

Search for the rare decays

$$B^0_s \rightarrow \mu^+ \mu^- \text{ and } B^0 \rightarrow \mu^+ \mu^-$$

1-Nov-2011

Hiroshi Nakano

Contents

1. Introduction

2. The LHCb detector

3. Analysis strategy

4. Event selection

5. Evaluation of the normalization factor

6. Signal and background likelihoods

7. Results & 8. Conclusions

1. Introduction

P1

Physics Letters B 698 (2011) 336–340

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$

LHCb Collaboration

ARTICLE INFO

Article history:
Received 13 March 2011
Received in revised form 12 April 2011
Accepted 13 April 2011
Available online 20 April 2011
Editor: W.O. Szafron

ABSTRACT

A search for the decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ is performed with about 37 pb⁻¹ of pp collisions at $\sqrt{s} = 7$ TeV collected by the LHCb experiment at the Large Hadron Collider at CERN. The observed numbers of events are consistent with the background expectations. The resulting upper limits on the branching ratios are $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 0.6 \times 10^{-8}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.5 \times 10^{-8}$ at 95% confidence level.

© 2011 CERN. Published by Elsevier B.V. All rights reserved.

Keywords:
LHC
b-factory
RHC
Rare decays
Lepton decay

1. Introduction

Within the Standard Model (SM) exclusive decays of the B^0 and B_s^0 mesons¹ are rare as they occur only via loop diagrams and are helicity suppressed. New Physics models, especially those with an extended Higgs sector, can significantly enhance the branching fractions, although in some models the rates are lower.

The amplitudes contributing to the branching ratio $\mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)$ (where $q = d, s$ for the B^0 and B_s^0 mesons respectively) can be expressed in terms of the scalar (c_s), pseudoscalar (c_p) and axial vector (c_a) Wilson coefficients in a completely general approach [1]. Within the SM, the contributions of c_s and c_p are negligible while c_a is calculated with an accuracy of a few percent [2]. The dominant contribution stems from an electromagnetic penguin with a Z^0 decaying into two muons. However, the accuracy of the SM prediction for $\mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)$ is limited by the knowledge of the decay constants of the B_q^0 mesons. This limitation can be reduced by normalizing to the well-measured mass differences of the B_q^0 mesons. Using this approach [3], the SM predictions are

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{SM} = (0.32 \pm 0.22) \times 10^{-8},$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)_{SM} = (0.070 \pm 0.001) \times 10^{-8}.$$

Many extensions to the SM predict a very different Higgs sector. For instance, within the Minimal Supersymmetric SM (MSSM)

in the large $\tan\beta$ approximation [4], $c_{SM} \propto \tan^2 \beta / M_h^2$, where M_h denotes the pseudoscalar Higgs mass and $\tan\beta$ the ratio of Higgs vacuum expectation values. The most restrictive limits on the search for $B_q^0 \rightarrow \mu^+ \mu^-$ have so far been achieved at the Tevatron, due to the large $b\bar{b}$ cross-section at hadron colliders. The best limits at 95% CL, published so far are obtained using 6.3 fb^{-1} by the D0 Collaboration [5], $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 5.1 \times 10^{-8}$, and using 2.8 fb^{-1} by the CDF Collaboration [6], $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 5.8 \times 10^{-8}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.8 \times 10^{-8}$. The CDF Collaboration has also presented preliminary results [7], that lower the limits to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 4.3 \times 10^{-8}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 0.78 \times 10^{-8}$.

The LHCb experiment is well suited for such searches due to its good invariant mass resolution, vertex resolution, muon identification and trigger acceptance. In addition, LHCb has a hadronic trigger capability which provides large samples of $B_q^0 \rightarrow \mu^+ \mu^-$ decays, where B and μ stand for a hadron (baryon or pion). These are used as control samples in order to reduce the dependence of the results on the simulation.

The measurements in this Letter use about 37 pb⁻¹ of integrated luminosity collected by LHCb between July and October 2010 at $\sqrt{s} = 7$ TeV. Assuming the SM branching ratios, about 0.7 (0.08) B_s^0 ($B^0 \rightarrow \mu^+ \mu^-$) are expected to be reconstructed using the $b\bar{b}$ cross-section, measured within the LHCb acceptance, of $75 \pm 14 \mu\text{b}$ [8].

2. The LHCb detector

The LHCb detector [9] is a single-arm forward spectrometer with an angular coverage from approximately 10 mrad to 300

¹ © CERN, for the benefit of the LHCb Collaboration.
² In this Letter the inclusion of charge-conjugate rates is implicit.

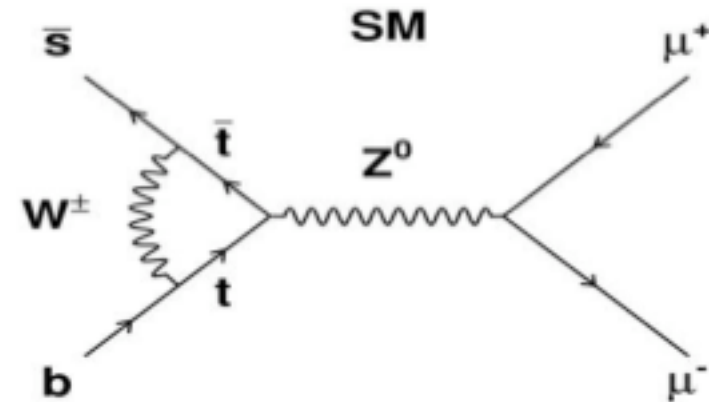
0168-8896/\$ – see front matter © 2011 CERN. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.physletb.2011.04.020

SM prediction and New Physics

Branching ratio of $B_{s,d}^0 \rightarrow \mu^+ \mu^-$ in SM

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{SM} = (0.32 \pm 0.02) \times 10^{-8},$$

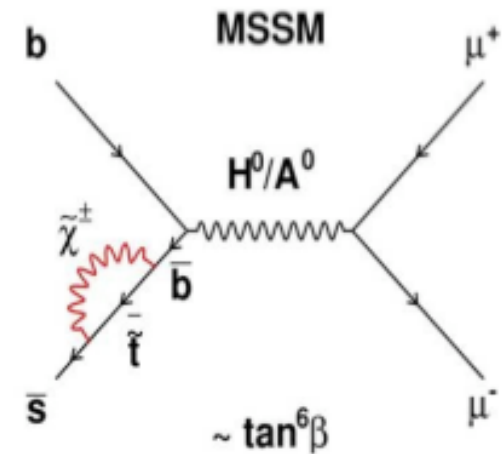
$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)_{SM} = (0.010 \pm 0.001) \times 10^{-8}.$$



Example: MSSM
(with R-parity conservation)

$$BR(B_s \rightarrow \mu^+ \mu^-) \propto \frac{\tan^6 \beta}{m_A^4}$$

→ limit or measurement of $B_{s,d} \rightarrow \mu\mu$
will strongly constrain $\tan\beta$ vs M_A plane



$B \rightarrow \mu\mu$ is one of the Golden mode for NP in LHCb

2. The LHCb detector

P1

Physics Letters B 599 (2015) 336–340

Comments lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$

LHCb Collaboration

ARTICLE INFO

ABSTRACT

Keywords: LHC; Higgs boson; Rare decay; Leptonic decay

1. Introduction

Within the Standard Model (SM) exclusive di-lepton decays of the B^0 and B_s^0 mesons¹ are rare as they occur only via loop diagrams and are helicity suppressed. New Physics models, especially those with an extended Higgs sector, can significantly enhance the branching fractions, although in some models the rates are low-energetic.

The amplitudes contributing to the branching ratio $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (where $q = d, s$ for the B^0 and B_s^0 mesons respectively) can be expressed in terms of the scalar (c_s), pseudoscalar (c_p) and axial vector (c_A) Wilson coefficients in a completely general approach [1]. Within the SM, the contributions of c_s and c_p are negligible, while c_A is calculated with an accuracy of a few percent [2]. The dominant contribution stems from an electromagnetic penguin with a Z^0 decaying into two muons. However, the accuracy of the SM prediction for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is limited by the knowledge of the decay constants of the B_s^0 meson. This limitation can be reduced by normalizing to the well-measured mass differences of the B_s^0 mesons. Using this approach [3], the SM prediction are:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{SM} = (0.32 \pm 0.02) \times 10^{-9},$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)_{SM} = (0.019 \pm 0.001) \times 10^{-9}.$$

Many extensions to the SM predict a very different Higgs sector. For instance, within the Minimal Supersymmetric SM (MSSM)

¹ © CERN, for the benefit of the LHCb Collaboration.
² In this letter the inclusion of charge-conjugate states is implied.
 0370-269X/© 2015 CERN. Published by Elsevier B.V. All rights reserved.
 doi:10.1016/j.physletb.2015.04.020

in the large $\tan\beta$ approximation [4], $\alpha_s^{(SM)} \times (\tan\beta)^2/M_W$, where M_W denotes the pseudoscalar Higgs mass and $\tan\beta$ the ratio of Higgs vacuum expectation values. The most restrictive limits on the search for $B_s^0 \rightarrow \mu^+ \mu^-$ have so far been achieved at the LHC, due to the large $b\bar{b}$ cross-section at hadron colliders. The best limits at 95% CL, published so far, are obtained using 5.1 fb^{-1} by the D0 Collaboration [5], $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 3.3 \times 10^{-7}$, and using 2 fb^{-1} by the CDF Collaboration [6], $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 5.8 \times 10^{-7}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.8 \times 10^{-7}$. The CDF Collaboration has also presented preliminary results [7] with 3.7 fb^{-1} , that lower the limits to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 4.3 \times 10^{-7}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 0.78 \times 10^{-7}$.

The LHCb experiment is well suited for such searches due to its good invariant mass resolution, vertex resolution, muon identification and trigger acceptance. In addition, LHCb has a hadronic trigger capability which provides large samples of $B_s^0 \rightarrow \mu^+ \mu^-$ decays, where b and \bar{b} stand for a b-hadron (baryon or meson). These are used as control samples in order to reduce the dependence of the results on the simulation.

The measurements in this Letter use about 37 fb^{-1} of integrated luminosity collected by LHCb between July and October 2010 at $\sqrt{s} = 7 \text{ TeV}$. Assuming the SM branching ratio, about 0.7 fb^{-1} $B_s^0 \rightarrow \mu^+ \mu^-$, ($B^0 \rightarrow \mu^+ \mu^-$) are expected to be reconstructed using the full cross-section, measured within the LHCb acceptance, of $75 \pm 14 \mu\text{b}$ [8].

2. The LHCb detector

The LHCb detector [9] is a single-arm forward spectrometer with an angular coverage from approximately 10° to 300°

P2

LHCb Collaboration / Physics Letters B 599 (2015) 336–340

331

starts of a vertex locator (VELO), a warm dipole magnet with a bending power of $\int B dl = 4 \text{ Tm}$, a tracking system, two ring-traiting Cherenkov detectors (RICH), a calorimeter system and a muon system. The VELO consists of a series of silicon modules, each providing a measure of the radial and azimuthal coordinates, with the sensitive area starting at 8 mm from the beam line during collisions. The tracking system comprises four layers of silicon sensors in the inner part and straw tubes in the outer part after the magnet. Track momenta are measured with a precision between $\delta p/p = 0.15\%$ at 5 GeV/c and $\delta p/p = 0.5\%$ at 100 GeV/c. The RICH system provides charged hadron identification in a momentum range 2–100 GeV/c. The calorimeter system consists of a pre-shower, a scintillating pad detector, an electromagnetic calorimeter and a hadronic calorimeter. It identifies high transverse energy (E_T) hadron, electron and photon candidates and provides information for the trigger. Five muon stations composed of MWTX (except in the highest rate region, where triple-GEMs are used) provide fast information for the trigger and muon identification capability.

LHCb has a two-level trigger system both for leptonic and purely hadronic final states. It exploits the finite lifetime and relatively large mass of charm and beauty hadrons to distinguish heavy flavour decays from the dominant light quark processes. The first trigger level (L1) is implemented in hardware and reduces the rate to a maximum of 1 MHz, the read-out rate of the whole detector. The second trigger level (High Level Trigger: HLT) is implemented in software running on an event filter CPU farm. In the first stage of the software trigger (HLT1) a partial event reconstruction is performed. The second stage (HLT2) performs a full event reconstruction to enhance the signal purity further.

The forward geometry of LHCb allows the first level trigger to collect events containing one or two muons with very low transverse momenta (p_T); more than 90% of the data were collected with a p_T threshold of 1.4 GeV/c for single muon triggers and $p_T(\mu_1) > 0.48 \text{ GeV/c}$ and $p_T(\mu_2) > 0.56 \text{ GeV/c}$ for di-muon triggers. The E_T threshold for the hadron trigger varied in the range 2.8 to 3.8 GeV. The single muon trigger line in the HLT requires either $p_T > 1.8 \text{ GeV/c}$ or includes a cut on the impact parameter ($|P|$) with respect to the primary vertex, which allows for a lower p_T requirement ($p_T > 0.8 \text{ GeV/c}$, $|P| > 0.11 \text{ mm}$). The di-muon trigger line requires muon pairs of opposite charge forming a common vertex and an invariant mass $M_{\mu\mu} > 4.7 \text{ GeV}/c^2$. A second trigger line, primarily to select J/ψ events, requires $2.95 < M_{\mu\mu} < 3.21 \text{ GeV}/c^2$. The remaining region of the di-muon invariant mass is also covered by trigger lines that in addition require the di-muon secondary vertex to be well separated from the primary vertex. Other HLT trigger lines select generic displaced vertices, providing a high efficiency for purely hadronic decays (the looser $B_s^0 \rightarrow \mu^+ \mu^-$).

3. Analysis strategy

An important feature of this analysis is its reliance on data as opposed to Monte Carlo simulation. Nevertheless, some Monte Carlo (MC) simulation has been used, based on the Pythia 6.4 generator [11] and the Geant4 package [12] for detector simulation. The first part of the analysis is the event selection (Section 4), which significantly reduces the size of the dataset by rejecting most of the background.

The second part consists of the study of three normalization channels with known branching ratios: $B^0 \rightarrow J/\psi \mu^+ \mu^-$, $B_s^0 \rightarrow J/\psi \mu^+ \mu^-$ and $B^0 \rightarrow K^* \mu^+ \mu^-$. Using each of these normalization channels, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ can be calculated

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \frac{\mathcal{N}(B_s^0 \rightarrow \mu^+ \mu^-)}{\mathcal{N}(B_s^0 \rightarrow J/\psi \mu^+ \mu^-)} \times \frac{\mathcal{B}(B_s^0 \rightarrow J/\psi \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)}$$

$$= \frac{\mathcal{N}(B_s^0 \rightarrow \mu^+ \mu^-)}{\mathcal{N}(B_s^0 \rightarrow J/\psi \mu^+ \mu^-)} \times \frac{\mathcal{N}(B_s^0 \rightarrow J/\psi \mu^+ \mu^-)}{\mathcal{N}(B_s^0 \rightarrow \mu^+ \mu^-)} \times \frac{\mathcal{B}(B_s^0 \rightarrow J/\psi \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)}$$

$$= \frac{\mathcal{N}(B_s^0 \rightarrow \mu^+ \mu^-)}{\mathcal{N}(B_s^0 \rightarrow J/\psi \mu^+ \mu^-)} \times \frac{\mathcal{N}(B_s^0 \rightarrow J/\psi \mu^+ \mu^-)}{\mathcal{N}(B_s^0 \rightarrow \mu^+ \mu^-)} \times \frac{\mathcal{B}(B_s^0 \rightarrow J/\psi \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)}$$

where $\mathcal{N}(B_s^0 \rightarrow \mu^+ \mu^-)$ denotes the normalization factor, $f_{B_s^0}$ denotes the probability that a b -quark fragments into a B_s^0 and f_{B^0} denotes the probability that a b -quark fragments into the B^0 -hadron relevant for the chosen normalization channel with branching fraction \mathcal{B}_{hadron} . The reconstruction efficiency $\epsilon^{(B_s^0)}$ includes the acceptance and particle identification, while $\epsilon^{(B^0)}$ denotes the selection efficiency on reconstructed events. The trigger efficiency on selected events is denoted by $\epsilon^{(HLT)}$. This normalization ensures that knowledge of the absolute luminosity and $b\bar{b}$ production cross-section are not needed, and that many systematic uncertainties cancel in the ratio of the efficiencies. The event selection for these channels is specifically designed to be as close as possible to the signal selection. The ratios of reconstruction and selection efficiencies are estimated from the simulations, while the ratios of trigger efficiencies on selected events are determined from data (Section 5).

In the third part of the analysis (Section 6) each selected event is given a probability to be signal or background in a two-dimensional probability space defined by the di-muon invariant mass and a geometrical likelihood (GL). The di-muon invariant mass and GL probability density functions for both signal and background are determined from data. This procedure ensures that even though the GL is defined using simulated events, the result will not be biased by discrepancies between data and simulation.

Section 7 describes the final measurement. In order to avoid uncertainties bias in the analysis, the invariant mass region for the signal ($M_{\mu\mu} \in [60 \text{ MeV}/c^2 \text{ and } M_{\mu\mu} \in [0 \text{ MeV}/c^2]$) was divided into the selection criteria and analysis procedure had been defined.

