Measurement of the Reference Channel $B^{\pm} \to J/\psi K^{\pm}$ for the Rare Decay $B_s^0 \to \mu^+\mu^-$ with the ATLAS Experiment

DISSERTATION

zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften

vorgelegt von
M.Sc. Bakul Gaur

eingereicht bei der
Naturwissenschaftlich-Technischen Fakultät der
Universität Siegen

September 2013

Dedicated to my mother, Dr. Usha Gaur.
Your lessons are the most invaluable asset.

Abstract

The ATLAS experiment is one of the four major experiments at the Large Hadron Collider (LHC) of CERN, Geneva. The experiment addresses the open questions of particle physics.

The decay of the B_s^0 meson to two muons is strongly suppressed in the Standard Model (SM). It is a rare decay with an expected branching fraction of 3.56×10^{-9}.

The $B_s^0 \to \mu^+\mu^-$ branching fraction is measured relative to a reference channel, $B^\pm \to J/\psi K^\pm$. The subject of this thesis is the measurement of the $B^\pm$ yield. The analysis uses $4.9\,\mathrm{fb}^{-1}$ of ATLAS data collected in the year 2011 at $7\,\mathrm{TeV}$ center-of-mass energy. The $B^\pm$ candidate selection is performed using a multivariate (MVA) technique. The yield is estimated using a two-dimensional unbinned maximum likelihood fit that accounts for the per-event uncertainties. The result is used for the computation of the upper limit on the $B_s^0 \to \mu^+\mu^-$ branching fraction published by the ATLAS collaboration [1, 2].

Zusammenfassung

Das ATLAS Experiment ist eines der vier großen Experimente des Large Hadron Collider (LHC) am CERN, Genf. Mit diesem Experiment sollen die noch offenen Fragen der Teilchenphysik geklärt werden.

Der Zerfall des B_s^0 Mesons in zwei Myonen ist ein stark unterdrückter Zerfallskanal des Standard Modells (SM). Es ist ein seltener Zerfall mit einem erwartetem Verzweigungsverhältnis von 3.56×10^{-9}.

Das $B_s^0 \to \mu^+\mu^-$-Verzweigungsverhältnis wird relativ zum Referenzkanal $B^\pm \to J/\psi K^\pm$ gemessen. Das Thema diese Arbeit ist die Messung der $B^\pm$-Ausbeute. Die Grundlage der Analyse sind $4.9\,\mathrm{fb}^{-1}$ der ATLAS Daten, welche im Jahr 2011 bei einer Schwerpunktsenergie von $7\,\mathrm{TeV}$ aufgenommen wurden. Die Auswahl der $B^\pm$-Ereignisse wurde unter Verwendung einer multivariaten Datenanalyse (MVA) durchgeführt. Die Ausbeute wird mittels eines zweidimensionalen ungebinten Maximum-Likelihood-Fits abgeschätzt, welcher die Pro-Ereignis-Unsicherheiten berücksichtigt. Mit Hilfe der Ergebnisse wird die obere Grenze auf das $B_s^0 \to \mu^+\mu^-$-Verzweigungsverhältnis berechnet, welche von der ATLAS-Kollaboration veröffentlicht wurde [1, 2].

Contents

Chapter 1

Introduction

The discovery of the electron in the late 19th century marked the beginning of a new era of understanding about the nature of matter. Since then the matter is known to be made up of fundamental particles. By now several other elementary particles have been discovered. It is remarkable that the diverse phenomena of the macroscopic world can be explained by interactions between a small set of elementary particles. Particle physics is a field of science that seeks to describe the properties of these particles and their interaction mechanisms. The current understanding is summarized in the Standard Model (SM) of particle physics.

The SM has been a very successful theory. However, it is incomplete. It describes how the three fundamental forces, electromagnetism, strong and weak nuclear forces, act on the particles, but it cannot yet explain the most commonly known force, the gravitation. Moreover, it describes only about 4-5% of the universe, the rest is dark matter and dark energy that are not yet fully understood. For reasons like these the world-wide particle physics community is seeking a theory that extends the SM. Several theories like Supersymmetry, Technicolor, Extra Dimensions, etc. have been proposed.

The Large Hadron Collider (LHC) situated in Geneva allows for experiments at the frontier of particle physics research. It is a proton accelerator designed to reach high luminosities at 14 TeV center-of-mass energy. It will provide the necessary data to perform precision measurements to check some of the SM predictions, and to search for hints of New Physics (NP).

The ATLAS detector is one of the four major particle detectors at the LHC. It is a sophisticated instrument that is used to measure the properties of the particles produced as a consequence of parton interactions in proton-proton collisions. It records the event data that is used later to perform physics analyses. The $B^0_{s,d} \to \mu^+\mu^-$ decays are sensitive probes of NP. The objective of one of the analyses in ATLAS is to measures the rate of the decay of $B^0_s \to \mu^+\mu^-$. It is measured relative to another decay, $B^\pm \to J/\psi K^\pm$. Both B^0_s and $B^\pm$ particles are produced in the proton-proton collisions. They have different production rates. This thesis describes the measurement of the $B^\pm$ yield in the context of the ATLAS $B^0_s \to \mu^+\mu^-$ analysis.

The thesis begins with an introduction of the ATLAS detector (Chapter 2) with emphasis on the sub-detectors that are important in the present analysis. Then a short overview of the data handling and event pre-processing is presented.

Chapter 3 discusses the theoretical aspects of the $B^0_s \to \mu^+\mu^-$ decay. It presents an introduction to the SM. The $B^0_s \to \mu^+\mu^-$ decay is a flavor changing interaction. In the SM it is forbidden at tree level. The factors responsible for the low decay rate of $B^0_s \to \mu^+\mu^-$ are

discussed. The chapter also presents a review of the latest predictions for its SM branching fraction.

It is a challenging task to measure the $B_s^0 \to \mu^+\mu^-$ branching fraction, because there are large numbers of background events that need to be suppressed in order to observe a small signal. The ATLAS $B_s^0 \to \mu^+\mu^-$ analysis strategy is discussed in Chapter 4. The chapter also explains the experimental procedure used to determine the branching fraction relative to the reference channel $B^\pm \to J/\psi K^\pm$.

The details of the measurement of the $B^\pm$ yield is presented in Chapter 5. The chapter describes the reconstruction of the events and the selection cuts applied to suppress the background events. The yield is determined using a two-dimensional unbinned maximum likelihood fit that accounts for the per-event uncertainties. The particulars of the fit and the systematic effects are discussed. The final results are summarized at the end.

A summary of the ATLAS $B_s^0 \to \mu^+\mu^-$ analysis is presented in Chapter 6. It also reports the latest $B_s^0 \to \mu^+\mu^-$ results from the other LHC experiments.

Chapter 2

The ATLAS Detector

The Large Hadron Collider (LHC) [3] situated in Geneva is the world's largest high energy particle accelerator designed to accelerate protons and lead-ions. It is built inside a circular tunnel, 27 km in circumference, approximately 100 m below the earth surface. The LHC can accelerate two beams of particles in opposite directions with precisely controlled orbits. The circulating particle beams are focused (to a RMS beam size of $16\,\mu$m [4]), and are allowed to interact at a well defined interaction point (IP) inside a detector.

Six experiments designed to study a wide variety of physics topics have been operating at the LHC. The major experiments are ALICE, ATLAS, CMS and LHCb. They are located inside four underground caverns (Figure 2.1). The other two experiments are LHCf and TOTEM located near the CMS and ATLAS detectors, respectively.

ATLAS is an acronym for A large Toroidal LHC ApparatuS. It is a general purpose detector, meaning that it is not dedicated to studying a particular physical process but a wide range of processes. This chapter provides an overview of the detector and describes some of its features relevant to the analysis presented in this thesis. ATLAS is a sophisticated detector consisting of many sub-detectors and utility systems. Reference [5] gives a detailed description of all its components.

2.1 Overview

The ATLAS detector is about 44 m long and 25 m in diameter. It has several concentric layers of detectors centered around the IP. The first is the Inner Detector (ID) for high-resolution tracking and vertexing. Then there are layers of electromagnetic and hadronic calorimeters that serve the purpose of particle energy measurement. Finally a system of muon detectors covers almost the entire region around the IP to track muons. Figure 2.2 shows the major detector components. This scheme of concentric detector layers provides symmetry in the $R - \phi$ plane (transverse to the beam direction), i.e. the material traversed by a particle in any radial director is almost independent of the azimuthal angle. Due to construction reasons the same is not possible in the longitudinal direction. The detector layers traversed by a particle in the z-direction have a disc-like geometry. The former region of the detector is called the barrel region, and the later the endcap. There are two endcaps, one on either side of the detector.

A schematic diagram of the ATLAS magnet system is shown in Figure 2.3. The ID is built
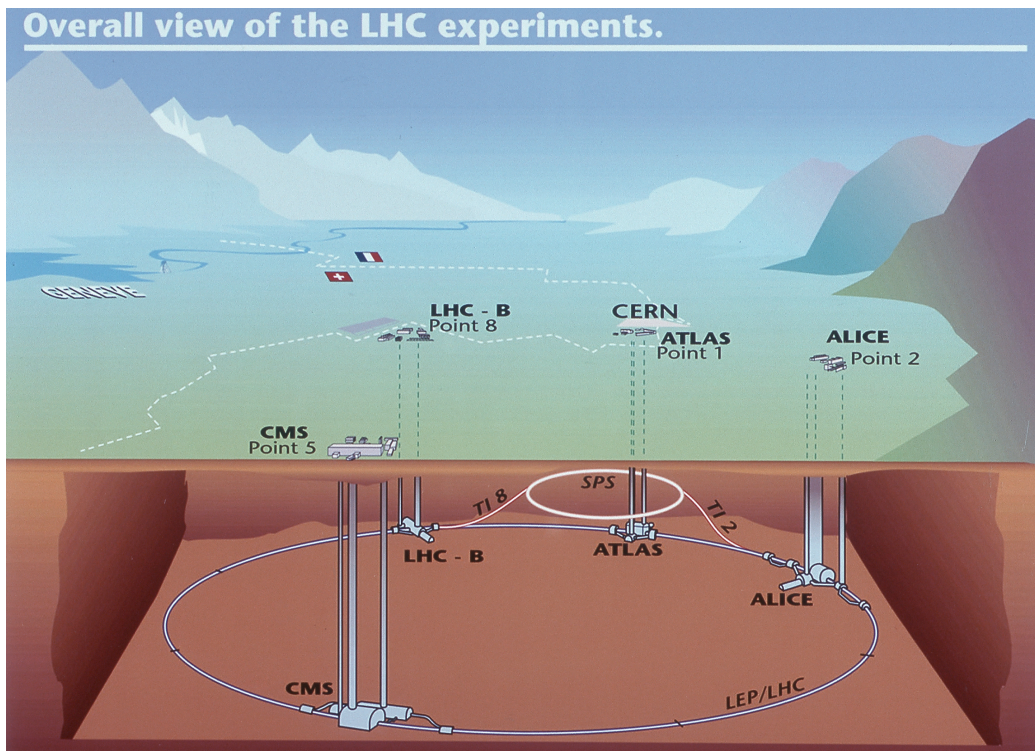
Figure 2.1: The location of the four major experiments (ALICE, ATLAS, CMS and LHCb) at the LHC [6].
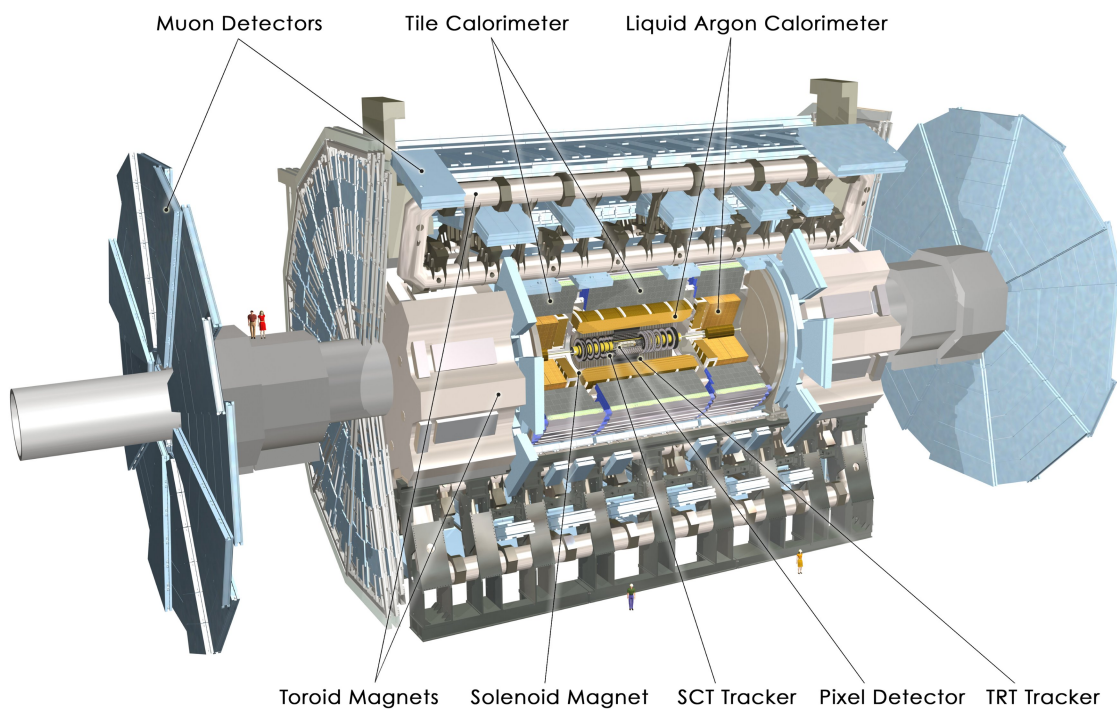

Figure 2.2: A cutaway drawing of the ATLAS detector [7] showing the major components. The detector is about 44 m long and 25 m in diameter.

inside a central superconducting solenoid (CS), 5.3 m long and 2.5 m in diameter, at the center of the detector. It provides a 2 T magnetic field for bending the charged particles to allow particle identification and momentum measurements. There is a system of large superconducting air-core toroid magnets for the Muon Spectrometer (MS). It consists of a set of eight large coils for the barrel, and two sets of eight small coils for the two endcaps. The outer diameter of the barrel toroid (BT) is 20.1 m, and it extends over a length of 25.3 m. Its bending power, $\int B dl$ (where B is the azimuthal magnetic field component integrated over the muon trajectory between the inner and outer radii of the toroid), is 1.5 to 5.5 Tm in the pseudorapidity range $|\eta| < 1.4$. The pseudorapidity is defined as:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right), \tag{2.1}$$

where θ is the polar angle. The endcap toroids (ECT) provide a bending power of 1 to 7.5 Tm in the pseudorapidity range $1.6 < |\eta| < 2.7$. In the transition region ($1.4 < |\eta| < 1.6$) the bending power is lower.
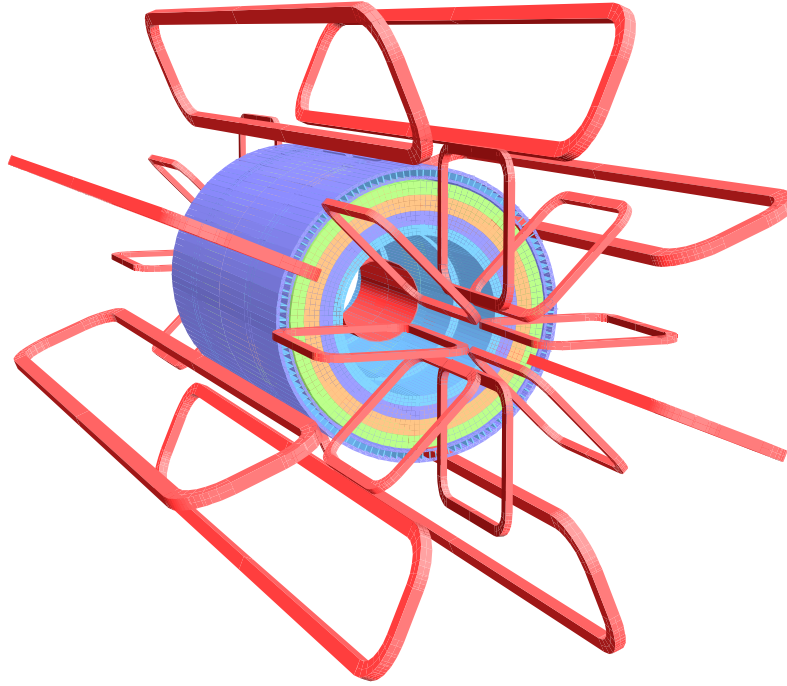
Figure 2.3: A drawing showing the spatial arrangement of the toroid coils (red rings). They extend over a length of 25.3 m. The outer edge of the coils is 10.1 m from the beam axis. The thick inner cylinders depict the calorimeters. The innermost cylinder (red cylinder) depicts the superconducting solenoid (1.3 m in radius), that covers the ID [5].

2.1.1 Inner Detector

The ATLAS ID is designed to work in a high particle flux environment. It employs radiation-hard semiconductors for tracking close to the beam pipe. It consists of a pixel detector to track short lived particles such as the *B* hadrons, and a Semiconductor Tracker (SCT) for high-precision tracking up to a radius of 51.4 cm from the IP. A Transition Radiation Tracker (TRT)

provides continuous tracking further up to a radius of 1.08 m. Figure 2.4 shows a schematic diagram of the ID.

The pixel detector has three barrel layers ($|\eta| < 1.7$), and three layers in each endcap ($1.7 < |\eta| < 2.5$). The layer closest to the beam pipe (B-layer) is about 5 cm away from the beam axis. There are a total of 1744 pixel sensors and 46080 readout channels per sensor. A pixel has a nominal area of $50 \times 400 \, \mu m^2$. The spatial resolution of the pixel detector is $10 \, \mu$m in the transverse plane, and $115 \, \mu$m along the z-direction. This high granularity of the sensor is very important to achieve a good vertex and track impact parameter resolution.

The SCT uses micro-strip silicon detectors. The readout strips are 6.356 cm long with a pitch of $80 \, \mu$m on $285 \, \mu$m thick sensor material. Each sensor has 768 strips. One SCT module has two sensor layers, and there are a total of 4088 modules. In the barrel region ($|\eta| < 1.4$) there are four layers of SCT modules providing four precision space points. The SCT has nine modules (or eighteen layers) in each endcap ($1.4 < |\eta| < 2.5$) with slightly different module geometry. The detector has a resolution of $17 \, \mu$m in the transverse plane, and $580 \, \mu$m in z. The high-resolution of the SCT contributes to the measurements of momentum, impact parameter and vertex position.

The TRT consists of several layers of drift tube (also called straw tube) detectors. It covers up to $|\eta| < 0.7$ and $|\eta| < 2.5$ in the barrel and endcap regions, respectively. The tubes in the barrel are parallel to the beam direction, and they provide $R - \phi$ position measurements of the tracks ($130 \, \mu$m resolution). In the endcaps the tubes are radially arranged in wheels. The ATLAS TRT has about 351000 straw tubes. They have a diameter of 4 mm and a $31 \, \mu$m tungsten anode wire for readout. The drift gas mixture used is 70% Xe, 27% CO_2 and 3% O_2. The tubes are made of layers of materials of varying dielectric constant. The passage of a charge particle produces transition radiations at the interfaces of the materials. The emitted photons are detected. Tracks typically produce a large number of hits (~ 36 hits per track) in the TRT. The photons emitted by electrons are in the X-ray region, and are hence used for particle identification.

2.1.2 Muon System

The ATLAS muon system consists of four sub-detectors. The Monitored Drift Tube (MDT) chambers form the outermost layer of the detector. They measure the coordinates of the muon tracks that have sufficient momentum to traverse through the calorimeters and reach the MS. These muons are deflected in the magnetic field generated by the barrel and endcap toroids. Muons with very low transverse momenta, $p_T < 0.5$ GeV, cannot reach the MS. Wherever high granularity is required Cathode Strip Chambers (CSC) are used. The MS also consists of Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). They are used for generating a trigger signal to report an event. Figure 2.5 shows the location of the four sub-detectors.

The Monitored Drift Tube chambers (MDTs) are precision detectors for momentum measurement and tracking. Their average spatial resolution is $35 \, \mu$m. The chambers consist of three to eight layers of drift tubes. There are three concentric layers of MDTs parallel to the beam axis in the barrel at radii 5 m, 7.5 m and 10 m. There are four large wheels in each endcap located at ± 7.4 m, ± 10.8 m, ± 14 m and ± 21.5 m from the IP. They cover a pseudorapidity range of $|\eta| < 2.7$. The drift tubes are made of aluminum. The tubes carry an argon-CO_2 gas mixture
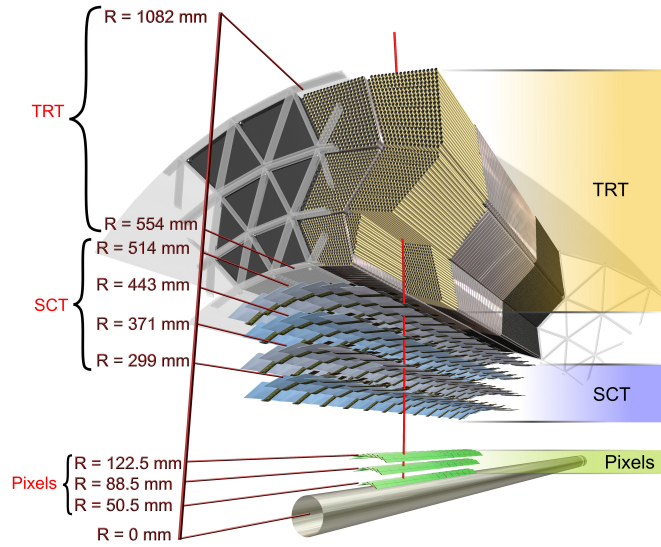
Figure 2.4: A schematic drawing showing the ID layers of the barrel section. There are three layers of pixel detector modules close to the beam pipe. The barrel SCT has eight layers of silicon micro-strip detectors (two layers per module). The TRT employs straw tube detectors for tracking. The barrel has about 50 000 straw tubes [5].

with a ratio of 93:7 at 3 bar pressure. Their thickness is $400\,\mu$m and they have a diameter of 3 cm. In the center of each tube there is a $50\,\mu$m tungsten-rhenium wire acting as anode. The signal is read out via current sensitive preamplifiers.

