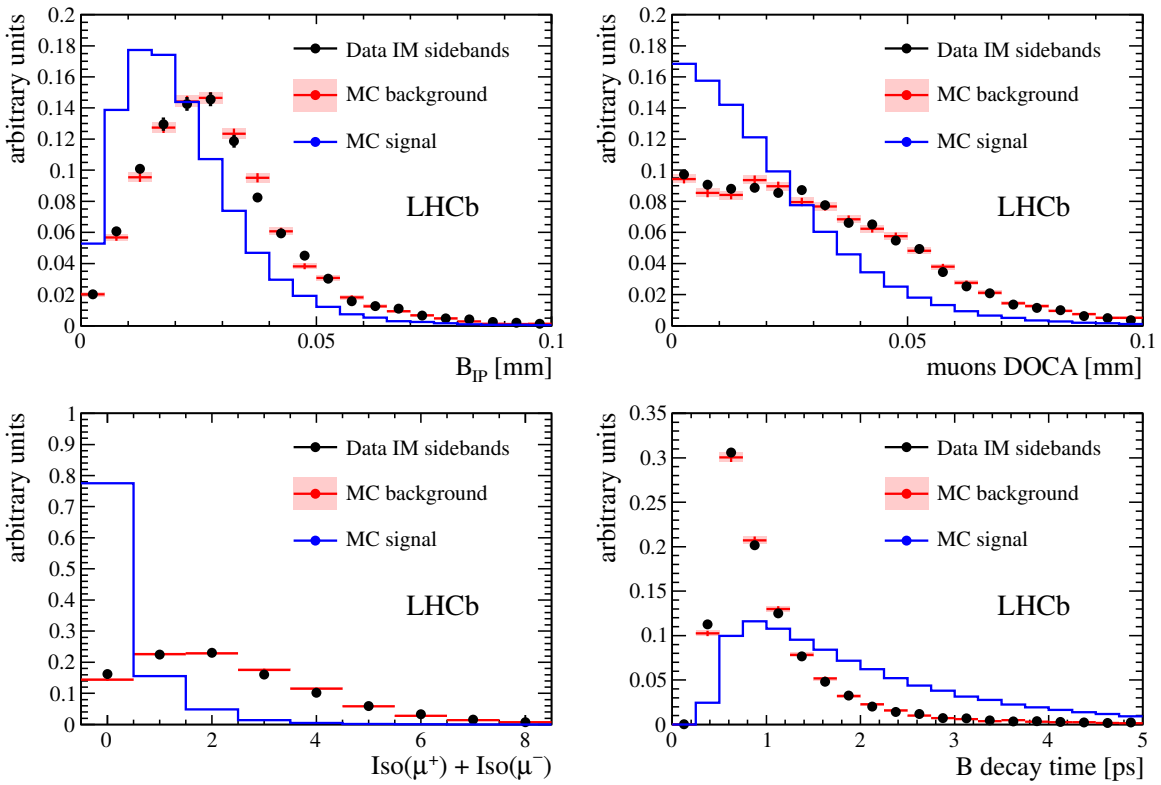

Figure 1: Distribution, for signal and background, of four out of the nine variables used in the BDT definition, from top left: $IP(B)$, DOCA, $I(\mu_1) + I(\mu_2)$, and $t(B)$. Blue is MC $B^0_{(s)} \to \mu^+\mu^-$ signal, red is MC $b\bar{b} \to \mu\mu X$, and black points are for $b\bar{b} \to \mu\mu X$ 2011 data.

Table 1: BDT PDF estimation for the 2012 dataset for signal and combinatorial background (see also Sect. 3.4).

BDT range	signal [%]	comb. background [%]
$[0.00, 0.25]$	$29.2^{+2.4}_{-2.4}$	$96.81^{+0.14}_{-0.14}$
$[0.25, 0.40]$	$16.9^{+1.8}_{-1.7}$	$2.34^{+0.12}_{-0.12}$
$[0.40, 0.50]$	$9.5^{+1.0}_{-0.9}$	$0.54^{+0.06}_{-0.06}$
$[0.50, 0.60]$	$9.1^{+0.7}_{-0.6}$	$0.18^{+0.04}_{-0.03}$
$[0.60, 0.70]$	$9.2^{+0.7}_{-0.7}$	$0.087^{+0.028}_{-0.023}$
$[0.70, 0.80]$	$9.1^{+0.9}_{-0.9}$	$0.014^{+0.017}_{-0.009}$
$[0.80, 1.00]$	$16.8^{+1.3}_{-1.2}$	$0.016^{+0.014}_{-0.009}$

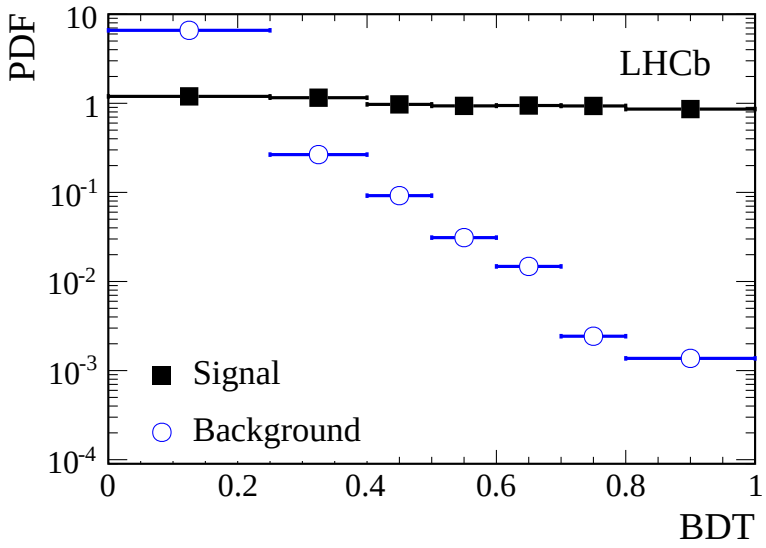

Figure 2: BDT PDF for the 2012 dataset, for the signal (black squares) and combinatorial background (blue open points) (see also Sect. 3.4). Values are normalized to the bin size.

mass hypotheses, as shown in Fig. 3. Table 2 summarizes the results. For m_{B^0} we use a weighted average between the $B^0 \to K\pi$ and $B^0 \to \pi^+\pi^-$ decay modes. The peak positions are about 0.1 % above the nominal PDG value for the B^0 and B^0_s mass, as in the 2011 analysis.

Table 2: Summary of Crystal Ball function parameters measured on data.

Channel	Peak position	Resolution
B^0	$(5284.36 \pm 0.26_{\rm stat} \pm 0.13_{\rm syst})$ MeV/c^2	$(24.63 \pm 0.13_{\rm stat} \pm 0.36_{\rm syst})$ MeV/c^2
B^0_s	$(5371.55 \pm 0.41_{\rm stat} \pm 0.16_{\rm syst})$ MeV/c^2	$(25.04 \pm 0.18_{\rm stat} \pm 0.36_{\rm syst})$ MeV/c^2

The B^0_s and B^0 invariant mass resolutions are estimated by two methods. The first, interpolating the resolution measured using dimuon resonances ($J/\psi, \psi(2S)$ and $\Upsilon(1S, 2S, 3S)$), the second measuring the resolution from fits to the invariant mass distributions of exclusive $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ decays. Figure 4 shows the charmonium and bottomonium resonances while Fig. 5 shows the interpolation of the invariant mass resolution.

The second methods uses the $B^0_{(s)} \to h^+h'^-$ sample and a technique quite similar to that used to calibrate the BDT. Combining the two results for the invariant mass resolution by taking the weighted average we get the values summarized in Table 2, which are used in the extraction of the results.