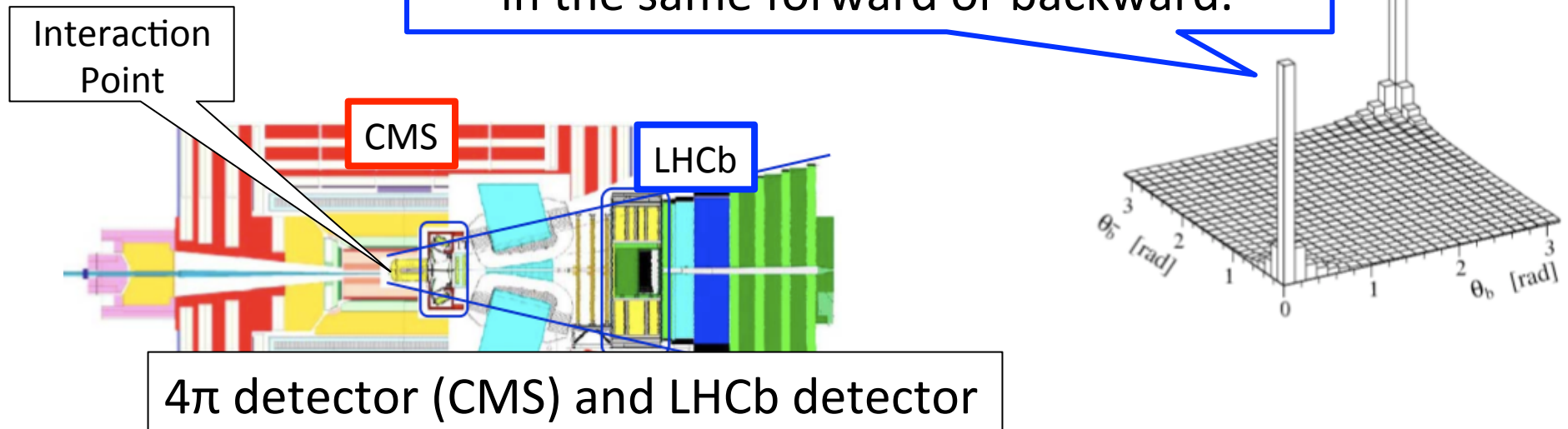
4. Event selection

The selection has been designed in order to reduce the data sample to a manageable level by simultaneously keeping the efficiency for the signals as high as possible and the selection between signal and background is kept as simple as possible. This last requirement is needed to minimize the systematic uncertainty in the ratio of the selection efficiencies. The optimal separation between signal and background is left to the likelihoods (Section 6). The basic cut of the selection have been defined on Monte Carlo simulation [10] and fit adapted to the data.

The data for the signal and all the normalization candidates are selected using either an inclusive two-body or a J/ψ selection. Tracks are first required to be of good quality ($\chi^2_{\text{ndf}} < 5$) and to be divided with respect to the closest primary vertex ($\chi^2_{\text{dof}} > 12.5$, where χ^2_{dof} is the difference between the χ^2 of the primary vertex built with and without the considered track). To reject bad combinations before performing the vertex fit, the two tracks are required to have a distance of closest approach of less than 0.3 mm. The secondary vertex is required to be well fitted ($\chi^2_{\text{ndf}} < 5$) and must be clearly separated from the primary vertex in the forward direction (vertex distance significance larger than 15). When more than one primary vertex is reconstructed, the one that gives the minimum impact parameter significance for the candidate is chosen. The reconstructed candidate has to point to the primary vertex ($\chi^2_{\text{ndf}} < 12.5$) in the case of the inclusive two-body selection; for all selections, the primary vertex is re-

LHCb

bb pairs are predominantly produced in the same forward or backward.



Rough estimate for B acceptance:
compare $B^\pm \rightarrow J/\psi K^\pm$ yield with CDF / D0

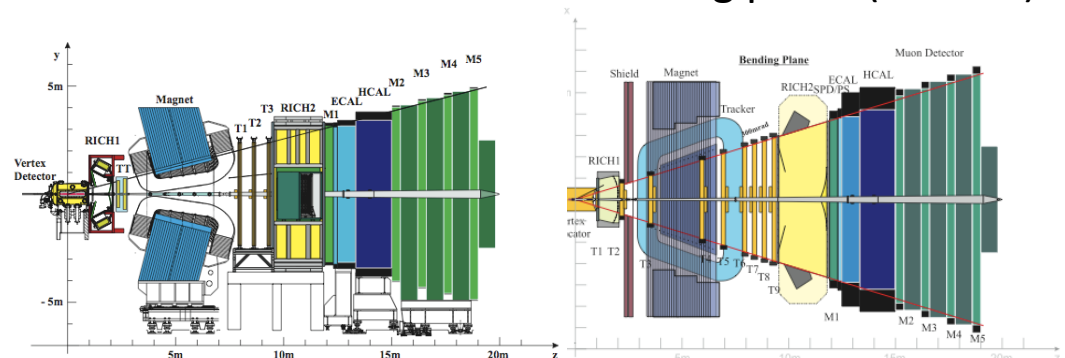
- LHCb

$N_{\text{signal}}: 12,366 \pm 403^{\text{stat+syst}} \quad (0.037 \text{fb}^{-1})$

- CDF (CMU-CM(U+X))

$N_{\text{signal}}: 19,762 \pm 203^{\text{stat+syst}} \quad (3.7 \text{fb}^{-1})$

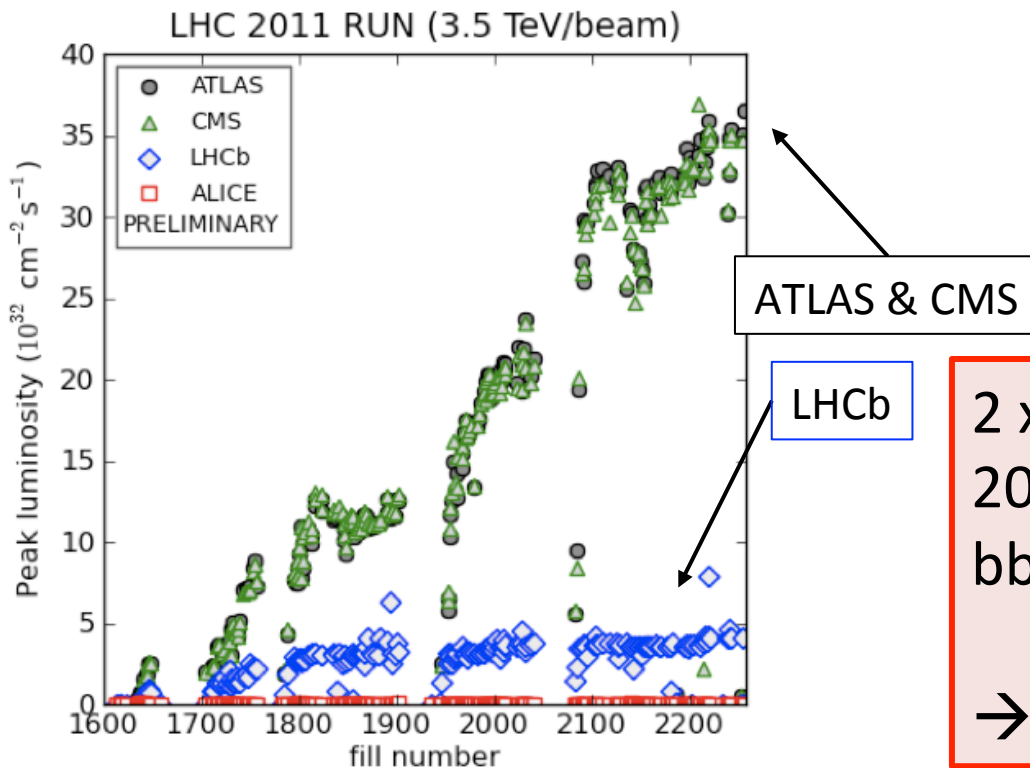
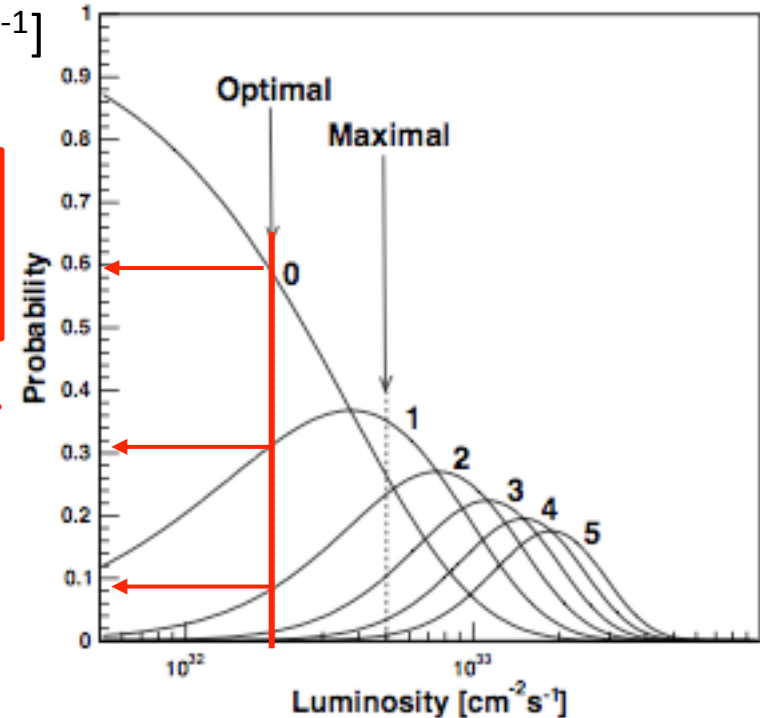
10-300mrad in the bending plane (horizontal)
10-250mrad in the non-bending plane (vertical)



Luminosity

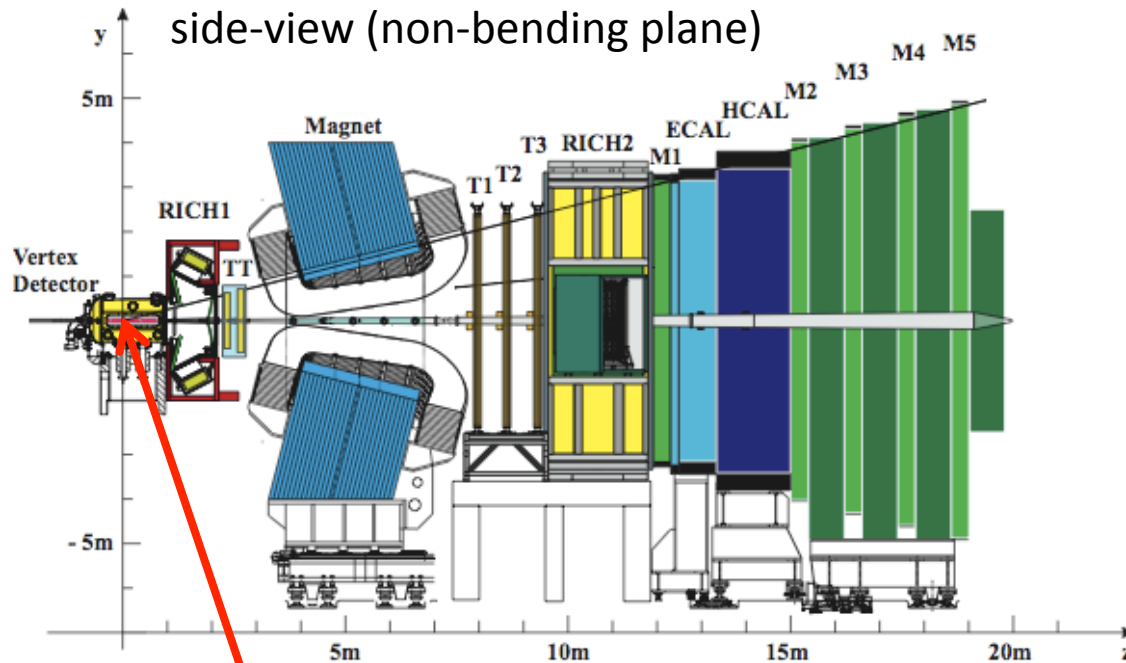
LHCb 's luminosity is tuned to $2 \times 10^{32} [\text{cm}^{-2}\text{s}^{-1}]$
(defocusing the beams)

0 [events/cross] (~60%)
1 [events/cross] (~30%)
2 [events/cross] (~10%)



$2 \times 10^{32} [\text{cm}^{-2}\text{s}^{-1}] = 200 [\mu\text{b}^{-1}\text{s}^{-1}]$
 $200[\mu\text{b}^{-1}\text{s}^{-1}] \times 10^7[\text{s}/\text{year}] = 2[\text{fb}^{-1}]$
 bb cross section @ 14 TeV $\sim 500[\mu\text{b}]$
 @ 7 TeV $\sim 300[\mu\text{b}]$
 $\rightarrow \sim 10^{12}[\text{bb}/\text{year}]$

Detector (1/4)



Vertex locator
(silicon micro strip)

reconstruct
primary vertex (b production)
& secondary vertex (b, c decay)

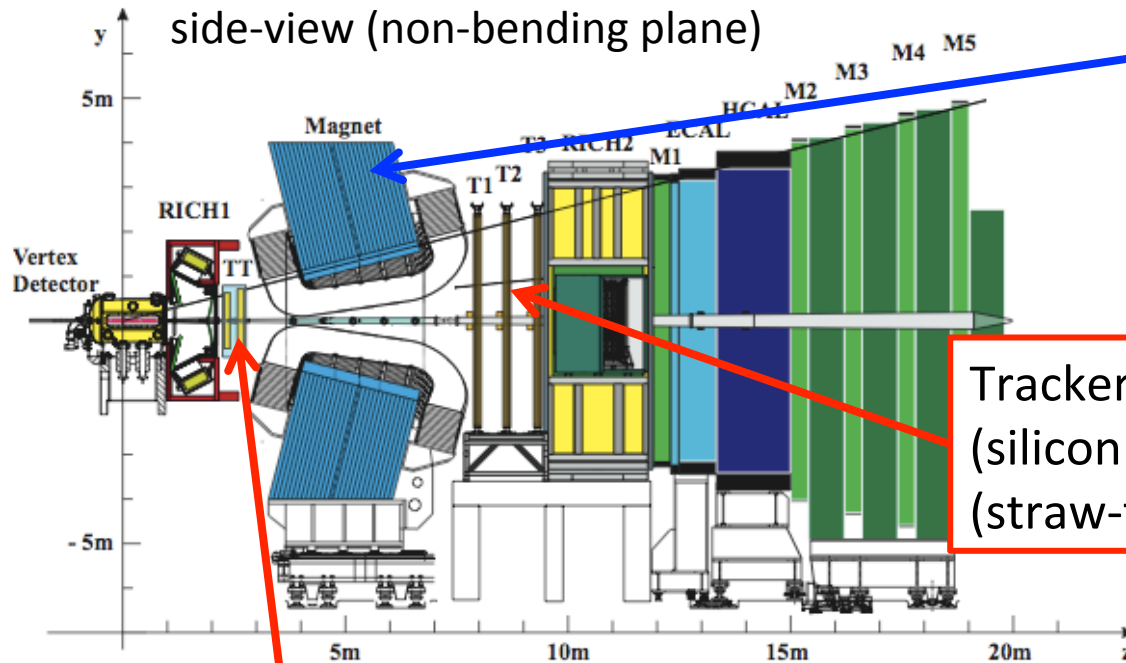
Primary vertex resolutions (25 tracks):

	LHCb [μm]	ATLAS [μm]	CMS [μm]
$\sigma(x)$	15.8	60	20-40
$\sigma(y)$	15.2	60	20-40
$\sigma(z)$	76	100	40-60

Impact parameter resolution (Belle)

$$\sigma_{xy} = 19 \oplus 50 / (p\beta \sin^{3/2} \theta) \mu\text{m} \quad \text{and} \quad \sigma_z = 36 \oplus 42 / (p\beta \sin^{5/2} \theta) \mu\text{m}.$$

Detector (2/4)



A warm dipole magnet

$$\int Bdl = 4Tm$$

Tracker stations (3 layers)
(silicon micro strip for small angle region)
(straw-tubes for large angle region)

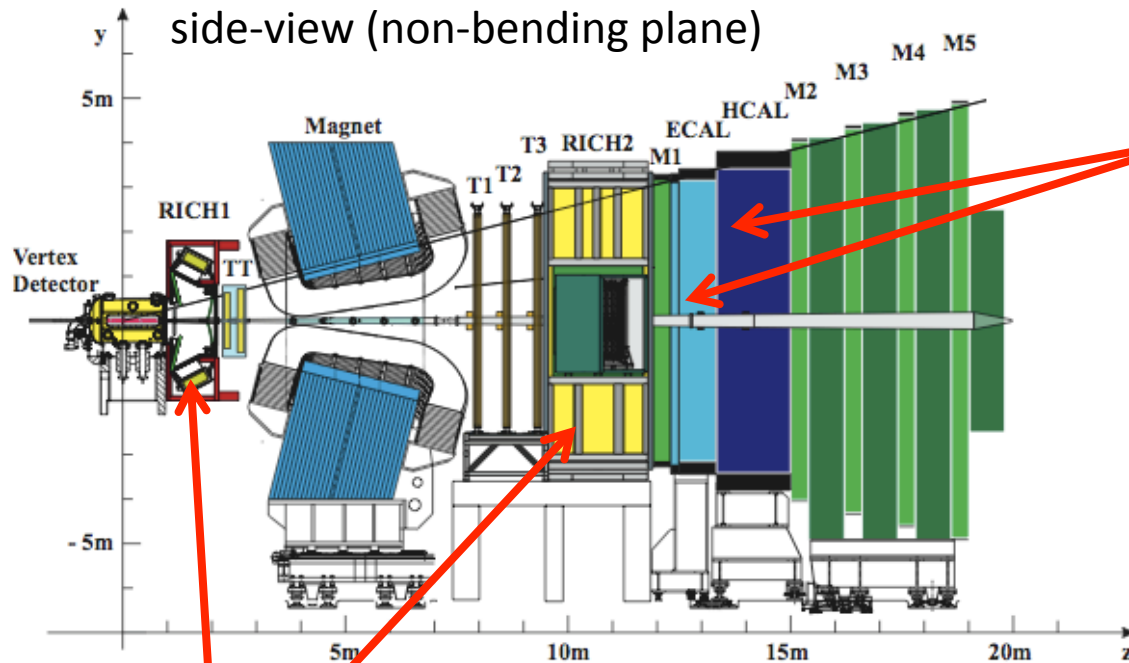
$$\begin{aligned} \delta p/p &= 0.35 \% (@5\text{GeV}/c) \\ &= 0.5 \% (@100\text{GeV}/c) \end{aligned}$$

Trigger Tracker
(4 layers of
silicon micro strip)

	momentum resolution	mass resolution $J/\psi \rightarrow \mu\mu$
LHCb	$\delta p/p = 0.4-0.6 \%$	13 MeV
CMS	$\delta p_t/p_t = 1-3 \%$	40 MeV
ATLAS	$\delta p_t/p_t = 5-6 \%$	71 MeV

Belle $\delta p_t/p_t = 0.2 p_t \oplus 0.3/\beta \%$

Detector (3/4)