The CSCs are multi-wire proportional chambers with segmented cathode planes. There is a set of four consecutive planes in each CSC. They provide two-coordinate information with a resolution of $40\,\mu$m in the bending plane and 5 mm in the transverse plane. They are located in the forward region of the detector ($2 < |\eta| < 2.7$), where there is a high flux of particles, and their fast response time is desirable for better tracking efficiency.

There are many physics processes with pure-leptonic and semi-leptonic final states. Muons play a very important role in event detection. There are many analyses in ATLAS that rely on triggers from muons, for example, the $B^0_s \to \mu^+\mu^-$ analysis. The MS encompasses fast trigger elements. The trigger system is described in Section 2.2. The RPCs provide trigger elements in the barrel region ($|\eta| < 1.05$) and TGCs in the endcaps ($1.05 < |\eta| < 2.4$).

A RPC has two resistive plates made of phenolic-melaminic plastic laminate. The two plates are kept parallel to each other and 2 mm apart by means of insulating spacers. The RPCs are operated at a high drift field, $4.9\,$kV$/$mm, to induce avalanches in the gas medium ($C_2H_2F_4/Iso-C_4H_{10}/SF_6$ gas mixture in ratio 94.7:5:0.3) when an ionizing particle traverses. The signal is read out on metallic strips mounted on the outer surface of the resistive plates. A coincidence signal from a system of RPCs is used as trigger. This rejects fake hits due to detector noise, and improves the trigger efficiency. The RPCs can trigger on tracks in a wide p_T range, from 6 to 35 GeV.

The TGCs are special multi-wire proportional chambers where the wires are closer (1.4 mm) to the cathode then to one another (1.8 mm). Using a high quenching gas mixture (CO_2 and $n - C_5H_{12}$) these chambers are able to reach quasi-saturation mode with a high gas gain of $\sim 3 \times 10^5$. The TGCs not only provide the trigger capability, but also a second azimuthal coordinate to compliment the MDTs in the endcap regions. The RPCs and TGCs have a typical response time of 15 to 25 ns, which suffice to trigger events at every 25 ns (LHC bunch spacing).
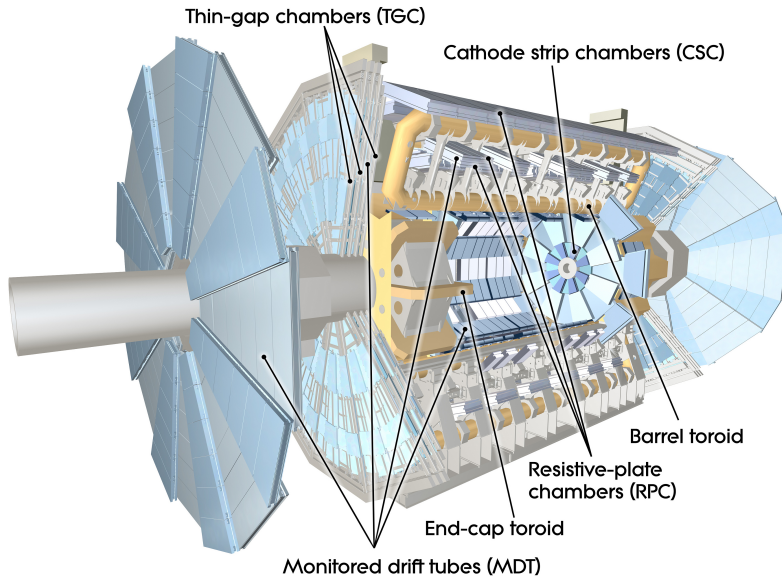
Figure 2.5: A cutaway drawing of the ATLAS detector showing the various components of the MS [5].

2.2 Event Triggering

One of the challenges faced by the experiments at the LHC is the large QCD background originating from lighter quarks. The trigger, that initiates the recoding of an event, should be very efficient in selecting events for physics analyses. The event rates can be as high as the proton bunch crossing frequency, 40 MHz. It is the task of the trigger to limit the event rates to something manageable by the data processing elements.

The ATLAS trigger system is a 3-level trigger system. A schematic diagram of the trigger system is shown in Figure 2.6. The Level 1 (L1) trigger is a hardware based trigger designed to make decisions in real-time. It is capable of reducing event rates to a few hundred kilohertz. The Level 2 (L2) trigger is a software based trigger. It uses partial event information and fast algorithms to identify useful events. It further reduces the event rates to a few kilohertz. The third trigger level, referred to as the Event Filter (EF), is a software trigger that uses the full event information to qualify the events. The L2 and EF trigger levels are together referred to as the High Level Trigger (HLT). The output rate after the HLT selection is about 400 Hz [8]. This trigger bandwidth is shared by various physics data streams.

For *B*-physics studies the events are selected based on muons. The L1 trigger is activated by a coincidence signal from the RPCs in the barrel region and the TGCs in the endcaps. It selects
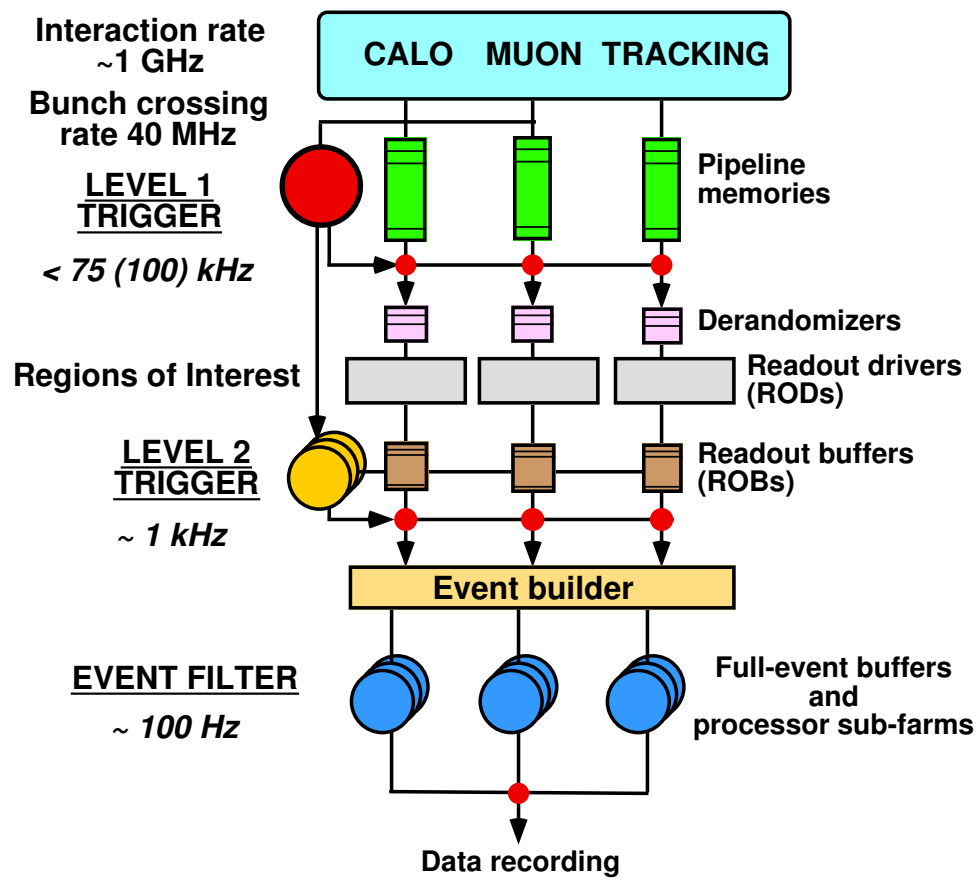
Figure 2.6: A schematic diagram of the ATLAS trigger and data acquisition systems [4].

muons above a certain p_T threshold. Figure 2.7 shows a quarter-section of the ATLAS muon system and the location of the various trigger elements. For each muon a region of interest (RoI) is defined based on its position information. The information is passed on to the second stage. At L2 a more detailed investigation is performed in the RoI. The muon tracks are reconstructed using fast algorithms. The analysis presented in this thesis uses di-muon candidates. A di-muon candidate is selected at L1, and L2 combines the two muon tracks to represent a resonance, e.g., the J/ψ resonance. The L2 trigger then selects candidates in a specific di-muon invariant mass range. This is referred to as the topological di-muon trigger. The decision of the L2 trigger is passed on to the EF, where a full reconstruction is performed using the complete event information. The EF refines the decision of the L2 trigger. References [8–10] present detailed information on the muon trigger system.
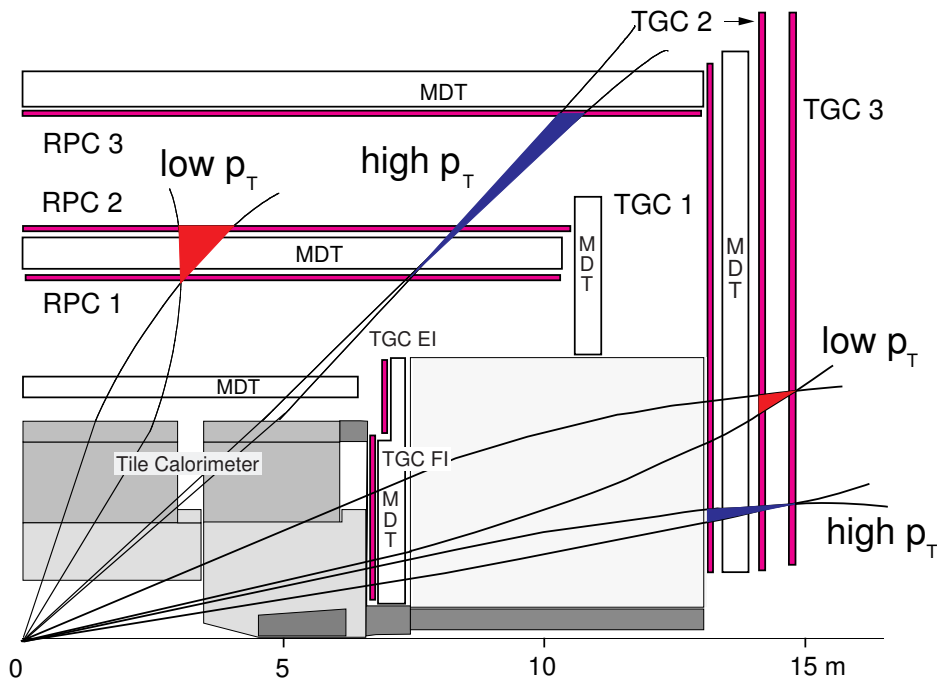
Figure 2.7: The L1 muon trigger scheme [8]. The diagram shows a quarter-section of the ATLAS muon system. The RPCs and TGCs are used to activate the L1 trigger.

The LHC is designed to operate at high luminosities. The amount of luminosity delivered to the experiments can vary from time to time. When the LHC is offline, the detector is used to record cosmic radiations, e.g., for the alignment and calibration of the sub-detectors. The trigger system is configurable to cope with these conditions. There are trigger menus for different luminosity conditions. The trigger can be pre-scaled to reject events to control the final output rates.

2.3 Data Handling

The data of an event selected by the L1 trigger is held in readout buffers (RoB) until the HLT rejects the event. If the event is accepted, the data acquisition system (DAQ) builds the event-

data structure and transfers it to a storage location. The EF uses the same event-data structure. After the event has been accepted, the EF classifies it into physics streams, and a corresponding tag is added to the event-data structure.

The data is processed shortly after its acquisition. A subset of the events is used to perform data quality and detector monitoring. The various sub-detector alignment and calibration constants are applied. A copy of the data is stored at CERN, and a replica is transferred to regional centers around the world, where the data can be accessed and analyzed.

In ATLAS the smallest unit of data is a luminosity block (LB). A LB is the shortest time interval (~ 2 minutes) for which the integrated luminosity can be determined. The data quality assessment is performed at the LB level, and if the LB does not meet the required quality criteria, then an appropriate quality flag is set. This information can be used later to reject events not suitable for the analysis.

2.4 The LHC Status and Plans

The LHC has been in operation since November 2009. It first began colliding proton beams of 450 GeV energy each (total center-of-mass energy, $\sqrt{s} = 900$ GeV). The beam energy has since been increased to 4 TeV in 2012. The machine reached an instantaneous luminosity of $\sim 7.73 \times 10^{33}$ cm^{-2}s^{-1}. In a typical LHC fill 1 374 proton-bunches[1] were accelerated, and collided at a rate of 40 MHz. There were also dedicated heavy-ion collisions runs (p-Pb and Pb-Pb) up to $\sqrt{s_{NN}} = 5$ TeV.

The ATLAS detector collected pp-collisions data worth 45 pb^{-1}, 5.25 fb^{-1} and 21.7 fb^{-1} in years 2010, 2011 and 2012, respectively [11]. A consequence of increase in instantaneous luminosity is the increased occupancy in the detector. A single bunch crossing can produce several interactions, or so called pile-up events, which can impose tighter constraints on analyses. Figure 2.8 shows the increase of pile-up in ATLAS as a function of time.
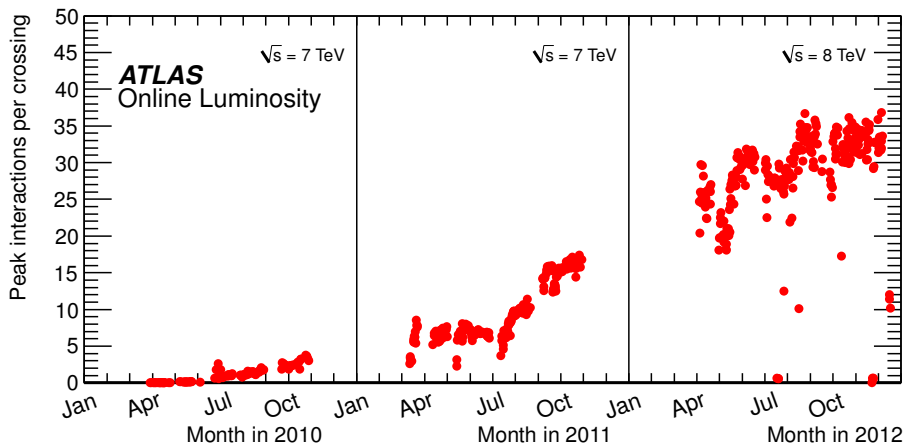

Figure 2.8: The maximum number of pile-up events in pp-collisions during stable beam periods in years 2010, 2011 and 2012 [11].

---

[1] 1 proton-bunch has about 1.15×10^{11} protons

Currently the LHC is undergoing maintenance and upgrade. Some of the ATLAS sub-detectors are also being upgraded. The machine is expected to regain operation in 2015. After the upgrade the accelerator is expected to run at 13 TeV center-of-mass energy, before reaching its design energy of 14 TeV, delivering an instantaneous luminosity of $10^{34}\,\mathrm{cm^{-2}s^{-1}}$ to the experiments.

There are proposals to upgrade the LHC, for example, to higher luminosities and energies [12, 13]. Higher occupancy and radiation levels imply modifications of the ATLAS detector as well. Studies to improve the ID, MS, calorimeters and the trigger system are in progress [14].

Chapter 3

Rare Decays of B Mesons

One of the prime objectives of particle physics is to find a model that describes how all sub-nuclear particles interact with each other. Any particle is subject to the four fundamental forces of nature: the strong force, the electromagnetic force, the weak force, and the gravitational force. Each force has a different propagation mechanism, strength and range. The strong force is the strongest of the four, but is only effective in the femtometer range. The electromagnetic force is about hundredth the strength of the strong force, and the weak force is several orders of magnitude weaker than that. The weakest is the gravity. Like the strong force, the weak force is short ranged, but the other two have an infinite range. The quest is to find one theory that explains all forms of matter and their interactions. The theory should also describe how the particles acquire their properties, e.g., the electric charge, mass, etc. Some of these issues are addressed by the Standard Model (SM), and some still need better theories.

An important subject in particle physics is the study of B mesons. They are short-lived particles (lifetime $\sim 10^{-12}$ s) produced as a result of particle collisions in high-energy particle accelerators. There are eight main members of the B meson species: $B_{(u)}^+$, $B_{(d)}^0$, B_c^+, B_s^0 and their anti-particles. The B mesons are used to perform precision measurements of theoretical parameters, and to look for New Physics (NP) effects. This chapter presents an introduction to the physics of the B_s^0 meson, and in particular, the rare decay of B_s^0 into two muons. It begins with an introduction to the SM and some of the essential topics in the context of B-physics. A short review of the current theoretical predictions of the $B_s^0 \to \mu^+\mu^-$ decay rates is presented at the end.

3.1   Standard Model

The Standard Model of particle physics is currently the best description of the nature of the elementary particles. It explains all experimental results so far. The elementary particles are classified, according to their characteristics, into two main categories. Particles with half-integer spins are called fermions. They obey the Pauli exclusion principle, and their characteristics are described by the Fermi–Dirac statistics. Particles with integer spins, called bosons, are described by the Bose–Einstein statistics.

Elementary matter particles like quarks and leptons are fermions. Experiments reveal they occur in three generations. The u- and d-quarks, the electron (e) and the electron-neutrino (ν_e)

form the stable first generation. The other two generations are unstable. Table 3.1 lists the elementary fermions by generation. Quarks and leptons can be identified based on their flavor quantum numbers and isospin (I_3). Quarks have a color-charge, which can take three values, red, green or blue. Hence, there are a total of 36 different quarks and anti-quarks. They exhibit a phenomenon called color confinement. They do not exist as free particles, but as color-neutral multi-quark bound states. For example, baryons are composite fermions consisting of 3 quarks. Baryons are assigned a baryon number, $B = 1$. Quarks have a baryon number, $B = \frac{1}{3}$. All elementary fermions have anti-particles with the same mass, but opposite charge and opposite quantum numbers.

	Generation			Charge	
	1st	2nd	3rd	$Q/|e|$	Spin
Quarks	u	c	t	$+2/3$	$1/2$
	d	s	b	$-1/3$	$1/2$
Leptons	e	μ	τ	-1	$1/2$
	ν_e	ν_μ	ν_τ	0	$1/2$

Table 3.1: The three generations of elementary fermions. The electric charge is expressed relative to the charge of an electron, $|e|$.

Quark–anti-quark bound states are called mesons. Mesons are composite bosons. For example, the B^+ meson consists of an anti-b-quark ($\bar{b}$) and a u-quark. Similarly, the B_s^0 meson is a $s\bar{b}$ system. All B mesons are unstable, and they spontaneously decay into other particles.

The SM describes the interaction mechanism between elementary fermions. The electromagnetic interactions between charged particles are summarized in the Quantum Electrodynamics (QED) theory [15]. The Glashow-Weinberg-Salam (GWS) theory [16–18] unifies the electromagnetic and weak interactions into a single theory of electroweak interactions. Lastly, the strong interactions are described in the theory of Quantum Chromodynamics (QCD) [19]. The interactions take place through exchange of mediator particles. There are mediators for each type of fundamental interaction. Table 3.2 lists the type of interactions[1] and the associated mediators. They all have a spin of 1, hence they are bosons. The gluons possess color-charges; a color and an anti-color. There are 8 gluon states: $r\bar{g}$, $g\bar{r}$, $r\bar{b}$, $b\bar{r}$, $g\bar{b}$, $b\bar{g}$, $(r\bar{r} - b\bar{b})/\sqrt{2}$, $(r\bar{r} + b\bar{b} - 2g\bar{g})/\sqrt{6}$. Gluons are not only mediators, but they can also interact with other colored particles (quarks and gluons).

Mathematically, the SM is a non-Abelian gauge theory with the gauge group $SU(3) \otimes SU(2) \otimes U(1)$. The electric charge and the spin quantum number are conserved in all interactions. The lepton and the baryon numbers are always conserved. The quark flavors are conserved in electromagnetic and strong interactions, but not in weak interactions. The various conservation laws require the theory to be symmetric under gauge transformations. The group $SU(3)$ describes the symmetry of the color-charge in strong interactions. The group

---

[1]Although gravitation is not yet included in the SM, the quantum field theory predicts the graviton, a particle of spin 2, that mediates gravitational interactions [20]. The general theory of relativity is a good model for gravitation, but there is no successful quantum theory of gravity yet.

Interaction	Mediator	
Strong	G	(gluon)
Electromagnetic	γ	(photon)
Weak	$W^{\pm}, Z^0$	

Table 3.2: Types of interactions and their mediators in the SM.

$SU(2) \otimes U(1)$ describes the electroweak interactions, where the $SU(2)$ group represents the weak isospin (I), and the $U(1)$ group represents the weak hyper-charge (Y). The hyper-charge is related to the electric charge, Q, and the third component of the weak isospin, I_3,

$$Y = 2(Q - I_3). \tag{3.1}$$

In the model the gauge fields were first introduced massless. This is contrary to the experiments, where $W^{\pm}$ and Z^0 bosons have been observed to have masses. The gauge fields acquire masses through spontaneous symmetry breaking from $SU(2) \otimes U(1) \to U(1)$. This is achieved by including a scalar field, ϕ (Higgs field), which has a non-zero vacuum expectation value (Higgs mechanism [21–23]). The SM predicts a Higgs boson, H^0, an electrically neutral particle of spin 0. It couples to fermions giving them masses proportional to the vacuum expectation value of the Higgs field ($\phi_{\min}$). The existence of a Higgs boson was recently confirmed at the LHC [24, 25].

Although the SM is very successful, there remain several unanswered questions. It does not explain why there exist only three generations of elementary fermions, and why there is a hierarchy in their masses. The cause of electroweak symmetry breaking is not understood. The gravitational interactions are not yet incorporated in the model. On the cosmological side, dark matter, dark energy and the baryon–anti-baryon asymmetry are not yet understood. All point to the fact that there exists physics beyond the SM.

Several theories have been proposed to extend the SM. Some examples of Beyond the Standard Model (BSM) theories are Supersymmetry, Technicolor and the Higgs Doublet Model. The LHC experiments are now providing new results and improved measurements, thereby constraining the parameters of these theories.