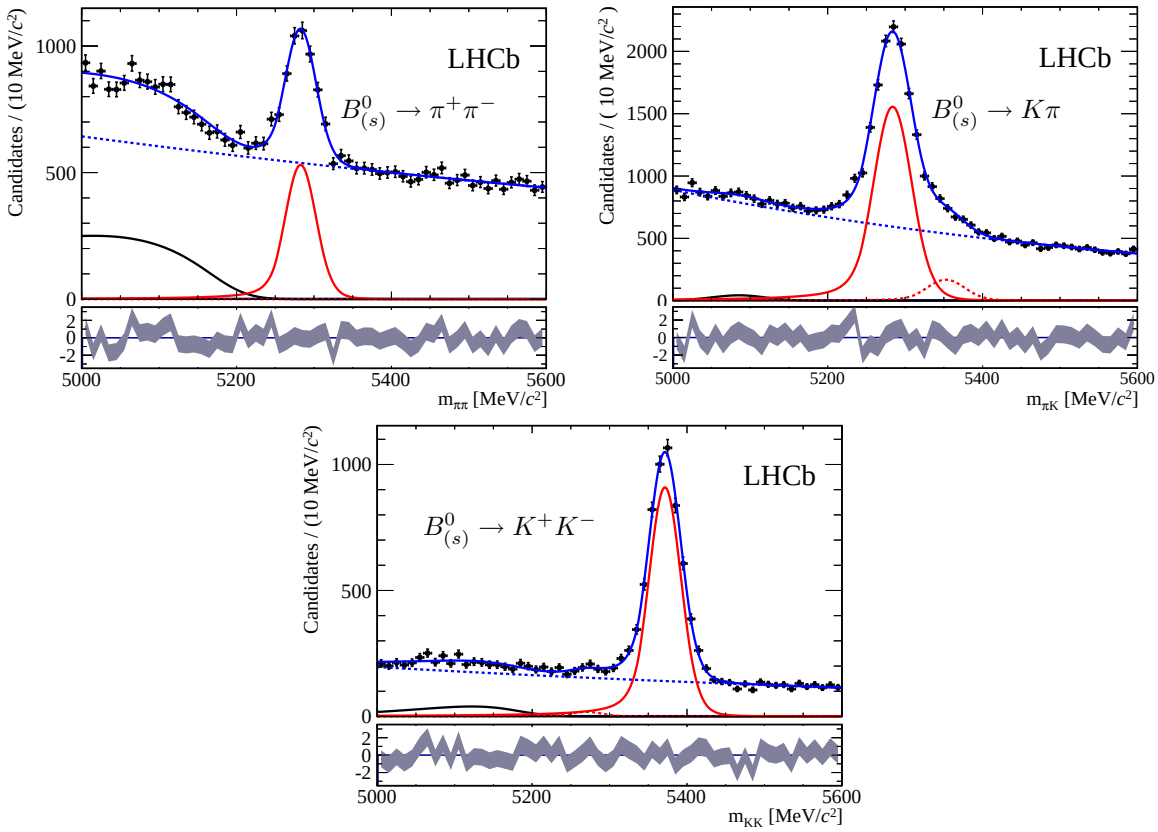

Figure 3: Invariant mass distribution of $B^0_{(s)} \to h^+h'^-$ separated into the different decay channels. The full red lines indicate the dominant signal model, the dashed red the sub-dominant (e.g. upper right: full is $B^0 \to K\pi$ and dashed is $B^0_s \to K\pi$). The black curve to the left indicates partially reconstructed background. The blue dotted curve describes the combinatorial background.

3.3 Peaking background estimate

The decay modes $B^0 \to K^+\pi^-$, $B^0 \to \pi^+\pi^-$, $B^0_s \to \pi^+K^-$ and $B^0_s \to K^+K^-$ can fake a signal if both hadrons are misidentified as muons. The hadron misidentification probability is minimised by combining information from the LHCb particle identification devices (the muon detector and the Ring Imaging Cherenkov detectors). The performance is illustrated in Fig. 6

This double misidentification (misID) probability for these decays has been estimated by convoluting the kaon and pion misID probability measured on data, with the momentum and $p_{\rm T}$ spectra of the two hadrons of Monte Carlo $B^0_{(s)} \to h^+h'^-$ decays.

As a result, we obtain a double misID probability $\epsilon_{hh\to\mu\mu} = (1.80 \pm 0.04_{\rm stat} \pm 0.08_{\rm syst}) \times 10^{-5}$. In 2011 data the same probability was $\epsilon_{hh\to\mu\mu} = (1.52 \pm 0.07_{\rm stat} \pm 0.07_{\rm syst}) \times 10^{-5}$.

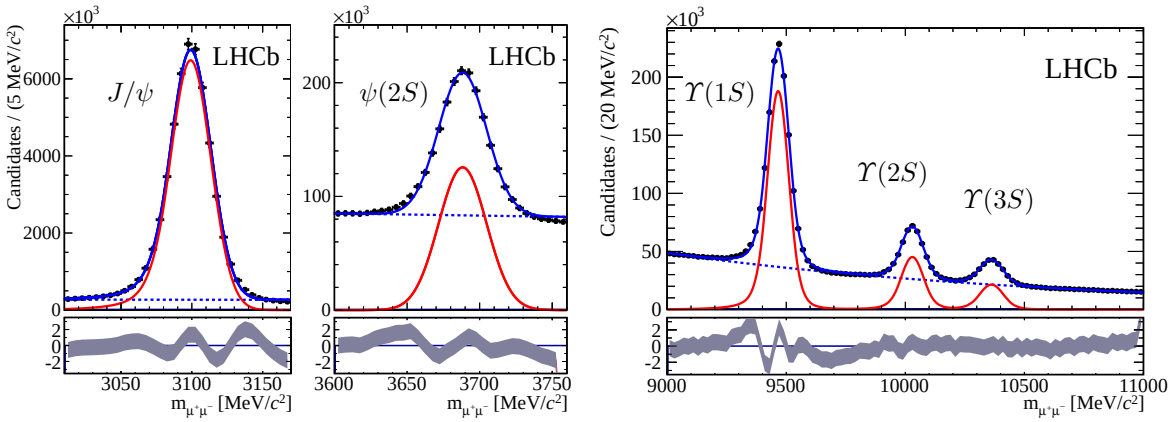

Figure 4: Dimuon invariant mass distribution for $J/\psi(1S) \to \mu^+\mu^-$, $\psi(2S) \to \mu^+\mu^-$ and $\Upsilon(1,2,3S) \to \mu^+\mu^-$. All resonances are described by double-sided Crystal Ball functions while the combinatorial background is described by an exponential function.

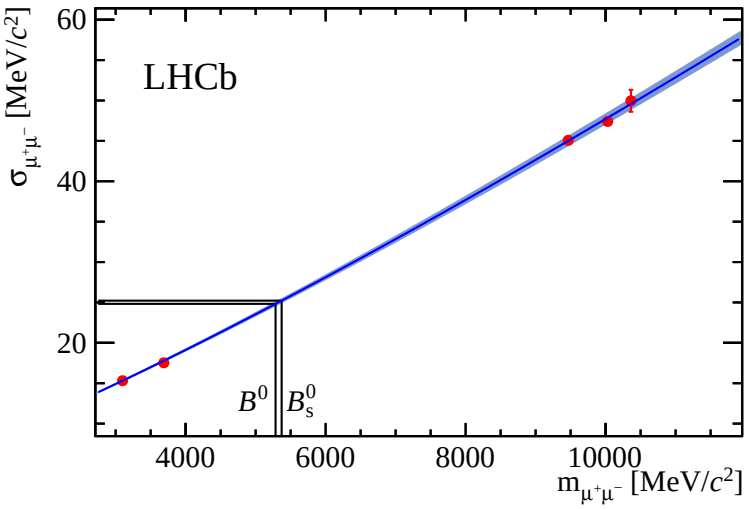

Figure 5: Interpolation of the invariant mass resolution between charmonium and bottomonium resonances to the mass of the B^0 and B_s^0 mesons.

The number of $B^0_{(s)} \to h^+h'^-$ double misidentified events is evaluated as:

$$N_{B^0_{(s)}\to h^+h^-\to\mu\mu} = \epsilon^{\rm TRIG}_{B^0_{(s)}\to\mu^+\mu^-}\, N_{hh}\, \epsilon_{hh\to\mu\mu}$$

where N_{hh} is the number of $B^0_{(s)} \to h^+h'^-$ events selected independently of the trigger and $\epsilon^{\rm TRIG}_{B^0_{(s)}\to\mu^+\mu^-}$ takes into account the trigger efficiency of $B^0_{(s)} \to \mu^+\mu^-$ events (as measured in

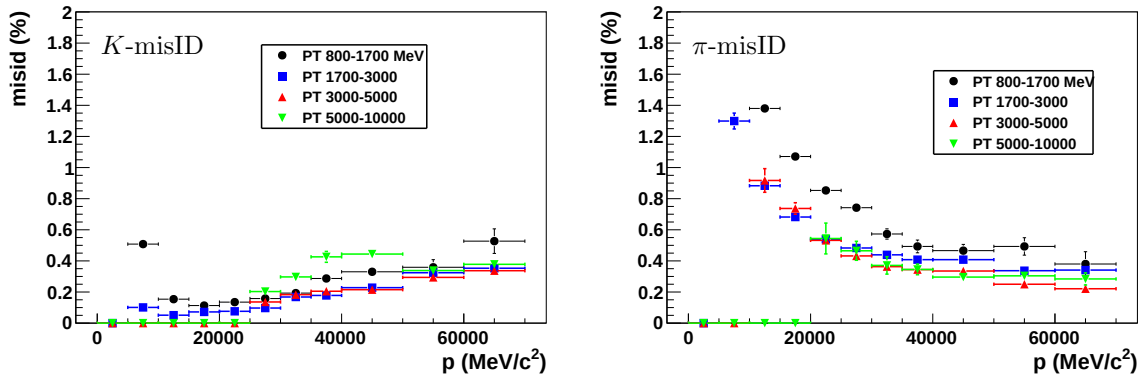

Figure 6: Kaon (left) and pion (right) misidentification probability measured from 2011 data.

simulation) as well as a correction measured from data for the efficiency of the $B^0_{(s)} \to h^+h'^-$ selection.

The number of $B^0_{(s)} \to h^+h'^-$ events are extracted by fitting the mass distributions of an inclusive $B^0_{(s)} \to h^+h'^-$ sample in the $K\pi, KK, \pi\pi$ mass hypotheses. In total, (8.6 ± 0.7) doubly misidentified $B^0_{(s)} \to h^+h'^-$ decays are expected in the [4900,6000] MeV/c^2 mass range, to be compared with the number obtained in 2011, $N(B^0_{(s)} \to h^+h'^- \to \mu\mu) = (5.4 \pm 0.7)$. The increase in the yield is both due to the increase in luminosity and cross-section in 2012 data and to the increase in double misidentification due to the higher pile-up of 2012 data. The previous yields in the [4900,6000] MeV/c^2 mass window are then evaluated for each BDT bin separately, and scaled to account for the acceptance of these decays in the signal mass window, $(8.8^{+3.0}_{-2.1})\%$ and $(48^{+20}_{-8})\%$ for B^0_s and B^0 signals respectively. The results are listed in Tables 7 and 8, for 2012 and 2011 data samples, respectively.

3.4 Combinatorial background estimate

The background yield leaking into B^0_s and B^0 mass windows is dominated by combinatorics of two muons coming from different B mesons, with the $B^0_{(s)} \to h^+h'^-$ misID background also playing a role, especially in the B^0 window. The yield of the combinatorial background in the signal window is evaluated by interpolating the mass sidebands, for each BDT bin separately.