Calorimeter system
 ECAL + HCAL
 ID high E_T hadron, e, gamma
 → “L0” trigger

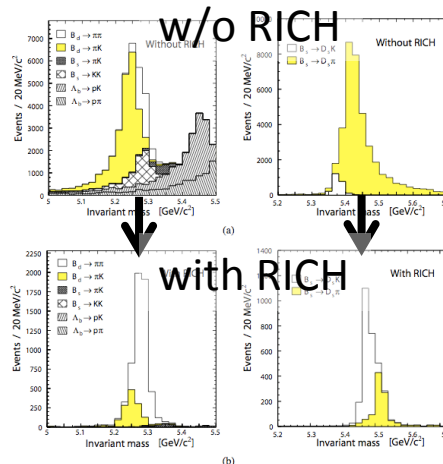
ECAL

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E[\text{GeV}]}} \oplus 1.5\%$$

HCAL

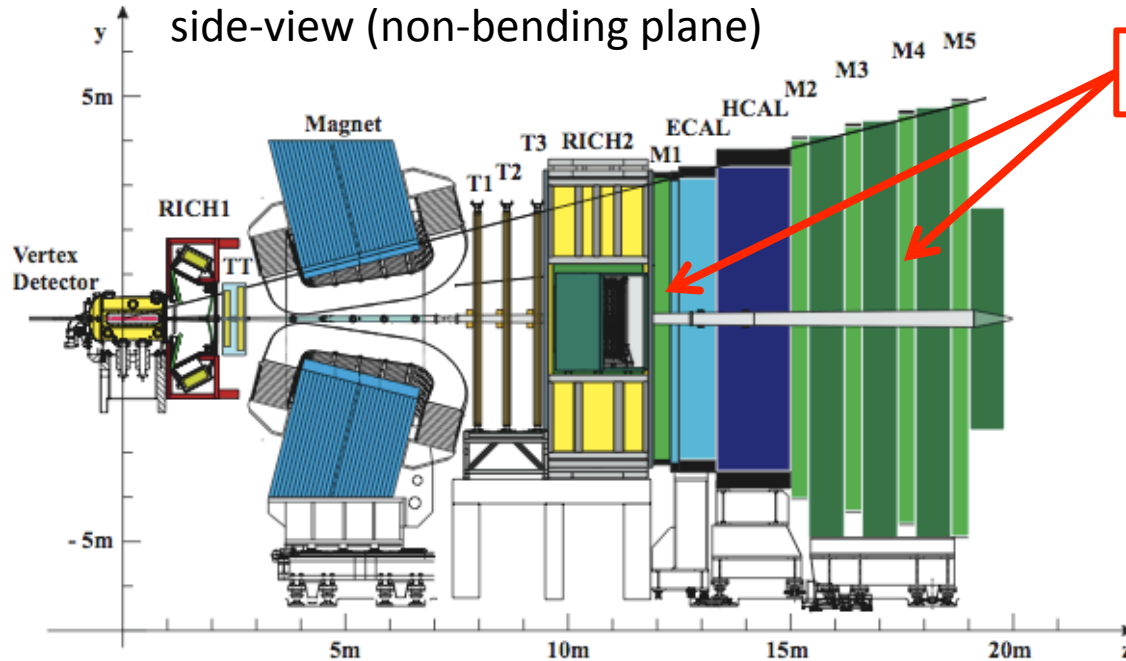
$$\frac{\sigma_E}{E} = \frac{80\%}{\sqrt{E[\text{GeV}]}} \oplus 10\%$$

RICH system
 Charged hadron ID
 (p range: 2 – 100 GeV)



	ECAL	HCAL
LHCb	25 X_0	----
	1.2 λ_1	5.6 λ_1
Belle	ECL	KLM
	16 X_0	----
	0.8 λ_1	3.9 λ_1

Detector (4/4)



Muon systems (1+ 4 layers)

Gold-plated CP sensitive decays

$$B_d^0 \rightarrow J/\psi(\mu^+ \mu^-) K_s^0$$

$$B_s^0 \rightarrow J/\psi(\mu^+ \mu^-) \phi$$

Golden modes for New Physics:

$$B \rightarrow \mu^+ \mu^-$$

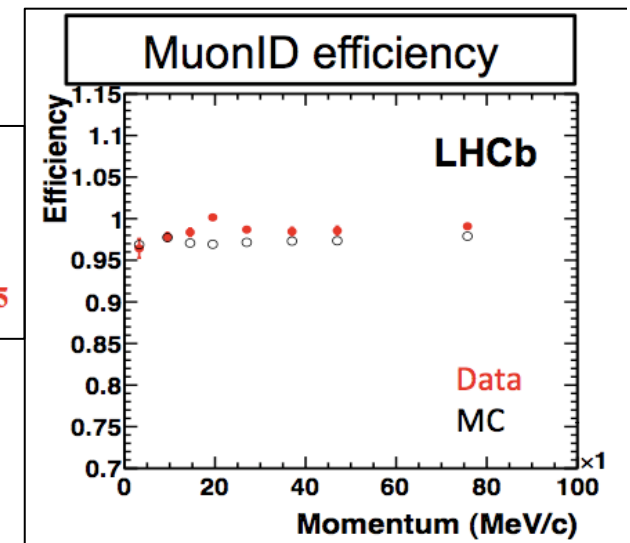
$$B \rightarrow K^* \mu^+ \mu^-$$

- Performance in the kinematic range of $B_s \rightarrow \mu\mu$:

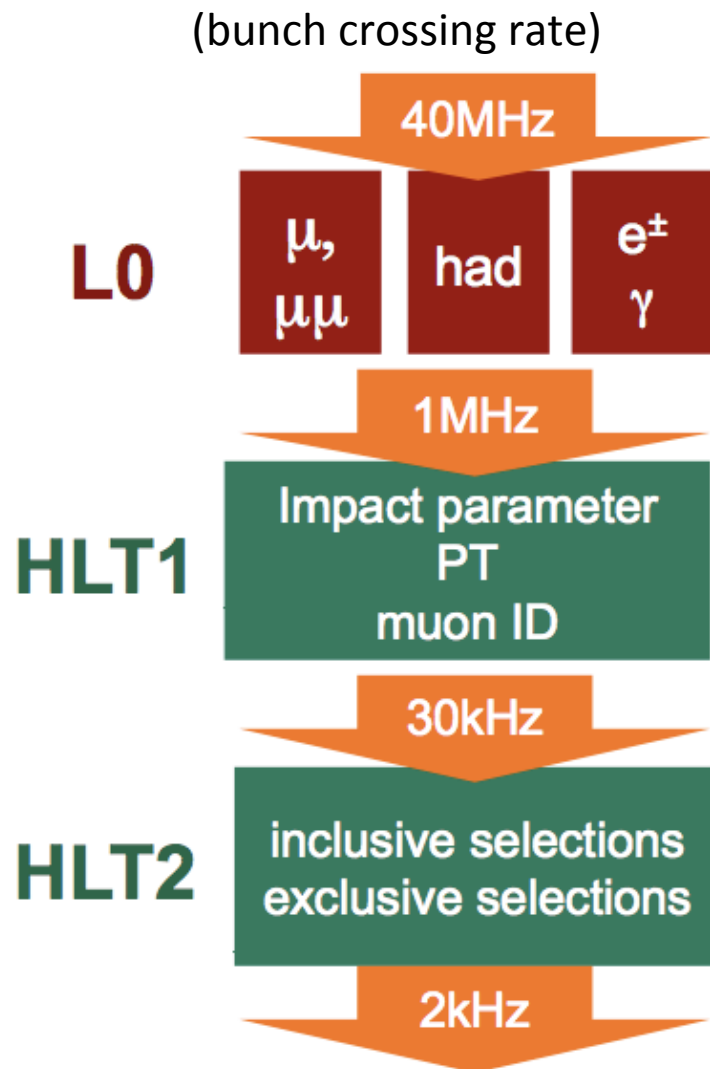
$$\epsilon(\mu \rightarrow \mu) \sim (97.1 \pm 1.3)\%$$

$$\epsilon(h \rightarrow \mu) \sim (7.1 \pm 0.5) 10^{-3}$$

$$\epsilon(hh \rightarrow \mu\mu) \sim (3.5 \pm 0.9) 10^{-5}$$



Trigger



L0 trigger

Pileup system \rightarrow discriminate single pp interactions
calorimeter trigger for gamma, e, hadron

muon trigger

(Single mu: $pt > 1.4 \text{ GeV}$
or di mu: $pt > 0.48 \text{ GeV}, 0.56 \text{ GeV}$)

High Level Trigger 1+2

using L0 information, tracking, vertexing

ex. for muon

$pt > 1.8 \text{ GeV}$
or $IP > 0.11 \text{ mm}, Pt > 0.8 \text{ GeV}$
for di muon
invariant mass is calculated

3. Analysis strategy

P2

LHC Collaboration (Phys. Rept. 896 (2017) 1-160)

[250] (rad) in the bending (non-bending) plane. The detector consists of a vertex locator (VTL), a warm dipole magnet with a bending power of $\int B dl = 4 \text{ Tm}$, a tracking system, two ring-imaging Cherenkov detectors (RICH), a calorimeter system and a muon system. The VTL consists of a series of silicon modules, each providing a measure of the radial and azimuthal coordinates, with the sensitive area starting at 8 mm from the beam line during collisions. The tracking system comprises four layers of silicon sensors before the magnet and three stations equipped with silicon sensors in the inner part and straw tubes in the outer part after the magnet. Track momenta are measured with a precision between $\delta p/p = 0.15\%$ at 5 GeV/c and $\delta p/p = 0.5\%$ at 100 GeV/c. The RICH system provides charged hadron identification in a momentum range 3–100 GeV/c. The calorimeter system consists of a protonium, a scintillating pad detector, an electromagnetic calorimeter and a hadronic calorimeter. It identifies high transverse energy (E_T) hadron, electron and photon candidates and provides information for the trigger. Five muon stations composed of MWPC (except in the highest rate region, where triple-GEMs are used) provide fast information for the trigger and muon identification capability.

LHCb has a two-level trigger system both for leptonic and purely hadronic final states. It exploits the finite lifetime and relatively large mass of charm and beauty hadrons to distinguish heavy flavour decays from the dominant light quark processes. The first trigger level (L1) is implemented in hardware and reduces the rate to a maximum of 1 MHz, the read-out rate of the whole detector. The second trigger level (High Level Trigger, HLT) is implemented in software running on an event filter CPU farm. In the first stage of the software trigger (HLT1) a partial event reconstruction is performed. The second stage (HLT2) performs a full event reconstruction to enhance the signal purity further.

The forward geometry of LHCb allows the first level trigger to collect events containing one or two muons with very low transverse momenta (p_T), more than 90% of the data were collected with a p_T threshold of 1.4 GeV/c for single muon triggers and $p_{T(\mu_1)} > 0.48 \text{ GeV/c}$ and $p_{T(\mu_2)} > 0.36 \text{ GeV/c}$ for dimuon triggers. The p_T threshold for the hadron trigger varied in the range 2.8 to 3.6 GeV. The single muon trigger line in the HLT requires either $p_T > 1.8 \text{ GeV/c}$ or includes a cut on the impact parameter (3σ) with respect to the primary vertex, which allows for a looser p_T requirement ($p_T > 0.8 \text{ GeV/c}$, $3\sigma > 0.11 \text{ mm}$). The dimuon trigger line requires muon pairs of opposite charge forming a common vertex and an invariant mass $M_{\mu\mu} > 4.7 \text{ GeV}/c^2$. A second trigger line, primarily to select J/ψ events, requires $2.90 < M_{\mu\mu} < 3.21 \text{ GeV}/c^2$. The remaining region of the dimuon invariant mass is also covered by trigger lines that in addition require the dimuon secondary vertex to be well separated from the primary vertex. Other HLT trigger lines select generic displaced vertices, providing a high efficiency for purely hadronic decays (for instance $B_s^0 \rightarrow K^+ K^-$).

3. Analysis strategy

An important feature of this analysis is to rely as much as possible on data and to restrict to a minimum the use of simulation. Nevertheless, some Monte Carlo (MC) simulation has been used, based on the PYTHIA 6.4 generator [11] and the GEANT4 package [12] for detector simulation. The first part of the analysis is the event selection (Section 4), which significantly reduces the size of the dataset by rejecting most of the background.

The second part consists of the study of three normalization channels with known branching ratios: $B^+ \rightarrow J/\psi(\mu^+ \mu^-) K^+$, $B^0 \rightarrow J/\psi(\mu^+ \mu^-) K^0$ and $B^0 \rightarrow K^{*0}(\mu^+ \mu^-) K^0$. In the case of the inclusive two-body selection, the $B\text{BR}[B^+ \rightarrow \mu^+ \mu^-]$ can be calculated

$$\begin{aligned}
 \text{BR}[B^+ \rightarrow \mu^+ \mu^-] &= \frac{\Gamma(B^+ \rightarrow \mu^+ \mu^-)}{\Gamma_{\text{tot}}(B^+)} \\
 &= \frac{\Gamma_{\text{charm}} \times \frac{\alpha_{\text{EM}}^2}{4\pi} \times \frac{|V_{cb}|^2 |V_{ub}|^2}{|V_{cb}|^2 |V_{ub}|^2} \times \frac{f_{\text{charm}}}{f_{B^+}} \times \frac{N_{B^+ \rightarrow \mu^+ \mu^-}}{N_{\text{charm}}}}{\Gamma_{\text{charm}} \times \frac{\alpha_{\text{EM}}^2}{4\pi} \times \frac{|V_{cb}|^2 |V_{ub}|^2}{|V_{cb}|^2 |V_{ub}|^2} \times \frac{f_{\text{charm}}}{f_{B^+}} \times \frac{N_{B^+ \rightarrow \mu^+ \mu^-}}{N_{\text{charm}}}} \\
 &= \theta_{B^+ \rightarrow \mu^+ \mu^-} \times N_{B^+ \rightarrow \mu^+ \mu^-}^{-1}
 \end{aligned}$$

where $\theta_{B^+ \rightarrow \mu^+ \mu^-}$ denotes the normalization factor, f_{B^+} denotes the probability that a b -quark fragments into a B^+ and f_{charm} denotes the probability that a b -quark fragments into the b -hadron relevant for the charm normalization channel with branching fraction BR_{charm} . The reconstruction efficiency $\epsilon^{\text{MC}}(B^+)$ includes the acceptance and particle identification, while $\epsilon^{\text{MC}}(B^+)$ denotes the selection efficiency on reconstructed events. The trigger efficiency on selected events is denoted by ϵ . This normalization scheme that knowledge of the absolute luminosity and $b\bar{b}$ production cross-section are not needed, and that many systematic uncertainties cancel in the ratio of the efficiencies. The event selection for these channels is specifically designed to be as close as possible to the signal selection. The ratios of reconstruction and selection efficiencies are estimated from the simulation, while the ratios of trigger efficiencies on selected events are determined from data (Section 5).

In the third part of the analysis (Section 6) each selected event is given a probability to be signal or background in a two-dimensional probability space defined by the dimuon invariant mass and a geometrical likelihood (GL). The dimuon invariant mass and GL probability density functions for both signal and background are determined from data. This procedure ensures that even though the GL is defined using simulated events, the result will not be biased by discrepancies between data and simulation.

Section 7 describes the final measurement. In order to avoid unnecessary bias in the analysis, the invariant mass region for the signal ($M_{\mu\mu} \in [60 \text{ MeV}/c^2 \text{ and } M_{\mu\mu} \in [80 \text{ MeV}/c^2]$) was blinded until the selection criteria and analysis procedure had been defined.

4. Event selection

The selection has been designed in order to reduce the data sample to a manageable level by simultaneously keeping the efficiency for the signal as high as possible and the selection between signal and control channels as similar as possible. This last requirement is needed to minimize the systematic uncertainty in the ratio of the selection efficiencies. The optimal separation between signal and background is left to the likelihood (Section 6). The basic cuts of the selection have been defined on Monte Carlo simulation [10] and then adapted to the data.

The data for the signal and all the normalization candidates are selected using either an inclusive two-body or a J/ψ selection. Tracks are first required to be of good quality ($\chi^2/\text{ndf} < 5$) and to be displaced with respect to the closest primary vertex ($|\chi^2/\text{ndf}| > 12.5$, where χ^2_{PV} is the difference between the χ^2 of the primary vertex both with and without the considered track). To reject bad combinations before performing the vertex fit, the two tracks are required to have a distance of closest approach of less than 0.5 cm. The secondary vertex is required to be well fitted ($\chi^2/\text{ndf} < 9$) and must be clearly separated from the primary in the forward direction (vertex distance significance larger than 1%). When more than one primary vertex is reconstructed, the one that gives the maximum impact parameter significance for the candidate is chosen. The reconstructed candidate has to point to the primary vertex ($|\chi^2/\text{ndf}| < 12.5$) in the case of the inclusive two-body selection. For all selections, the primary vertex is self-

4. Event selection

5. Evaluation of the normalization factor

6. Signal and background likelihoods

4. Event selection

P2

LHC Collaboration / Physics Letters B 695 (2011) 236–240

311

[250] mrad in the bending (non-bending) plane. The detector consists of a vertex locator (VELO), a warm dipole magnet with a bending power of $fB\ell = 4.7$ Tm, a tracking system, two ring-imaging Cherenkov detectors (RICH), a calorimeter system and a muon system. The VELO consists of a series of silicon modules, each providing a measure of the radial and azimuthal coordinates, with the sensitive area starting at 8 cm from the beam line during collisions. The tracking system comprises four layers of silicon sensors in the inner part and straw tubes in the outer part after the magnet. Track momenta are measured with a precision of $\delta p/p = 0.33\%$ at 5 GeV/c and $\delta p/p = 0.5\%$ at 100 GeV/c. The RICH system provides charged hadron identification in a momentum range 2–100 GeV/c. The calorimeter system consists of a preshower, a scintillating pad detector, an electromagnetic calorimeter and a hadronic calorimeter. It identifies high-transverse energy (E_T) hadrons, electrons and photons candidates and provides information for the trigger. Five muon stations composed of MWPC (except in the highest rate region, where triple-GEMs are used) provide fast information for the trigger and muon identification capability.

LHCb has a two-level trigger system both for leptonic and purely hadronic final states. It exploits the fast detector and relatively large mass of charm and beauty hadrons to distinguish heavy flavour decays from the dominant light quark processes. The first trigger level (L1) is implemented in hardware and reduces the rate to a maximum of 1 MHz, the read-out rate of the whole detector. The second trigger level (High Level Trigger, HLT) is implemented in software running on an event filter CPU farm. In the first stage of the software trigger (HLT1) a partial event reconstruction is performed. The second stage (HLT2) performs a full event reconstruction to enhance the signal purity further.