3.2 CKM Matrix and the Unitary Triangle

The weak interactions are not necessarily flavor conserving, i.e. a quark may decay into another quark, even across its generation. Processes like $K_L^0 \to \pi^+ e^- \nu_e$, where a s-quark decays to a u-quark by exchange of $W^{\pm}$, are possible. These are flavor changing changed-current (weak) interactions.

The change of flavor in weak interactions is allowed by the Glashow-Iliopoulos-Maiani (GIM) mechanism [26], where a weak interaction takes place between a flavor eigenstate, for example, u, and a rotated eigenstate, d'. The weak eigenstate d' can be expressed as a linear combination of the pure flavor eigenstates d, s, and b,

$$d' = V_{ud}d + V_{us}s + V_{ub}b. \tag{3.2}$$

The factors V_{ud}, V_{us} and V_{ub} are coupling strengths of the u-quark to the d, s and b quarks, respectively. The decay probabilities of quarks in flavor changing weak interactions are represented by a 3×3 unitary matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [27, 28]:

$$
\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}
$$
$$
= V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}.
$$
(3.3)

The SM offers no handle to compute the elements of this matrix. They are empirically measured from experiments [29]:

$$
V_{\text{CKM}} = \begin{pmatrix} 0.9743 & 0.2253 & 0.0035 \\ 0.2252 & 0.9734 & 0.0412 \\ 0.0087 & 0.0404 & 0.9991 \end{pmatrix}.
$$
(3.4)

Since flavor changing weak interactions exist across quark generations, the non-diagonal elements of this matrix are non-zero, but their probabilities are small. The probability of a quark conversion within the same generation (diagonal elements) is nearly 1.

The CKM matrix is unitary. Therefore, the following relation holds:

$$
V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0,
$$
(3.5)

where V_{ij}^* is the complex conjugate of the matrix element V_{ij}. It is useful to parametrize the matrix in terms of the 4 Wolfenstein parameters, λ, A, ρ and η [30]:

$$
V_{\text{CKM}} \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}.
$$
(3.6)

From Equations 3.5 and 3.6 it follows that the matrix can be represented by a triangle in a complex plane of $\bar{\rho}$ and $\bar{\eta}$, with sides of length $|V_{ud}V_{ub}^*/V_{cd}V_{cb}^*|$, $|V_{td}V_{tb}^*/V_{cd}V_{cb}^*|$ and 1. This is called the unitarity triangle, and is shown in Figure 3.1. In Figure 3.2 several measurements of the CKM matrix parameters are compared. The final values of the parameters ($\lambda = 0.2253 \pm 0.0006$, $A = 0.811^{+0.022}_{-0.012}$, $\rho = 0.131^{+0.026}_{-0.013}$, $\eta = 0.345^{+0.013}_{-0.014}$) are obtained using a global fit to all measurements imposing theoretical constraints and the unitarity condition. The details of the fit are described in [29, 31, 32].

3.3 Flavor Changing Neutral Currents

In the SM, weak processes like $b \rightarrow s + Z^0$ are forbidden due to flavor conservation laws and the GIM mechanism. They can only proceed indirectly via the exchange of $W^\pm$, for example, $b \rightarrow q + W^- \rightarrow s + Z^0$, where q is one of the u, c, and t quarks. Since these processes depend on the weak charged current interaction of the quarks, they are governed by the CKM parameters. These decays are highly suppressed in the SM.
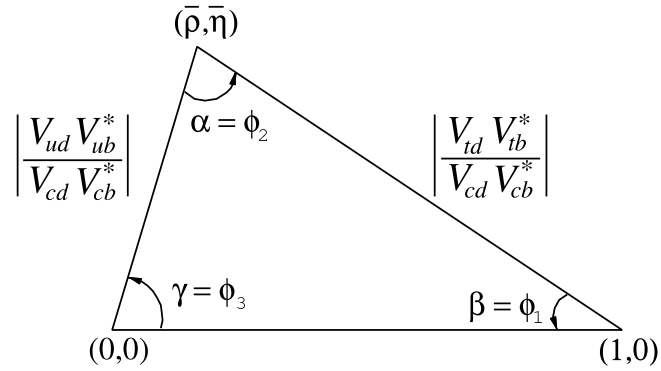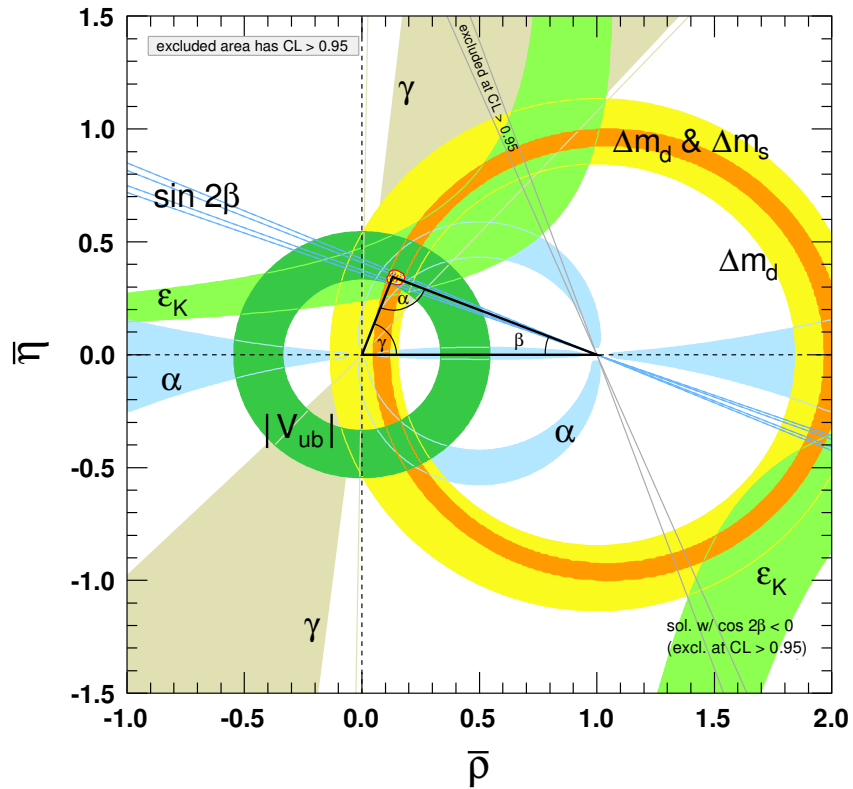
Figure 3.1: The unitarity triangle [29].

Figure 3.2: The unitarity triangle fit. The various shaded regions are the results from different measurements [29] at 95% Confidence Level (CL). The global fit to all measurements is depicted by the solid triangle. The fit allows precise determination of the CKM matrix parameters.

The decays $B^0_{s,d} \rightarrow l^+ l^-$, where l can be one of the three leptons, are examples of flavor changing neutral current (FCNC) processes. The lowest order Feynman diagrams, that contribute to these processes, are shown in Figure 3.3(a). They are referred to as penguin (left) and box (right) diagrams. In the SM, the processes depicted by the Z^0-penguin are the dominating ones. A neutral current process can also be mediated by a photon, but due to vector current conservation there is no contribution from a photon penguin diagram. Since B^0_s and B^0_d are spinless particles, the final state leptons are either both left-handed or right-handed. This leads to a helicity suppression proportional to m^2_l/m^2_B, where m_l and m_B are the masses of the lepton and the B meson, respectively. The $B^0_d \rightarrow l^+ l^-$ decays are further suppressed by a factor of V^2_{td}/V^2_{ts} compared to $B^0_s \rightarrow l^+ l^-$.

(a) Standard Model
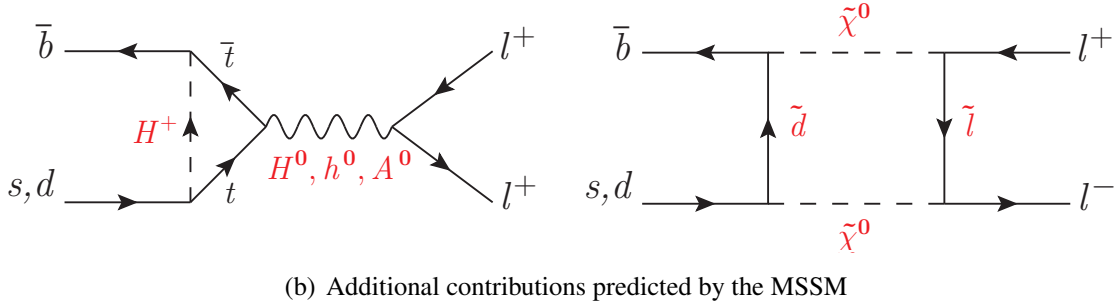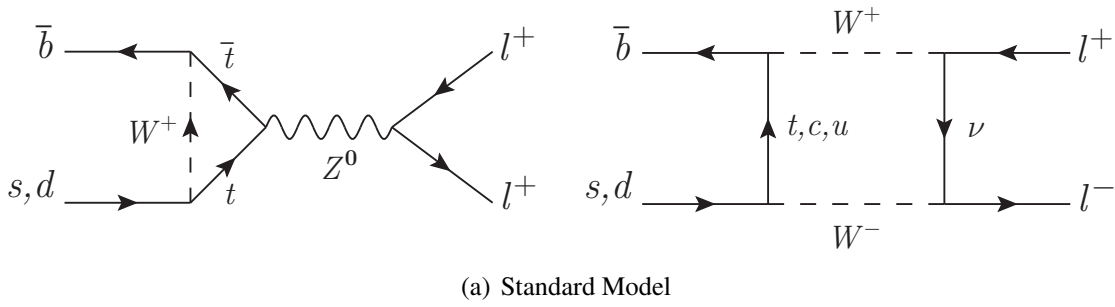

(b) Additional contributions predicted by the MSSM

Figure 3.3: $B^0_{s,d} \rightarrow l^+ l^-$ Feynman diagrams.

The $B^0_s \rightarrow \mu^+ \mu^-$ decay is an important probe for NP. There are BSM models where its amplitude can be enhanced by additional contributions from other mediator particles. For example, in the Minimal Supersymmetric Standard Model (MSSM) there can be contributions from the neutral and the charged MSSM Higgs bosons, H^0, h^0, A^0, $H^\pm$, and from supersymmetric particles like $\widetilde{\chi}^0$. Examples of MSSM penguin and box diagrams contributing to the $B^0_s \rightarrow \mu^+ \mu^-$ decay are shown in Figure 3.3(b). In the MSSM, the decay rate is proportional to $\tan^6 \beta$, where $\tan \beta$ is a MSSM parameter equal to the ratio of the vacuum expectation values of the two Higgs doublets. A review of relevant BSM models is presented in reference [33].

The $B^0_s \rightarrow \mu^+ \mu^-$ channel is not only sensitive to NP, but with just two muons in the final state, it has the advantage of being theoretically clean. Experimentally, the muons are easier to detect and analyze compared to quark induced jets. The experimental aspects of the decay are discussed in Chapter 4.

3.4 $B_s^0 \to \mu^+\mu^-$ Branching Fraction

The branching fraction is defined as the probability that a particle will decay through a specific decay channel. The sum of the branching fractions of all possible decay modes of that particle is 1. The decay amplitudes are obtained from an effective low-energy Hamiltonian describing the weak interaction of the quarks. They are calculated using the operator product expansion (OPE) method [34].

The effective Hamiltonian for the $B_s^0 \to l^+l^-$ decays can be expressed as:

$$\mathscr{H}_{\text{eff}} = -\frac{G_F \alpha}{\sqrt{2}\pi} \left\{ V_{tb}^* V_{ts} \sum_i^{10,S,P} \left(C_i(\mu)\mathscr{O}_i(\mu) + C_i'(\mu)\mathscr{O}_i'(\mu) \right) + h.c. \right\}, \quad (3.7)$$

where, G_F is proportional to the weak coupling constant, and α is the electromagnetic coupling constant. The Wilson coefficients (C_i, C_i') describe the perturbative short-distance physics, and $\mathscr{O}_i$, $\mathscr{O}_i'$ are local operators that describe the non-perturbative long-distance effects. The Wilson coefficients and the local operators both depend on the renormalization scale μ (chosen to be of the order of the t-quark mass $\sim 173\,\text{GeV}$ [35]).

In the SM the most important contribution comes from the term $C_{10}\mathscr{O}_{10}$. The coefficient C_{10} is related to the t-quark mass, $m_t(\mu)$:

$$C_{10}(\mu) = -\frac{\eta_Y Y_0(x_t(\mu))}{\sin^2\theta_W} = -4.134, \quad (3.8)$$

where, $x_t(\mu) = m_t^2(\mu)/m_W^2$, η_Y is a QCD correction factor for the t-quark contribution ($\eta_Y = 1.012$) [36], and θ_W is the weak mixing angle. The loop function Y_0 has the form [36]:

$$Y_0(x_t) = \frac{x_t}{8}\left\{ \frac{4-x_t}{1-x_t} + \frac{3x_t}{(1-x_t)^2}\ln x_t \right\}. \quad (3.9)$$

The other terms in Equation 3.7 represent scalar (S) and pseudoscalar (P) couplings to leptons. In the SM their contributions are negligible. The details of the various operators and Wilson coefficients in Equation 3.7 are described in [33–36].

3.4.1 Theoretical Value

The non-radiative branching fraction of $B_s^0 \to \mu^+\mu^-$ is given by the following expression [35, 37],

$$\mathscr{B}(B_s^0 \to \mu^+\mu^-)_{\text{SM}} = \frac{G_F^4 m_W^4 \sin^4\theta_W}{8\pi^5} \tau_{B_s^0} f_{B_s^0}^2 m_{B_s^0} m_\mu^2 \sqrt{1 - \frac{4m_\mu^2}{m_{B_s^0}^2}} |C_{10}V_{tb}^* V_{ts}|^2, \quad (3.10)$$

where, $\tau_{B_s^0} = (1.503 \pm 0.010)\,\text{ps}$ [38] is the B_s^0 lifetime, and $f_{B_s^0} = (225 \pm 3)\,\text{MeV}$ [39] is its decay constant. Using the central values of m_t, V_{tb} and V_{ts}, the SM branching fraction is predicted to be:

$$\mathscr{B}(B_s^0 \to \mu^+\mu^-)_{\text{SM}} = 3.25 \times 10^{-9} \left[\frac{m_t}{173.2\,\text{GeV}}\right]^{3.07} \left[\frac{f_{B_s^0}}{225\,\text{MeV}}\right]^2 \left[\frac{\tau_{B_s^0}}{1.500\,\text{ps}}\right] \left|\frac{V_{tb}^* V_{ts}}{0.0405}\right|^2$$

$$= (3.25 \pm 0.17) \times 10^{-9}. \quad (3.11)$$

This value includes next-to-leading-order QCD corrections. Its dependence on the deviations in m_t, $f_{B_s^0}$, $\tau_{B_s^0}$, $V_{tb}^* V_{ts}$ is indicated by the exponents on the brackets. The uncertainty in the branching fraction is dominated by the uncertainties on the CKM factor $|V_{tb}^* V_{ts}|$ (4%) and the decay constant $f_{B_s^0}$ (2.7%).

3.4.2 Effective Value

The B_s^0 meson is not a mass eigenstate, but a mixture of two states. B_s^0 can either decay directly, or it can decay after transforming into its anti-particle, $\bar{B}_s^0$. This gives rise to a time-dependent B_s^0–$\bar{B}_s^0$ oscillation of frequency $\Delta m_s = 17.768\,\mathrm{ps}^{-1}$ [40].

The decay width difference $\Delta\Gamma_s = \Delta\Gamma_L - \Delta\Gamma_H = (0.116 \pm 0.018 \pm 0.006)\,\mathrm{ps}^{-1}$ [41] of the light and heavy mass eigenstates, is significant. Due to this non-vanishing width difference the measured B_s^0 branching fraction is a time integrated value [42, 43]:

$$\overline{\mathscr{B}}(B_s^0 \to \mu^+\mu^-)_{\mathrm{SM}} = \frac{1}{2} \int_0^\infty \langle \Gamma\left(B_s^0(t) \to \mu^+\mu^-\right)\rangle. \tag{3.12}$$

At $t = 0$ the B_s^0–$\bar{B}_s^0$ mixing effects are absent, and the above expression is represented by Equation 3.10. The theoretical value (Equation 3.11) is related to the effective branching fraction, $\overline{\mathscr{B}}(B_s^0 \to \mu^+\mu^-)$:

$$\mathscr{B}(B_s^0 \to \mu^+\mu^-)_{\mathrm{SM}} = \left[\frac{1 - y_s^2}{1 + \mathscr{A}_{\Delta\Gamma}^{\mu\mu} y_s}\right] \overline{\mathscr{B}}(B_s^0 \to \mu^+\mu^-)_{\mathrm{SM}}, \tag{3.13}$$

where, the decay width parameter $y_s = \Delta\Gamma_s/2\Gamma_s = 0.087 \pm 0.014$ is taken from other measurements [41], and $\mathscr{A}_{\Delta\Gamma}^{\mu\mu} \in [-1, +1]$ is an observable. In the SM $\mathscr{A}_{\Delta\Gamma}^{\mu\mu} = +1$. Therefore, the time-integrated SM branching fraction is [33, 42]:

$$\begin{aligned}\overline{\mathscr{B}}(B_s^0 \to \mu^+\mu^-)_{\mathrm{SM}} &= \frac{1}{1 - y_s} \mathscr{B}(B_s^0 \to \mu^+\mu^-)_{\mathrm{SM}} \\ &= (3.56 \pm 0.18) \times 10^{-9}.\end{aligned} \tag{3.14}$$

The effective branching fraction given by Equation 3.14 can be compared with experimental results. A deviation from this value can indicate the presence of NP effects. Both the theoretical branching fraction (Equation 3.10) and $\mathscr{A}_{\Delta\Gamma}^{\mu\mu}$ are experimental observables, and are sensitive to NP effects. The former requires large statistics with flavor-tagging information to measure the time-dependent decay rate of the B_s^0 meson. The untagged B_s^0, $\bar{B}_s^0$ events can be used to determine $\mathscr{A}_{\Delta\Gamma}^{\mu\mu}$. The branching fraction has also to be further corrected for radiative $B_s^0 \to \mu^+\mu^-$ transitions, where a photon may be emitted directly or via bremsstrahlung. However, the radiative correction factor is not influenced by NP [35].

The latest $\mathscr{B}(B_s^0 \to \mu^+\mu^-)$ results from various experiments are discussed in Chapter 6.

Chapter 4

Analysis Strategy for the $B_s^0 \to \mu^+\mu^-$ Measurement

The $B_s^0 \to \mu^+\mu^-$ decay is a rare decay with a very low expected branching fraction, $O(10^{-9})$ (Chapter 3). It is a demanding task to collect sufficient statistics to perform this measurement. Besides the low decay rate there are other experimental challenges:

- Similar event topologies of other decay modes, e.g., $B_{s,d}^0 \to h^+ h'^-$ (where h is a hadron), contribute to an irreducible background.

- The $B_s^0 \to \mu^+\mu^-$ invariant mass spectrum is expected to have a $B_d^0 \to \mu^+\mu^-$ component, because in the SM their predicted branching fractions[1] differ only by one order of magnitude [35]:

$$\mathscr{B}(B_d^0 \to \mu^+\mu^-)_{\text{SM}} = \frac{1}{30.35}\mathscr{B}(B_s^0 \to \mu^+\mu^-)_{\text{SM}}. \qquad (4.1)$$

  The mass difference between B_s^0 and B_d^0 mesons, $m_{B_s^0} - m_{B_d^0} = 87.35 \pm 0.23\,\text{MeV}$ [29], is not large enough compared to the di-muon mass resolution of the ATLAS detector. The expected mass resolution near the B_s^0 mass is $60\,\text{MeV}$ in the barrel region, and $110\,\text{MeV}$ in the endcap [1]. This makes it difficult to resolve the two decays.

- Due to low statistics, the event selection optimization is driven by Monte Carlo (MC) simulations. The residual data-MC discrepancies have to be accounted for as systematic uncertainties.

Therefore it is a challenging task to measure the $B_s^0 \to \mu^+\mu^-$ branching fraction. This rare decay has been a topic of research at several particle collider experiments at Tevatron and LHC.