To this purpose, the invariant mass distribution has been fitted in the range $4900 < m_{\mu\mu} < 5224$ MeV/c^2 and $5432 < m_{\mu\mu} < 5966$ MeV/c^2, assuming a single exponential function for the combinatorial background. The lower boundary at 4900 MeV/c^2 was chosen to exclude background sources other than combinatorial, mainly cascading $b \to c\mu X \to \mu\mu X$, while the intermediate mass values are set at m(B^0)-60 MeV/c^2 and m(B^0_s)$+60$ MeV/c^2 respectively.

In the present version of the analysis many exclusive B decay channels have been considered which may pollute the mass sidebands, and thus affect the combinatorial background estimate. The results of our estimates are shown in Table 3, which summarizes the expected yields in the mass range [4900-6000] MeV/c^2, and for BDT above 0.8. The $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^{+(0)} \to \pi^{+(0)}\mu^+\mu^-$ decays, together with $B^0_{(s)} \to h^+h'^-$ misID events, represent ∼85% of the exclusive background yield in this region, and for this reason they are included in the combinatorial background interpolation as separate PDFs.

Table 3: Number of events expected in the 2011 and 2012 data samples for all the dominant exclusive background sources estimated in the mass range [4900-6000] MeV/c^2, for BDT above 0.8.

	2011	2012
$B^0 \to \pi^-\mu^+\nu_\mu$	3.51 ± 0.25	4.04 ± 0.28
$B^0_{(s)} \to h^+h'^-$ misID	0.91 ± 0.12	1.37 ± 0.11
$B^{+(0)} \to \pi^{+(0)}\mu^+\mu^-$	1.12 ± 0.35	1.32 ± 0.39

For $B^0 \to \pi^-\mu^+\nu_\mu$ and $B \to \pi\mu^+\mu^-$ decays, the mass and BDT PDF are obtained from MC, with corrections applied from data control samples; the absolute normalization is instead evaluated relative to the $B^+ \to J/\psi K^+$ decays. For $B^0 \to \pi^-\mu^+\nu_\mu$, $B \to \pi\mu^+\mu^-$, and $B^0_{(s)} \to h^+h'^-$ misID, the normalizations and the parameters describing the mass and BDT PDFs are fluctuated in the fit according to their total uncertainties. The combinatorial background is parameterized with an exponential function, the parameters of which are free to vary in the fit.

The observed yield of events in the dimuon sidebands, the union of $4900 < m_{\mu\mu} < 5224$ MeV/c^2 and $5432 < m_{\mu\mu} < 5966$ MeV/c^2 is measured for each BDT bin for 2012 data.

The fits to the mass spectra of the 7 BDT bins of 2012 data are shown in Fig. 7. From these fits the BDT PDF for the combinatorial background shown in Fig. 2 and listed in Table 1 is evaluated. The expected numbers of combinatorial background events in B^0_s and B^0 mass windows are listed in Table 7. Due to the change in the treatment of the background composition with respect to Ref. [1], the combinatorial background estimate has been repeated for the 2011 data sample, as shown in Fig. 8 and Table 8.

4 Normalization

To estimate the signal branching ratio, we normalize the number of observed signal events to the numbers of events in two complementary channels: $B^+ \to J/\psi(\mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$. The first decay has similar trigger and muon identification efficiencies as the signal but different number of particles in the final state, while the second channel has a similar topology but a different trigger selection.

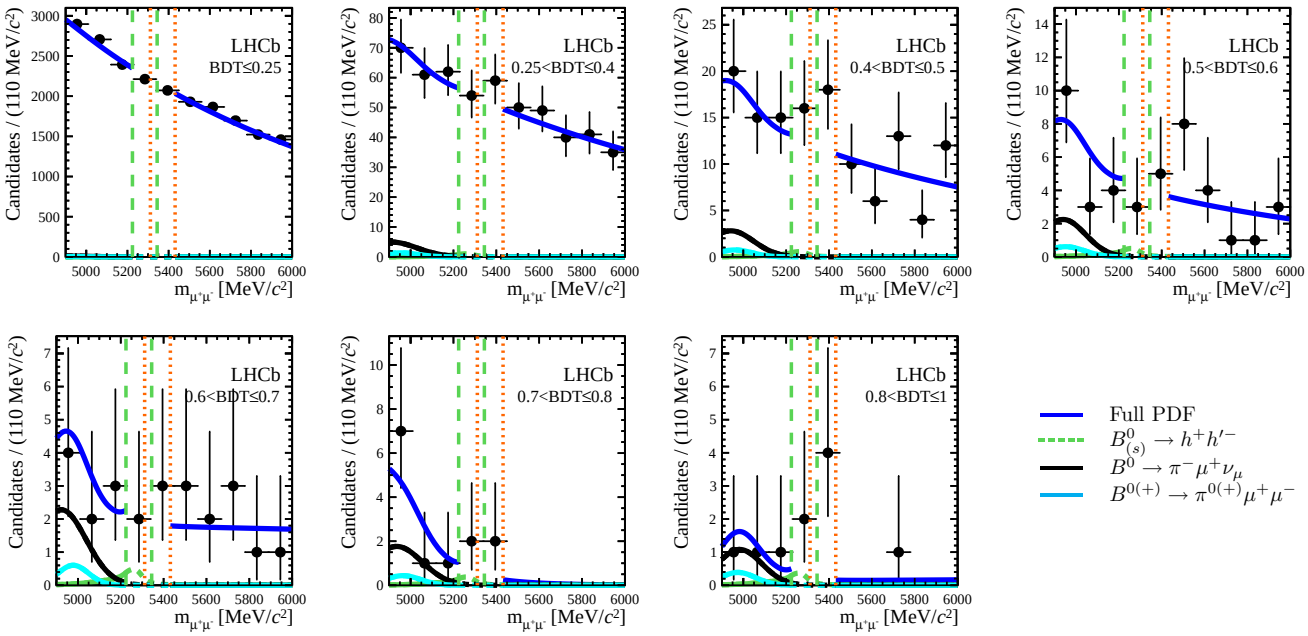

Figure 7: 2012 data: fit to the dimuon mass sidebands in 7 bins of BDT. Dots are data, black line is $B^0 \to \pi^-\mu^+\nu_\mu$, cyan is $B^{+(0)} \to \pi^{+(0)}\mu^+\mu^-$, green is $B^0_{(s)} \to h^+h'^-$ misID and blue is total fit. Vertical orange (green) dashed lines indicate the $B^0_s \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) search windows excluded from the background estimation fit.

To translate the number of observed events into a branching ratio we use the following equation:

$$\mathcal{B} = \mathcal{B}_{\rm norm} \times \frac{\epsilon_{\rm norm}}{\epsilon_{\rm sig}} \times \frac{f_{\rm norm}}{f_{d(s)}} \times \frac{N_{B^0_{(s)}\to\mu^+\mu^-}}{N_{\rm norm}} = \alpha^{\rm norm}_{B^0_{(s)}\to\mu^+\mu^-} \times N_{B^0_{(s)}\to\mu^+\mu^-},$$

where $\alpha_{\rm norm}$ is the normalization factor (or single event sensitivity) and $f_{d(s)}$ and $f_{\rm norm}$ are the probabilities that a b quark fragments into a $B^0_{(s)}$ and into the b-hadron relevant in the chosen normalization mode respectively. With $\mathcal{B}_{\rm norm}$ we indicate the visible branching fraction and with $N_{\rm norm}$ the number of signal events in the normalization channel obtained from a fit to the invariant mass distribution. The branching ratios of the considered normalisation channels are respectively $\mathcal{B}(B^+ \to J/\psi K^+) = (6.01 \pm 0.21) \times 10^{-5}$ (where the $J/\psi \to \mu^+\mu^-$ branching fraction is also included) and $\mathcal{B}(B^0 \to K^+\pi^-) = (1.94 \pm 0.06) \times 10^{-5}$, respectively [4]. The efficiency $\epsilon_{\rm sig(norm)}$ for the signal (normalization channel) is the product of the reconstruction efficiency of all the final state particles of the decay including the geometric acceptance of the detector ($\epsilon^{\rm rec}$), the selection efficiency for reconstructed events ($\epsilon^{\rm sel|rec}$), and the trigger efficiency for reconstructed and selected events ($\epsilon^{\rm trg|sel}$). The values of these efficiencies are reported in Table 4 separately for the two normalization channels.