The forward geometry of LHCb allows the first level trigger to collect events containing one or two muons with very low transverse momenta (p_T), more than 90% of the data were collected with a p_T threshold of 1.4 GeV/c for single muon triggers and $p_T(\mu_1) > 0.48$ GeV/c and $p_T(\mu_2) > 0.56$ GeV/c for double muon triggers. The E_T threshold for the hadron trigger varied in the range 3.8 to 38 GeV. The single muon trigger line in the HLT requires either $p_T > 1.8$ GeV/c or inclusion a cut on the impact parameter $|P|$ with respect to the primary vertex, which allows for a lower p_T requirement ($p_T > 0.8$ GeV/c, $|P| > 0.11$ mm). The di-muon trigger line requires muon pair of opposite charge forming a common vertex and an invariant mass $M_{\mu\mu} > 4.7$ GeV/c². A second trigger line, primarily to select J/ψ events, requires $2.00 < M_{\mu\mu} < 3.20$ GeV/c². The remaining region of the invariant mass is also covered by trigger lines that in addition require the di-muon secondary vertex to be well separated from the primary vertex. Other HLT trigger lines select generic displaced vertices, providing a high efficiency for purely hadronic decays (the invariant $M^2 < 8^2$ GeV²).

3. Analysis strategy

An important feature of this analysis is to rely as much as possible on data and to restrict to a minimum the use of simulation. Nevertheless, some Monte Carlo (MC) simulation has been used, based on the PYTHIA 6.4 generator [11] and the GEANT4 package [12] for detector simulation. The first part of the analysis is the event selection (Section 4), which significantly reduces the size of the dataset by rejecting most of the background.

The second part consists of the study of three normalization channels with known branching ratios: $B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+$, $B^0 \rightarrow J/\psi(\mu^+\mu^-)K^0$ and $B^0 \rightarrow \mu^+\mu^-$. Using each of these normalization channels, $B\mathcal{B}[B^+ \rightarrow \mu^+\mu^-]$ can be calculated

$$B\mathcal{B}[B^+ \rightarrow \mu^+\mu^-] = \frac{N_{\text{obs}}(B^+ \rightarrow \mu^+\mu^-)}{N_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-)} \times \frac{f_{\text{tag}}}{f_{\text{tag}}^{\text{MC}}} \times \frac{N_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-)}{N_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-)}$$

where $N_{\text{obs}}(B^+ \rightarrow \mu^+\mu^-)$ denotes the number of $B^+ \rightarrow \mu^+\mu^-$ events, f_{tag} denotes the probability that a b -quark fragments into a B^+ and $f_{\text{tag}}^{\text{MC}}$ denotes the probability that a b -quark fragments into the b -hadron relevant for the chosen normalization channel with branching fraction $\mathcal{B}_{\text{norm}}$. The reconstruction efficiency $\epsilon^{\text{MC}}(B^+)$ includes the acceptance and particle identification while $\epsilon^{\text{MC}}(B^+)$ denotes the selection efficiency on reconstructed events. The trigger efficiency on selected events is denoted by $\epsilon^{\text{TRIG}}(B^+)$. This normalization ensures that knowledge of the absolute luminosity and b production cross-section are not needed, and that many systematic uncertainties cancel in the ratio of the efficiencies. The event selection for these channels is specifically designed to be as close as possible to the signal selection. The ratios of reconstruction and selection efficiencies are estimated from the simulation, while the ratios of trigger efficiencies on selected events are determined from data (Section 5).

In the third part of the analysis (Section 6) each selected event is given a probability to be signal or background in a two-dimensional probability space defined by the di-muon invariant mass and a geometrical likelihood [12]. The di-muon invariant mass and CL probability density functions for both signal and background are determined from data. This procedure ensures that even though the CL is defined using simulated events, the result will not be biased by discrepancies between data and simulation.

Section 7 describes the final measurement. In order to avoid uncertainties bias in the analysis, the invariant mass region for the signal ($M_{\mu\mu} > 60$ MeV/c² and $M_{\mu\mu} < 100$ MeV/c²) was blinded and the selection criteria and analysis procedure had been defined.

4. Event selection

The selection has been designed in order to reduce the data sample to a manageable level by simultaneously keeping the efficiency for the signal as high as possible and the selection between signals and control channels as similar as possible. This requirement is needed to minimize the systematic uncertainty in the ratio of the selection efficiencies. The optimal separation between signal and background is left to the kinematics (Section 6). The basic cut of the selection has been defined on Monte Carlo simulation [10] and then adapted to the data.

The data for the signal and all the normalization candidates are selected using either an inclusive two-body or a J/ψ selection. Tracks are first required to be of good quality ($\chi^2/\text{ndf} < 5$) and to be displaced with respect to the closest primary vertex ($\chi^2_{\text{rel}}/\text{ndf} > 12.5$, where χ^2_{rel} is the difference between the χ^2 of the primary vertex both with and without the considered track, its respect has contribution before performing the vertex fit, the two tracks are required to have a distance of closest approach of less than 0.3 mm. The secondary vertex is required to be well fitted ($\chi^2/\text{ndf} < 9$) and must be clearly separated from the primary vertex (the forward direction vertex distance significance larger than 15). When more than one primary vertex is reconstructed, the one that gives the minimum impact parameter significance to the candidate is chosen. The reconstructed candidate has to point to the primary vertex ($\chi^2_{\text{rel}}/\text{ndf} < 12.5$) in the case of the inclusive two-body selection, for all selections, the primary vertex is selected

P3

LHC Collaboration / Physics Letters B 695 (2011) 236–240

311

Table 1
Summary of the factors and their uncertainties needed to calculate the normalization factors ($\mathcal{B}_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-)$) for the three normalization channels considered. The branching ratios are taken from Ref. [13,16]. The trigger efficiency and number of B^+ candidates (intercepted) in only 25 events, as described in the text.

	$B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+$	$B^0 \rightarrow J/\psi(\mu^+\mu^-)K^0$	$B^0 \rightarrow \mu^+\mu^-$
$\mathcal{B}_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-)$	5.98 ± 0.23	0.40 ± 0.02	0.96 ± 0.05
$\mathcal{B}_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-) \times \epsilon^{\text{TRIG}}(B^+)$	3.6 ± 0.9	0.25 ± 0.02	0.96 ± 0.05
$\mathcal{B}_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-) \times \epsilon^{\text{TRIG}}(B^+) \times f_{\text{tag}}^{\text{MC}}$	1.94 ± 0.26	0.82 ± 0.06	0.02 ± 0.02
$N_{\text{obs}}(B^+ \rightarrow \mu^+\mu^-)$	$12,366 \pm 403$	8.4 ± 1.3	$2,27 \pm 0.8$
$\epsilon_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-)$	1.03 ± 2.8	1.03 ± 2.8	1.03 ± 2.8
$\epsilon_{\text{MC}}(B^+ \rightarrow \mu^+\mu^-) \times \epsilon^{\text{TRIG}}(B^+)$	1.03 ± 2.8	1.03 ± 2.8	1.03 ± 2.8

and excluding the signal tracks before calculating the $\chi^2_{\text{rel}}/\text{ndf}$ and vertex distance significance of the candidate. Tracks are defined as muons if they have at least one hit in two to four of the last four muon stations depending on the momentum. In the inclusive J/ψ selection both tracks must be identified as muons and have an invariant mass within 60 MeV/c² of the nominal J/ψ mass. The efficiency of the muon identification requirement has been measured using an inclusive sample of J/ψ events where one of the tracks does not use any information from the muon chambers. The efficiency measured with data agrees with MC expectations as a function of momentum within 2%, and the residual differences are taken into account in the systematic uncertainties.

Events passing the two-body selection are considered $B^+ \rightarrow \mu^+\mu^-$ candidates if both tracks pass the muon identification criteria, and their invariant mass lies within 60 MeV/c² of the nominal $B^+ \rightarrow \mu^+\mu^-$ mass. The invariant mass of the $B^+ \rightarrow \mu^+\mu^-$ candidates has to be within 900 MeV/c² of the nominal B^+ mass. As the acceptance of the tracking station is larger than the muon chamber, the selected $B^+ \rightarrow \mu^+\mu^-$ candidates are required to have both tracks within the muon chamber acceptance to minimize the differences with $B^+ \rightarrow \mu^+\mu^-$. The total efficiency including acceptance, reconstruction and selection criteria on MC $B^+ \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ events are 5.5% and 4.5% respectively; the main difference is due to material interaction. Assuming the SM branching ratio, 0.3 $B^+ \rightarrow \mu^+\mu^-$ and 0.04 $B^0 \rightarrow \mu^+\mu^-$ events are expected after all selection requirements. There are 143 (242) $B^+ \rightarrow \mu^+\mu^-$ candidates identified from data in the B^+ (B^0) mass window.

The dominant background after the $B^+ \rightarrow \mu^+\mu^-$ selection is expected to be $b\bar{b} \rightarrow \mu\mu X$ [14,15]. This is confirmed by comparing the kinematical distributions of the sideband data with a $b\bar{b} \rightarrow \mu\mu X$ MC sample. The muon identification probability as a function of momentum obtained from data using $B^+ \rightarrow \mu^+\mu^-$, $B^0 \rightarrow \mu^+\mu^-$ and $\phi \rightarrow K^+K^-$ decays is in good agreement with MC expectations. An estimate of the background arising from misidentified hadrons is obtained by reweighting the hadron misidentification probability using the momentum spectrum of the background in the invariant mass sidebands. The single hadron average misidentification probability is measured to be $(7.1 \pm 0.3) \times 10^{-3}$ and the double hadron misidentification probability is $(3.3 \pm 0.3) \times 10^{-3}$ when the correlation between the momenta of the two hadrons is taken into account. About 10% of the background is due to pairs consisting of one real muon and a hadron misidentified as muon, mostly from decay in flight. The contribution from double misidentified hadrons is negligible. The number of expected $B^+ \rightarrow \mu^+\mu^-$ candidates misidentified as $B^+ \rightarrow \mu^+\mu^-$ within the search window of ± 60 MeV/c² around the B^+ (B^0) mass is less than 0.1 (0.2).

For the $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow J/\psi K^0$ normalization channels some additional cuts are required. In the former case, the K^+ candidates are required to pass the same track quality and impact parameter cuts as the muons from the J/ψ . For $B^0 \rightarrow J/\psi K^0$ normalization, the K^+K^- invariant mass is required to be within ± 16 MeV/c² of the ϕ mass [14]. The B vertex has to be of good

quality, $\chi^2/\text{ndf} < 25$. The requirements on $\chi^2_{\text{rel}}/\text{ndf}$ and vertex separation significance for the B candidate are the same as those for the signal selection. The total efficiencies including acceptance, reconstruction and selection criteria for MC $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow J/\psi K^0$ events are 2.8% and 1.3% respectively.

5. Evaluation of the normalization factor

The branching factors of the three normalization channels, $B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+$, $B^0 \rightarrow J/\psi(\mu^+\mu^-)K^0$ and $B^0 \rightarrow \mu^+\mu^-$, are shown in Table 1. The first two decays have similar trigger and muon identification efficiency to the signal but a different number of particles in the final state, while the third channel has the same two-body topology but is selected with the hadronic trigger. The branching ratio of the $B^+ \rightarrow J/\psi$ decay is not known precisely ($\sim 25\%$) but has the advantage that the normalization of $B^+ \rightarrow \mu^+\mu^-$ with a B^+ decay does not require the knowledge of the ratio of fragmentation fractions, which has an uncertainty of $\sim 15\%$ [13].

5.1. Ratio of reconstruction and selection efficiencies

The accuracy of the simulation of the reconstruction efficiency ϵ^{MC} relies on the knowledge of the detector geometrical acceptance, the material interactions and the tracking efficiency. The uncertainty on the tracking efficiency is taken to be 4% per track [9] and this is the dominant source of the systematic uncertainty in the ratio with the two normalization channels involving J/ψ mesons. The ratios $\epsilon_{\text{MC}}^{\text{MC}}(B^+ \rightarrow \mu^+\mu^-) / \epsilon_{\text{MC}}^{\text{MC}}(B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+)$ and $\epsilon_{\text{MC}}^{\text{MC}}(B^+ \rightarrow \mu^+\mu^-) / \epsilon_{\text{MC}}^{\text{MC}}(B^0 \rightarrow \mu^+\mu^-)$ are shown in Ref. [11]. The measurement from data is 0.59 ± 0.04 , in good agreement with the estimate from MC simulation (0.58 \pm 0.02).

The accuracy of the simulation of $\epsilon^{\text{MC}}(B^+)$ relies on how well the MC describes the variations arising the selection. Of those only the IP distributions show a significant discrepancy: the data are measured to have $\sim 10\%$ worse resolution than the simulation, lowering the track parameters in MC to reproduce the IP distribution in data changes the selection efficiencies by 5–7% depending on the channel, however the ratios of $\epsilon_{\text{MC}}^{\text{MC}}(B^+)$ remain unchanged within the MC statistical uncertainties. The ratio predicted by the MC are 0.85 ± 0.01 , $B^+ \rightarrow J/\psi K^+$, 0.63 ± 0.05 , $B^0 \rightarrow J/\psi K^0$ and 1.09 ± 0.01 , $B^+ \rightarrow \mu^+\mu^-$, where the uncertainties correspond to the MC statistical uncertainties. The largest contribution to the difference in the selection efficiencies for $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow J/\psi K^0$ compared to the signal comes from the additional $\chi^2_{\text{rel}}/\text{ndf}$ requirements on the extra tracks in the normalization channels. For the $B^+ \rightarrow K^+K^-$ normalization channel, the selection

Purpose :

Soft selection to reduce size of dataset.

(signal efficiency \rightarrow as high as possible

sample efficiency \rightarrow as similar as possible to signal efficiency)

μ ID using “muon stations”.

Good quality events (tracks, vertex) are selected.

Condition for signal candidate:

$$\Delta M < 60 \text{ MeV}/c^2$$

After all selections ...

343 (342) B_s (B_d) $\rightarrow \mu^+\mu^-$ candidates are remain.
In those candidates, 0.3 (0.04) events are expected.

Dominant background :

~90% $bb \rightarrow \mu\mu X$

~10% fake + μ

5. Evaluation of the normalization factor

P3

LHC Collaboration / Physics Letters B 695 (2010) 130–140

Table 1
Summary of the factors and their uncertainties needed to calculate the normalization factors $\lambda_{\text{sig}}^{\text{MC}}(\mu^{\pm})$ for the three normalization channels considered. The branching ratios are taken from Ref. [13,15]. The trigger efficiency and number of $B^0 \rightarrow K^* \pi^0$ candidates (averaged) in only TS events, as described in the text.

	$B^0 \rightarrow \mu^+ \mu^-$	$B^0 \rightarrow \mu^+ \mu^-$	$B^0 \rightarrow \mu^+ \mu^-$	N_{MC}	$\lambda_{\text{sig}}^{\text{MC}}(\mu^{\pm}) \times 10^3$	$\lambda_{\text{sig}}^{\text{MC}}(\mu^{\pm}) \times 10^3$
$B^0 \rightarrow J/\psi \mu^+ \mu^-$	5.94 ± 0.23	0.40 ± 0.02	0.36 ± 0.05	$12,366 \pm 403$	8.4 ± 1.3	2.77 ± 0.18
$B^0 \rightarrow J/\psi \mu^+ \mu^-$	5.4 ± 0.5	0.40 ± 0.02	0.36 ± 0.05	702 ± 71	10.3 ± 2.9	2.81 ± 0.80
$B^0 \rightarrow K^* \pi^0$	1.34 ± 0.06	0.82 ± 0.06	0.012 ± 0.010	378 ± 34	3.3 ± 1.8	1.39 ± 0.40

ted excluding the signal tracks before calculating the χ^2_{rel} and vertex separation significance for the B candidate are the same as those for the signal selection. The final efficiencies including acceptance, reconstruction and selection criteria for MC $B^0 \rightarrow J/\psi K^*$ and $B^0 \rightarrow J/\psi \mu^+ \mu^-$ events are 2.08 and 1.38 respectively.

5. Evaluation of the normalization factor

The branching fractions of the three normalization channels, $B^0 \rightarrow J/\psi(\mu^+ \mu^-) K^*$, $B^0 \rightarrow J/\psi(\mu^+ \mu^-) \mu^+ \mu^-$ and $B^0 \rightarrow K^* \pi^0$, are shown in Table 1. The first two decays have similar trigger and event identification efficiency to the signal but a different number of particles in the final state, while the third channel has the same two-body topology but is selected with the hadronic trigger. The branching ratio of the $B^0 \rightarrow J/\psi \mu^+ \mu^-$ decay is not known precisely ($\sim 2\%$) but has the advantage that the normalization $B^0 \rightarrow \mu^+ \mu^-$, with a B^0 decay rate not requiring the knowledge of the ratio of fragmentation fractions, which has an uncertainty of $\sim 13\%$ [15].

5.1. Ratio of reconstruction and selection efficiencies

The accuracy of the simulation of the reconstruction efficiency ϵ^{MC} relies on the knowledge of the detector geometrical acceptance, the material interactions and the tracking efficiency. The uncertainty on the tracking efficiency is taken to be 4% for track [8] and this is the dominant source of the systematic uncertainty in the ratio with the two normalization channels involving J/ψ mesons. The ratio $\frac{\epsilon^{\text{MC}}(B^0 \rightarrow \mu^+ \mu^-)}{\epsilon^{\text{MC}}(B^0 \rightarrow K^* \pi^0)}$ predicted by the simulation are 0.38 ± 0.03 ($B^0 \rightarrow J/\psi K^*$), 0.39 ± 0.03 ($B^0 \rightarrow J/\psi \mu^+ \mu^-$) and 0.73 ± 0.03 ($B^0 \rightarrow K^* \pi^0$).

The effect of an extra particle on the ratio of ϵ^{MC} is cross-checked in data using the decay $B^0 \rightarrow J/\psi(\mu^+ \mu^-) K^* \pi^0$. Selecting B^0 and K^* in a similar phase-space region, the ratio $B^0 \rightarrow J/\psi K^* \pi^0$ and $B^0 \rightarrow J/\psi K^*$ yields (corrected for the ratio of branching ratios) is a good measure of the ratio of ϵ^{MC} between $B^0 \rightarrow J/\psi K^* \pi^0$ and $B^0 \rightarrow J/\psi K^*$, as shown in Ref. [16]. The measurement from data is 0.59 ± 0.04 , in good agreement with the estimate from MC simulation (0.38 ± 0.03).