The total number of $b\bar{b}$ pairs, N_b, produced in particle collisions is proportional to the integrated luminosity, $\mathscr{L}$:

$$N_b = \sigma_b \mathscr{L}, \qquad (4.2)$$

where σ_b is the $b\bar{b}$ production cross-section. Figure 4.1 shows the predicted cross-sections of the various physics processes as a function of the center-of-mass energy of the proton beams. Hadron colliders like the LHC have a large b-quark production cross-section of $\sim 30\,\text{mb}$. This

---

[1]For the production ratio see Section 4.1.

is an advantage. The experiments at the LHC are able to reach the required sensitivity to not only search for the $B_s^0 \to \mu^+\mu^-$ decay, but also to perform a measurement. The first evidence of the decay was recently observed by LHCb [44,45] (Chapter 6).
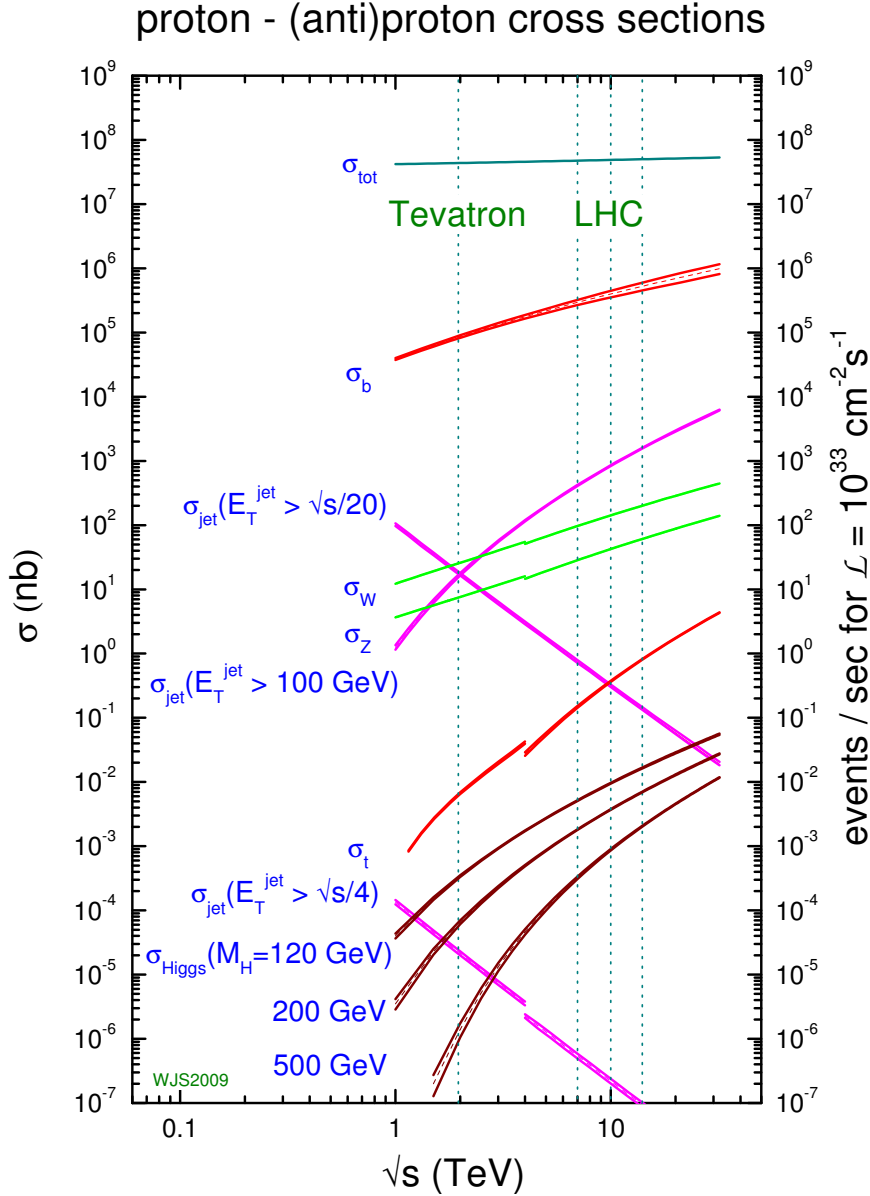

Figure 4.1: Predicted production cross-sections of the various physics processes (curved lines) as a function of the proton-proton center-of-mass energy [46].

On the downside, hadron colliders have a large QCD background originating from lighter quarks. Thanks to the sophisticated trigger system and the muon identification algorithms (Section 5.2), ATLAS is able to select interesting event candidates. The di-muon topological trigger efficiency is about 33% (4 GeV threshold) [9].

This chapter focuses on the ATLAS $B_s^0 \to \mu^+\mu^-$ analysis. The decay $B^\pm \to J/\psi K^\pm$ plays an important role in the analysis. It serves as the reference channel relative to which the $B_s^0 \to \mu^+\mu^-$ branching fraction is measured. The next sections present a discussion on the

computation of the experimental branching fraction and its measurement strategy. An overview of the $B_s^0 \to \mu^+\mu^-$ analysis is presented in Section 4.2. The results are summarized in Chapter 6.

4.1   Computation of $B_s^0 \to \mu^+\mu^-$ Branching Fraction

The $B_s^0 \to \mu^+\mu^-$ branching fraction is the ratio of the number of observed $B_s^0 \to \mu^+\mu^-$ events, $N_{B_s^0}$, and the total number of B_s^0 mesons produced, $N_{B_s^0}^{\text{prod}}$:

$$\mathscr{B}(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s^0}}{N_{B_s^0}^{\text{prod}}}. \tag{4.3}$$

The later can be computed using Equation 4.2. In ATLAS, $\mathscr{L}$ is currently measured to a precision of $\pm 1.8\%$ [47]. However, to determine σ_b, absolute values of certain quantities, like di-muon trigger efficiency, kinematic acceptance, etc., need to be determined precisely. An alternative is to measure $\mathscr{B}(B_s^0 \to \mu^+\mu^-)$ relative to the branching fraction of a reference channel, which is known to a good precision from other measurements. This is the preferred strategy at the LHC [2, 45, 48]. The branching fraction is hence obtained from the observed number of B mesons in the two channels:

$$\frac{\mathscr{B}(B_s^0 \to \mu^+\mu^-)}{\mathscr{B}_{\text{ref}}} \propto \frac{N_{B_s^0}}{N_{\text{ref}}}. \tag{4.4}$$

The event candidates in the two channels must have the same selection criteria to keep systematic effects as low as possible. With this approach the measurement is independent of luminosity variations and the uncertainty of the $b\bar{b}$ cross-section measurement.

4.1.1   Reference Channel

The decay $B^\pm \to J/\psi K^\pm \to \mu^+\mu^- K^\pm$ serves as the reference channel. The B^+ (B^-) meson first decays into a J/ψ meson and a K^+ (K^-) meson. Then the J/ψ particle decays into two muons. A schematic diagram depicting the decay process is shown in Figure 4.2(a). Like the signal channel (Figure 4.2(b)) it has two muons in the final state.

The main advantage of choosing the $B^\pm \to J/\psi K^\pm \to \mu^+\mu^- K^\pm$ decay as the reference channel is its large branching fraction [29]:

$$\begin{aligned}
\mathscr{B}(B^\pm \to J/\psi K^\pm \to \mu^+\mu^- K^\pm) &= \mathscr{B}(B^\pm \to J/\psi K^\pm) \times \mathscr{B}(J/\psi \to \mu^+\mu^-) \\
&= (6.02 \pm 0.20) \times 10^{-5}.
\end{aligned} \tag{4.5}$$

Therefore a large statistics data sample is available for analysis. This allows for a precise determination of the $B^\pm$ yield, $N_{B^\pm}$. The measurement of the reference channel yield is the main topic of this thesis. The details are presented in the next chapter.
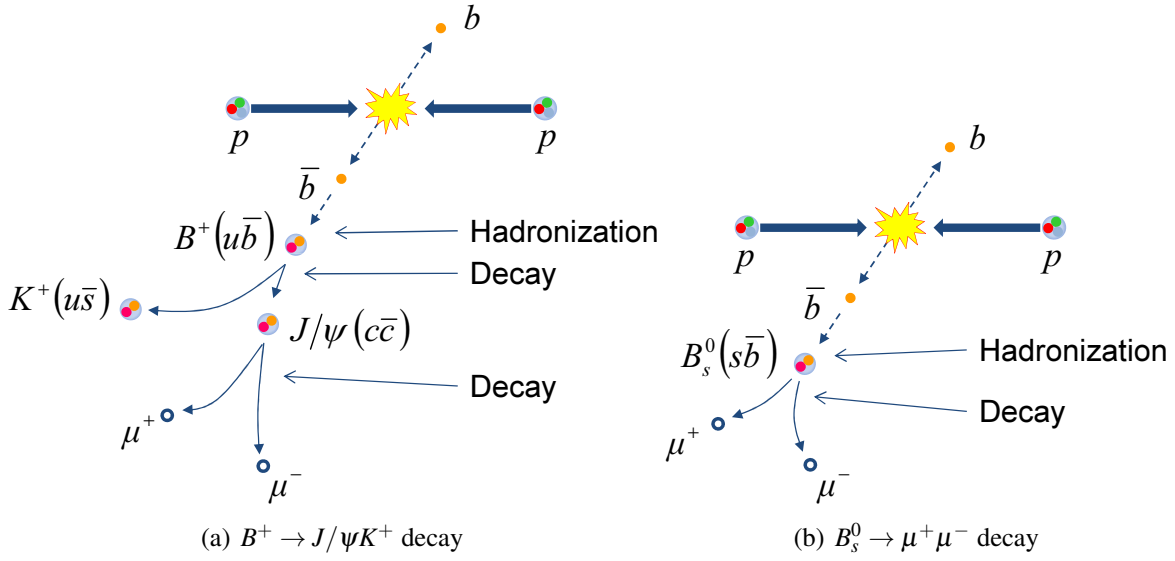
(a) $B^+ \to J/\psi K^+$ decay   (b) $B_s^0 \to \mu^+\mu^-$ decay

Figure 4.2: Schematic diagrams showing the decay processes of the signal and the reference channel.

4.1.2 Baseline Formula

The B_s^0 and B^+ mesons differ in their quark content (Figure 4.2). After a b-quark is produced it goes through a process called hadronization, where it combines with one or more quarks to form a composite particle. The b-quark hadronizes with the u- and s-quarks with different probabilities, f_u and f_s[2]. The branching fraction formula in Equation 4.4 has to be corrected for the different hadronization probabilities.

Due to the rareness of the $B_s^0 \to \mu^+\mu^-$ decay it is very important to suppress the backgrounds. The signal channel is dominated by a combinatorial background from $b\bar{b} \to \mu\mu X$ events, where X can be any other particle. This type of background results from random combinations of muons coming from two different semi-leptonic b decays. It is a continuous background, and it spans the whole $B_s^0 \to \mu^+\mu^-$ invariant mass spectrum. There is also a small but irreducible $B_{s,d}^0 \to h^+ h'^-$ background due to hadrons that are misidentified as muons. This is a peaking background. It is estimated from MC studies. In the analysis of the signal channel several selection cuts are applied in order to enhance the signal-to-background ratio. They mainly help to suppress the combinatorial background component. The efficiency of selection (ε) is taken into account in the branching fraction calculation. The trigger and reconstruction efficiencies are also included in ε. The selection cuts are discussed in detail in Section 5.3. Since the sub-detectors do not cover the total solid angle around the IP, the event counts in the two channels need to be corrected for the reduced acceptance of the detector (A). The ratio of acceptances and efficiencies, $R_{A\varepsilon} = \frac{A_{B^\pm}\varepsilon_{B^\pm}}{A_{B_s^0}\varepsilon_{B_s^0}}$, is estimated using signal and reference channel MC samples. This factor corrects the event yields in the branching fraction calculation.

The $B_s^0 \to \mu^+\mu^-$ branching fraction is computed using the following baseline formula:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s^0}}{N_{B^\pm}} \times \mathcal{B}(B^\pm \to J/\psi K^\pm \to \mu^+\mu^- K^\pm) \times R_{A\varepsilon} \times \frac{f_u}{f_s}, \qquad (4.6)$$

---

[2]Also called fragmentation fractions

where $N_{B_s^0}$ and $N_{B^\pm}$ are the number of observed signal and reference channel events, respectively. The two channels have similar final state particles (di-muon). Thus, the change in $A\varepsilon$ is highly correlated between the two channels [2]. If similar selection cuts are applied in the two channels, the systematic uncertainties due the selection should partially cancel out in the numerator and the denominator of Equation 4.6. The residual data-MC discrepancies are taken as a systematic uncertainty on $R_{A\varepsilon}$. Finally, there are the uncertainties due to the third kaon track, which need to be included separately (Table 6.1). The ratio of the hadronization probabilities, $f_s/f_u = 0.256 \pm 0.020$ [49] (assuming $f_u = f_d$), is taken from other measurements. The factor f_s/f_u is the one of the main factors that contributes to the uncertainty (8%) in the final result.

It is convenient to express Equation 4.6 in terms of the Single Event Sensitivity (SES):

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = N_{B_s^0} \times \text{SES}. \tag{4.7}$$

For a single observed signal event ($N_{B_s^0} = 1$), the branching fraction is equal to the SES. It is the inverse of an effective efficiency, and is a useful measure of the performance of the analysis prior to the analysis of the actual $B_s^0 \to \mu^+\mu^-$ data.

4.2   Analysis Flow

The $B_s^0 \to \mu^+\mu^-$ analysis is performed in three main steps. The first step involves determining the optimal selection cuts to achieve the best signal-to-background ratio. In the second step the optimized cuts are used to determine the expected background in the signal channel, the $B^\pm$ yield in the reference channel and $R_{A\varepsilon}$. In the last step the number of events in the signal region are counted, and are compared with the estimated continuum background. A flowchart of the $B_s^0 \to \mu^+\mu^-$ analysis is shown in Figure 4.3.

The ATLAS trigger system selects interesting events, that qualify a topological di-muon trigger condition, during the online data taking process. If both muons have $p_\text{T} > 4\,\text{GeV}$, the event is retained.

The analyses involving the measurement of small signals are sensitive to selection cuts. A prior knowledge of the effects of the cuts on the signal sample can bias the cut optimization. Hence, the common strategy is to 'blind' the signal region until the selection cuts are finalized. The background events from the sidebands can still provide an estimate of the background in the signal region. This is performed by interpolation. Thus, in the analysis, the invariant mass range $m_{\mu^+\mu^-} \in [5066, 5666]\text{MeV}$ is hidden until the final step.

The events in the sidebands ($m_{\mu^+\mu^-} \in [4766, 5066]\text{MeV}$, $m_{\mu^+\mu^-} \in [5666, 5966]\text{MeV}$) are split into two sets. To avoid statistical biases, one set is used for cut optimization (odd numbered events in data), and the other is used to estimate the background in the signal region (even numbered events in data).

The event selection cuts are optimized using a multivariate technique. The method uses Boosted Decision Trees (BDT) [50, 51] with 13 input variables to compute an event classifier, Q. The BDT selection is described in Section 5.3.

The signal search window $\Delta m_{\mu^+\mu^-}$, in the $B_s^0 \to \mu^+\mu^-$ invariant mass spectrum, is optimized together with the BDT classifier cut Q_cut. The method determines the optimal $\Delta m_{\mu^+\mu^-}$
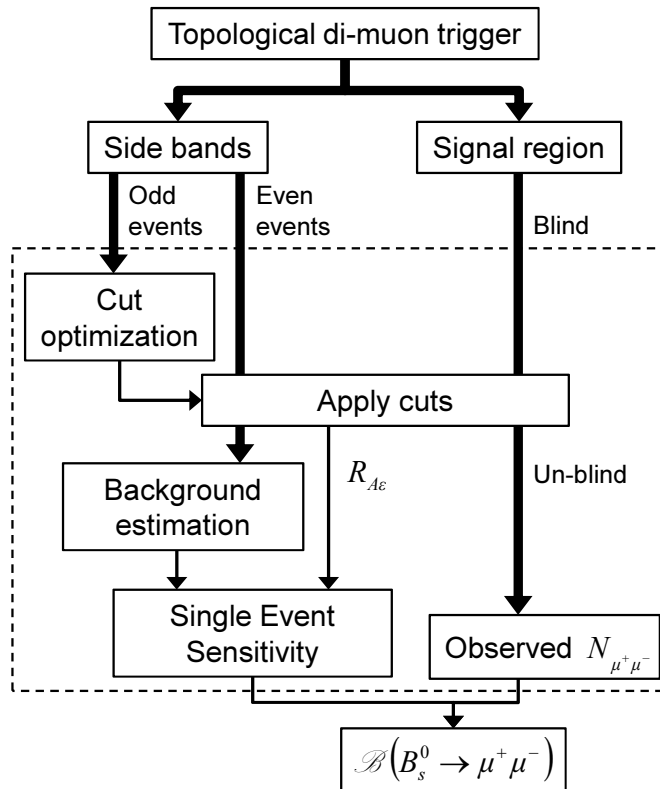
Figure 4.3: A flowchart showing the $B_s^0 \to \mu^+\mu^-$ analysis flow. The cut optimization process and the computation of $R_{A\varepsilon}$ rely on the signal and reference channel MC samples. The rest of the analysis uses pp collision data.

and Q_{cut} to obtain the maximum value for the estimator [52]:

$$\mathscr{P}(Q_{\text{cut}}, \Delta m_{\mu^+\mu^-}) = \frac{\varepsilon_{\text{sig}}}{1 + \sqrt{N_{\text{bkg}}}}, \tag{4.8}$$

where ε_{sig} is the signal selection efficiency and N_{bkg} is the continuum background interpolated from the sidebands (two times the number of odd events). The same classifier cut is then used to compute the acceptance and efficiency ratio ($R_{A\varepsilon}$) from the MC, and the reference channel yield using data.

The SES is computed assuming a background-only hypothesis in the signal region. The lower the SES, the easier it is to observe a rare decay. After un-blinding the dataset, the event count in the search window, $N^{\text{obs}}_{\mu^+\mu^-}$, is used to establish an upper limit on the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction (Chapter 6).

The data can be split into certain mass resolution categories distinguished by the maximum pseudorapidity ($|\eta^\mu|_{\text{max}}$) of the muon tracks, and the analysis can be performed separately for each case. The part of the analysis indicated by the dashed lines in Figure 4.3 can be repeated for each category. Finally, the results from all the categories can be combined to set an upper limit. This allows to take advantage of the better invariant mass resolution, e.g., in the barrel region. However, the trade off is reduced event statistics per category. The analysis is performed using both strategies: a single bin for all data, and multiple bins of $|\eta^\mu|_{\text{max}}$. Although Reference [2] reports the single bin study, parts of this thesis, e.g., the $B^\pm$ yield measurement described in the next chapter, report results from both strategies.

Chapter 5

Measurement of $B^{\pm}$ Yield

The $B^{\pm} \to J/\psi K^{\pm}$ decay is an important channel for several measurements performed with the ATLAS detector [1, 53]. References [54, 55] describe dedicated analyses of this channel. In the $B_s^0 \to \mu^+ \mu^-$ analysis it is used as a reference channel. The selection criteria imposed here are not optimal for the measurement of the $B^{\pm}$ yield, but they are dictated by what is applied in the main $B_s^0 \to \mu^+ \mu^-$ analysis. The $B^{\pm}$ yield is an important ingredient for the $\mathscr{B}(B_s^0 \to \mu^+ \mu^-)$ calculation (Chapter 4). The extraction of the reference channel yield from pp collision data is described here.

This chapter begins with a description of the various datasets used in the reference channel analysis (Section 5.1). First the events in the datasets are reconstructed (Section 5.2), then the selection cuts are applied. A note on the cut optimization in the $B_s^0 \to \mu^+ \mu^-$ analysis is presented in Section 5.3. The reference channel invariant mass spectrum and its various background components are explained in Section 5.4. Section 5.5 describes the details of the $B^{\pm}$ yield measurement. A discussion of the results is presented in Section 5.6.

5.1 Data Samples

The analysis presented in this thesis takes its data from two sources: from pp collisions and from MC simulations. The data samples are described below.

5.1.1 Collision Data

This analysis uses data from pp collisions at $\sqrt{s} = 7\,\text{TeV}$. The data was recorded by the ATLAS detector between March 22 – October 30, 2011. It corresponds to a total integrated luminosity of $4.9\,\text{fb}^{-1}$ with an uncertainty of 1.8% [47].

The ATLAS data acquisition and trigger systems are described in Chapter 2. During the data acquisition process the various detector systems are closely monitored, and their status information is recorded. This is used to determine the quality of the data. The luminosity blocks (LB), that do not qualify for physics analyses, are rejected, e.g., when the solenoid and toroid magnetic fields were not available. A LB is also rejected if the proton beams were not stable during data acquisition.

The collision data is also used to check the performance of the sub-detectors by studying well-known physics channels and good quality ID tracks. It allows to perform better alignment and calibration of the sub-detectors. Recalculated alignment and calibration constants are then applied to the raw ATLAS data to improve the data quality for physics analyses.

5.1.2 Monte Carlo

Chapter 4 introduced the use of MC simulations in the $B_s^0 \to \mu^+\mu^-$ analysis. The MC events are generated using algorithms that simulate the relevant physics phenomena. The algorithms encode the current understanding of the underlying physics processes: pp scattering, flavor production, hadronization, etc.. These algorithms have several parameters [56]. They are tuned to the latest SM parameters and experimental constraints, as of year 2011 [57, 58]. One of the main simulation parameters is the energy of the colliding beams. The samples used in this analysis were generated for protons colliding at 7 TeV center-of-mass energy. The generated events are retained, if they have at least one decay of interest (a pre-specified decay channel) per event. For example, an event of the $b\bar{b} \to \mu\mu X$ sample will contain at least two muons.

After the generation the events are simulated by a full detector simulation software. The software contains detailed information on the geometry and the material composition of the detector. It simulates the passage of the particles through the detector material in conditions similar to the actual data-taking, e.g., electromagnetic fields and their distortions, shifts in alignment and calibration, known defects in the detector, etc.. The software calculates the energy loss per unit length (dE/dx) as the particle track traverses through the various materials. It simulates the response of the sub-detectors, hits in the pixel and tracking detectors, energy absorbed in the calorimeters, etc.. The behavior of the front-end electronics is simulated in a process called digitization, and the complete event data structure is built. The digitization overlays hits from several events from multiple pp collisions to simulate event pile-up. The event data is stored in a format similar to the one used by the ATLAS data acquisition system, so that the events can be reconstructed using a common event reconstruction algorithm for both the data and the MC samples.

The analysis relies on the MC for several needs. The reference channel analysis presented in this chapter uses the following MC samples: $B^+ \to J/\psi K^+$, $B^+ \to J/\psi \pi^+$ and inclusive $b\bar{b} \to \mu\mu X$. Some of their properties are listed in Table 5.1. The $B^+ \to J/\psi \pi^+$ MC is used for the estimation of the mis-reconstructed $B^{\pm} \to J/\psi \pi^{\pm}$ decays that enter the $B^{\pm}$ invariant mass spectrum. The $b\bar{b} \to \mu\mu X$ MC is used for the estimation of the combinatorial and partially-reconstructed-decays background. These are discussed further in Section 5.4.

The $B^+ \to J/\psi K^+$ and $B^+ \to J/\psi \pi^+$ MC samples are generated using PYTHIA 6.4 [56], and the ATLAS detector is simulated in the GEANT4 framework [59]. A large sample is required for the inclusive $b\bar{b} \to \mu\mu X$ decays. It is generated using PYTHIA, and then simulated in ATLFAST-II [60], where the most important detector systems are simulated (e.g., ID and MS) while the response of the others is parametrized for fast simulation.

The MC generators are configured in order to improve the computational performance. A cut is applied on the perpendicular component of the momentum, $\hat{p}_{\perp}$[1], of all the partons produced in the hard scattering process (Figure 5.1). The hard scattered partons with $\hat{p}_{\perp} > 5$ GeV

---

[1] In literature the variable is also referred to as the `CKIN(3)` variable [56].