The ratios of acceptance and reconstruction efficiencies are computed using Monte Carlo simulation. Potential differences between simulation and data are included as systematic uncertainties. The selection efficiencies are determined using Monte Carlo

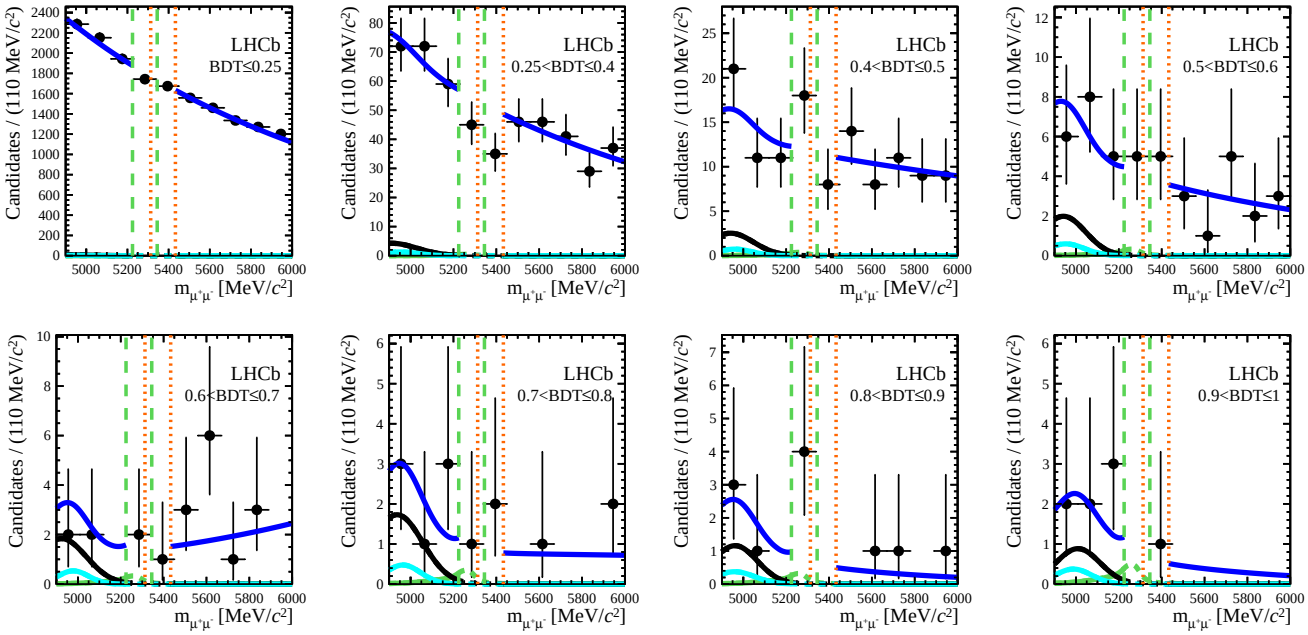

Figure 8: 2011 data: fit to the dimuon mass sidebands in 8 bins of BDT. Dots are data, black line is $B^0 \to \pi^- \mu^+ \nu_\mu$, cyan is $B^{+(0)} \to \pi^{+(0)} \mu^+ \mu^-$, green is $B^0_{(s)} \to h^+ h'^-$ misID and blue is total fit. Vertical orange (green) dashed lines indicate the $B^0_s \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) search windows excluded from the background estimation fit. The legend is the same as in Fig. 7

simulation and cross-checked with data. Reweighting techniques have been used for all Monte Carlo distributions that do not match those from data. The trigger efficiency is evaluated directly from data with events triggered independently of the normalisation channels.

For the fragmentation fraction f_s/f_d we use the value determined by LHCb in two different ways: from the relative abundance of $B^0_s \to D^-_s \pi^+$, $B^0 \to D^- K^+$ and $B^0 \to D^- \pi^+$ [10] and from semileptonic $B \to DX$ decays [11]. In the previous analysis, we used the average of these two LHCb results: $f_s/f_d = 0.267^{+0.021}_{-0.020}$ [11]. Recently, the measurement based on hadronic decays has been updated by LHCb [12] and the new combination with the semileptonic measurement is

$$f_s/f_d = 0.256 \pm 0.020\,. \tag{1}$$

This update shows also a slight dependence of f_s/f_d on the B meson $p_{\rm T}$. Since the BDT is correlated with the $p_{\rm T}$ of the B meson, we checked the f_s/f_d variation across the BDT range. We found a variation comparable to the uncertainty in f_s/f_d (Eq. 1) and so we use this value without further corrections.

Figure 9, left, shows the invariant mass distribution of the events passing the $B^+ \to J/\psi K^+$ selection together with the fit to determine the number of $B^+ \to J/\psi K^+$ events. The $B^+ \to J/\psi K^+$ signal distribution is modelled with a double Gaussian function, while

the background is modelled with two functions: an exponential for the combinatorial background, and a Gaussian for the mis-identified $B^+ \to J/\psi\pi^+$ decay. The number of signal events after the selection is $N(B^+ \to J/\psi K^+) = 424\,200 \pm 1500$. We have assigned a systematic error of 0.3% due to the differences between the result obtained with the fit and the number of candidates obtained after background subtraction.

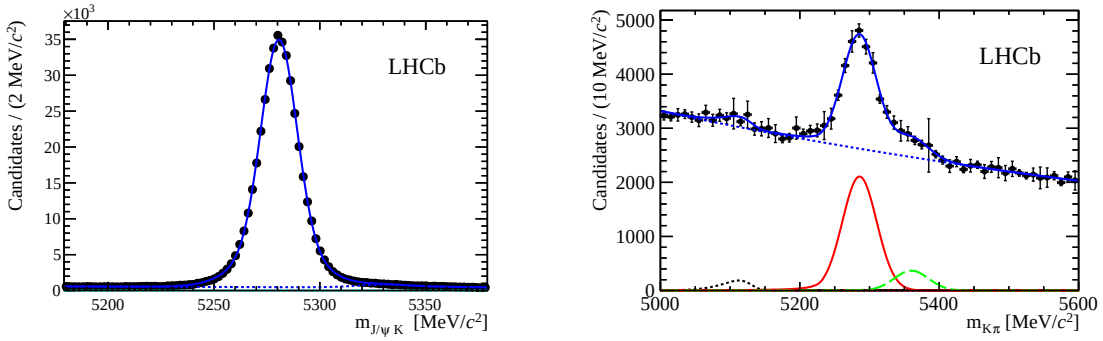

Figure 9: Invariant mass distribution of the $B^+ \to J/\psi K^+$ (left) and $B^0 \to K^+\pi^-$ (right) candidates after the stripping selection. The fit to data is superimposed in blue while the blue dotted line is the combinatorial background. In the fit to the $B^0 \to K^+\pi^-$ data, the $B^0 \to K^+\pi^-$ component (red line), the $B^0_{(s)} \to K\pi$ component (green dashed line), and the partially reconstructed background (black dotted line) are also visible.

The second normalization channel used in this analysis is $B^0 \to K^+\pi^-$. To separate the $B^0 \to K^+\pi^-$ candidates from the inclusive $B^0_{(s)} \to h^+h'^-$ sample, a PID selection is applied (using $\Delta LL_{K-\pi}$). The same reweighting procedure is then applied to the events as in the BDT calibration to account for their PID selection. The distribution of events is shown in Fig. 9, right, where the error bars are dominated by the uncertainty on the $\Delta LL_{K-\pi}$ weights. The total number of $B^0 \to K^+\pi^-$ candidates was found to be $N(B^0 \to K^+\pi^-) = 14\,600 \pm 1100$, which contains also the systematic uncertainties due to the reweighting procedures. The stability of the $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ yields has been checked throught the data taking and found consistent with the expected variations from trigger configurations.

The normalization factor $\alpha_{\rm norm}$ for the different control channels and the factors that enter in its calculation are listed in Table 4. A weighted average assuming the uncertainty on f_d/f_s to be correlated between the $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ channels, gives:

$$\alpha_{B^0_s \to \mu^+\mu^-} = (2.80 \pm 0.25) \times 10^{-10}\,,$$
$$\alpha_{B^0 \to \mu^+\mu^-} = (7.16 \pm 0.34) \times 10^{-11}\,,$$

which are the normalization factors used in the computation of the limits.

The normalization factors discussed above assume a signal mass window $\Delta m < 60\,\text{MeV}$ around the nominal B^0_s (B^0) masses, which reflects the choice performed for limit evaluation.

In addition a maximum likelihood fit in a wider mass range has been performed to determine the branching fraction. The normalization factors in this case have been updated to:

$$\alpha^{\rm noMW}_{B^0_s\to\mu^+\mu^-} = (2.52 \pm 0.23) \times 10^{-10}\,,$$
$$\alpha^{\rm noMW}_{B^0\to\mu^+\mu^-} = (6.45 \pm 0.30) \times 10^{-11}\,.$$

Given these measured normalization factors and assuming the SM rates, the $1.1\,\text{fb}^{-1}$ 2012 data sample is expected to contain 14.1 $B^0_s \to \mu^+\mu^-$ and 1.7 $B^0 \to \mu^+\mu^-$ decays.

5 Results

The unblinded distribution of events in the invariant mass–BDT plane is shown in Fig. 10 separately for 2012 and 2011 data sets.

The compatibility of the distribution of the observed events with that expected for a given branching fraction hypothesis is computed using the CL_s method [7]. This provides $\text{CL}_\text{s+b}$, a measure of the compatibility of the observed distribution with the signal plus background hypothesis, CL_b, a measure of the compatibility with the background-only hypothesis, and $\text{CL}_\text{s} = \text{CL}_\text{s+b}/\text{CL}_\text{b}$. As input to this computation we use the expected number of combinatorial background events, peaking background events, and signal events assuming the SM branching fractions, together with the number of observed events for each of the 63 bins, 7 bins in BDT and 9 bins in invariant mass, of the 2012 signal regions.

The systematic uncertainties on the signal predictions in each bin are computed by fluctuating the mass parameters, the BDT fractional yields and the normalization factors within Gaussian distributions defined by their associated uncertainties. The systematic uncertainty on the estimated number of combinatorial background events is computed by fluctuating within a Poissonian distribution the number of events measured in the sidebands, and by varying within 1 σ the value of the exponent.

The above signal and background expectations, integrated over all mass bins, are listed for 2012 data in Table 7, for each of the 7 BDT bins separately, and for both B^0_s and B^0 signal mass regions. The same information is also reported for 2011 data in Table 8 (where 8 BDT bins are considered).

Table 4: Summary of the factors and their uncertainty entering in the normalization for the two normalization channels considered. The value of $f_d/f_s = 0.256 \pm 0.020$ is used.