The accuracy of the simulation of ϵ^{MC} relies on how well the MC describes the variables entering the selection. Of these only the D distributions show a significant discrepancy. The data are measured to have $\sim 10\%$ worse resolution than the simulation. Smearing the track parameters in MC to reproduce the D distribution in data changes the selection efficiencies by 5–7% depending on the channel. However, the ratio of $\epsilon^{\text{MC}}(B^0 \rightarrow \mu^+ \mu^-)$ remains unchanged within the MC statistical uncertainty. The ratios predicted by the MC are 0.45 ± 0.01 ($B^0 \rightarrow J/\psi K^*$), 0.40 ± 0.01 ($B^0 \rightarrow J/\psi \mu^+ \mu^-$) and 1.09 ± 0.01 ($B^0 \rightarrow K^* \pi^0$), where the largest contribution to the difference in the selection efficiencies for $B^0 \rightarrow J/\psi K^*$ and $B^0 \rightarrow J/\psi \mu^+ \mu^-$ compared to the signal comes from the additional χ^2_{rel} requirements on the extra tracks in the normalization channels. For the $B^0 \rightarrow K^* \pi^0$ normalization channel, the selection

quality, $\chi^2_{\text{rel}} < 25$. The requirements on χ^2_{rel} and vertex separation significance for the B candidate are the same as those for the signal selection. The final efficiencies including acceptance, reconstruction and selection criteria for MC $B^0 \rightarrow J/\psi K^*$ and $B^0 \rightarrow J/\psi \mu^+ \mu^-$ events are 2.08 and 1.38 respectively.

P4

LHC Collaboration / Physics Letters B 695 (2010) 130–140

333

efficiency is higher than for the signal as the tight ($40\% \text{MeV}/c^2$) mass window is not applied to the $B^0 \rightarrow K^* \pi^0$ channel. The ratios of efficiencies including acceptance, reconstruction and selection between normalization and signal decays are shown in Table 1.

5.2. Ratio of trigger efficiencies

The trigger efficiency ϵ^{TRIGG} can be estimated from data as described in Ref. [16]. Events that would have triggered even without the presence of the decay products of the signal under study are tagged as TS events (Trigger Independent Signal). TS events are mostly triggered by the decay products of the other B which can be in the acceptance given the forward geometry of LHCb.

If the presence of the signal under study alone is sufficient to trigger events are tagged as TOS (Trigger On-Signal). An event can also be TS and TOS simultaneously (TOS/TOS). The overall trigger efficiency on selected events can then be expressed as:

$$\epsilon^{\text{TRIGG}} = \frac{\epsilon^{\text{TS}} + \epsilon^{\text{TOS}}}{\epsilon^{\text{TS}} + \epsilon^{\text{TOS}}} = \epsilon^{\text{TS}} \frac{\epsilon^{\text{TOS}}}{\epsilon^{\text{TS}}} \quad (2)$$

where ϵ^{TS} is not directly observable as it corresponds to a sample of selected events for a fully efficient trigger. The TS efficiency (ϵ^{TS}) can however be measured directly on data using the ratio $\epsilon^{\text{TOS}}/\epsilon^{\text{TS}}$. Therefore ϵ^{TRIGG} can be expressed in terms of fully observable quantities.

The trigger efficiency for selecting $B^0 \rightarrow J/\psi K^*$ and $B^0 \rightarrow J/\psi \mu^+ \mu^-$ is obtained from a large inclusive sample of J/ψ events using Eq. (2). The result is $\epsilon^{\text{TRIGG}}(B^0 \rightarrow \mu^+ \mu^-) = 0.93 \pm 0.04$ and 0.94 ± 0.04 , where the systematic uncertainty reflects the approximation of the method as seen in the simulation. This efficiency is parameterized as $\epsilon^{\text{TRIGG}} = \epsilon^{\text{TS}} \frac{\epsilon^{\text{TOS}}}{\epsilon^{\text{TS}}} = 0.93 \pm 0.04$, where the systematic uncertainty is increased to account for the limitations of using only two variables (the largest p_T and D of the muons in the final state) to parameterize the trigger response.

In the case of the $B^0 \rightarrow K^* \pi^0$ normalization channel, the trigger efficiency is computed using the same events that are used for the normalization in Eq. (1). Therefore, combining Eqs. (1) and (2) results in an expression equivalent to a normalization which uses only TS events, the total number of these events after the first trigger steps (10 and HLT1) is 578, accepting all HLT2 triggers, which does not allow for a possible measurement of ϵ^{TRIGG} . Instead, this efficiency can be measured using another control channel, $B^0 \rightarrow J/\psi K^*$, with the result: $\epsilon^{\text{TRIGG}}(B^0 \rightarrow HLT1) = 0.93 \pm 0.04$. The small correction due to the HLT2 trigger efficiency on selected $B^0 \rightarrow K^* \pi^0$ candidates is taken from the trigger simulation. The ratio $\frac{\epsilon^{\text{TRIGG}}(B^0 \rightarrow \mu^+ \mu^-)}{\epsilon^{\text{TRIGG}}(B^0 \rightarrow K^* \pi^0)}$ for the three normalization channels are given in Table 1.

5.3. Overall normalization factor

The yields needed to evaluate the normalization factor for the two channels containing a J/ψ in the final state are obtained from a Gaussian fit to the invariant mass distribution. The number of candidates can be seen in Table 1, where the uncertainty is dominated by the differences observed using different fitting models. In the case of the $B^0 \rightarrow K^* \pi^0$ decay, the SICH particle identification and mass information are used to extract the fitting of $K^* \pi^0$ events from the selected inclusive $B^0 \rightarrow K^* \pi^0$ sample. The efficiency of the same and pion identification requirements is not needed since their ratio is extracted from the known ratio of $B^0 \rightarrow \pi^+ \pi^-$ and $B^0 \rightarrow K^* \pi^0$ branching ratios as described

in Ref. [16]. The number of TS $B^0 \rightarrow K^* \pi^0$ events observed is shown in Table 1.

As can be seen in Table 1, the normalization factors calculated using the three complementary channels give compatible results. The final normalization factor is a weighted average which takes into account all the sources of correlation, in particular the dominant one coming from the uncertainty on $\lambda_{\text{sig}}^{\text{MC}}(\mu^{\pm}) = 3.71 \pm 0.47$ [15], with the result:

$$\lambda_{\text{sig}}^{\text{MC}}(\mu^{\pm}) = (0.6 \pm 1.1) \times 10^{-3},$$

$$\lambda_{\text{sig}}^{\text{MC}}(K^* \pi^0) = (2.24 \pm 0.16) \times 10^{-3}.$$

6. Signal and background likelihoods

After the selection described in Section 4 the signal purity attaining the 3M branching ratio is still about 10^{-3} for $B^0 \rightarrow \mu^+ \mu^-$ and 10^{-4} for $B^0 \rightarrow K^* \pi^0$. Further discrimination is achieved through the combination of two independent variables: the multi-variate analysis discriminant likelihood, GL, combining information that is largely based on the topology of the event, and the invariant mass. The GL is defined using the statistical method described in Ref. [13,17]. The GL is defined to have a flat distribution between zero and one for signal candidates, and to cluster around zero for background candidates. The geometrical variables included in the definition of the GL are intended to be a complete set describing the properties of the decay, and the transverse momentum of the B candidate is also included, which is correlated with the invariant mass. The variables used in the definition of the GL are:

- **Distance of the B candidate.** This variable is computed using the distance between the secondary vertex and primary vertex, and the reconstructed momentum of the B candidate. When more than one primary vertex is reconstructed, the one that gives the minimum B -impact parameter significance is chosen.
- **Mass impact parameter χ^2 .** This is the lowest impact parameter χ^2 of the two muon candidates with respect to any primary vertex reconstructed in the event.
- **Impact parameter of the B candidate.**
- **Distance of closest approach between the two muon candidates.**
- **Isolation.** For each of the muon candidates, a search is performed for other tracks that can make a good vertex with the muon candidate, as in Ref. [10]. The number of compatible tracks is used as the discriminant variable.
- **Transverse momentum of the B candidate.**

The analysis is performed in two-dimensional bins of invariant mass and GL. The invariant mass in the signal region ($> 500 \text{ MeV}/c^2$ around the B^0 and the B^0 masses) is divided into six bins of equal width, and the GL into four bins of equal width distributed between zero and one. A probability to be signal or background is assigned to events falling in each bin.

6.1. Signal geometrical likelihood

Although the GL variable described above was defined using MC events, the probability that a signal event has a given value of GL is obtained from data using inclusive $B^0 \rightarrow K^* \pi^0$ events. Studies with large samples of MC events show that after reconstruction and selection the GL distributions obtained from $B^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^* \pi^0$ signal events agree within uncertainties [3X]. On the other hand, the two distributions are different after the trigger is simulated. This bias can be removed if only TS $B^0 \rightarrow K^* \pi^0$ events are used in the evaluation of the GL distribution. However, the total number of TS $B^0 \rightarrow K^* \pi^0$ events after all trigger steps [10, HLT1

Concept of normalization

$$BR_{\text{signal}} = \frac{N'_{\text{signal}}}{N'_{\text{total}}} = \frac{N_{\text{signal}} / \epsilon_{\text{signal}}}{N'_{\text{total}}}$$

$$N_{\text{signal}} = \epsilon_{\text{signal}} * N'_{\text{signal}}$$

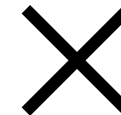
N : detected #
N' : occurred #

sample:
normalization with
well known events.

$$N'_{\text{total}} = L * \sigma \quad \text{or} \quad N'_{\text{total}} = \frac{N_{\text{sample}} / \epsilon_{\text{sample}}}{BR_{\text{sample}}}$$

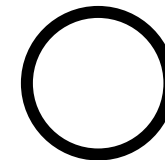
$$= \frac{N_{\text{signal}} / \epsilon_{\text{signal}}}{L * \sigma}$$

- 1) Rely on MC
- 2) Many systematics



$$= BR_{\text{sample}} * \frac{N_{\text{signal}} / \epsilon_{\text{signal}}}{N_{\text{sample}} / \epsilon_{\text{sample}}}$$

- 1) Rely on data > MC
- 2) Many systematics cancel



$$\mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)$$

exact formula of Br

$$= \mathcal{B}_{\text{norm}} \times \frac{\epsilon_{\text{norm}}^{\text{REC}} \epsilon_{\text{norm}}^{\text{SEL|REC}} \epsilon_{\text{norm}}^{\text{TRIG|SEL}}}{\epsilon_{\text{sig}}^{\text{REC}} \epsilon_{\text{sig}}^{\text{SEL|REC}} \epsilon_{\text{sig}}^{\text{TRIG|SEL}}} \times \frac{f_{\text{norm}}}{f_{B_q^0}} \times \frac{N_{B_q^0 \rightarrow \mu^+ \mu^-}}{N_{\text{norm}}}$$

HFAG average of LEP/Tevatron value
 $f_d/f_s = 3.71 \pm 0.47$

Three independent normalization channels used:

$B^\pm \rightarrow J/\psi(\mu\mu) K^\pm$	$B_s \rightarrow J/\psi(\mu\mu) \phi(KK)$	$B^0 \rightarrow K^+ \pi^-$
BR = 5.98×10^{-5} ($\pm 3.7\%$)	BR = 3.35×10^{-5} ($\pm 26\%$)	BR = 1.94×10^{-5} ($\pm 3.1\%$)
<ul style="list-style-type: none"> • Similar trigger and PID • Tracking efficiency (+1track) dominates error on efficiency ratio • f_d/f_s dominates overall uncertainty 	<ul style="list-style-type: none"> • Similar trigger and PID • Tracking efficiency (+2tracks) dominates error on efficiency ratio • BR dominates overall uncertainty 	<ul style="list-style-type: none"> • Different trigger \rightarrow use events triggered independent of signal • Identical topology • Uncertainty from f_d/f_s, trigger, mass fit

Results

	$\mathcal{B} (\times 10^{-5})$	$\frac{\epsilon_{\text{norm}}^{\text{REC}} \epsilon_{\text{SEL}}^{\text{REC}}}{\epsilon_{\text{sig}}^{\text{REC}} \epsilon_{\text{SEL}}^{\text{REC}}}$	$\frac{\epsilon_{\text{norm}}^{\text{TRIG}} \epsilon_{\text{SEL}}^{\text{TRIG}}}{\epsilon_{\text{sig}}^{\text{TRIG}} \epsilon_{\text{SEL}}^{\text{TRIG}}}$	N_{norm}	$\alpha_{B_S^0 \rightarrow \mu^+ \mu^-} (\times 10^{-9})$	$\alpha_{B^0 \rightarrow \mu^+ \mu^-} (\times 10^{-9})$
$B^+ \rightarrow J/\psi(\mu^+ \mu^-)K^+$	5.98 ± 0.22	0.49 ± 0.02	0.96 ± 0.05	$12,366 \pm 403$	8.4 ± 1.3	2.27 ± 0.18
$B_S^0 \rightarrow J/\psi(\mu^+ \mu^-)\phi(K^+ K^-)$	3.4 ± 0.9	0.25 ± 0.02	0.96 ± 0.05	760 ± 71	10.5 ± 2.9	2.83 ± 0.86
$B^0 \rightarrow K^+ \pi^-$	1.94 ± 0.06	0.82 ± 0.06	0.072 ± 0.010	578 ± 74	7.3 ± 1.8	1.99 ± 0.40

- Normalization factors from three channels consistent
 → take the **weighted average**

$$\alpha_{B_S \rightarrow \mu\mu} = 8.6 \pm 1.1 \times 10^{-9}$$

$$\alpha_{B^0 \rightarrow \mu\mu} = 2.24 \pm 0.16 \times 10^{-9}$$

(dominated by $B^\pm \rightarrow J/\psi K^\pm$)

→ used for expectation of
of signal events

6. Signal and background likelihoods

P4

efficiency is higher than for the signal as the tight ($460 \text{ MeV}/c^2$) mass window is not applied to the $B^0 \rightarrow \pi^+ \pi^-$ channel. The ratios of efficiencies including acceptance, reconstruction and selection are tagged as TS events (Trigger Independent Signal). TS events are mostly triggered by the decay products of the other B which can be in the acceptance given the forward geometry of LHCb.

If the presence of the signal under study alone is sufficient to trigger events are tagged as TOS (Trigger On Signal). An event can also be TS and TOS simultaneously (TS&TOS). The overall trigger efficiency on selected events can then be expressed as:

$$\epsilon^{\text{TS&TOS}} = \frac{\epsilon^{\text{TS}} \epsilon^{\text{TOS}}}{\epsilon^{\text{TS}} + \epsilon^{\text{TOS}} - \epsilon^{\text{TS}} \epsilon^{\text{TOS}}}$$

where ϵ^{TS} is not directly observable as it corresponds to a sample of selected events for a fully efficient trigger. The TS efficiency (ϵ^{TS}) can however be measured directly on data using the ratio $\epsilon^{\text{TS&TOS}}/\epsilon^{\text{TOS}}$. Therefore $\epsilon^{\text{TS&TOS}}$ can be expressed in terms of fully observable quantities.

The trigger efficiency for selecting $B^0 \rightarrow J/\psi K^0$ and $B^0 \rightarrow J/\psi \pi^0$ is obtained from a large inclusive sample of J/ψ events using Eq. (2). The result is $\epsilon^{\text{TS}}(J/\psi) = (83.9 \pm 0.5_{\text{stat}} \pm 2.0_{\text{sys}})\%$, where the systematic uncertainty reflects the approximation of the method as seen in the simulation. This efficiency is parameterized as a function of the largest p_T and the largest P of the two muons. Using the phase space of the $B^0 \rightarrow \mu^+ \mu^-$ decay in three muon variables, the trigger efficiency for the signal is evaluated to be $\epsilon^{\text{TS}}(B^0) = (99.9 \pm 0.4_{\text{stat}} \pm 4.0_{\text{sys}})\%$, where the systematic uncertainty is increased to account for the limitations of using only two variables (the largest p_T and P of the muon in the final state) to parameterize the trigger response.

In the case of the $B^0 \rightarrow K^+ K^-$ normalization channel, the trigger efficiency is computed using the same events that are used for the normalization in Eq. (1). Therefore, combining Eqs. (1) and (2) results in an expression equivalent to a normalization which uses only TS events. The total number of these events after the first trigger steps (ID and HLT) is 538, accepting all HLT triggers, which does not allow for a precise measurement of ϵ^{TS} . Instead, this efficiency can be measured using another control channel, $B^0 \rightarrow J/\psi K^0$, with the result: $\epsilon^{\text{TS}}(J/\psi) = (83.9 \pm 0.6)\%$. The small correction due to the HLT trigger inefficiency on selected $B^0 \rightarrow K^+ K^-$ candidates is taken from the trigger simulation. The ratios $\epsilon^{\text{TS&TOS}}/\epsilon^{\text{TOS}}$ for the three normalization channels are given in Table 1.

5.3. Overall normalization factor

The yields needed to evaluate the normalization factor for the two channels containing a J/ψ in the final state are obtained from a Gaussian fit to the invariant mass distribution. The number of candidates can be seen in Table 1, where the uncertainty is dominated by the differences observed using different fitting models. In the case of the $B^0 \rightarrow K^+ K^-$ decay, the RCI particle identification and mass information are used to extract the fraction of $K^+ K^-$ events from the selected inclusive $B^0 \rightarrow \pi^+ \pi^-$ sample. The efficiency of the same and pair identification requirements is not needed since their ratio is extracted from the known ratio of $B^0 \rightarrow \pi^+ \pi^-$ and $B^0 \rightarrow K^+ K^-$ branching ratios as described

in Ref. [10]. The number of TS $B^0 \rightarrow K^+ K^-$ events observed is shown in Table 1.