Channel	Generated events	Simulation tool	Description
$B^+ \to J/\psi K^+$	5×10^6	GEANT4	reference channel
$B^+ \to J/\psi \pi^+$	1×10^6	GEANT4	mis-reconstructed background
$b\bar{b} \to \mu\mu X$	2×10^8	ATLFAST-II	inclusive $b\bar{b} \to \mu\mu X$ decays

Table 5.1: Monte Carlo datasets used in the analysis. The events are generated using PYTHIA 6.4. The second column summarizes the available event statistics after the generator level selection cuts.

are retained. In the $b\bar{b} \to \mu\mu X$ channel this cut is $\hat{p}_\perp > 7\,\text{GeV}$. The other generator level selections include:

- Selection of decay products: two final state muons[2], and a third particle, which is a kaon or a pion depending on the channel. In the $b\bar{b} \to \mu\mu X$ channel only the muons are required.

- Cuts on the transverse momenta, p_T, of the decay products: muons in the $B^+ \to J/\psi K^+$ and $B^+ \to J/\psi \pi^+$ samples are required to have $p_T > 2.5\,\text{GeV}$, and in the $b\bar{b} \to \mu\mu X$ sample $p_T > 3.5\,\text{GeV}$. The kaon and muon particles are required to have $p_T > 0.5\,\text{GeV}$.

- Fiducial cuts on the pseudorapidity of the decay products: particles within $|\eta| < 2.5$ are retained.

Note that the MC samples are produced for only one charge conjugate.
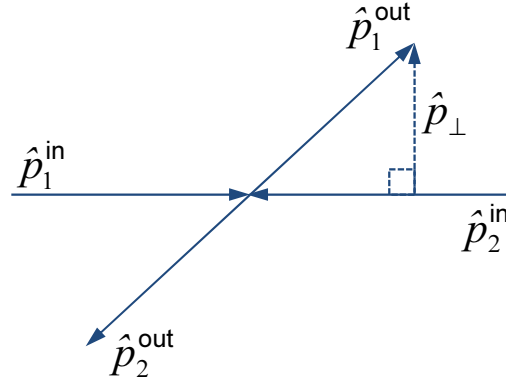
Figure 5.1: A 2-particle $\to$ 2-particle hard scattering process in the center-of-mass frame of the incoming particles. $\hat{p}_1^{\text{in}}$, $\hat{p}_2^{\text{in}}$ are the momenta of the incoming particles, and $\hat{p}_1^{\text{out}}$, $\hat{p}_2^{\text{out}}$ are momenta of the outgoing ones. $\hat{p}_\perp$ is the perpendicular component of the momentum of the outgoing particle w.r.t. the direction of the incoming particles.

The reference channel analysis also indirectly depends on the following MC datasets used in the main $B_s^0 \to \mu^+\mu^-$ analysis: $B_s^0 \to \mu^+\mu^-$, $B_s^0 \to J/\psi\phi$ (where the J/ψ further decays

---

[2]The selection is performed using the particle identification information specified in the 'Monte Carlo particle numbering scheme' [29].

into $\mu^+\mu^-$ and the ϕ decays into K^+K^-) and inclusive $B^0_{s,d} \to h^+ h'^-$ decays (where 4 decay modes are simulated, $B^0_{s,d} \to K^+K^-$, $B^0_{s,d} \to \pi^+\pi^-$, $B^0_{s,d} \to K^+\pi^-$ and $B^0_{s,d} \to \pi^+K^-$). The $B^0_s \to J/\psi\phi$ decay is used as a control channel. The $B^0_{s,d} \to h^+ h'^-$ MC is used to estimate the size of the resonant background in the main channel. More information on these MC channels can be found in References [1, 2].

5.2   Event Reconstruction

An event has to be reconstructed before any selection can be applied. This involves the reconstruction of particle tracks and decay vertices. It starts with the decay products and builds up all objects up to the primary vertex. Figure 5.2 shows a flowchart of the various steps involved in the reconstruction of an event.

The $B^{\pm} \to J/\psi K^{\pm}$ analysis (as also the signal channel analysis) is based on muons tracks. The muons lose little energy in the ID and calorimeter layers. Those with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ are able to reach the MS, producing hits in both ID and MS. In contrast, the kaons either decay, or are captured in the material of the inner layers. They are stopped by the calorimeters. They produce tracks only in the ID. The ID is, however, unable to distinguish kaons from other hadrons or charged leptons.

The reconstruction of the muon and the ID tracks is described in the next sections. Once the tracks are reconstructed, the reconstruction of the primary and secondary vertices is carried out (Section 5.2.2). The $B^0_s \to \mu^+\mu^-$ event reconstruction is performed in a similar manner [2].

5.2.1   Track Reconstruction

The reconstruction of the muon tracks is first performed at the sub-detector level (ID and MS), then other detector information is included to reconstruct the full track. The muon tracks reconstructed in the MS are called stand-alone muons. ATLAS uses special track extrapolation software [61] to correctly account for the effects of the various detector materials traversed by the particles. The combined muon tracks are reconstructed using the ID and MS track segments.

5.2.1.1   Stand-alone Muons

ATLAS uses a stand-alone reconstruction algorithm [62] to identify track segments in the MS. The algorithm uses RPC/TGC information to find hits in the muon stations (MDTs/CSCs) to first build local track segments. Then it builds up muon track candidates from combinations of the local track segments, taking the magnetic field effects into account. It performs a global fit for each track candidate using individual hit space-points. Finally, the track candidates are extrapolated to the IP taking into account the energy losses due to the material crossed by the particle in the calorimeters. In the last step the covariance matrices of the fits are recomputed.

The stand-alone reconstruction has an advantage of a large coverage, up to $|\eta| < 2.7$. However, it is difficult to reconstruct low momentum muons ($p_{\mathrm{T}} \sim 1 - 3\,\mathrm{GeV}$), that cannot reach the outer stations of the MS.
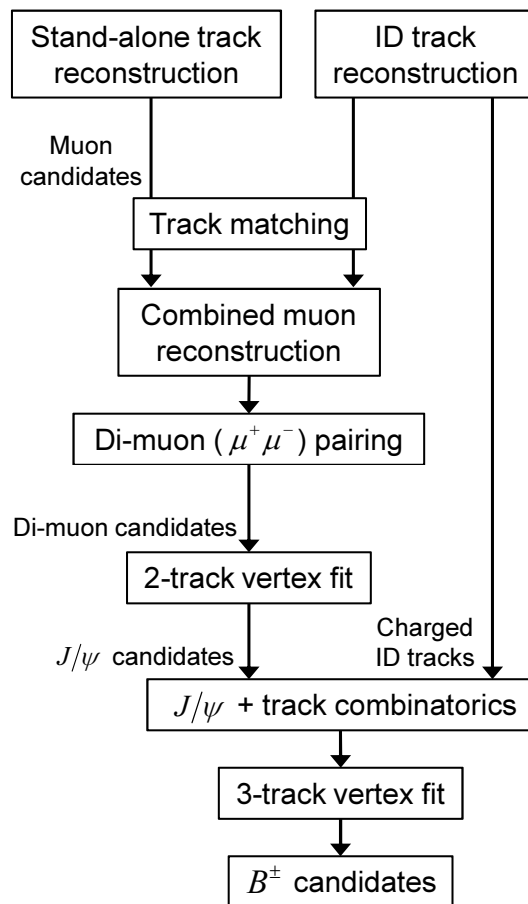
Figure 5.2: An outline of the $B^{\pm} \to J/\psi K^{\pm}$ event reconstruction.

5.2.1.2 ID Track Reconstruction

The reconstruction of tracks in the ID [63] is performed using information from all three ID sub-detectors. The raw data from the pixel detector and the SCT are converted into clusters, which define the hit space-points. These space-points are used to form track candidates. A fit is performed on each track candidate, and any outliers are discarded. Then certain track quality cuts are applied. These include cuts on the pixel and SCT hits, and the number of holes per track. A hole is a live silicon sensor crossed by a track without generating any clusters. The physics analysis also applies further track quality cuts. These cuts are listed in Section 5.3.

In case of the TRT, the arrival time information of the ionization is converted into so called 'drift circles'. The track candidates selected in the previous step are extended into the TRT region, and they are associated with the TRT drift circles. A final fit is performed using the full information from the sub-detectors.

The reconstruction of ID tracks is limited to $|\eta| < 2.5$. However, high p_T muons can be reconstructed with an efficiency $> 99.5\%$ in the whole pseudorapidity range. In the barrel region, the efficiency falls to about 98% for low p_T muon ($p_\text{T} \geq 1\,\text{GeV}$) tracks [9].

5.2.1.3 Combined Muons

The combined reconstruction algorithm (called STACO) [62] matches the ID track candidates with those reconstructed in the MS using a χ^2 selection criterion. Then it statistically combines the two independent measurements of the track parameters. The covariance matrix of the combined track is computed from the two sets of track parameters and their covariance matrices. Due to the limited ID coverage the track combinations can only be performed for tracks with $|\eta| < 2.5$.

The combined tracks have an improved momentum resolution in a wide momentum range, $6 < p_\text{T} < 100\,\text{GeV}$. The reconstruction method also provides good rejection against muons originating from secondary interactions and pion/kaon decays.

5.2.2 Vertex Reconstruction

The muon track candidates are reconstructed as described in the previous section. The analysis uses only combined muon candidates. An algorithm is used to reconstruct J/ψ candidates. It builds a set of di-muon candidates from the combined muon track collection by performing all possible track combinations. A di-muon candidate is selected if the two muons are oppositely charged. Each di-muon candidate is fitted to a 2-track vertex using a Kalman filter based fitter [64]. The two tracks are assigned the muon mass, $m_{\mu^\pm} = 105.658\,\text{MeV}$ [29], and the vertex invariant mass is computed using the track parameters. The J/ψ candidate is retained if it qualifies the selection cuts listed in Section 5.3.

The $B^\pm \to J/\psi K^\pm$ decay is effectively a three prong decay. The J/ψ particle has a very short lifetime, $O(10^{-21}\,\text{s})$, and in the reconstruction its decay vertex is almost inseparable from that of the $B^\pm$ decay. The three tracks appear to originate from a common vertex. A reconstructed J/ψ candidate is combined with a charged ID track. This object is referred to as a $B^\pm$ candidate. The third track is not one of the two ID tracks associated with the J/ψ candidate. A cascade vertex fit is performed on the three tracks using the full information of the tracks. In this

step, the individual tracks are refitted to obtain a better vertex fit. The fit has a low value of χ^2 if the three tracks originate from two cascade decay vertices. The third track is assigned the kaon mass, $m_{K^\pm} = 493.677\,\text{MeV}$, and the di-muon is assigned the J/ψ mass, $m_{J/\psi} = 3.097\,\text{GeV}$. Using the refitted track parameters and their covariance matrices the $B^\pm$ invariant mass and its uncertainty are calculated. Similarly, the J/ψ candidate is combined with other ID track candidates, that are not identified as muons, each time a vertex fits is performed. The $B^\pm$ candidate with the best vertex fit χ^2/NDF is selected. The whole procedure is repeated for the other J/ψ candidates.

The primary vertex (PV) is reconstructed using an iterative algorithm [65]. A first pass PV reconstruction is performed independent of the $B^\pm$ event reconstruction. First, certain reconstructed tracks are chosen, based on some selection criteria [65], that are compatible with originating from the interaction region. Then vertex seeds are found based on the maximum of the z-distribution of the reconstructed tracks. An iterative χ^2 fit is performed using the seeds and the nearby tracks. The outlying tracks are down-weighted in the fit χ^2 computation. Tracks that are more than 7σ away from the vertex are taken as new seeds to find other vertices, and the procedure is repeated. The reconstructed tracks are associated with the PV candidates. Once the $B^\pm$ candidates are reconstructed, the three tracks of the $B^\pm$ candidates are removed from the collection used in the PV reconstruction, and the PV is refitted.

Choosing the correct PV candidate is important for the calculation of some of the analysis variables, e.g., the track impact parameter d_0. The analysis uses primary or pile-up vertices that are closest in the z-direction to the secondary vertex (SV) of the $B^\pm$ candidate.

5.3 Event Selection

This section summarizes all selection cuts used in the reference channel analysis. These cuts are applied to the pp collision data as well as the MC datasets after[3] the event reconstruction.

As explained in Section 4.2, the cuts are optimized to achieve the best signal to background separation in the signal channel. The same set of baseline cuts are applied in both the channels, except the cuts on the kaon track. The next section describes these cuts in detail. Then a multivariate (MVA) selection is applied. It is highly efficient in rejecting combinatorial background events. Section 5.3.2 describes how the MVA selection is carried out.

5.3.1 Baseline Cuts

The analysis uses events that are selected by the topological di-muon trigger (Section 2.2). The p_T threshold for the two muons in the HLT is $4\,\text{GeV}$. The events with a di-muon invariant mass that is not compatible with the J/ψ mass, $m_{\mu^+\mu^-} \notin [2.5, 4.3]\,\text{GeV}$, are rejected.

To select good quality tracks for the reconstruction of the J/ψ candidates, a set of preselection cuts [66] are applied before the baseline selection cuts. These cuts are applied (sequentially) to both the muon tracks, and can be described as follows:

- A B-layer sensor can be masked during the data acquisition if it has low efficiency, or if it is dead. If B-layer hits were expected for a track, the track is required to have at least 1

---
[3]Some cuts are applied during the event reconstruction for better computational efficiency.

B-layer hit. However, the track is selected in case no B-layer hit was expected.

- If the number of pixel hits plus the number of crossed dead pixel sensors is > 1, the track is selected.

- If the number of SCT hits plus the number of crossed dead SCT sensors is ≥ 6, the track is selected.

- If the number of pixel holes plus the number of SCT holes is < 2, the track is selected.

- For $N = N_{\text{hits}}^{\text{TRT}} + N_{\text{outliers}}^{\text{TRT}}$, where $N_{\text{hits}}^{\text{TRT}}$ and $N_{\text{outliers}}^{\text{TRT}}$ are the number of track TRT hits and TRT outliers, respectively,

  - If the track has $|\eta| < 1.9$, then the track is selected only if $N > 5$ and $N_{\text{outliers}}^{\text{TRT}} < 0.9 N$.
  - If the track has $|\eta| \geq 1.9$ and $N > 5$, then the track is selected only if $N_{\text{outliers}}^{\text{TRT}} < 0.9 N$ (note in this case tracks with $N \leq 5$ are selected anyway).

In addition, both tracks are required to have, separately, at least 1 pixel hit, 6 SCT hits, and 1 TRT hit.

The di-muon track combinatorics in the J/ψ reconstruction is performed on muons tracks compatible with the triggered muons. Therefore, the two muons tracks are required to have $p_{\text{T}} > 4 \, \text{GeV}$. There is an implicit cut on the pseudorapidity of the two muon tracks, since the combined muon reconstruction is performed for tracks with $|\eta| < 2.5$. A cut on the vertex fit quality is applied to reject combinatorics of fake track candidates. Vertex fits with $\chi^2/\text{NDF} > 10$ are rejected. A tighter invariant mass cut is applied on the reconstructed J/ψ candidates. Only candidates in the invariant mass range of $\pm 3\sigma$ around the mean J/ψ mass, $m_{\mu^+\mu^-} \in [2.915, 3.275] \, \text{GeV}$, are accepted.

The third track in the $B^\pm$ vertex fit is required to have $p_{\text{T}} > 2.5 \, \text{GeV}$ and $|\eta| < 2.5$. Kaon tracking efficiency estimation (Chapter 6) requires well reconstructed kaon tracks. This imposes stringent requirements on the transverse (d_0) and longitudinal (z_0) impact parameters of the kaon tracks. Kaon track candidates with $|d_0| < 1.5 \, \text{mm}$ and $|z_0 \sin \theta| < 1.5 \, \text{mm}$ are selected, where θ is the polar angle of the track measured w.r.t. the beam axis. Exceptionally, here the d_0 of the kaon track is measured w.r.t. the beam spot and not the PV, and the z_0 is corrected using the refitted PV position vector. The kaon candidates are also required to have at least 1 B-layer hit, if B-layer hits were expected for that track. The cuts on pixel, SCT and TRT hits are 1, 6 and 1, respectively, as in the case of the muon tracks.

The $B^\pm$ candidates are required to have a good vertex fit. Candidates with vertex fit χ^2/NDF above 6 are rejected. A fiducial cut is applied on the pseudorapidity of the reconstructed $B^\pm$ candidates, $|\eta^B|_{\text{max}} < 2.5$. Finally, candidates with $p_{\text{T}} > 8 \, \text{GeV}$ and the invariant mass $m_{J/\psi K^\pm} \in [4.930, 5.630] \, \text{GeV}$ are retained for the analysis. The number of $B^\pm$ candidates in each data sample, that survive the baseline selection cuts, are listed in Table 5.2. There are on an average 3.3 $B^\pm$ candidates per event in the data before event selection, which reduces to approximately 1 candidate per event after the baseline selection cuts are applied.

The barrel region of the detector provides a better invariant mass resolution than the endcaps. The resolution influences the measurement uncertainties. Therefore, it is useful to split the analysis in bins of pseudorapidity. Following Equation 4.6 the splitting applies to both the

Dataset	Selected candidates
Data	641 356
$B^+ \to J/\psi K^+$ MC	175 979
$B^+ \to J/\psi \pi^+$ MC	31 631
$b\bar{b} \to \mu\mu X$ MC	21 046

Table 5.2: $B^\pm$ candidates after the baseline selection cuts. All datasets are processed using the $B^\pm \to J/\psi K^\pm$ reconstruction algorithm. In the $b\bar{b} \to \mu\mu X$ MC sample the $B^\pm \to J/\psi K^\pm$ and $B^\pm \to J/\psi \pi^\pm$ decays were vetoed on MC truth level.

signal and the reference channel. The datasets are split into three mass resolution categories: barrel-barrel (BB), barrel-transition (BT), and anywhere-endcap (AE). The splitting is based on the pseudorapidity of the two muons tracks $|\eta^{\mu_1}|$ and $|\eta^{\mu_2}|$. If $|\eta^\mu|_{max}$ is the larger of the two muon momenta, the three categories are defined as follows:

- Barrel-barrel: $B^\pm$ candidates with both muon tracks passing through the barrel region, i.e $|\eta^\mu|_{max} < 1.0$.

- Barrel-transition: $B^\pm$ candidates with either or both muon tracks passing through the transition region between the barrel and the endcaps, i.e. $1.0 \leq |\eta^\mu|_{max} < 1.5$.

- Anywhere-endcap: $B^\pm$ candidates with at least one muon track passing through the endcap region, i.e $1.5 \leq |\eta^\mu|_{max} < 2.5$.

The analysis is also performed on the complete dataset, i.e. the merge of all the three categories. It is called the single-bin analysis. The number of $B^\pm$ candidates in each mass resolution category are listed in Table 5.3.

Dataset	Single-bin	BB	BT	AE
Data 2011	641 356	356 493	129 971	154 892
$B^+ \to J/\psi K^+$ MC	175 979	93 007	38 712	44 260
$B^+ \to J/\psi \pi^+$ MC	31 631	17 103	6 804	7 724
$b\bar{b} \to \mu\mu X$ MC	21 046	11 290	4 689	5 067

Table 5.3: $B^\pm$ candidates by mass resolution category. In the $b\bar{b} \to \mu\mu X$ MC sample the $B^\pm \to J/\psi K^\pm$ and $B^\pm \to J/\psi \pi^\pm$ decays were vetoed on MC truth level.

5.3.2 Multivariate Selection

An important goal of the analysis is to achieve the best signal-to-background ratio. One approach is to apply selection cuts on a set of analysis variables, one at a time, and determine the signal yield. Examples of such variables are shown in Figure 5.3. For each variable, an optimal cut is determined by studying the signal and background MC distributions, to obtain the best
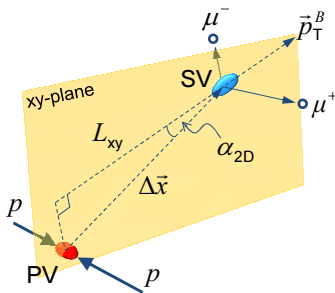
final selection efficiency and signal purity. This is also referred to as a cut-and-count analysis (e.g., the analysis presented in [67]). This approach can be very inefficient if the variables are correlated non-linearly.

The signal channel analysis uses a Boosted Decision Tree (BDT) algorithm[4] to achieve a better rejection of the background compared to the cut-and-count approach. A schematic diagram of a decision tree is shown in Figure 5.4. It uses successive decision nodes. A node categorizes an event as signal-like (S) or background-like (B) depending on the value of a discriminating variable, e.g., x_i. Each node is associated with a different variable. The terminal nodes (S/B) are called leaves. They are bins that store events after they have been classified during a training. This algorithm uses 6000 decision trees with a maximum node depth of 2 [68].

The selection criteria applied at each node, c_1, c_2, c_3, etc., are determined during the BDT training process. A leaf is marked as signal-leaf or background-leaf during the training. The decision trees are boosted by assigning weights to the events in the background-leaves, so the next tree is trained using the re-weighted events. Similarly, the third tree is trained, and so on. The training algorithm is described in References [50, 51].

The training needs a signal sample and a background sample. The two samples are derived from the $B_s^0 \to \mu^+\mu^-$ MC and the odd numbered sideband events (Section 4.2) in collision data. The samples are reconstructed in the $B_s^0 \to \mu^+\mu^-$ mode, and the baseline selection is applied before the training. Only half of the available odd numbered sidebands events in the data are used for the training, i.e. approximately $1/4^\text{th}$ of the available event statistics. Similarly, only half of the MC events are used for the training. The second halves of the two samples are used for testing the BDT performance and cut optimization.

The algorithm uses 13 analysis variables for event classification. The variables are listed in the Table 5.4 sorted by their discriminating power (rank).

Variable	Rank	Description
$L_\text{xy} = \frac{\Delta\vec{x}\cdot\vec{p}_\text{T}^B}{|\vec{p}_\text{T}^B|}$	1	Transverse decay length: it is defined as the projection of $\Delta\vec{x}$ on $\vec{p}_\text{T}^B$, where $\Delta\vec{x}$ is the difference between PV and SV position vectors, and $\vec{p}_\text{T}^B$ is the B transverse momentum vector.

(continued on next page)

[4]A part of the TMVA package [51].