	$\mathcal{B}$	$\frac{\epsilon^{\rm rec}_{\rm norm}\epsilon^{\rm sel|rec}_{\rm norm}}{\epsilon^{\rm rec}_{\rm sig}\epsilon^{\rm sel|rec}_{\rm sig}}$	$\frac{\epsilon^{\rm trg|sel}_{\rm norm}}{\epsilon^{\rm trg|sel}_{\rm sig}}$	$N_{\rm norm}$	$\alpha^{\rm norm}_{B^0\to\mu^+\mu^-}$	$\alpha^{\rm norm}_{B^0_s\to\mu^+\mu^-}$
	$(\times 10^{-5})$				$(\times 10^{-11})$	$(\times 10^{-10})$
$B^+ \to J/\psi K^+$	6.01 ± 0.21	0.548 ± 0.018	0.932 ± 0.012	$424\,200 \pm 1500$	7.24 ± 0.39	2.83 ± 0.27
$B^0 \to K^+\pi^-$	1.94 ± 0.06	0.908 ± 0.031	0.057 ± 0.002	$14\,600 \pm 1100$	6.93 ± 0.67	2.71 ± 0.34

The expected and measured limits for $B^0 \to \mu^+\mu^-$ branching fraction at 90 % and 95 % C.L. for the 2012 data sample are shown in Table 5. The expected limits are computed allowing the presence of $B^0_{(s)} \to \mu^+\mu^-$ events according to the SM branching fractions, including cross-feed between the two modes.

Table 5: 2012 results: expected and observed limits on the $B^0 \to \mu^+\mu^-$ branching fraction.

Mode	Limit	at 90 % C.L.	at 95 % C.L.
$B^0 \to \mu^+\mu^-$	Exp. bkg+SM	8.5×10^{-10}	10.5×10^{-10}
	Exp. bkg	7.6×10^{-10}	9.6×10^{-10}
	Observed	10.5×10^{-10}	12.5×10^{-10}

The comparison of the distributions of observed events and expected background events results in a p-value ($1 - \mathrm{CL_b}$) of 9×10^{-4} for the $B^0_s \to \mu^+\mu^-$ and 16 % for the $B^0 \to \mu^+\mu^-$ decay, where the $\mathrm{CL_b}$ are computed at the branching fraction values corresponding to $\mathrm{CL_{s+b}} = 0.5$. We observe an excess of $B^0_s \to \mu^+\mu^-$ candidates with respect to background expectations with a significance of about 3.3 standard deviations.

A simultaneous unbinned maximum-likelihood fit to the mass projections in the 7 BDT bins has been performed on the full mass range to extract the $B^0_s \to \mu^+\mu^-$ branching ratio. This fit uses the same PDFs as in the combinatorial background interpolation, with the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions as additional free parameters. The fit gives:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (5.1\,^{+2.3}_{-1.9}(\mathrm{stat})\,^{+0.7}_{-0.4}(\mathrm{syst})) \times 10^{-9},$$

where the central value is extracted from the maximum of the logarithm of the profile likelihood. The statistical uncertainty reflects the interval corresponding to a change of 0.5 with respect to the maximum of the likelihood after fixing all the fit parameters to their expected values except the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions and the slope and normalization of the combinatorial background. The systematic uncertainty is obtained by subtracting in quadrature the statistical uncertainty from the total uncertainty obtained from the likelihood with all nuisance parameters left to vary according to their uncertainties. An additional systematic uncertainty of 0.16×10^{-9} reflects the impact on the result of the change in the parameterization of the combinatorial background from a single to a double exponential, and is added in quadrature.

Following the change in the combinatorial background determination with respect to Ref. [1], the upper limits and branching fractions have been also updated for the 2011 data. We obtain $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 5.1 \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 13 \times 10^{-10}$ at 95 % CL to be compared to the published limits $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 4.5 \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 10.3 \times 10^{-10}$ at 95 % CL [1], respectively. The p-value for $B^0_s \to \mu^+\mu^-$ changes from 18 % to 11 % and the $B^0_s \to \mu^+\mu^-$ branching fraction increases by $\sim 0.3\,\sigma$ from $(0.8\,^{+1.8}_{-1.3}) \times 10^{-9}$ to $(1.4\,^{+1.7}_{-1.3}) \times 10^{-9}$. This shift is compatible with the systematic uncertainty previously

assigned to the background shape [1]. The values of the $B^0_s \to \mu^+\mu^-$ branching fraction obtained with the 2011 and 2012 datasets are compatible within 1.5σ.

The 2011 and 2012 results are then combined by computing the CL_s and performing the maximum likelihood fit simultaneously to the eight and seven BDT bins of the 2011 and 2012 datasets, respectively. The parameters that are considered 100% correlated between the two datasets are the f_s/f_d value, the $\mathcal{B}(B^+ \to J/\psi\, K^+)$ and $\mathcal{B}(B^0 \to K^+\pi^-)$, the transition point of the Crystal Ball function that describes the signal mass lineshape, the mass distribution of the $B^0_{(s)} \to h^+h'^-$ peaking background, the BDT and the mass distributions of the $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^{+(0)} \to \pi^{+(0)}\mu^+\mu^-$ backgrounds and the SM predictions of the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions. The results for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays, integrated over all mass bins, for the simultaneous fit to the 2012 and 2011 dataset are summarized in Table 7 and Table 8.

The expected and observed CL_s values as a function of the branching fraction for the 2011+2012 combination are shown in Fig 11. The expected and observed upper limits for the $B^0 \to \mu^+\mu^-$ channel obtained from the combined 2011+2012 datasets are summarized in Table 6. The observed CL_b value at $CL_{s+b} = 0.5$ is 89 %. The probability that background processes can produce the observed number of $B^0_s \to \mu^+\mu^-$ candidates or more is 5×10^{-4} and corresponds to a statistical significance of about 3.5 standard deviations.

Table 6: Combination of 2011 and 2012 results: expected and observed limits on the $B^0 \to \mu^+\mu^-$ branching fraction.

Mode	Limit	at 90 % C.L.	at 95 % C.L.
$B^0 \to \mu^+\mu^-$	Exp. bkg+SM	5.8×10^{-10}	7.1×10^{-10}
	Exp. bkg	5.0×10^{-10}	6.0×10^{-10}
	Observed	8.0×10^{-10}	9.4×10^{-10}

From the simultaneous unbinned maximum likelihood fit on the full mass range to the mass projections on the 8 and 7 BDT bins of the 2011 and 2012 datasets, shown in Fig. 12, the $B^0_s \to \mu^+\mu^-$ branching ratio has been measured. The value of the $B^0_s \to \mu^+\mu^-$ branching fraction extracted from the fit is

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.2\,^{+1.4}_{-1.2}(\text{stat})\,^{+0.5}_{-0.3}(\text{syst})) \times 10^{-9},$$

where the statistical and the systematic uncertainties have the same meaning as before. The value is in agreement with the SM expectation. The log likelihood profile for the combined 2011-2012 fit is shown in Fig. 13

The distributions of the invariant mass of $B^0_s \to \mu^+\mu^-$ candidates for the combined 2011 and 2012 dataset and for three different BDT cuts (BDT>0.5, >0.7, and >0.8) are shown in Fig 14. In Fig. 15 one $B^0_s \to \mu^+\mu^-$ candidate falling into the BDT>0.8 sensitive

region is shown with the LHCb Event Display; the zoom of the $z-y$ projection of the same event is shown in Fig. 16.

Given the high probability of finding a signal, two-sided limits on the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ have been set using $\mathrm{CL_{s+b}}$. The true value of the $B_s^0 \to \mu^+\mu^-$ branching fraction is contained in the interval $[1.3, 5.8] \times 10^{-9}$($[1.1, 6.4] \times 10^{-9}$) at 90 % C.L. (95 % C.L.), where the lower and upper limit are the branching fractions evaluated at $\mathrm{CL_{s+b}} = 0.95$ ($\mathrm{CL_{s+b}} = 0.975$) and $\mathrm{CL_{s+b}} = 0.05$ ($\mathrm{CL_{s+b}} = 0.025$), respectively. These results are in good agreement with the lower and upper limits derived from integrating the profile likelihood obtained from the unbinned fit.

6 Conclusions

A search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ is performed using $1.0\,\mathrm{fb}^{-1}$ and $1.1\,\mathrm{fb}^{-1}$ of pp collision data collected at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, respectively. The data in the B^0 search window are consistent with the background expectation and the world's best upper limit of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 9.4 \times 10^{-10}$ at 95 % CL is obtained. The data in the B_s^0 search window show an excess of events with respect to the background-only prediction with a statistical significance of 3.5σ. A fit to the data leads to $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.2\,^{+1.5}_{-1.2}) \times 10^{-9}$ which is in agreement with the SM prediction.

References

[1] LHCb collaboration, R. Aaij *et al.*, *Strong constraints on the rare decays $B_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$*, Phys. Rev. Lett. **108** (2012) 231801, arXiv:1203.4493.

[2] P. Speckmayer, A. Hocker, J. Stelzer, and H. Voss, *The toolkit for multivariate data analysis, TMVA 4*, J. Phys. Conf. Ser. **219** (2010) 032057.

[3] H.-J. Yang, B. P. Roe, and J. Zhu, *Studies of boosted decision trees for MiniBooNE particle identification*, Nucl. Instrum. Meth. **A555** (2005) 370, arXiv:physics/0508045.

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

[5] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays $B/B_s \to (\pi, K)(l^+l^-, l\nu, \nu\bar{\nu})$ in the perturbative QCD approach beyond the leading-order*, arXiv:1207.0265.

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

[7] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. **G28** (2002) 2693.

Table 7: Expected combinatorial background, $B^0_{(s)} \to h^+h'^-$ background, cross-feed, and signal events assuming SM predictions, together with the number of observed events in the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ mass signal regions, in bins of BDT for the 2012 data sample.