As can be seen in Table 1, the normalization factors calculated using the three complementary channels give compatible results. The final normalization factor is a weighted average which takes into account all the sources of correlations, in particular the dominant one coming from the uncertainty on $f_{B^0}/f_{B^+} = 3.71 \pm 0.47$ [15], with the result:

$$\alpha_{B^0 \rightarrow \pi^+ \pi^-} = (6.6 \pm 1.1) \times 10^{-9},$$
$$\alpha_{B^0 \rightarrow \mu^+ \mu^-} = (2.24 \pm 0.16) \times 10^{-9}.$$

5. Signal and background likelihoods

After the selection described in Section 4 the signal purity assuming the SM branching ratio is still about 10^{-3} for $B^0 \rightarrow \mu^+ \mu^-$ and 10^{-4} for $B^0 \rightarrow \pi^+ \pi^-$. Further discrimination is achieved through the combination of two independent variables: the multivariate analysis discriminant likelihood, GL, combining information that is largely based on the topology of the event, and the invariant mass. The GL is defined using the statistical method described in Refs. [13,17]. The GL is defined to have a flat distribution between zero and one for signal candidates, and to cluster around zero for background candidates. The geometrical variables included in the definition of the GL are introduced to be a complete set describing the properties of the decay, and the transverse momentum of the B candidate is also included, which is uncorrelated with the invariant mass. The variables used in the definition of the GL are:

- Lifetime of the B candidate. This variable is computed using the distance between the secondary vertex and primary vertex, and the reconstructed momentum of the B candidate. When more than one primary vertex is reconstructed, the one that gives the minimum B impact parameter significance is chosen.
- Mass impact parameter λ^* . This is the lowest impact parameter λ^* of the two mass candidates with respect to any primary vertex reconstructed in the event.
- Impact parameter of the B candidate.
- Distance/correlation approach between the two mass candidates.
- Isolation. For each of the muon candidates, a search is performed for other tracks that can make a good vertex with the muon candidate, as in Ref. [10]. The number of compatible tracks is used as the discriminant variable.
- Transverse momentum of the B candidate.

The analysis is performed in two-dimensional bins of invariant mass and GL. The invariant mass in the signal regions ($\pm 30 \text{ MeV}/c^2$ around the B^0 and the B^+ masses) is divided into six bins of equal width, and the GL into four bins of equal width distributed between zero and one. A probability to be signal or background is assigned to events falling in each bin.

5.1. Signal geometrical likelihood

Although the GL variable described above was defined using MC events, the probability that a signal event has a given value of GL is obtained from data using inclusive $B^0 \rightarrow \pi^+ \pi^-$ events. Studies with large samples of MC events show that after reconstruction and selection the GL distributions obtained from $B^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \pi^+ \pi^-$ signal events agree within uncertainties (3%). On the other hand, the two distributions are different after the trigger is simulated. This bias can be removed if only $B^0 \rightarrow \pi^+ \pi^-$ events are used in the evaluation of the GL distribution. However, the total number of TS $B^0 \rightarrow \pi^+ \pi^-$ events after all trigger steps (ID, HLT

P5

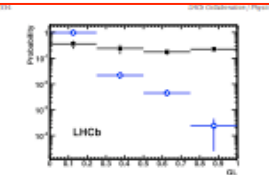


Fig. 1. Probability of signal events in bins of GL obtained from the inclusive sample of TS $B^0 \rightarrow \pi^+ \pi^-$ events (red solid squares); the background probability (open circles) is obtained from the events in the sidebands of the muon invariant mass distribution in the B^0 mass window.

Table 2. Probability of signal events in bins of GL obtained from the inclusive sample of TS $B^0 \rightarrow \pi^+ \pi^-$ events. The background probability in the B^0 mass window is obtained from the events in the sidebands of the muon invariant mass distribution.

GL bin	Signal prob.	Background prob.
0.0-0.25	0.360 ± 0.130	0.071 ± 0.044
0.25-0.5	0.239 ± 0.086	0.027 ± 0.014
0.5-0.75	0.176 ± 0.046	0.004 ± 0.002
0.75-1.0	0.225 ± 0.036	0.000 ± 0.000

and HLT) is 152 which is insufficient. Instead, for the $B^0 \rightarrow \pi^+ \pi^-$ events, the first two trigger steps (ID and HLT) are required to be TS while at the HLT step any of the HLT triggers are accepted. This yields 355 events. The GL distribution obtained using these events is corrected for the small bias ($\pm 3\%$) introduced at the HLT step using the trigger simulation. Detailed checks with a large sample of $D^0 \rightarrow K^+ \pi^-$ decays have validated this procedure. The number of TS $B^0 \rightarrow \pi^+ \pi^-$ events in each GL bin is obtained from a fit to the inclusive mass distribution [18] assigning the muon mass to the two particles. The measured functions in each GL bin can be seen in Fig. 1 and are quoted in Table 2. The systematic uncertainties are included, estimated by comparing the results from the inclusive $B^0 \rightarrow \pi^+ \pi^-$ fit model with those obtained using a double Crystal Ball function [73] and a simple background subtraction. The measured GL distribution obtained from TS $B^0 \rightarrow \pi^+ \pi^-$ events is compatible with a flat distribution, as expected if the simulation reproduces correctly the data.

5.2. Signal invariant mass likelihood

The signal mass likelihood is parameterized using a Crystal Ball function [73]. Two methods have been used to estimate the $B^0 \rightarrow \pi^+ \pi^-$ mass resolution from data. The first of these methods uses an interpolation between the measured resolutions for π^0 resonances ($J/\psi, \psi(2S)$ and $\psi(3S)$ resonances [71(15), 71(25), 71(35)] decaying into two muons. It has been observed that over this mass range the muon invariant mass resolution depends linearly on the invariant mass of the muon pair to good approximation. Events selected in the mass ranges around the $\psi(2S)$ and $\psi(3S)$ resonances were weighted such that the momentum spectra of these resonances reproduce the expected momentum spectrum of the B hadron in the decay $B^0 \rightarrow \mu^+ \mu^-$. The mass resolutions of the $\psi(2S)$ and $\psi(3S)$ resonances were then determined fitting a Crystal Ball ($J/\psi, 71(15)$) or a Gaussian ($\psi(2S), 71(25)$ and $\psi(3S)$) over exponential back-

The mass resolution is defined as the σ of the Crystal Ball when there is sufficient data to perform a fit with the Crystal Ball function (J/ψ and $\psi(3S)$). Otherwise a Gaussian fit is made ($\psi(2S), 71(25), 71(35)$) and the σ of the Gaussian is used as an estimate of the σ of the Crystal Ball. For the Crystal Ball function, the parameters describing the radiative tail are in good agreement between data and the Monte Carlo simulation. No systematic shifts in the resolution has been found by using a Crystal Ball or a Gaussian above the transition point.

Interpolating linearly between the two fixed resolutions to $M_{\mu\mu}$ an invariant mass resolution of $\sigma = 26.83 \pm 0.14 \text{ MeV}/c^2$ was estimated for $B^0 \rightarrow \mu^+ \mu^-$. The systematic uncertainty is estimated to be $1 \text{ MeV}/c^2$ mainly due to the reweighting of the momentum spectrum of the muon resonances and the variation of the resolution over the width of the $B^0 \rightarrow \mu^+ \mu^-$ signal region.

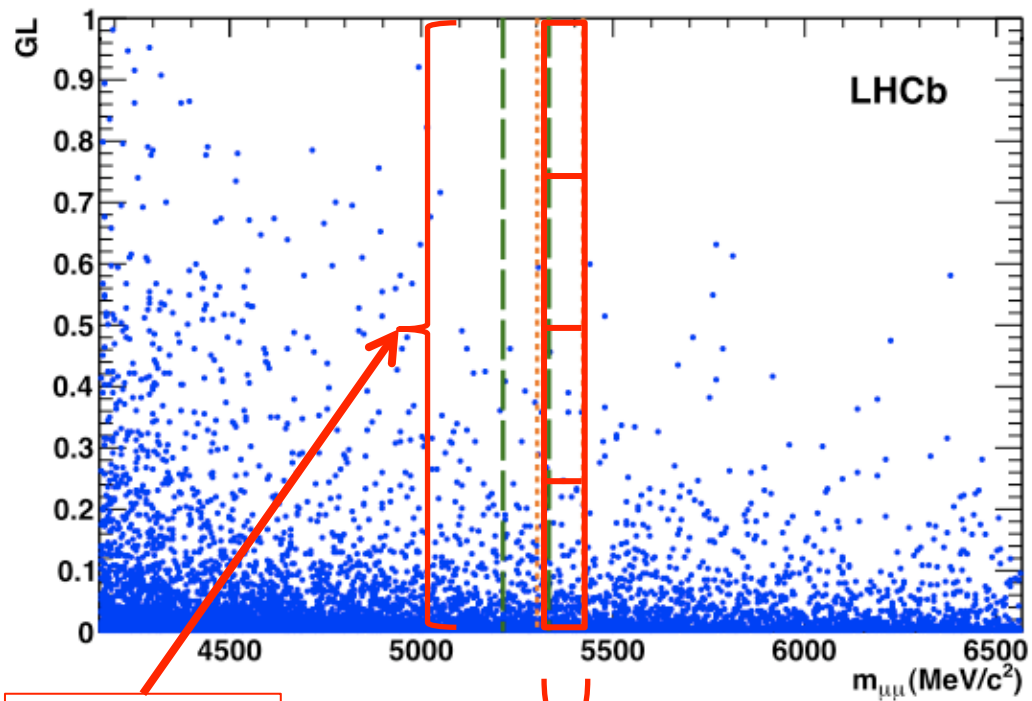
The second method that was used to estimate the invariant mass resolution from data is to use the inclusive $B^0 \rightarrow \pi^+ \pi^-$ sample. The particle identification requirement would modify the invariant and transverse momentum spectrum of pions and kaons, and thus the mass resolution. Therefore, the fit is performed to the inclusive $B^0 \rightarrow \pi^+ \pi^-$ sample without requiring particle identification and assigning the muon mass to the decay products. The invariant mass resolution from data is in the range $25.5 \pm 1.0 \text{ MeV}/c^2$, as shown in Fig. 2. The fitted parameters are: the mass resolution, the B^0 and B^+ masses, the signal yield, the combinatorial background yields, as well as the fraction of radiative tail and the parameters that describe the combinatorial background. The relative contributions of B^0 and B^+ decays are fixed to their known values. The result of the fit for the mass resolution, $\sigma = 25.8 \pm 1.0 \text{ MeV}/c^2$, is consistent with the value obtained from the interpolation method. However, by varying the assumptions made for the parameters describing the partially reconstructed three-body b -hadron decays (physical background), the estimate obtained for the resolution can change by up to $2.7 \text{ MeV}/c^2$. This is assigned as systematic uncertainty for this method.

The weighted average of the two methods, $\sigma = 26.7 \pm 0.9 \text{ MeV}/c^2$, is taken as the invariant mass resolution and considered to be the same for B^0 and B^+ decays. The mean values of the masses obtained from the inclusive $B^0 \rightarrow \pi^+ \pi^-$ fit are consistent with, but not as precise as, the values obtained using the exclusive decay modes $B^0 \rightarrow K^+ K^-$ and $B^0 \rightarrow K^+ \pi^-$ obtained using the RCI particle identification: $M_{B^0} = 5275.0 \pm 1.0 \text{ MeV}/c^2$ and $M_{B^+} = 5363.1 \pm 1.5 \text{ MeV}/c^2$, which are used in the evaluation of the invariant mass likelihood. The mean values of the masses are $\sim 0.3\%$ below the known values [14] which is attributed to a small residual miscalibration of the magnetic field map. However this bias has no impact on the analysis provided that the search windows are centred around the measured values.

5.3. Background likelihood

The mass sidebands are defined in the range between $M_{\mu\mu} \pm 600$ ($1200 \text{ MeV}/c^2$) for the lower (upper) two GL bins, including the two search windows ($M_{\mu\mu} \pm 50 \text{ MeV}/c^2$). The background in the mass sidebands is fitted with an exponential function, $f(M) = A e^{-kM}$. The value of the exponential index k is fixed independently in each GL bin, in order to account for potentially different background composition. The distribution of the invariant mass for each GL bin is shown in Fig. 3, and the predictions for the numbers of events in the signal region can be seen in Table 3 and 4. The background probability in the B^0 mass window is a function of GL is shown in Fig. 1 and in Table 2. The results have been checked by fitting the exponential index k to be the same in all GL bins, using a double exponential, or using a simple linear

Events are classified in two dimensional plane.
invariant mass
geometrical likelihood (GL)



GL \rightarrow 4 bins

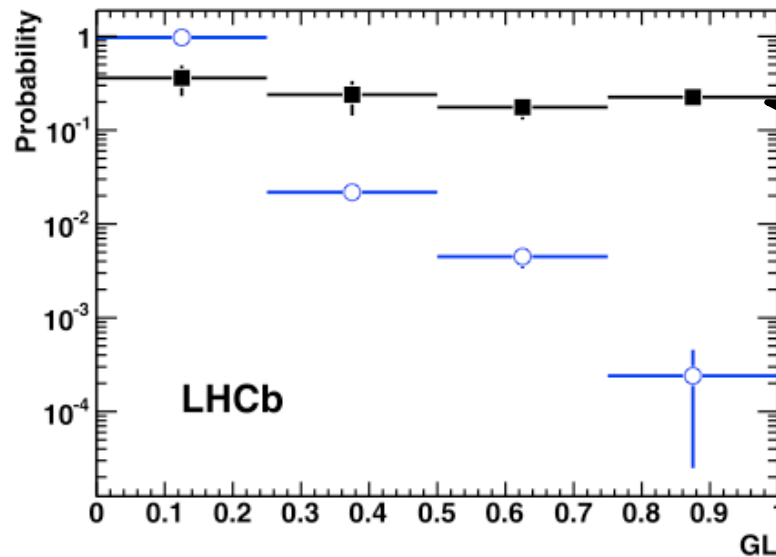
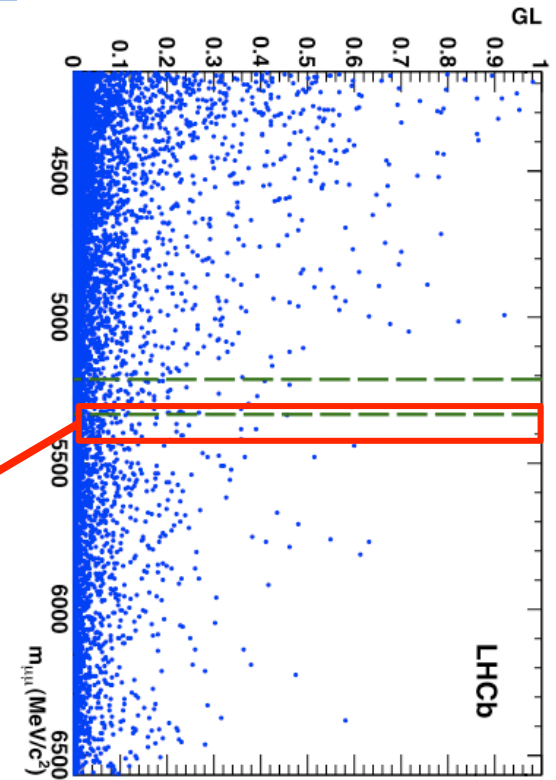
$\Delta m \rightarrow$ 6 bins

expected signals & BGs are calculated

GL:
 Obtained from data using inclusive $B^0 \rightarrow h^+ h'^-$
 Agree with distribution obtained from $B^0 \rightarrow \mu^+ \mu^-$

variables used for GL:
 Lifetime & pt of the B
 Impact parameter of μ , B
 distance of $\mu\mu$... etc

background probability in the B_s^0 mass window



Signal
 ~25% for each

BG
 cluster around 0

→ used for expectation of
 # of signal events

Mass resolution are calculated in two ways

1) interpolation $J/\psi, \psi(2S), Y(1S), Y(2S), Y(3S) \rightarrow \mu^+\mu^-$

CB function fit

Gaussian fit

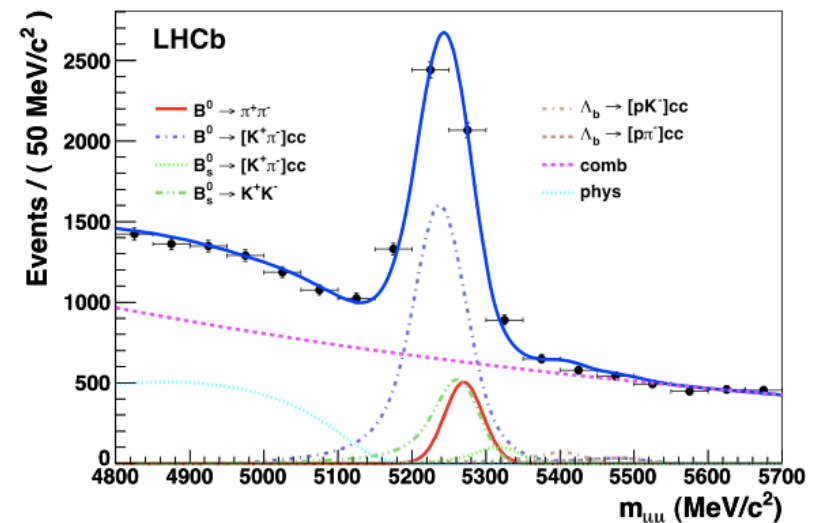
$\rightarrow \sigma = 26.83 \pm 0.14 \text{ MeV}$

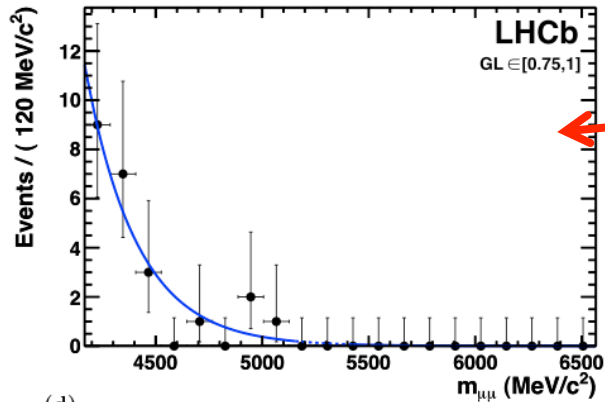
2) inclusive $B_q^0 \rightarrow h^+h'^-$ reconstruction