Variable	Rank	Description
$I_{0.7} = \frac{p_T^B}{p_T^B + \Sigma p_T^i}$	2	Isolation of the B candidate: it is defined as the ratio of p_T^B to the sum of p_T^B and p_T of all tracks (excluding tracks from the B vertex), with $p_T > 0.5\,\text{GeV}$, within a cone $\Delta R < 0.7$. Here $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, where $\Delta\eta$ and $\Delta\phi$ are differences of the pseudorapidity and azimuthal angles of a track w.r.t. those of the B candidate. A signal-like B candidate is expected to be more isolated, i.e. to have larger $I_{0.7}$ (Figure 5.3(a)).
$\|\alpha_{2D}\|$	3	2D pointing angle: it is the angle between $\Delta\vec{x}$ and $\vec{p}_T^B$. The momentum vector of a signal-like B candidate is more aligned to $\Delta\vec{x}$ than a combinatorics background-like B candidate, hence the former has a smaller pointing angle (Figure 5.3(b)).
p_L^{min}	4	It is the minimum of the two muon momenta projected on the B transverse momentum vector, $\vec{p}_T^B$.
p_T^B	5	Magnitude of the B transverse momentum.
ct/σ	6	Proper decay length significance: is defined as the proper decay length, $ct = \frac{L_{xy} m_B}{p_T^B}$, divided by its uncertainty. m_B is the PDG mass [29] of the B meson. This variable is highly correlated with L_{xy}.
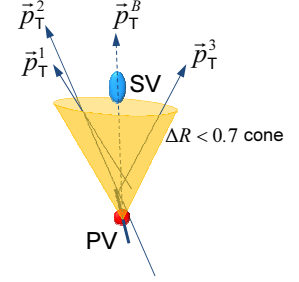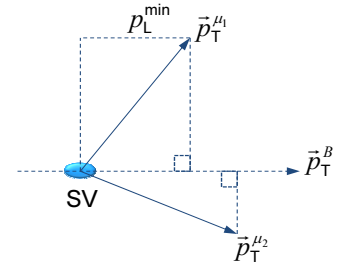
(*continued on next page*)
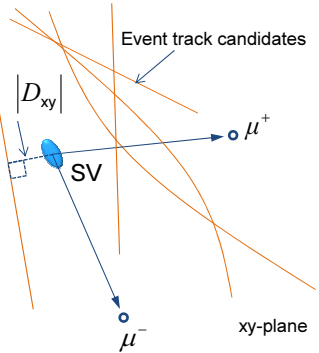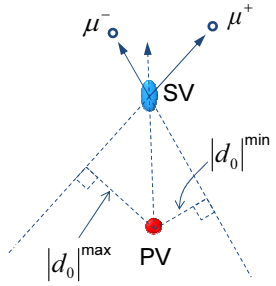
Variable	Rank	Description
$\ln \chi_z^2$, $\ln \chi_{xy}^2$	7, 13	Significance of the separation between PV and SV in z direction and in the transverse plane (xy), respectively. It is computed from the position of the two vertices, $\vec{x}_{PV}$, $\vec{x}_{SV}$ and the covariance matrices from their vertex fits, Σ_{PV}, Σ_{SV}. The χ^2 in z and xy is computed using the expression: $$\chi^2 = \Delta\vec{x}^T (\Sigma_{PV} + \Sigma_{SV})^{-1}\Delta\vec{x}, \qquad (5.1)$$ where $\Delta\vec{x} = \vec{x}_{PV} - \vec{x}_{SV}$. The variables help in distinguishing B candidates with clearly separated primary and secondary vertices (signal like) from those with overlapping vertices (background like).
$\|D_{xy}\|$, $\|D_z\|$	8, 11	Absolute values of the minimum distance of closest approach (DCA) of tracks in the xy-plane and along z, respectively. The variables are similar to the track impact parameter, but they are defined w.r.t. the SV. First the DCA is computed for each track candidate, then the minimum value is selected, i.e. corresponding to the track closest to the SV of the B candidate. This helps in rejecting B candidates with more number of tracks than the number that originate from a genuine B decay.
$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$	9	3D pointing angle between $\Delta\vec{x}$ and $\vec{p}^B$ computed from their pseudorapidities and azimuthal angles.
$\|d_0\|^{max}$, $\|d_0\|^{min}$	10, 12	The maximum and minimum values of the transverse impact parameter of the two muon tracks. It is the distance of closest approach of the tracks w.r.t. the PV.

Table 5.4: BDT variables ordered by their discriminating power (rank).
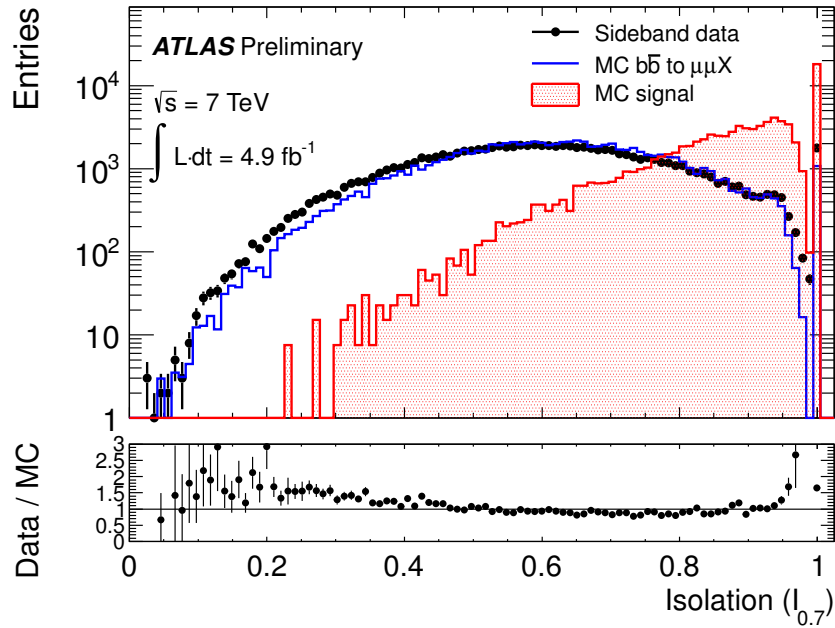
(a) B_s^0 isolation
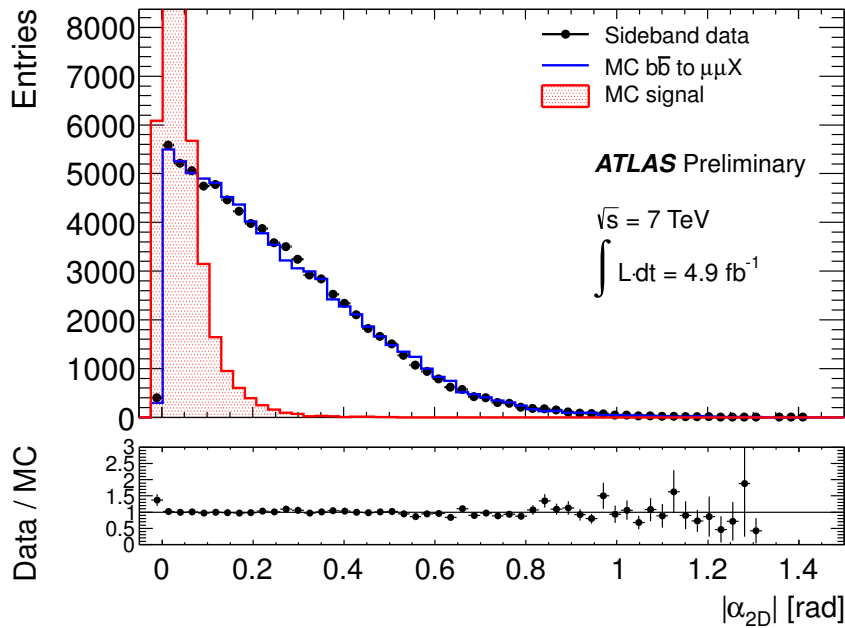
(b) 2D pointing angle

Figure 5.3: The plots show isolation and 2D pointing angle distributions of the B_s^0 candidates after the baseline selection, and selecting $L_{xy} > 0.2\,\text{mm}$ [2]. They are the two most important discriminating variables after L_{xy}. In each case the shaded red histogram shows the $B_s^0 \to \mu^+\mu^-$ MC distribution after the re-weighting, the blue histogram shows the combinatorial background distribution from the $b\bar{b} \to \mu\mu X$ MC sample, and the solid black dots show the sideband distribution in collision data. The Data/MC plots show the agreement between the two background distributions.
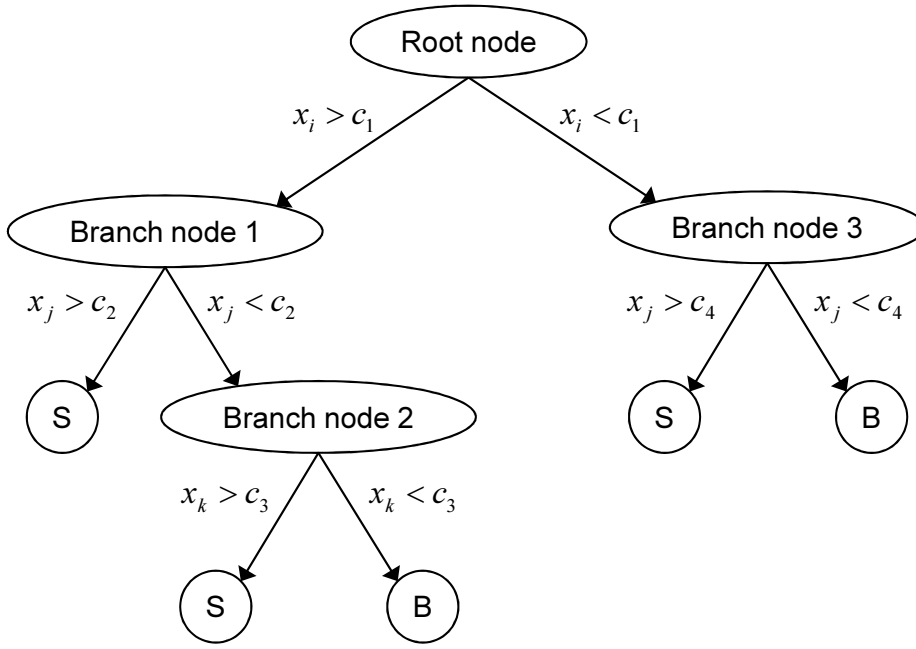
Figure 5.4: A schematic diagram of a decision tree.

These variables are similarly defined for the $B^{\pm}$ channels. Figure 5.3 shows the $I_{0.7}$ and $|\alpha_{2D}|$ distributions for the two test samples described above. The distributions for the B_s^0 candidates reconstructed from the $b\bar{b} \to \mu\mu X$ background MC sample are also overlaid. The plots for the other discriminating variables can be found in Appendix A.

The BDT analysis algorithm computes an event classifier, $Q \in [-1,1]$, for each event in the test samples. The value of Q is closer to $+1$ for signal-like events. It is closer to -1 for background-like events. The Q-distributions for the two test samples are shown in Figure 5.5. A cut (Q_{cut}) is applied on the classifier variable to select signal events. In contrast to the cut-and-count approach, the BDT selection does not apply a hard cut on any individual discriminating variable for the whole sample. A region of phase space is selected based on the value of Q. The optimization of the cut on the classifier variable is briefly discussed in Section 5.3.4.

5.3.3 Monte Carlo Re-weighting

The Q_{cut} is optimized using the MC data. It is necessary that the MC describes the known physics reasonably well. The agreement of the collision data with the MC data can be tested in the reference channel. In practice there are differences between the two owing to the generator level (GL) selections in the MC production, and due to general discrepancies in detector modeling and event reconstruction. It is also necessary to correct for effects, e.g., the polarization of the J/ψ particle, which are not simulated in PYTHIA. The corrections are applied on event-by-event basis. A weight is assigned to each MC event, which is thereafter considered in the various calculations performed using the MC, and also the training of the BDT.

The GL corrections are performed by re-weighting the MC samples using specially produced unbiased $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ MC samples. Since the two channels have different selection cuts, the GL weights are determined for each channel separately. The unbi-
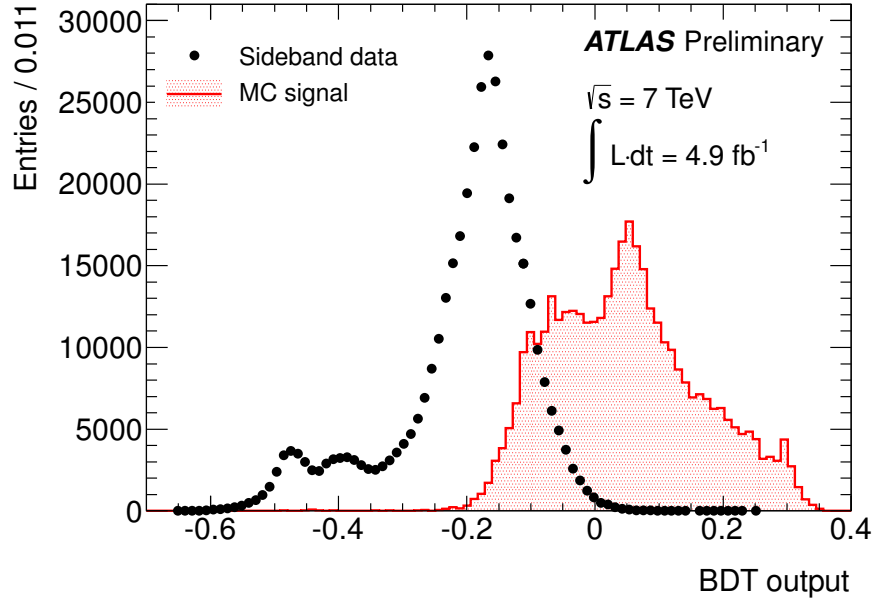
Figure 5.5: Distributions of the BDT classifier Q (plotted on the abscissa axis) for signal and background events [2]. The shaded red histogram shows the $B_s^0 \to \mu^+\mu^-$ signal MC distribution, and the solid black dots show the distribution of the odd numbered sideband events in collision data (background).

ased MC samples are produced without the kinematic selection cuts on the final state particles described in Section 5.1.2, and with relaxed requirements on the *b*-quark kinematics: $|\eta^b| < 4$ and $p_T^b > 2.5\,\text{GeV}$. The samples are binned in two-dimensional bins of η^B and p_T^B of the *B* mesons. For each bin the GL selection efficiency is determined, and the inverse of the efficiency is taken as a weight. These weights are called GL weights. The events in the biased MC datasets are re-weighted according to their η^B and p_T^B values using the GL weights.

In the $B^+ \to J/\psi K^+$ channel, PYTHIA produces a flat angular distribution of the K^+ w.r.t. the two muons from the J/ψ. The J/ψ meson is a spin 1 particle. The angular distribution of the produced K^+, implied by the zero helicity state of the J/ψ in the B^+ rest frame, can be parameterized as:

$$\frac{dN_{B^+}}{d\cos\theta^*} \propto 1 - \cos^2\theta^* = \sin^2\theta^*, \tag{5.2}$$

where θ^* is the angle of the K^+ w.r.t. μ^+ in the B^+ rest frame. The $B^+ \to J/\psi K^+$ MC sample is shaped using the above expression to reproduce the angular distribution.

The discrepancies between the collision data and MC are corrected using data driven (DD) weights. These are calculated in the reference channel, and then applied to both the signal and the reference channel MC. The p_T^B and η^B distributions of the $B^\pm$ signal are obtained from the odd numbered $B^\pm$ candidates in collision data by statistical subtraction of the background. The distributions are compared to the GL-corrected $B^+ \to J/\psi K^+$ MC. The DD weights are factorized as a product of p_T^B and η^B weights:

$$W_{ij}(p_T^B, \eta^B) = w_i(p_T^B) \times w_j(\eta^B), \tag{5.3}$$

where, i and j are bins in the p_T^B and η^B distributions, and $w_x = N_x^{\text{data}}/N_x^{\text{MC}}$ is the ratio of the number of data/MC entries in the bin x. The weight W_{ij} is applied to the MC, and the calculation is repeated until the weights converge.

(a) L_{xy}

(b) $I_{0.7}$

(c) $|\alpha_{2D}|$

(d) p_L^{min}
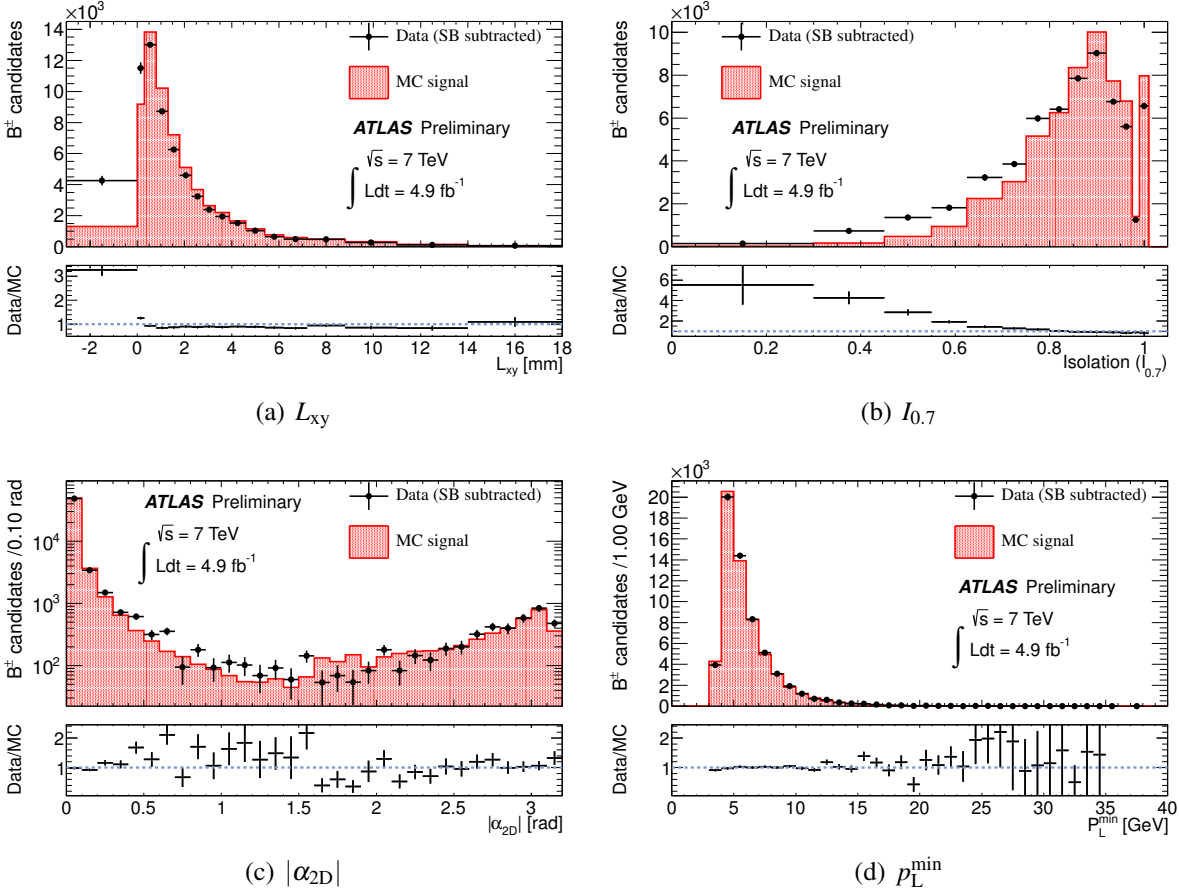
Figure 5.6: Collision data to MC comparison for the four most important discriminating variables [2]. All distributions shown are for the $B^\pm \to J/\psi K^\pm$ channel. The shaded red histograms show the re-weighted MC distributions. The overlaid solid black dots with error bars show the distribution of the background subtracted collision data. The Data/MC plots show the agreement between the data and the re-weighted MC.

The agreement of the collision data with the re-weighted MC is checked for all analysis variables. Figure 5.6 shows the comparison for the four most important variables. The residual data-MC discrepancies are taken as systematic uncertainties (Chapter 6).

5.3.4 Cut Optimization

The optimization of Q_{cut} [2] is performed by maximizing the value of the estimator given in Equation 4.8. The Q_{cut} is optimized together with the $B_s^0 \to \mu^+\mu^-$ signal search window half width $\Delta m_{\mu^+\mu^-}$. The procedure uses odd numbered sideband events in the collision data and the $B_s^0 \to \mu^+\mu^-$ MC sample as explained in Section 5.3.2.

The estimator $\mathscr{P}$ was evaluated at different $Q_{\text{cut}}, \Delta m_{\mu^+\mu^-}$ points by performing a 2D cut-scan. The distribution of the estimator as a function of the two optimized variables is shown in

Figure 5.7. The distribution has a maximum at $Q_{\text{cut}} = 0.1177$ and $\Delta m_{\mu^+\mu^-} = 121\,\text{MeV}$. This Q_{cut} is used for the selection of the B_s^0 candidates as well as the $B^\pm$ candidates. Table 5.5 lists the surviving $B^\pm$ candidates in each dataset after the Q_{cut} is applied.
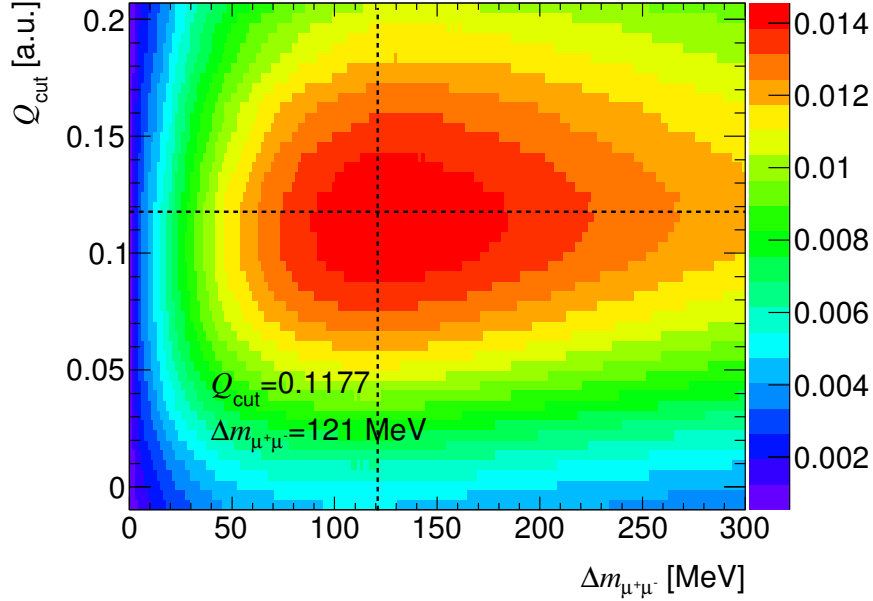
Figure 5.7: $\mathscr{P}$ estimator distribution for the single-bin cut optimization [69]. The value of the estimator (color scale) is shown as a function of the Q_{cut} and the width of the signal search window (plotted on the abscissa axis). The dashed lines show the maximum of the $\mathscr{P}$ distribution, which corresponds to the best signal separation. The optimization yields $Q_{\text{cut}} = 0.1177$ and the width of signal search window as $\Delta m_{\mu^+\mu^-} = 121\,\text{MeV}$.