Mode	BDT bin	0.0 – 0.25	0.25 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 1.0
$B^0_s \to \mu^+\mu^-$	Exp. comb. bkg	2345^{+40}_{-40}	$56.7^{+3.0}_{-2.9}$	$13.1^{+1.5}_{-1.4}$	$4.42^{+0.91}_{-0.81}$	$2.10^{+0.67}_{-0.56}$	$0.35^{+0.42}_{-0.22}$	$0.39^{+0.33}_{-0.21}$
	Exp. peak. bkg	$0.250^{+0.083}_{-0.068}$	$0.145^{+0.049}_{-0.040}$	$0.081^{+0.027}_{-0.023}$	$0.075^{+0.024}_{-0.020}$	$0.071^{+0.023}_{-0.019}$	$0.062^{+0.021}_{-0.017}$	$0.104^{+0.034}_{-0.028}$
	Exp. signal	$3.69^{+0.59}_{-0.52}$	$2.14^{+0.37}_{-0.33}$	$1.20^{+0.21}_{-0.18}$	$1.16^{+0.18}_{-0.16}$	$1.17^{+0.18}_{-0.16}$	$1.15^{+0.19}_{-0.17}$	$2.13^{+0.33}_{-0.29}$
	Observation	2274	65	19	5	3	1	3
$B^0 \to \mu^+\mu^-$	Exp. comb. bkg	2491^{+42}_{-42}	$59.5^{+3.3}_{-3.2}$	$13.9^{+1.6}_{-1.5}$	$4.74^{+1.00}_{-0.89}$	$2.10^{+0.74}_{-0.61}$	$0.55^{+0.50}_{-0.31}$	$0.29^{+0.34}_{-0.19}$
	Exp. peak. bkg	$1.49^{+0.50}_{-0.36}$	$0.86^{+0.29}_{-0.22}$	$0.48^{+0.16}_{-0.12}$	$0.44^{+0.15}_{-0.11}$	$0.42^{+0.14}_{-0.10}$	$0.369^{+0.126}_{-0.093}$	$0.62^{+0.21}_{-0.15}$
	Exp. cross-feed	$0.627^{+0.104}_{-0.091}$	$0.363^{+0.066}_{-0.057}$	$0.204^{+0.036}_{-0.032}$	$0.197^{+0.032}_{-0.027}$	$0.199^{+0.032}_{-0.028}$	$0.196^{+0.034}_{-0.030}$	$0.362^{+0.058}_{-0.051}$
	Exp. signal	$0.442^{+0.062}_{-0.057}$	$0.256^{+0.040}_{-0.036}$	$0.144^{+0.022}_{-0.020}$	$0.139^{+0.019}_{-0.017}$	$0.140^{+0.019}_{-0.018}$	$0.138^{+0.021}_{-0.019}$	$0.255^{+0.035}_{-0.031}$
	Observation	2433	59	19	3	2	2	2

Table 8: Expected combinatorial background, $B^0_{(s)} \to h^+h'^-$ background, cross-feed, and signal events assuming SM predictions, together with the number of observed events in the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ mass signal regions, in bins of BDT for the 2011 data sample.

Mode	BDT bin	0.0 – 0.25	0.25 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 0.9	0.9 – 1
$B^0_s \to \mu^+\mu^-$	Exp. comb. bkg	1880^{+33}_{-33}	$55.5^{+3.0}_{-2.9}$	$12.1^{+1.4}_{-1.3}$	$4.16^{+0.88}_{-0.79}$	$1.81^{+0.62}_{-0.51}$	$0.77^{+0.52}_{-0.38}$	$0.47^{+0.48}_{-0.36}$	$0.24^{+0.44}_{-0.20}$
	Exp. peak. bkg	$0.129^{+0.066}_{-0.050}$	$0.066^{+0.024}_{-0.019}$	$0.052^{+0.018}_{-0.015}$	$0.047^{+0.015}_{-0.013}$	$0.053^{+0.017}_{-0.014}$	$0.050^{+0.016}_{-0.013}$	$0.052^{+0.017}_{-0.014}$	$0.049^{+0.018}_{-0.014}$
	Exp. signal	$2.70^{+0.81}_{-0.80}$	$1.30^{+0.27}_{-0.23}$	$1.03^{+0.20}_{-0.17}$	$0.92^{+0.15}_{-0.13}$	$1.06^{+0.17}_{-0.15}$	$1.10^{+0.17}_{-0.15}$	$1.26^{+0.20}_{-0.17}$	$1.31^{+0.28}_{-0.25}$
	Observation	1818	39	12	6	1	2	1	1
$B^0 \to \mu^+\mu^-$	Exp. comb. bkg	1995^{+34}_{-34}	$59.2^{+3.3}_{-3.2}$	$12.6^{+1.6}_{-1.5}$	$4.44^{+0.99}_{-0.86}$	$1.67^{+0.66}_{-0.54}$	$0.75^{+0.58}_{-0.40}$	$0.44^{+0.57}_{-0.38}$	$0.22^{+0.48}_{-0.20}$
	Exp. peak. bkg	$0.78^{+0.38}_{-0.29}$	$0.40^{+0.14}_{-0.10}$	$0.311^{+0.107}_{-0.079}$	$0.280^{+0.092}_{-0.068}$	$0.314^{+0.103}_{-0.076}$	$0.297^{+0.096}_{-0.071}$	$0.309^{+0.101}_{-0.075}$	$0.296^{+0.107}_{-0.079}$
	Exp. cross-feed	$0.43^{+0.13}_{-0.13}$	$0.205^{+0.044}_{-0.037}$	$0.163^{+0.032}_{-0.027}$	$0.145^{+0.025}_{-0.021}$	$0.168^{+0.029}_{-0.025}$	$0.174^{+0.029}_{-0.024}$	$0.199^{+0.033}_{-0.029}$	$0.206^{+0.046}_{-0.040}$
	Exp. signal	$0.328^{+0.096}_{-0.097}$	$0.158^{+0.030}_{-0.027}$	$0.125^{+0.022}_{-0.019}$	$0.112^{+0.016}_{-0.015}$	$0.129^{+0.019}_{-0.017}$	$0.134^{+0.018}_{-0.016}$	$0.153^{+0.022}_{-0.019}$	$0.159^{+0.032}_{-0.029}$
	Observation	1904	50	20	5	2	1	4	1

[8] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$*, Phys. Lett. **B708** (2012) 55, arXiv:1112.1600.

[9] CDF Collaboration, A. Abulencia *et al.*, *Search for $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ decays in $p\bar{p}$ collisions with CDF II*, Phys. Rev. Lett. **95** (2005) 221805, arXiv:hep-ex/0508036.

[10] LHCb Collaboration, R. Aaij *et al.*, *Determination of f_s/f_d for* 7 TeV *pp collisions and a measurement of the branching fraction of the decay $B_d \to D^-K^+$*, Phys. Rev. Lett. **107** (2011) 211801, arXiv:1106.4435.

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

[12] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of fragmentation fractions f_s/f_d and dependence on B meson kinematics*, LHCb-PAPER-2012-037 (2012), in preparation.

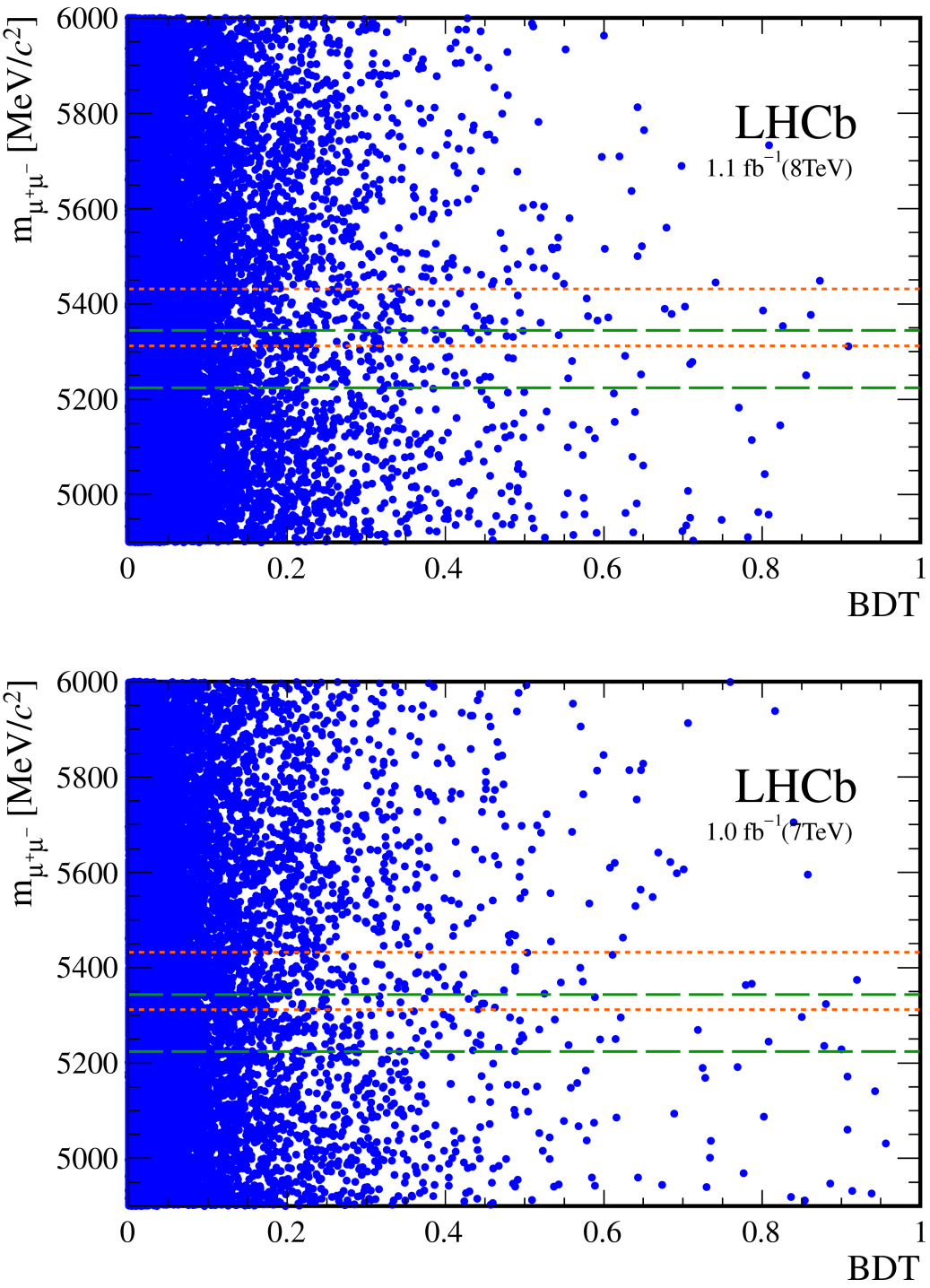

Figure 10: Dimuon mass versus BDT for selected candidates; orange short-dashed (green long-dashed) lines indicate the $\pm 60\,\mathrm{MeV}/c^2$ search window around the B_s^0 (B^0) mass. Top: unblinded 2012 data. Bottom: unblinded 2011 data.