$\rightarrow \sigma = 25.8 \pm 1.0 \text{ MeV}$

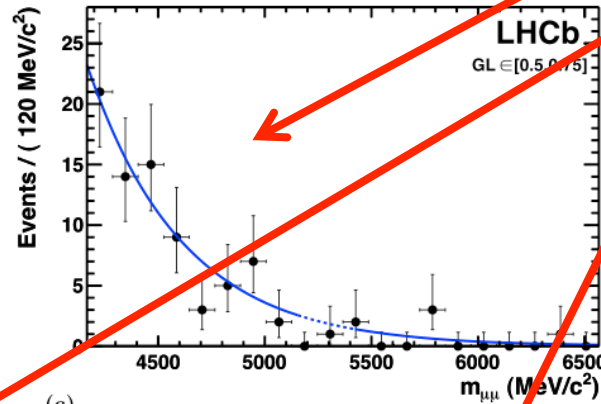
$\sigma = 26.7 \pm 0.9 \text{ MeV}$

\rightarrow used for expectation of
of signal events

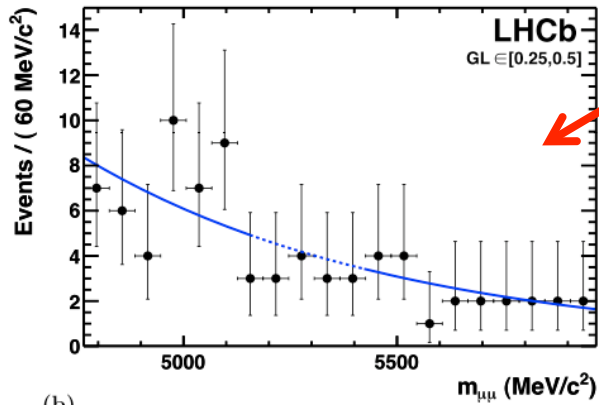




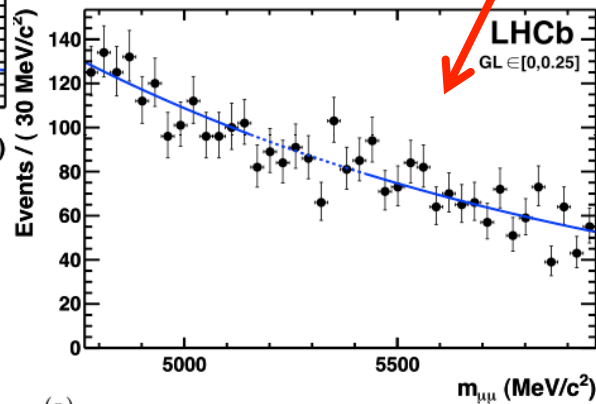
(d)



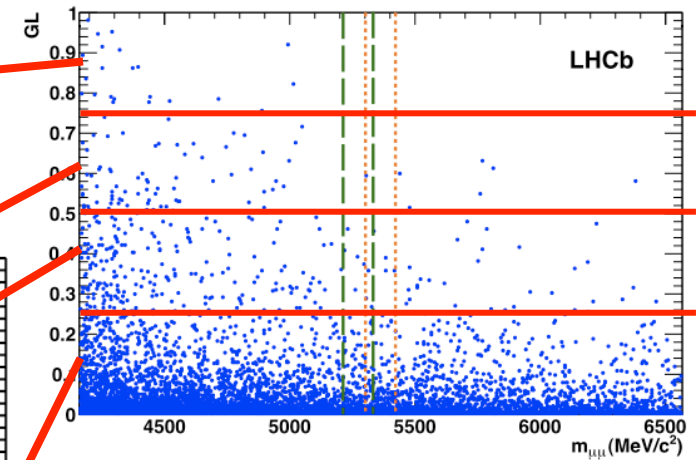
(c)



(b)



(a)



Sideband region

$\pm 1200 \text{ MeV}$ for upper GL bins
 $\pm 600 \text{ MeV}$ for lower GL bins
 (fitted with $A * \exp[-k * x]$)

→ used for expectation of # of BG events

B^0_s prediction for the number of events in the signal region

Invariant mass bin (MeV/c^2)		GL bin			
		[0, 0.25]	[0.25, 0.5]	[0.5, 0.75]	[0.75, 1]
[-60, -40]	Exp. bkg.	$56.9^{+1.1}_{-1.1}$	$1.31^{+0.19}_{-0.17}$	$0.282^{+0.076}_{-0.065}$	$0.016^{+0.021}_{-0.010}$
	Exp. sig.	$0.0076^{+0.0034}_{-0.0030}$	$0.0050^{+0.0027}_{-0.0020}$	$0.0037^{+0.0015}_{-0.0011}$	$0.0047^{+0.0015}_{-0.0010}$
	Observed	39	2	1	0
[-40, -20]	Exp. bkg.	$56.1^{+1.1}_{-1.1}$	$1.28^{+0.18}_{-0.17}$	$0.269^{+0.072}_{-0.062}$	$0.0151^{+0.0195}_{-0.0094}$
	Exp. sig.	$0.0220^{+0.0084}_{-0.0081}$	$0.0146^{+0.0067}_{-0.0054}$	$0.0107^{+0.0036}_{-0.0027}$	$0.0138^{+0.0035}_{-0.0025}$
	Observed	55	2	0	0
[-20, 0]	Exp. bkg.	$55.3^{+1.1}_{-1.1}$	$1.24^{+0.17}_{-0.16}$	$0.257^{+0.069}_{-0.059}$	$0.0139^{+0.0179}_{-0.0086}$
	Exp. sig.	$0.038^{+0.015}_{-0.015}$	$0.025^{+0.012}_{-0.010}$	$0.0183^{+0.0063}_{-0.0047}$	$0.0235^{+0.0060}_{-0.0044}$
	Observed	73	0	0	0
[0, 20]	Exp. bkg.	$54.4^{+1.1}_{-1.1}$	$1.21^{+0.17}_{-0.16}$	$0.246^{+0.066}_{-0.057}$	$0.0128^{+0.0165}_{-0.0080}$
	Exp. sig.	$0.038^{+0.015}_{-0.015}$	$0.025^{+0.012}_{-0.010}$	$0.0183^{+0.0063}_{-0.0047}$	$0.0235^{+0.0060}_{-0.0044}$
	Observed	60	0	0	0
[20, 40]	Exp. bkg.	$53.6^{+1.1}_{-1.0}$	$1.18^{+0.17}_{-0.15}$	$0.235^{+0.063}_{-0.054}$	$0.0118^{+0.0152}_{-0.0073}$
	Exp. sig.	$0.0220^{+0.0084}_{-0.0081}$	$0.0146^{+0.0067}_{-0.0054}$	$0.0107^{+0.0036}_{-0.0027}$	$0.0138^{+0.0035}_{-0.0025}$
	Observed	53	2	0	0
[40, 60]	Exp. bkg.	$52.8^{+1.0}_{-1.0}$	$1.14^{+0.16}_{-0.15}$	$0.224^{+0.060}_{-0.052}$	$0.0108^{+0.0140}_{-0.0068}$
	Exp. sig.	$0.0076^{+0.0031}_{-0.0027}$	$0.0050^{+0.0025}_{-0.0019}$	$0.0037^{+0.0013}_{-0.0010}$	$0.0047^{+0.0013}_{-0.0010}$
	Observed	55	1	0	0

B^0_d prediction for the number of events in the signal region

Invariant mass bin (MeV/c ²)		GL bin			
		[0, 0.25]	[0.25, 0.5]	[0.5, 0.75]	[0.75, 1]
[-60, -40]	Exp. bkg.	60.8 ^{+1.2} _{-1.1}	1.48 ^{+0.19} _{-0.18}	0.345 ^{+0.084} _{-0.073}	0.024 ^{+0.027} _{-0.014}
	Exp. sig.	0.00090 ^{+0.00036} _{-0.00035}	0.00060 ^{+0.00029} _{-0.00023}	0.00044 ^{+0.00016} _{-0.00012}	0.00056 ^{+0.00015} _{-0.00011}
	Observed	59	2	0	0
[-40, -20]	Exp. bkg.	59.9 ^{+1.1} _{-1.1}	1.44 ^{+0.19} _{-0.17}	0.329 ^{+0.080} _{-0.070}	0.022 ^{+0.024} _{-0.013}
	Exp. sig.	0.00263 ^{+0.00093} _{-0.00093}	0.00174 ^{+0.00076} _{-0.00061}	0.00128 ^{+0.00038} _{-0.00030}	0.00164 ^{+0.00035} _{-0.00025}
	Observed	67	0	0	0
[-20, 0]	Exp. bkg.	59.0 ^{+1.1} _{-1.1}	1.40 ^{+0.18} _{-0.17}	0.315 ^{+0.077} _{-0.067}	0.020 ^{+0.022} _{-0.012}
	Exp. sig.	0.0045 ^{+0.0017} _{-0.0017}	0.0030 ^{+0.0014} _{-0.0011}	0.00219 ^{+0.00067} _{-0.00054}	0.00280 ^{+0.00060} _{-0.00045}
	Observed	56	2	0	0
[0, 20]	Exp. bkg.	58.1 ^{+1.1} _{-1.1}	1.36 ^{+0.18} _{-0.16}	0.300 ^{+0.073} _{-0.064}	0.019 ^{+0.021} _{-0.011}
	Exp. sig.	0.0045 ^{+0.0017} _{-0.0017}	0.0030 ^{+0.0014} _{-0.0011}	0.00219 ^{+0.00067} _{-0.00054}	0.00280 ^{+0.00060} _{-0.00045}
	Observed	60	0	0	0
[20, 40]	Exp. bkg.	57.3 ^{+1.1} _{-1.1}	1.33 ^{+0.17} _{-0.16}	0.287 ^{+0.070} _{-0.061}	0.017 ^{+0.019} _{-0.010}
	Exp. sig.	0.00263 ^{+0.00093} _{-0.00093}	0.00174 ^{+0.00076} _{-0.00061}	0.00128 ^{+0.00038} _{-0.00030}	0.00164 ^{+0.00035} _{-0.00025}
	Observed	42	2	1	0
[40, 60]	Exp. bkg.	56.4 ^{+1.1} _{-1.1}	1.29 ^{+0.17} _{-0.16}	0.274 ^{+0.067} _{-0.058}	0.0158 ^{+0.0175} _{-0.0094}
	Exp. sig.	0.00090 ^{+0.00033} _{-0.00032}	0.00060 ^{+0.00027} _{-0.00021}	0.00044 ^{+0.00014} _{-0.00011}	0.00056 ^{+0.00013} _{-0.00010}
	Observed	49	2	0	0

Results

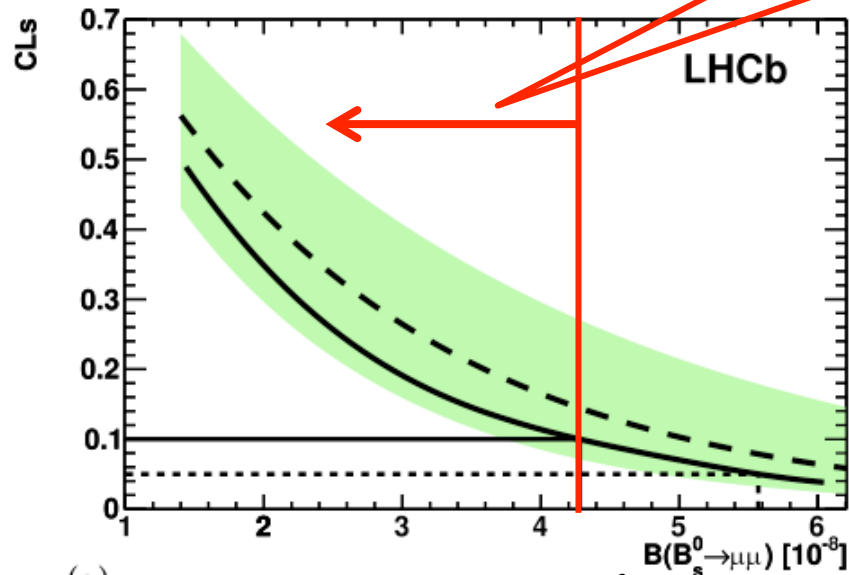
Green shaded area

CL_b : BG only hypothesis
compatible within 1σ

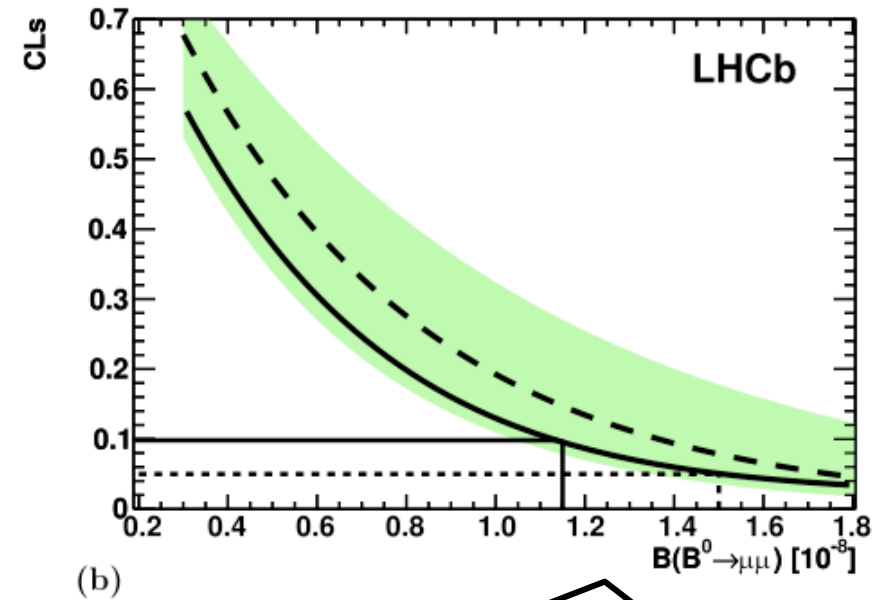
Solid (Dashed) line

CL_s : Signal + BG hypothesis
Observed (Expected)

CL_s is higher than 0.1 at $BR < 4.3 \times 10^{-8}$
→ The area $BR > 4.3 \times 10^{-8}$ is excluded at CL 90%.



$BR B_s^0 \rightarrow \mu\mu$
 $< 4.3(5.6) \times 10^{-8}$ @ 90(95)%CL



$BR B_d^0 \rightarrow \mu\mu$
 $< 1.2(1.5) \times 10^{-8}$ @ 90(95)%CL

7. Results & 8. Conclusions

P6

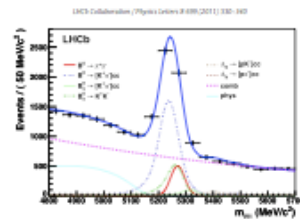


Fig. 2. Fit of the invariant mass distribution for $B^0 \rightarrow K^{*0} \pi^0$ candidates in the GL using the CL method [20]. The pink dashed curve is the combinatorial background component, while the physical background is shown with a light blue dotted curve. The remaining contributions are from the two-body decays of the B^0 , K^* and π^0 . (For interpretation of the reference to colour in this figure legend, the reader is referred to the web version of this paper.)

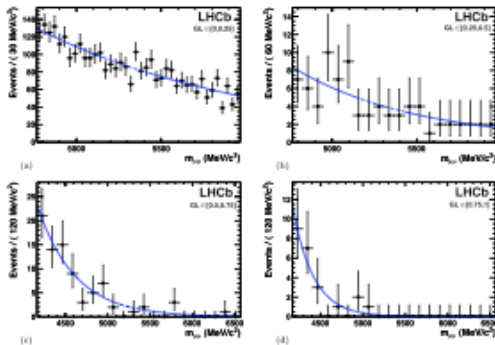


Fig. 3. Distribution of the μ^+ invariant mass for different GL bins: (a) (5, 25), (b) (25, 50), (c) (50, 75), (d) (75, 100). The blue solid lines show the integration result used and the dashed line shows the result of the integration in the search window. The interpretation of the references to colour in this figure legend is referred to the web version of this paper.

fit in the region around the signal window. In all cases the predicted background is consistent with the result of the exponential fit with different λ values, although the quality of the fit is significantly worse when λ is forced to be the same for all bins.

7. Results

For each of the 24 bins (4 bins in GL and 6 bins in mass) the expected number of background events is computed from the fit

to the invariant mass sidebands described in Section 6.3. The results are shown in Tables 3 and 4. The expected numbers of signal events are computed using the normalisation factors from Section 5, and the signal sidebands computed in Section 6.3 and Section 6.2 for a given value of $B[0] \rightarrow \mu^+ \mu^-$. The expected numbers of signal events for the SM branching ratios are shown in Tables 3 and 4. The distribution of observed events in the GL vs invariant mass plane can be seen in Fig. 4, and the observed number of events in each bin are given in Tables 3 and 4. The

P7

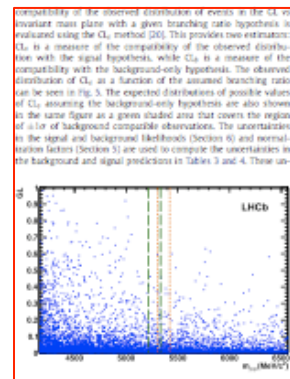


Fig. 4. Observed distribution of selected dilepton events in the GL vs invariant mass plane. The orange short-dashed (green long-dashed) lines indicate the 400 MeV/c² search window around the B^0 (μ^+), the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.