		Single-bin	BB	BT	AE
Cuts	Q_{cut}	0.1177	0.1092	0.1135	0.1262
	$\Delta m_{\mu^+\mu^-}$	121 MeV	109 MeV	145 MeV	153 MeV
Datasets	Data 2011	44 479	26 274	10 127	9 636
	$B^+ \to J/\psi K^+$ MC	48 042	27 289	11 419	10 897
	$B^+ \to J/\psi \pi^+$ MC	8 487	4 899	1 989	1 907
	$b\bar{b} \to \mu\mu X$ MC	3 582	2 171	901	695

Table 5.5: Optimized cuts and the surviving $B^\pm$ candidates after applying the Q_{cut} in different mass resolution categories. The $\Delta m_{\mu^+\mu^-}$ selection is only applied in the $B_s^0 \to \mu^+\mu^-$ channel (Chapter 6). In the $b\bar{b} \to \mu\mu X$ MC sample the $B^\pm \to J/\psi K^\pm$ and $B^\pm \to J/\psi \pi^\pm$ decays were vetoed on MC truth level.

The optimization for the three mass resolution categories is performed simultaneously using the estimator:

$$\mathscr{P} = \frac{\Sigma \varepsilon_{\text{sig}}^i}{1 + \sqrt{\Sigma N_{\text{bkg}}^i}}, \tag{5.4}$$

where, $\Sigma \varepsilon_{\mathrm{sig}}^{i}$, with $i = $ BB, BT, AE, is the sum of the weights of the $B_s^0 \to \mu^+ \mu^-$ MC events passing a certain Q_{cut} in a certain range $\Delta m_{\mu^+ \mu^-}$ in different mass resolution categories divided by the total sum of the weights after the baseline selection cuts. $\Sigma N_{\mathrm{bkg}}^{i}$ is the sum of the expected background events in the signal region in each category. N_{bkg}^{i} are obtained by linear interpolation of the sidebands in the B_s^0 invariant mass distribution, and integrating the Q-distributions of the background events with $Q > Q_{\mathrm{cut}}$. The results for the three categories are summarized in Table 5.5.

5.4 $B^{\pm}$ Invariant Mass Spectrum

The $B^{\pm}$ invariant mass spectrum after applying the baseline selection cuts on the collision data is shown in Figure 5.8. The background is dominated by $b\bar{b} \to \mu\mu X$ combinatorics across the whole spectrum. Besides the combinatorial background there are significant contributions from partially reconstructed decays (e.g. $B^{+(0)} \to J/\psi K^{*+(0)}$, $B^+ \to \chi_{c1,2}K^+$) to the left of the $B^{\pm} \to J/\psi K^{\pm}$ peak, and the decay of $B^{\pm} \to J/\psi \pi^{\pm}$ to the right, where the pion tracks are assigned a kaon mass in the three-track vertex fit. As there is no particle identification information available such decays cannot be separated from the $B^{\pm} \to J/\psi K^{\pm}$ signal. The fraction of $B^{\pm} \to J/\psi \pi^{\pm}$ background relative to the $B^{\pm} \to J/\psi K^{\pm}$ signal is about 4.8% [29].
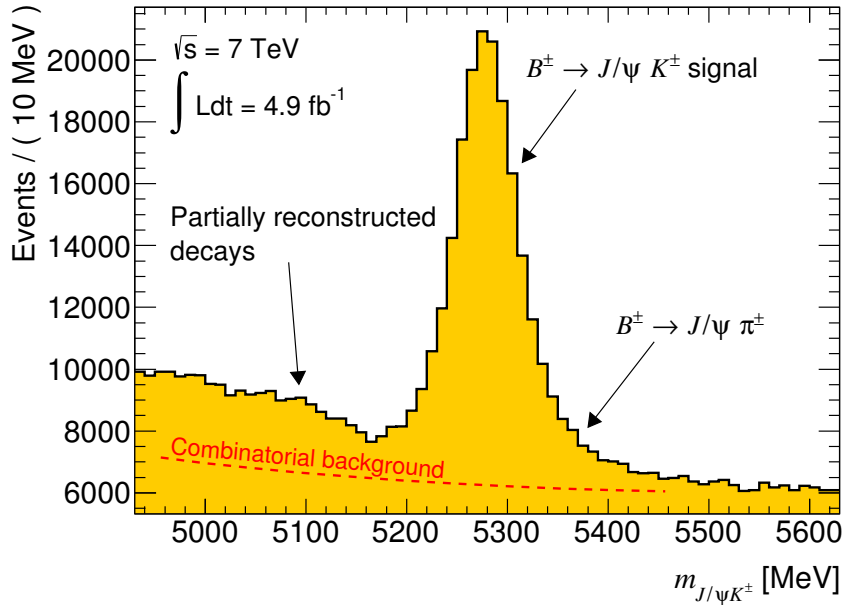
Figure 5.8: The $B^{\pm}$ invariant mass spectrum after applying the baseline selection cuts on the collision data. The dashed red line roughly indicates the extent of the combinatorial background. Partially reconstructed decays form a shoulder to the left of the $B^{\pm} \to J/\psi K^{\pm}$ peak. The $B^{\pm} \to J/\psi \pi^{\pm}$ background is located around 5360 MeV enhancing the right tail of the peak.

The BDT selection is very efficient in rejecting the combinatorial background. Figure 5.9(a) shows the $B^{\pm}$ invariant mass distribution, in the single-bin case, after applying the Q_{cut} (Table 5.5). However, the Q_{cut} has little effect on the resonant backgrounds due to their signal like

decay kinematics.
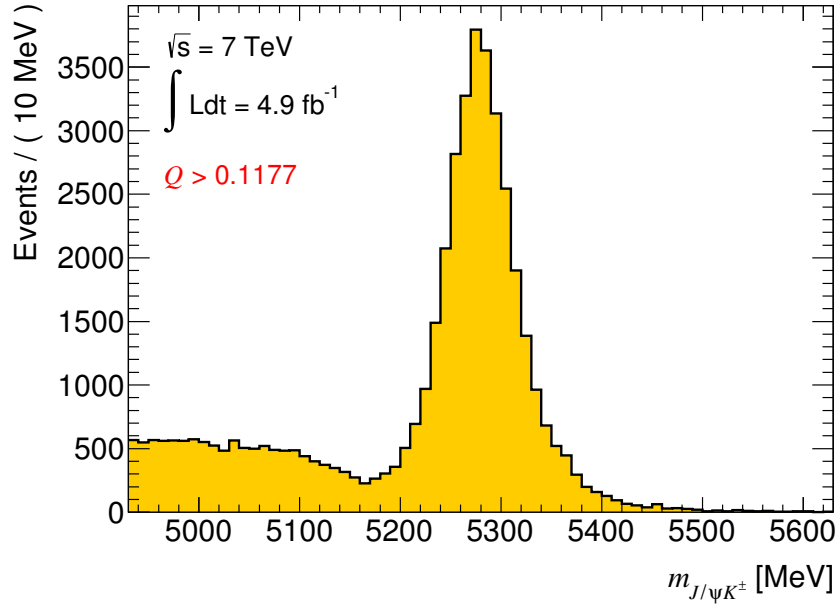
(a) Collision data
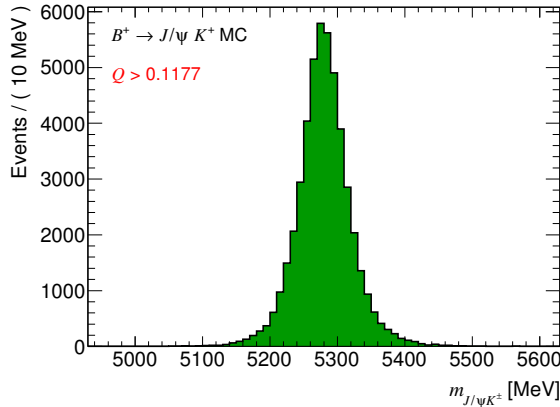
(b) $B^+ \to J/\psi K^+$ MC
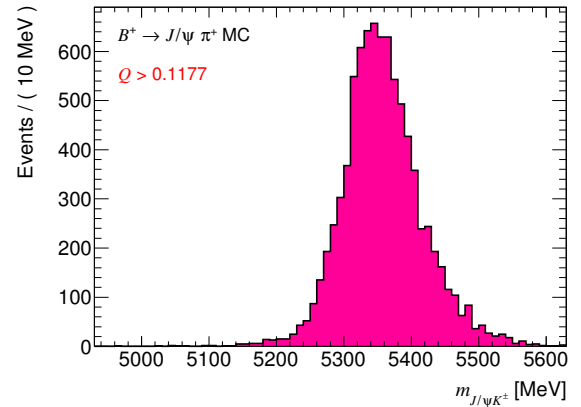
(c) $B^+ \to J/\psi \pi^+$ MC

Figure 5.9: $B^{\pm}$ invariant mass spectra after applying the Q_{cut} on collision data, $B^+ \to J/\psi K^+$ MC and $B^+ \to J/\psi \pi^+$ MC samples (single-bin case).

The partially reconstructed decays include several decay channels. These can be broadly classified into three categories:

- The decay of B^+ or B_d^0 to a J/ψ and an excited state of the kaon: $B^+ \to J/\psi K^{*+}$, $B_d^0 \to J/\psi K^{*0}$. The latter is the most prominent decay contributing to this type of background (Table 5.6). The excited kaon further decays into a kaon and a pion. The pion track is missed in the three-track vertex fit, and hence the B invariant mass shifts towards the lower side of the spectrum.

- Four prong decay of B^+ or B_d^0 to a J/ψ, a kaon and a pion: $B^+ \to J/\psi K^+ \pi^0$, $B_d^0 \to$

$J/\psi K^+\pi^-$. The pion is either captured in the detector material (the case of π^-) or its track is not included in the three-track vertex fit (the case of π^0).

- Decays involving the production of charmonium states of χ_{c1} and χ_{c2}, $B^+ \rightarrow \chi_{c1,2}K^+$, where the $\chi_{c1,2}$ particles decay into a J/ψ and a photon, which is not reconstructed.

These decays have been analyzed [70] using the $b\bar{b} \rightarrow \mu\mu X$ MC sample. There are also contributions from $c\bar{c}$ decays, but they are assumed to be negligible. The relative yields of the various decay modes from the $b\bar{b} \rightarrow \mu\mu X$ MC sample are listed in Table 5.6. The invariant mass distributions of some of these decays are compared in Figure 5.10.
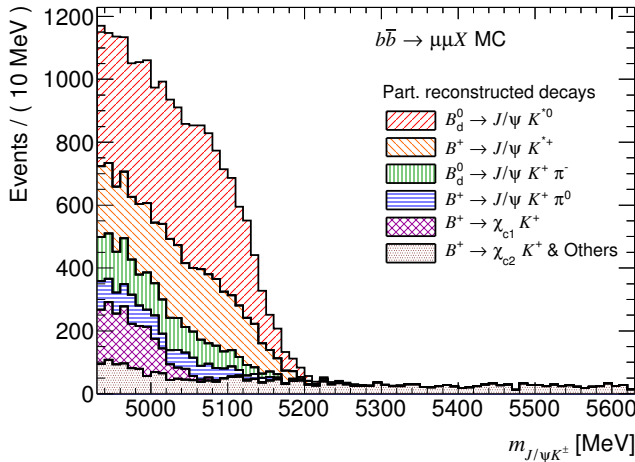
Mode	Yield (%)
$B_d^0 \rightarrow J/\psi K^{*0}$	44
$B^+ \rightarrow J/\psi K^{*+}$	22
$B_d^0 \rightarrow J/\psi K^+\pi^-$	9
$B^+ \rightarrow J/\psi K^+\pi^0$	5
$B^+ \rightarrow \chi_{c1}K^+$	7
$B^+ \rightarrow \chi_{c2}K^+$ & others	13

Figure 5.10: The invariant mass spectrum of the partially reconstructed decays after the baseline selection (single-bin case) extracted from the $b\bar{b} \rightarrow \mu\mu X$ MC sample.

Table 5.6: Partially reconstructed decay modes and their relative yields after the baseline selection (single-bin case) extracted from the $b\bar{b} \rightarrow \mu\mu X$ MC sample [70].

5.4.1 Invariant Mass Resolution Dependence on Pseudorapidity

With the detector the best di-muon invariant mass resolution is achieved in the barrel region. In the region $1.0 \lesssim |\eta| \lesssim 2.5$ the resolution deteriorates due to overlapping sub-detector layers crossed by the tracks. This effect is shown in Figures 5.11(a) and 5.11(c).

The effect is also observed in the mass uncertainty vs. $|\eta^B|$ distribution shown in Figure 5.11(b). In the barrel region the mass uncertainty distribution (Figure 5.11(d)) shows Gaussian-like per-event uncertainties, whereas in the other mass resolution categories the distribution is clearly a superposition of different shapes. Since these shapes are unknown, in the present analysis a kernel estimation technique [71] is used instead to model these distributions (e.g. Section 5.5.1.5).

5.4.2 Comparison of Data and Monte Carlo

The MC samples used in the analysis are kinematically re-weighted to improve the agreement between the collision data and MC. The re-weighing procedure is discussed in Section 5.3.3.

(a) $m_{J/\psi K^\pm}$ vs. η^B

(b) $\delta m_{J/\psi K^\pm}$ vs. η^B

(c) $m_{J/\psi K^\pm}$ distributions

(d) $\delta m_{J/\psi K^\pm}$ distributions
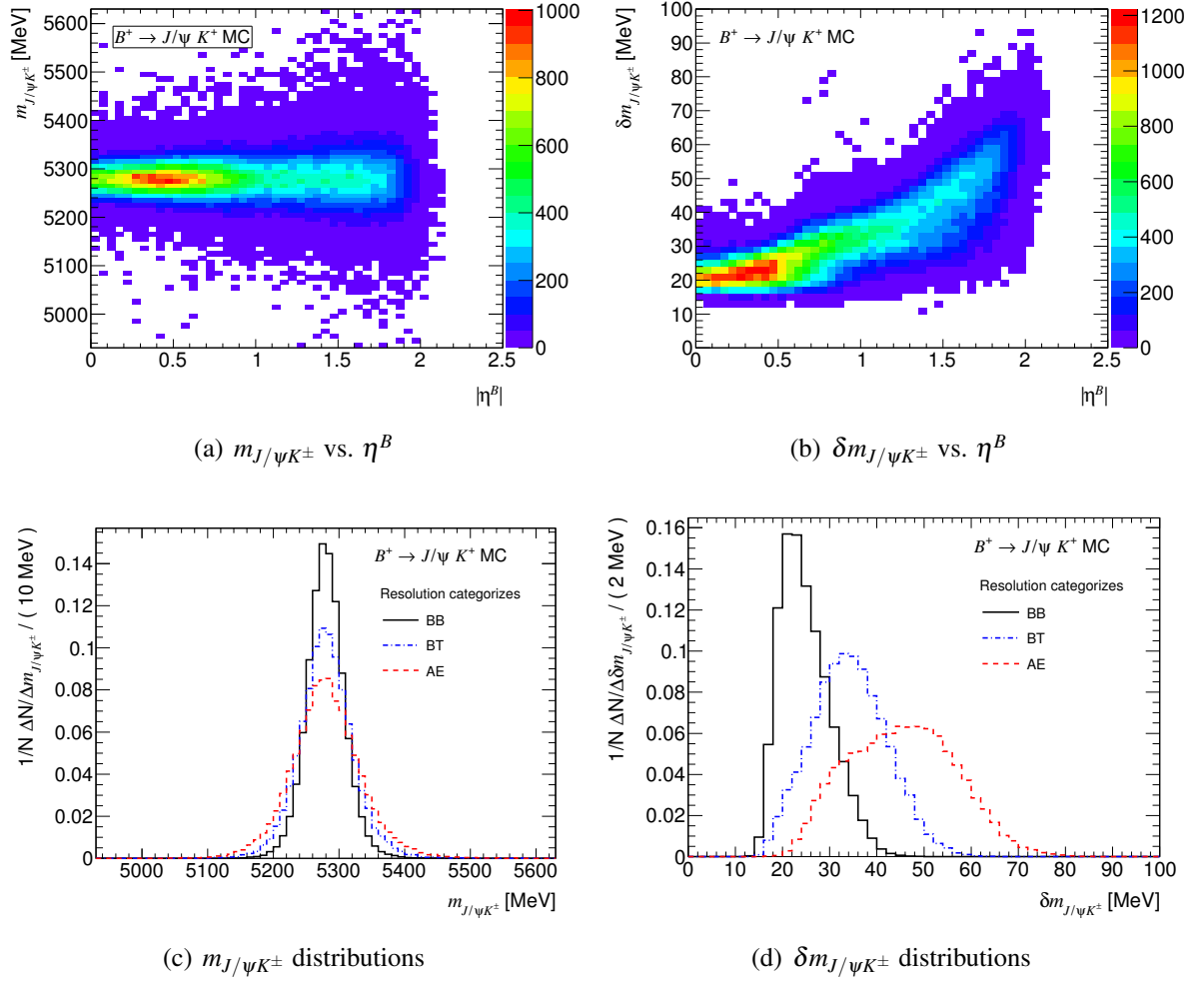
Figure 5.11: Dependence of the $B^\pm$ invariant mass and its uncertainty on pseudorapidity. The plots are extracted from the $B^+ \to J/\psi K^+$ MC sample: the top row shows the variations in the whole pseudorapidity range ($|\eta^B|$), and the bottom row compares the distributions (normalized to unit area) in the three mass resolution categorizes (based on $|\eta^{\mu_{1/2}}|$).

The extraction of the $B^\pm$ yield uses mass uncertainty PDFs constructed from the $B^+ \to J/\psi K^+$ and $B^+ \to J/\psi \pi^+$ MC samples. Figure 5.12(a) compares the mass uncertainty distribution in the $B^+ \to J/\psi K^+$ sample before and after the re-weighting. The agreement of the re-weighted MC with collision data is shown in Figure 5.12(b). The Kolmogorov–Smirnov test is used to compare the two distributions. It shows, with a p-value of 0.2259, a reasonable agreement between the two.
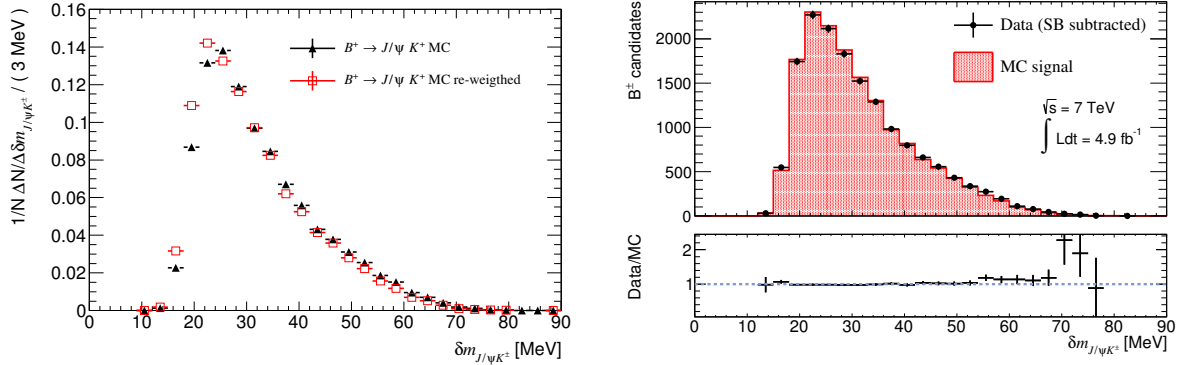
(a) Re-weighted MC against MC with no weights applied

(b) Collision data against re-weighted MC

Figure 5.12: Comparison of the re-weighted MC with collision data in the $\delta m_{J/\psi K^\pm}$ variable [72].

5.5 $B^\pm$ Yield Extraction

The $B^\pm$ yield is determined by performing an unbinned maximum likelihood fit to the $B^\pm$ invariant mass spectrum. The yield is computed from even numbered events in collision data following the strategy in the $B_s^0 \to \mu^+\mu^-$ channel (Section 4.2). Section 5.5.1 describes the construction of the likelihood function and the fit inputs. A discussion of the actual fit results and the performance of the fit is presented in Section 5.5.2. The computation of the $B^\pm$ yield from the fit results and its uncertainties is described in Section 5.5.3.

5.5.1 Construction of the Likelihood

The unbinned fit, for the extraction of the $B^\pm$ yield, uses per-event information so that the final result is sensitive to the quality of the vertex fit. A good vertex fit implies lower uncertainty on the $B^\pm$ invariant mass, and such events should influence the final result more than those with higher uncertainties.