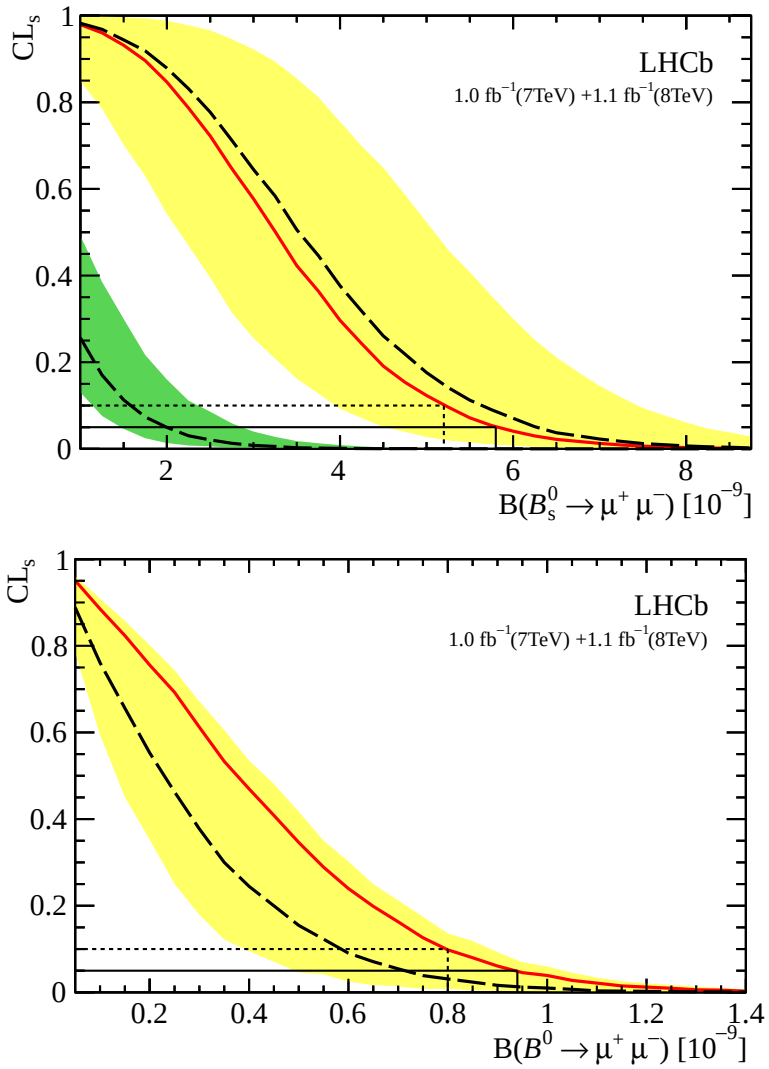

Figure 11: CL_s as a function of the assumed $\mathcal{B}$ for $B_s^0 \to \mu^+\mu^-$ (upper panel) and $B^0 \to \mu^+\mu^-$ (lower panel) decays for the combined 2011+2012 dataset. The long dashed gray curves are the medians of the expected CL_s distributions if background and SM signal were observed. The yellow area covers, for each $\mathcal{B}$, 34 % of the expected CL_s distribution on each side of its median. The solid red curves are the observed CL_s. For the $B_s^0 \to \mu^+\mu^-$ (upper panel), the long dashed gray curve in the green area is the expected CL_s distribution if background only was observed; the green area covers 34 % of the expected CL_s distribution on each side of the median.

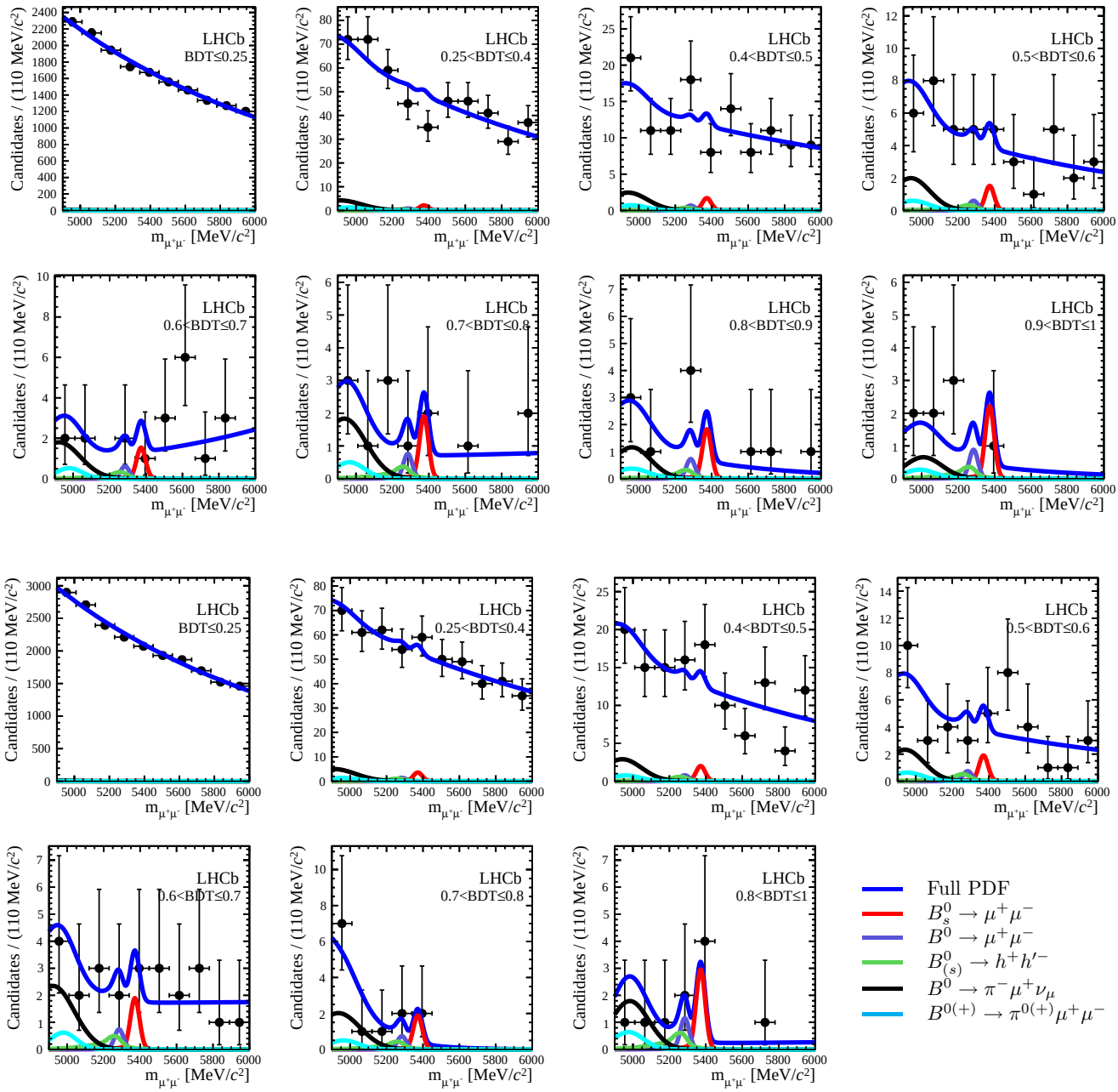

Figure 12: Simultaneous fit of the invariant mass distribution in the 8 BDT bins of 2011 (top) and 7 BDT bins of 2012 (bottom) data; black line is $B^0 \to \pi^-\mu^+\nu_\mu$, cyan is $B^{+(0)} \to \pi^{+(0)}\mu^+\mu^-$, green dashed is $B^0_{(s)} \to h^+h'^-$ misID, red is $B^0_s \to \mu^+\mu^-$ and, purple is $B^0 \to \mu^+\mu^-$. The fit result is superimposed in blue.