Table 3

Expected background, expected SM signal and observed number of events in bins of GL and invariant mass, in the 400 MeV/c ² mass window around the B^0 mass central value of 5033.3 MeV/c ² .		GL bin			
Invariant mass [GeV]		(0, 25)	(25, 50)	(50, 75)	(75, 100)
[−40, −40]	Exp. Bkg	58.0 ^{+1.1}	1.40 ^{+0.03}	0.40 ^{+0.01}	0.01 ^{+0.00}
	Observed	39	2	1	0
[−40, −20]	Exp. Bkg	58.1 ^{+1.1}	1.28 ^{+0.03}	0.29 ^{+0.01}	0.01 ^{+0.00}
	Observed	33	2	0	0
[−20, 0]	Exp. Bkg	53.3 ^{+1.1}	1.42 ^{+0.03}	0.20 ^{+0.01}	0.01 ^{+0.00}
	Observed	73	0	0	0
[0, 20]	Exp. Bkg	54.4 ^{+1.1}	1.20 ^{+0.03}	0.26 ^{+0.01}	0.01 ^{+0.00}
	Observed	43	0	0	0
[20, 40]	Exp. Bkg	53.8 ^{+1.1}	1.18 ^{+0.03}	0.23 ^{+0.01}	0.01 ^{+0.00}
	Observed	33	0	0	0
[40, 60]	Exp. Bkg	52.1 ^{+1.1}	1.14 ^{+0.03}	0.22 ^{+0.01}	0.01 ^{+0.00}
	Observed	33	1	0	0

compatibility of the observed distribution of events in the GL vs invariant mass plane with a given branching ratio hypothesis is evaluated using the CL_s method [20]. This provides two estimates: CL_s is a measure of the compatibility of the observed distribution with the signal hypothesis, while CL_b is a measure of the compatibility with the background-only hypothesis. The observed distribution of CL_s as a function of the assumed branching ratio can be seen in Fig. 5. The expected distributions of possible values of CL_s assuming the background-only hypothesis are also shown in the same figure as a given shaded area that covers the region of a lot of background compatible observations. The uncertainties in the signal and background likelihoods (Section 6) and normalisation factors (Section 5) are used to compute the uncertainties in the background and signal predictions in Tables 3 and 4. These un-

certainties are the main source of systematic uncertainty and they are included in the CL_s using the techniques described in Ref. [20]. Given the specific pattern of the observed events, the systematic uncertainty on the background prediction has a negligible effect on the quoted limits. The effect of the uncertainty on the signal prediction increases the quoted limits by less than 2%. The evaluation of CL_s [20] gives a probability of about 20% for the compatibility with the background-only hypothesis for both of the B^0 and B^+ decays. This low value can be attributed to the slight deficit of observed events in the most sensitive bins, as can be seen in Tables 3 and 4. As no significant deviation from the background-only hypothesis is observed, upper limits are computed using the CL_b distribution in Fig. 5 with the results

$$B[0] \rightarrow \mu^+ \mu^- < 4.3 (5.6) \times 10^{-8} \text{ at } 90\% (95\%) \text{ C.L.}$$

$$B[B^+] \rightarrow \mu^+ \mu^- < 1.2 (1.3) \times 10^{-8} \text{ at } 90\% (95\%) \text{ C.L.}$$

while the expected values of the limits are $B[0] \rightarrow \mu^+ \mu^- < 3.1 (3.5) \times 10^{-8}$ and $B[B^+] \rightarrow \mu^+ \mu^- < 1.4 (1.8) \times 10^{-8}$ at 90% (95%) C.L. The limits observed are similar to the best published limits [5] for the decay $B^0 \rightarrow \mu^+ \mu^-$ and more restrictive for the decay $B^+ \rightarrow \mu^+ \mu^-$ [5].

8. Conclusions

With about 37 pb⁻¹ of integrated luminosity, LHCb has searched for the rare decays $B^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow \mu^+ \mu^-$ and reached sensitivities similar to the existing limits from the Tevatron. This could be achieved due to the large acceptance and trigger efficiency of LHCb, as well as the large ab cross-section in pp collisions at $\sqrt{s} = 7$ TeV. The observed events are compatible with the background expectations, and the upper limits are evaluated to be

$$B[0] \rightarrow \mu^+ \mu^- < 5.6 \times 10^{-8} \text{ at } 95\% \text{ C.L.}$$

$$B[B^+] \rightarrow \mu^+ \mu^- < 1.5 \times 10^{-8} \text{ at } 95\% \text{ C.L.}$$

P8

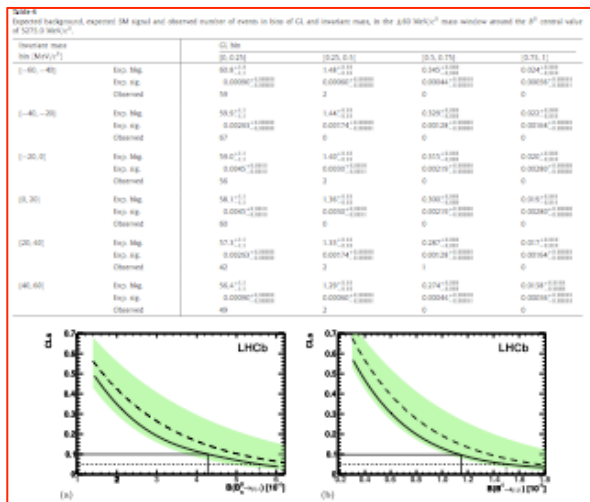


Fig. 5. (a) Observed (bold curve) and expected (dashed curve) CL values as a function of $B[0] \rightarrow \mu^+ \mu^-$. The green shaded area contains the 1σ interval of possible results compatible with the expected value when only background is observed. The 90% (95%) CL observed value is identified by the solid (dashed) line. (b) The same for $B[B^+] \rightarrow \mu^+ \mu^-$. (For interpretation of the reference to colour in this figure legend, the reader is referred to the web version of this paper.)

while the expected values of the limits are $B[0] \rightarrow \mu^+ \mu^- < 6.5 \times 10^{-8}$ and $B[B^+] \rightarrow \mu^+ \mu^- < 1.8 \times 10^{-8}$ at 95% C.L. The LHC is expected to deliver a much larger sample of pp collisions in 2021. Given the low level of background in the most sensitive bins shown in Tables 3 and 4, LHCb should be able to explore the branching region of branching ratios at the 10^{-8} level in the near future.

Acknowledgements

We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at CERN and at the LHCb institutes, and acknowledge support from the National Agency: CNRS, INFN, INFN, INFN and INFN (Italy); CLAS, NSRF (China); CNRS/IN2P3 (France); INFN, INFN, INFN and INFN (Germany); INFN (Italy); INFN (Italy); INFN and INFN (Netherlands);

Open access

This article is published Open Access at sciencedirect.com. It is distributed under the terms of the Creative Commons Attribution License (CC BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

References

[1] C. Adami, et al., Phys. Rev. D 64 (2001) 044014.

- First LHCb result (0.037 fb⁻¹)

$$\mathbf{BR(B_s \rightarrow \mu^+ \mu^-) < 4.3 (5.6) 10^{-8} @ 90 (95\% CL)}$$

$$\mathbf{BR(B^0 \rightarrow \mu^+ \mu^-) < 1.2 (1.5) 10^{-8} @ 90 (95\% CL)}$$

latest result (at EPS July 2011)

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{SM} = (0.32 \pm 0.02) \times 10^{-8},$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)_{SM} = (0.010 \pm 0.001) \times 10^{-8}.$$

CDF 7fb-1 @ $\sqrt{s}=1.96\text{TeV}$

$$0.46 \times 10^{-8} < \text{Br}(B_s \rightarrow \mu\mu) < 3.9 \times 10^{-8} \text{ @ 90\% C.L.}$$

$$\text{Br}(B_d \rightarrow \mu\mu) < 0.6 \times 10^{-8} \text{ @ 95\%C.L.}$$

Events
observed !?

CMS 1.14fb-1 @ $\sqrt{s}=7\text{TeV}$

$$\text{Br}(B_s \rightarrow \mu\mu) < 1.6(1.9) \times 10^{-8} \text{ @ 90(95)\%C.L.}$$

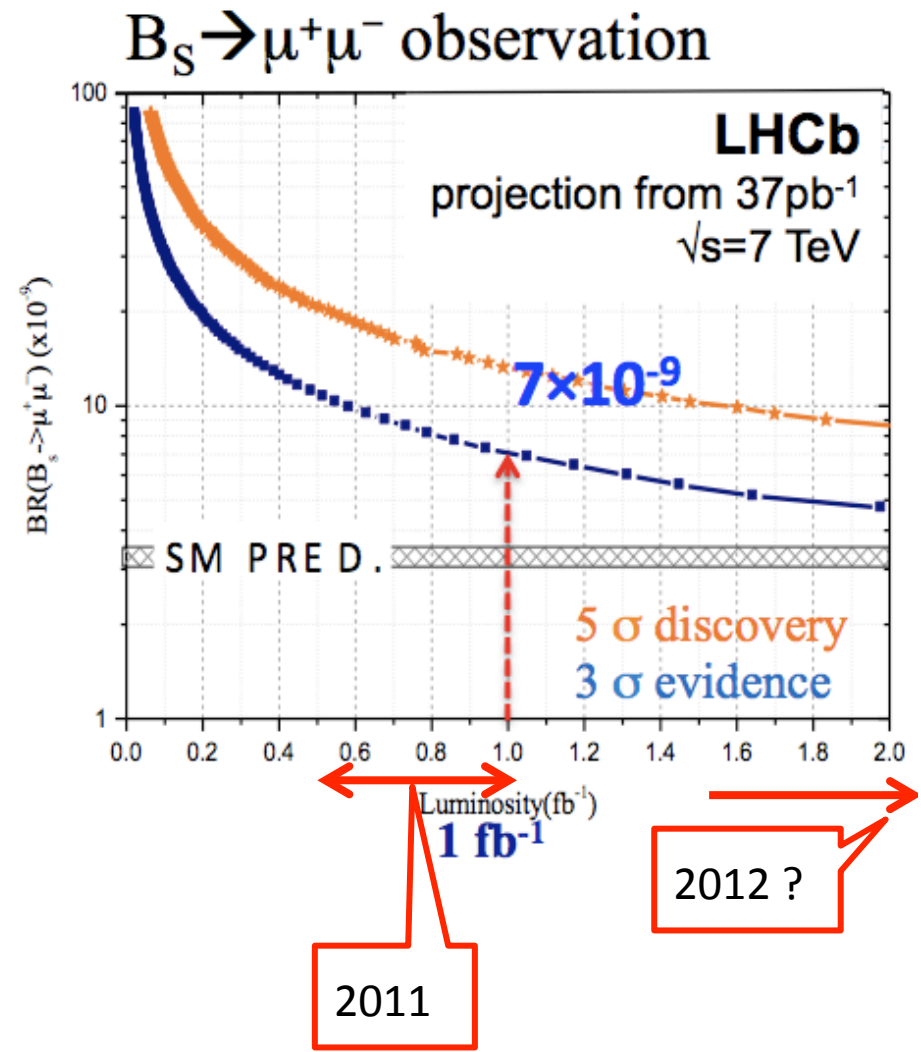
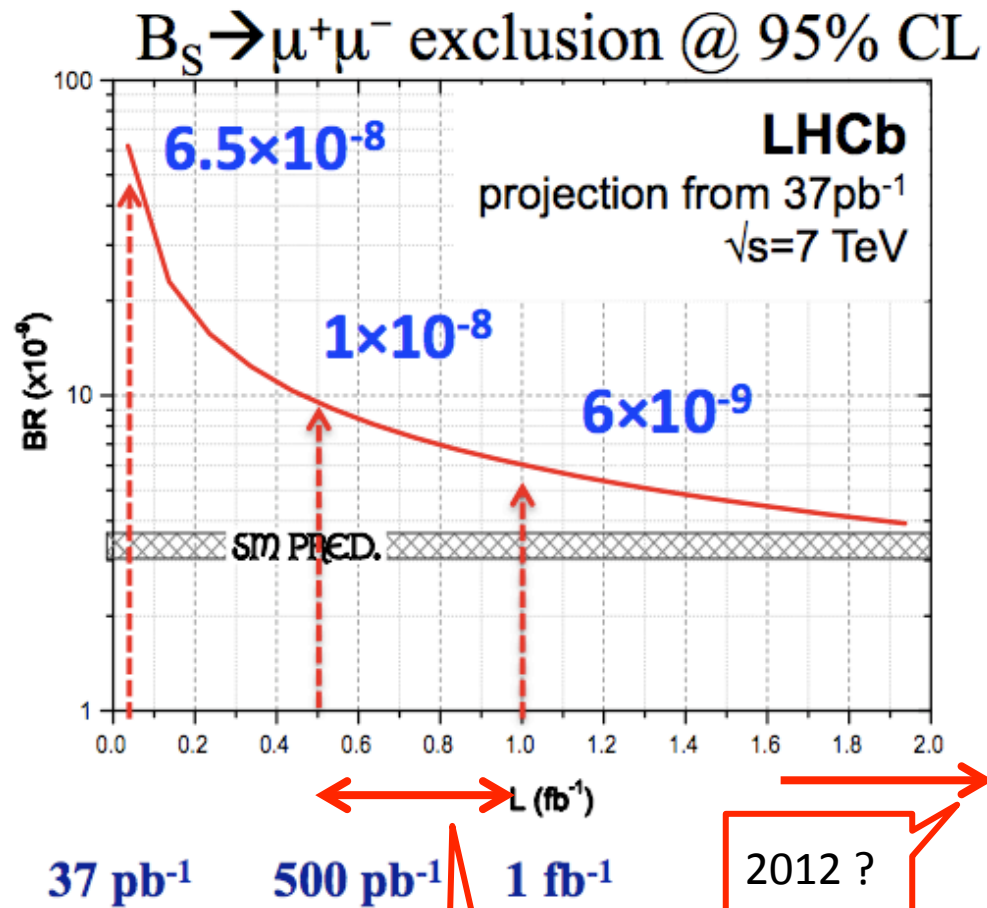
$$\text{Br}(B_d \rightarrow \mu\mu) < 0.37(0.46) \times 10^{-8} \text{ @ 90(95)\%C.L.}$$

LHCb 0.3fb-1 @ $\sqrt{s}=7\text{TeV}$

$$\text{Br}(B_s \rightarrow \mu\mu) < 1.3(1.6) \times 10^{-8} \text{ @ 90(95)\%C.L.}$$

$$\text{Br}(B_d \rightarrow \mu\mu) < 0.42(0.52) \times 10^{-8} \text{ @ 90(95)\%C.L.}$$

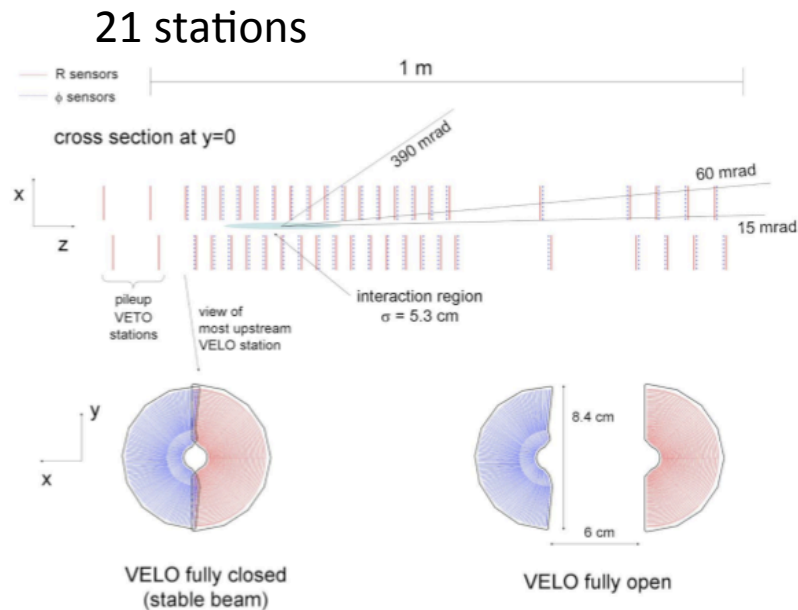
Future prospects



Back up

VELO

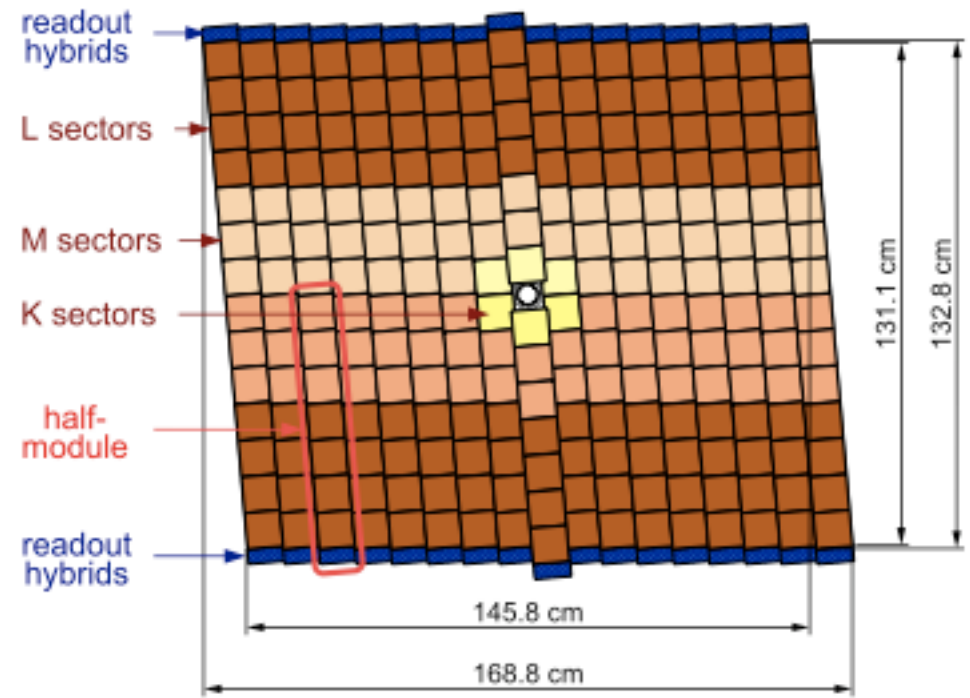
Table 5.1: Principal characteristics of VELO sensors.



	R sensor	ϕ -sensor
number of sensors	42 + 4 (VETO)	42
readout channels per sensor	2048	2048
sensor thickness	300 μm	300 μm
smallest pitch	40 μm	38 μm
largest pitch	102 μm	97 μm
length of shortest strip	3.8 mm	5.9 mm
length of longest strip	33.8 mm	24.9 mm
inner radius of active area	8.2 mm	8.2 mm
outer radius of active area	42 mm	42 mm
angular coverage	182 deg	\approx 182 deg
stereo angle	-	10–20 deg
double metal layer	yes	yes
average occupancy	1.1%	1.1/0.7% inner/outer

the *pile-up veto system* and are described in section 7.1. The VELO sensors are placed at a radial distance from the beam which is smaller than the aperture required by the LHC during injection and must therefore be retractable. The detectors are mounted in a vessel that maintains vacuum around

TT



bibliography

The LHCb Detector at the LHC
The LHCb Collaboration

Search for the very rare decays $B_{s/d} \rightarrow \mu^+ \mu^-$ at LHCb
Justine Serrano

Search for $B_{s(d)}^0 \rightarrow \mu^+ \mu^-$ with CMS
Urs Langenegger

Updated Search for $B_{s/Bd} \rightarrow \mu^+ \mu^-$ at CDF
Thomas Kuhr

<http://lpc.web.cern.ch/lpc/lumiplots.htm>

(Many materials from this slide)

Search for the rare decays $B_s \rightarrow \mu^+ \mu^-$ and $B_0 \rightarrow \mu^+ \mu^-$ with the LHCb Experiment
Johannes Albrecht

Panning for the Golden Decay $B_s^0 \rightarrow J/\psi \phi$ at LHCb
Jan Amoraal