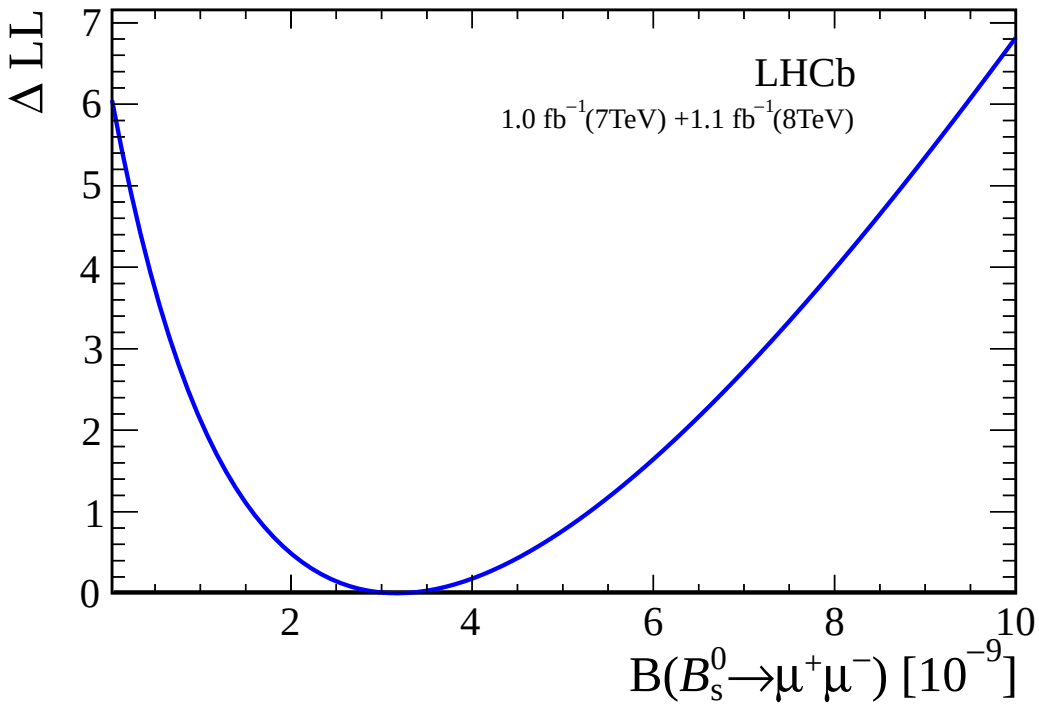

Figure 13: Likelihood as a function of the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ value for the simultaneous unnbinned fit on the full mass range of the 2011 and 2012 datasets.

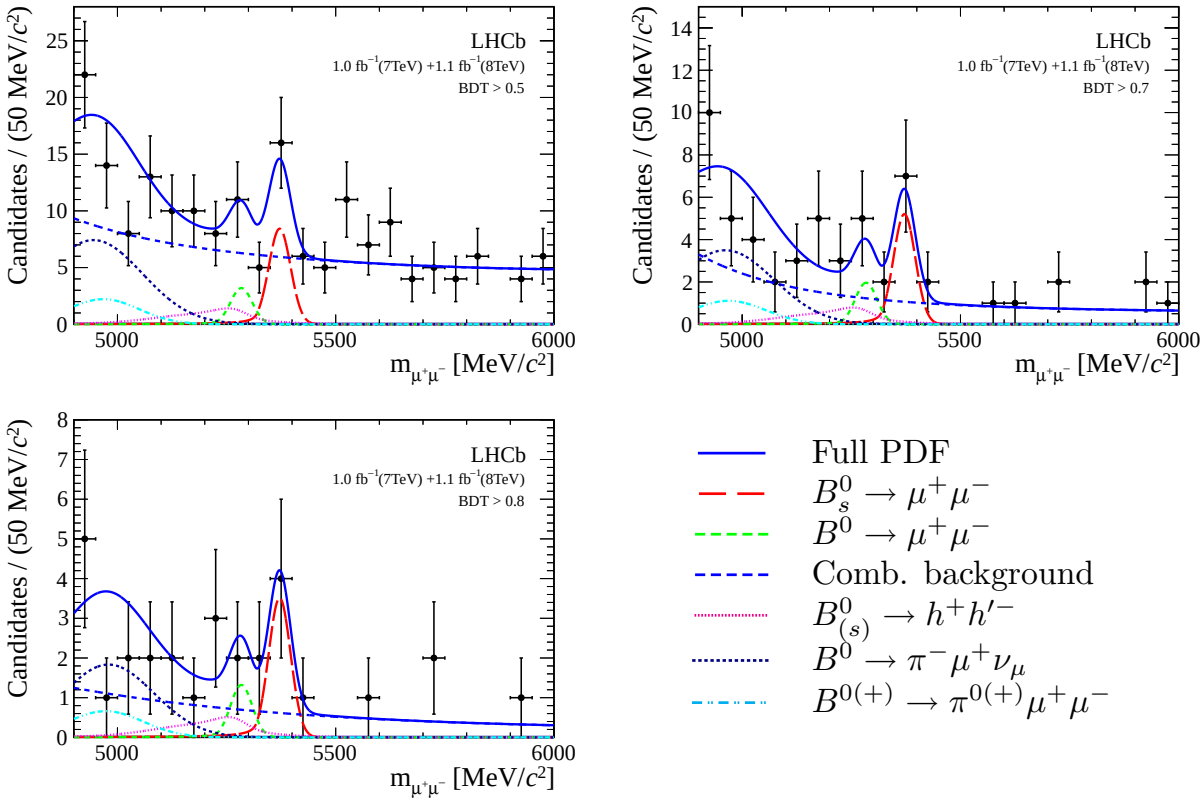

Figure 14: Invariant mass distribution of selected $B^0_s \to \mu^+\mu^-$ candidates (black points) for combined 2011 and 2012 dataset and for BDT>0.5, BDT>0.7, and BDT>0.8.

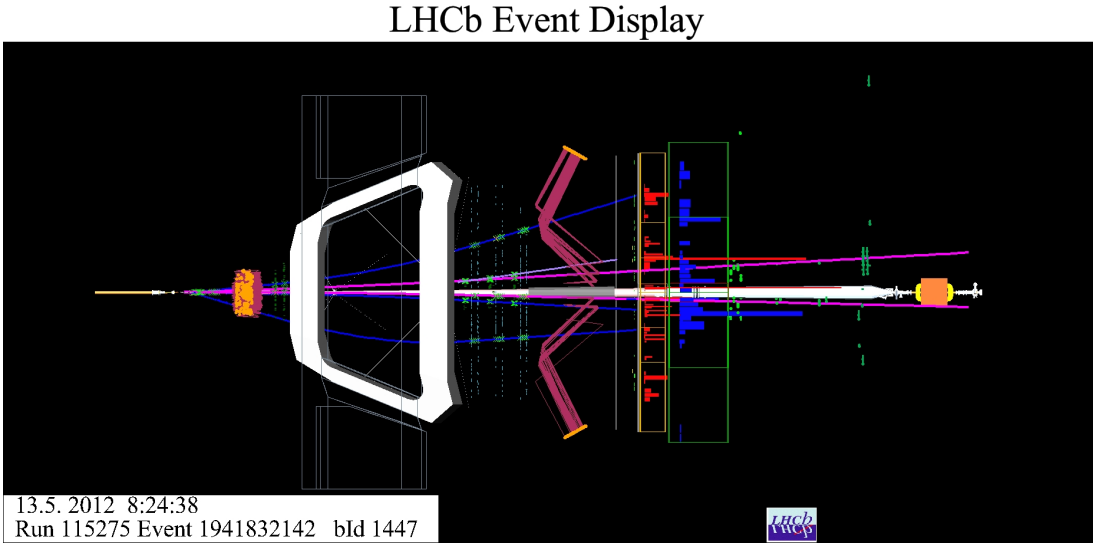

Figure 15: Event display of a $B^0_s \to \mu^+\mu^-$ candidate falling into the B^0_s most sensitive region in the invariant mass versus BDT plane: BDT=0.826 and $m_{\mu\mu} = 5353\,\mathrm{MeV}/c^2$.

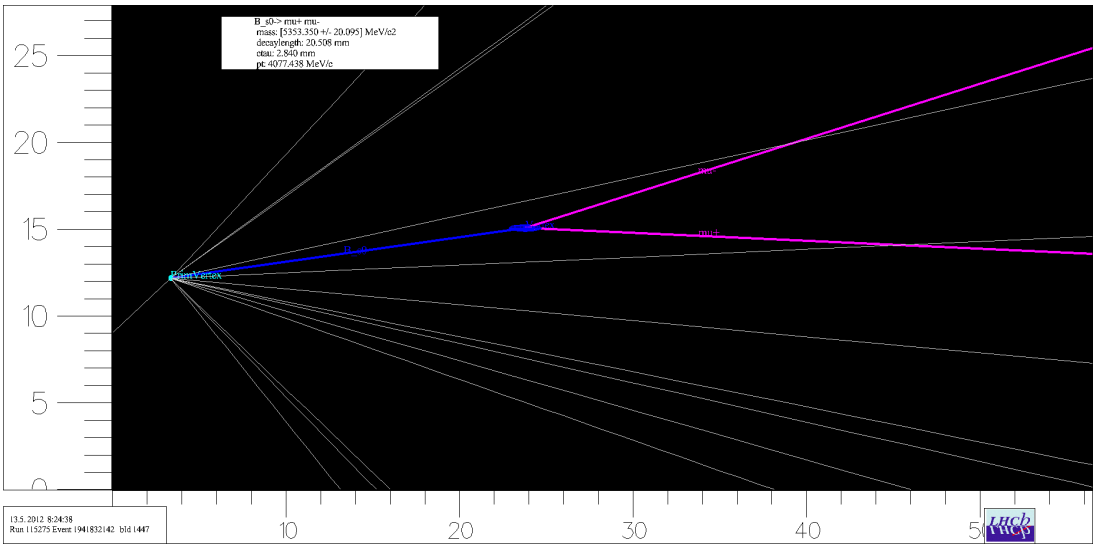

Figure 16: Zoom of the $z - y$ projection of the same event shown in Fig. 15.