An extended likelihood fit is preferred to obtain a better estimate of the uncertainty on the $B^\pm$ yield. The following likelihood function is used:

$$\mathscr{L} = \prod_i^N M(m^i_{J/\psi K^\pm}, \delta m^i_{J/\psi K^\pm}) \, \text{Poisson}(N_{\text{expected}}, N_{\text{observed}}), \tag{5.5}$$

where N is the number of events in the dataset, and $m^i_{J/\psi K^\pm}$ and $\delta m^i_{J/\psi K^\pm}$ are event-wise $B^\pm$ invariant mass values and their uncertainties. The Poissonian function describes the probability of observing the actual number of events, N_{observed}, given the number of expected event, N_{expected}. The latter is a function of the fit parameters. The two-dimensional likelihood model, M, is a sum of the probability density functions (PDF) used to model the $B^\pm$ signal and the individual background distributions. It is defined as:

$$M(m_{J/\psi K^\pm}, \delta m_{J/\psi K^\pm}) = \frac{N_{\text{sig}} F_{\text{sig}} + N_{\text{C}} F_{\text{C}} + N_{\text{S}} F_{\text{S}} + N_{\text{P}} F_{\text{P}}}{N_{\text{sig}} + N_{\text{C}} + N_{\text{S}} + N_{\text{P}}}, \tag{5.6}$$

where, F_{sig} is the PDF used to model the $B^\pm$ signal, and N_{sig} is a normalization factor representing the number of signal events. N_{sig} is the most important fit parameter. Similarly, the other terms represent the combinatorial background (C), the partially reconstructed decays (S) and the $B^\pm \to J/\psi \pi^\pm$ background (P). N_{C}, N_{S}, and N_{P} are also fit parameters. The fit to this composite model is hereafter referred to as the main fit. The final fit[5] is performed by minimizing[6] $-\log \mathscr{L}$ (Section 5.5.2).

The four PDFs, F_{sig}, F_{C}, F_{S} and F_{P}, are products of a (normalized) function and a mass uncertainty PDF:

$$F_{\text{sig}}(m_{J/\psi K^\pm}, \delta m_{J/\psi K^\pm}) = G(m_{J/\psi K^\pm} | \delta m_{J/\psi K^\pm}) f_{B^+ \to J/\psi K^+}(\delta m_{J/\psi K^\pm}), \tag{5.7}$$

$$F_{\text{C}}(m_{J/\psi K^\pm}, \delta m_{J/\psi K^\pm}) = C(m_{J/\psi K^\pm}) f_{\text{RB}}(\delta m_{J/\psi K^\pm}), \tag{5.8}$$

$$F_{\text{S}}(m_{J/\psi K^\pm}, \delta m_{J/\psi K^\pm}) = S(m_{J/\psi K^\pm}) f_{\text{LB+RB}}(\delta m_{J/\psi K^\pm}), \tag{5.9}$$

$$F_{\text{P}}(m_{J/\psi K^\pm}, \delta m_{J/\psi K^\pm}) = P(m_{J/\psi K^\pm}) f_{B^+ \to J/\psi \pi^+}(\delta m_{J/\psi K^\pm}). \tag{5.10}$$

The functions, G, C, S and P, are defined in Equations 5.12, 5.14, 5.15 and 5.16. They are discussed in the next sections. The mass uncertainty PDFs are described in Section 5.5.1.5.

5.5.1.1 Signal Model

The $B^\pm \to J/\psi K^\pm$ peak in the $B^\pm$ invariant mass spectrum (Figure 5.9(a)) is modeled by a Gaussian distribution:

$$G'(m_{J/\psi K^\pm}, \delta m_{J/\psi K^\pm}) = \exp\left(-\frac{(m_{J/\psi K^\pm} - m_{B^\pm})^2}{2(s\,\delta m_{J/\psi K^\pm})^2} \right), \tag{5.11}$$

where $m_{B^\pm}$ is the mean $B^\pm$ mass. It is one of the fit parameters. The scale factor s is an additional parameter, which allows for variations in the per-event vertex fit uncertainties, $\delta m_{J/\psi K^\pm}$, to adjust the overall width of the Gaussian distribution. This approach offers a way to take per-event uncertainties into account, an advantage not offered by binned fits, and it provides a more precise estimate of the statistical uncertainty on the $B^\pm$ yield.

The per-event uncertainties of the invariant mass values are somewhat underestimated in the three-track vertex fit. The scale factor compensates for this offset. The offset itself is independent of the $\delta m_{J/\psi K^\pm}$ variations. A study to check this relationship is documented in

---

[5]The fit is performed using the tools offered by the RooFit toolkit [73].

[6]The minimization is performed using the MINUIT package [74].

Appendix B. Hence, it is justified to choose a linear relationship between the width of the Gaussian and the scale factor.

The PDF thus constructed is an unnormalized two-dimensional PDF in $m_{J/\psi K^\pm}$ and $\delta m_{J/\psi K^\pm}$. The version used in Equation 5.7 is a conditional PDF obtained by adjusting its normalization for a given $\delta m_{J/\psi K^\pm}$:

$$G(m_{J/\psi K^\pm}|\delta m_{J/\psi K^\pm}) = \frac{G'(m_{J/\psi K^\pm}, \delta m_{J/\psi K^\pm})}{\int G'(m_{J/\psi K^\pm}, \delta m_{J/\psi K^\pm})\, dm_{J/\psi K^\pm}}. \tag{5.12}$$

The integral of $G(m_{J/\psi K^\pm}|\delta m_{J/\psi K^\pm})$ has different values for different $\delta m_{J/\psi K^\pm}$, effectively weighing the events in the likelihood computation. The product in Equation 5.7 is still normalized, because the PDF of $\delta m_{J/\psi K^\pm}$ is normalized to 1:

$$\int f_{B^+ \to J/\psi K^+}(\delta m_{J/\psi K^\pm})\, d\delta m_{J/\psi K^\pm} = 1. \tag{5.13}$$

A one-dimensional fit with a single Gaussian model does not describe the distribution well (poor χ^2/NDF). As shown in Figure 5.11(c) the mass resolution deteriorates with increasing pseudorapidity. A double-Gaussian or triple-Gaussian model with a common centroid fits the distribution better. However, these models are not used, because the increased number of free parameters in the fit produce inaccurate covariance matrices.

5.5.1.2   Model for Combinatorial Background

The combinatorial background (C) spans the entire $B^\pm$ invariant mass spectrum. The distribution drops smoothly (with a changing slope) towards the higher side of the spectrum range forming a long tail. In the spectral range it is modeled as an exponential PDF:

$$C(m_{J/\psi K^\pm}) = e^{\lambda\, m_{J/\psi K^\pm}}, \tag{5.14}$$

where, $m_{J/\psi K^\pm}$ is the measured $B^\pm$ invariant mass, and λ (≤ 0) is a free parameter.

An estimate of the λ parameter can be obtained by fitting the model (Equation 5.14) to a subset of the $b\bar{b} \to \mu\mu X$ MC sample, after the $B^\pm \to J/\psi K^\pm$ signal events, the partially reconstructed decay modes, and the $B^\pm \to J/\psi\pi^\pm$ events have been vetoed. Alternatively, selecting events in the collision data with BDT classifier $Q < -0.3$ produces a sample rich in combinatorial background events, which can be used to estimate λ (Figure 5.13(a)). This value is taken as the initial value in the final fit.

5.5.1.3   Model for Partially Reconstructed Decays Background

The partially reconstructed decay background (S) has a complex shape affected by several decay modes (Figure 5.10). The overall shape can be approximated by a step like function. The simplest model for this background is provided by the complementary error function distribution:

$$S(m_{J/\psi K^\pm}) = \text{erfc}\left(\frac{m_{J/\psi K^\pm} - \mu_S}{\sigma_S}\right), \tag{5.15}$$
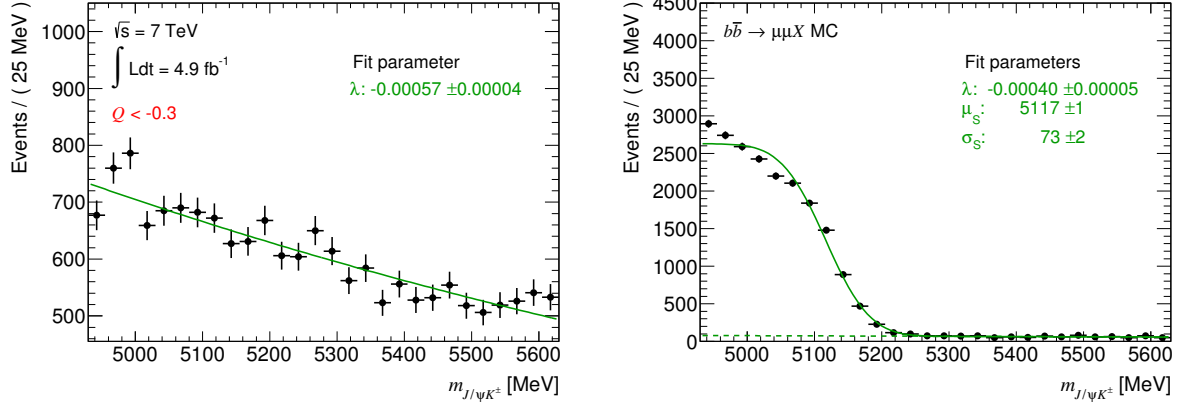
(a) An exponential fit (Equation 5.14) to the continuum background in collision data (obtained after a BDT cut. $Q < -0.3$).

(b) A fit using a sum of Equations 5.14 and 5.15 (variable fractions) to the $b\bar{b} \to \mu\mu X$ sample. The dashed curve shows the fraction of the exponential background. The solid curve is the sum of both components.

Figure 5.13: Estimation of λ, μ_S and σ_S parameters for the main fit.

where, $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$ is the complementary error function, μ_S and σ_S are its mean and standard deviation.

The background is better described by adding two complementary error functions with different means. However, with the large number of fit parameters and low event statistics the main fit performs poorly. Therefore, Equation 5.15 is used as the model for this background.

A preliminarily fit to truth-matched partially reconstructed decays in the $b\bar{b} \to \mu\mu X$ MC sample using Equation 5.15 is shown in Figure 5.13(b). The values of the two parameters from this fit are used as initial values of these parameters in the main fit. The parameters are left free to vary in the main fit.

5.5.1.4   Model for $B^\pm \to J/\psi \pi^\pm$ Background

The shape of the $B^\pm \to J/\psi \pi^\pm$ background (P) is best described by the Crystal Ball (CB) function [75–77]:

$$P(m_{J/\psi K^\pm}) = \begin{cases} e^{-t^2/2} & \text{if } t \geq -|\alpha| \\ \left(\frac{k}{|\alpha|}\right)^k \times e^{-|\alpha|^2/2} \times \left(\frac{k}{|\alpha|} - |\alpha| - t\right)^{-k} & \text{if } t < -|\alpha| \end{cases} \qquad (5.16)$$

where $t = (m_{J/\psi K^\pm} - \mu_P)/\sigma_P$. A part of the CB shape is described by a Gaussian distribution, and the other by an exponential. With $\alpha < 0$ the exponential tail is formed on the right side of the PDF. The parameters μ_P and σ_P are mean and standard deviation of the Gaussian part. α and k are also parameters of the CB function.

The $B^\pm \to J/\psi \pi^\pm$ background cannot be suppressed by analysis cuts because it kinematically resembles the $B^\pm \to J/\psi K^\pm$ signal, and it is also located close to the signal peak. The mean μ_P is highly correlated with the mean $m_{B^\pm}$ of the signal Gaussian, and it has to be constrained by external means. These factors make this background the most difficult component to

parameterize. In the main fit, all four parameters of the CB function are fixed. They are obtained from a fit to the $B^+ \to J/\psi\pi^+$ MC invariant mass distribution (Figure 5.14 and Table 5.7).
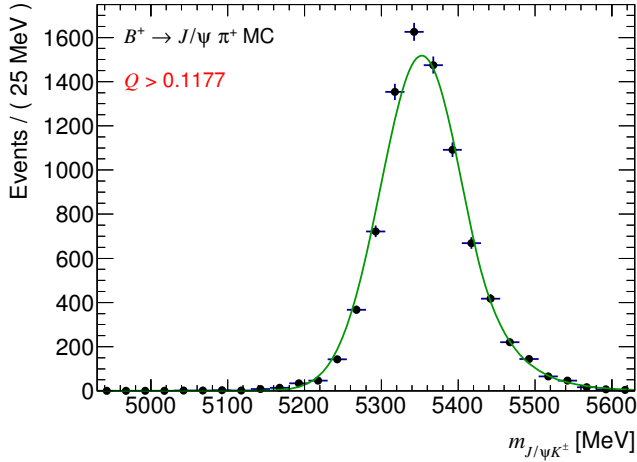
Parameter	Value
μ_P	5353 MeV
σ_P	53.8 MeV
α	-1.28
k	145

Figure 5.14: A fit to the $B^+ \to J/\psi\pi^+$ MC invariant mass distribution (single-bin case) using Equation 5.16. The fit parameters are listed in the adjacent table.

Table 5.7: The values of the CB parameters from the $B^+ \to J/\psi\pi^+$ MC fit.

A bifurcated Gaussian[7] model was tried as an alternative to the CB function, but its agreement with the invariant mass distribution is not as good as the CB function.

5.5.1.5   Construction of Mass Uncertainty PDFs

The maximum likelihood fit model described above uses a one-dimensional mass uncertainty PDF for each model component. The need arises because the sample on which the fits is performed has mixed types of events, i.e. the signal and the various types of background events, where each event class has a different $\delta m_{J/\psi K^\pm}$ distribution. In this case, if a conditional PDF is used to model the distribution of an observable ($m_{J/\psi K^\pm}$), using a single PDF that represents its uncertainty distribution can give biased fit results. Reference [78] points out the pitfalls of such template based fits.

The mass uncertainty PDFs have to represent un-binned data. Here a kernel estimation technique is used to construct the PDFs as described in reference [71]. They are constructed by superposing Gaussian kernels[8] representing per-event invariant mass uncertainties. If the PDF is constructed from a MC dataset, each kernel is weighted using the MC weights described in Section 5.3.3. It can be shown that the PDF describes the un-binned data by projecting the PDF on the histogram created from the original sample. Figure 5.15(a) shows a mass uncertainty PDF overlaid on the original $\delta m_{J/\psi K^\pm}$ histogram. Although the projection follows the shape of the histogram, it is not smooth, and it is affected by the statistical fluctuations in that histogram. This can be controlled by applying a smoothing transformation. The smoothing parameter [73] $\rho = 2$ gives a reasonable smoothing without altering much the overall shape of the PDF. The

---

[7]Defined as: $f(x) = \exp\left(\frac{-(x-\mu)^2}{2\sigma_{L,R}^2 + \alpha_{L,R}(x-\mu)^2}\right)$, where the parameters $\sigma_{L,R}$ and $\alpha_{L,R}$ take different values on either sides (left, right) of the mean μ. It is sometimes referred to as the Cruijff distribution.

[8]In RooFit this is implemented in class `RooKeysPdf`.

projection of the PDF after applying the smoothing is shown in Figure 5.15(b). This procedure is followed for all mass uncertainty PDFs used in the fit.

(a) Default smoothing $\rho = 1$ (b) Smoothing with $\rho = 2$
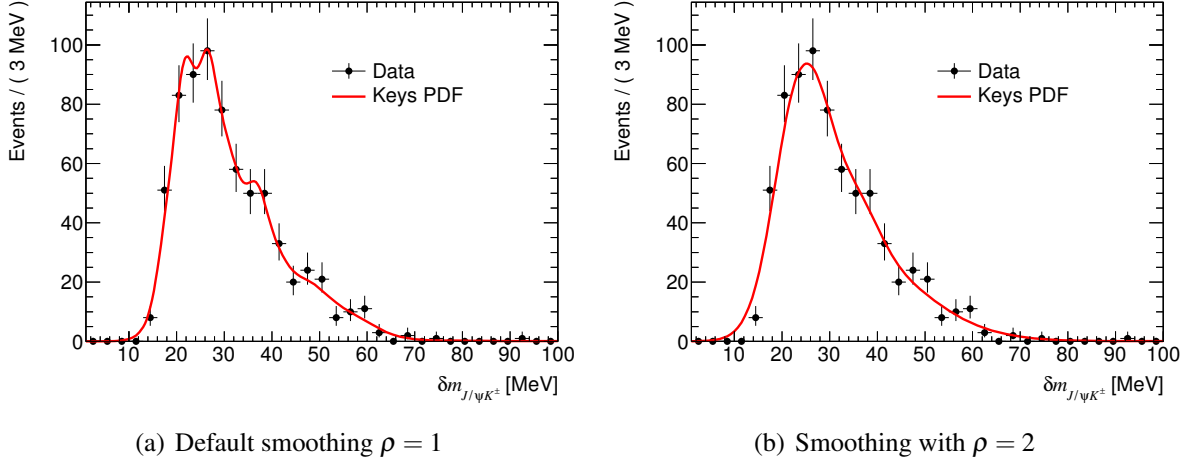
Figure 5.15: Examples of mass uncertainty PDFs overlaid on top of the original $\delta m_{J/\psi K^{\pm}}$ histogram. The PDFs have been normalized to the area of the histogram. The figure demonstrates the effects of choosing different values of the smoothing parameter ρ on the shape of the resulting PDF.

In Equation 5.7, $f_{B^+ \to J/\psi K^+}$ is the mass uncertainty PDF that describes the $B^{\pm}$ invariant mass uncertainty distribution. Since the signal and background events are difficult to separate in collision data, this PDF is constructed using the $B^+ \to J/\psi K^+$ MC dataset instead. Similarly, $f_{B^+ \to J/\psi \pi^+}$ (Equation 5.10) is constructed from the $B^+ \to J/\psi \pi^+$ MC dataset. The mass uncertainty PDFs for the other two background components can also be obtained from the MC. However, the size of the available event statistics in the $b\bar{b} \to \mu\mu X$ sample after the selection cuts is small for those background components. In order to keep the statistical uncertainty on the final result low, the mass uncertainty PDFs for the partially reconstructed decays and the combinatorial background, f_{RB} and f_{LB+RB} (Equations 5.8 and 5.9), are constructed from the collision data. The left and right sidebands in the $B^{\pm} \to J/\psi K^{\pm}$ channel are defined as $[4930, 5130]$ MeV (LB) and $[5430, 5630]$ MeV (RB). The population of the combinatorial background events can be best extracted from the RB. Hence, f_{RB} is constructed from the RB events. The number of partially reconstructed decay events in the RB can be ignored in some mass resolution categories, but for the sake of consistency, events from both the sidebands are combined to construct f_{LB+RB}. Only odd numbered events in collision data are used to construct the mass uncertainty PDFs to have a statistically independent sample from that used in the yield extraction. A comparison of the four mass uncertainty PDFs is shown in Figure 5.16.

5.5.2 Final Fit to Collision Data

The final fit is performed in the $B^{\pm}$ invariant mass range $4930 - 5630$ MeV. The four parameters of the $B^{\pm} \to J/\psi \pi^{\pm}$ background are fixed to the values listed in Table 5.7. Its amplitude, N_P, is left as a free parameter. All other model parameters are also free parameters.
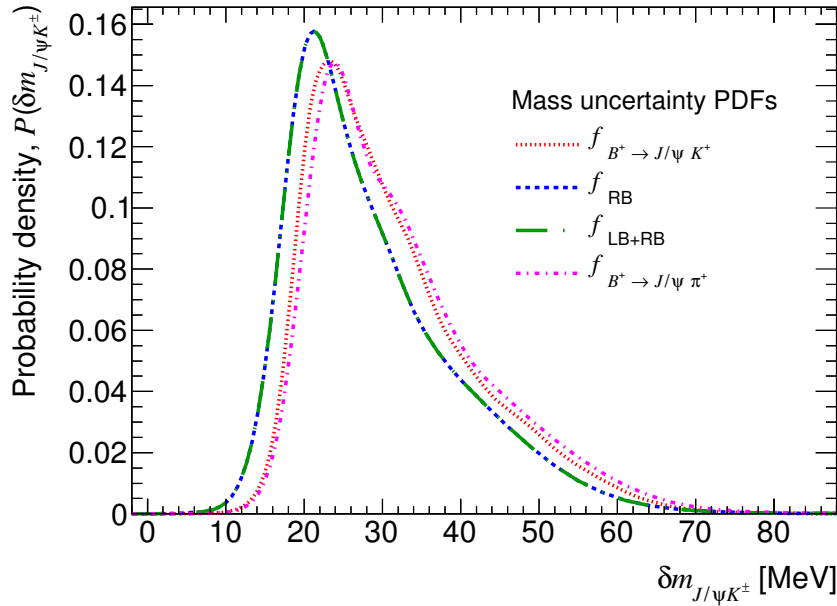
Figure 5.16: Comparison of mass uncertainty PDFs used in Equations 5.7, 5.8, 5.9 and 5.10 (single-bin case).

The fitting algorithm determines the minimum value of $-\log\mathscr{L}$ in a multi-dimensional parameter space. The derivative of the log-likelihood, $\frac{d\log\mathscr{L}}{dp_{i}}$, is computed for each fit parameter, p_{i}, keeping the other parameters constant, and the value of p_{i} is estimated at which the derivative approaches zero. The procedure is iterative, and the algorithm stops when it reaches a minimum in the multi-dimensional parameter space. The uncertainties of the parameters and their covariances are estimated at that point. The next section comments on the performance of the minimization.

The fit to the even numbered events in the collision data for the single-bin case is shown in Figure 5.17(a). The figure shows the overlay of the fit function (projected onto the $m_{J/\psi K^{\pm}}$ dimension) on top of the binned data distribution. Figure 5.17(b) shows the same in the $\delta m_{J/\psi K^{\pm}}$ dimension. The final values of the fit parameters are listed in Table 5.8, and their correlation matrix is shown in Figure 5.18.

The ratio of the number of $B^{\pm} \to J/\psi\pi^{\pm}$ events to the number of $B^{\pm} \to J/\psi K^{\pm}$ events estimated from the fit, $\frac{N_{\mathrm{P}}}{N_{\mathrm{sig}}} = (5.9 \pm 0.4)\%$, is slightly larger than expected (4.8% [29]). The two channels can have differences in their reconstruction efficiencies, and the misidentification probability of the $B^{\pm} \to J/\psi\pi^{\pm}$ events is unknown in collision data. It can be estimated from the MC samples, and a Gaussian constraint could be applied on N_{P}. Using $B^{+} \to J/\psi\pi^{+}$ and $B^{+} \to J/\psi K^{+}$ MC cross-sections as reported by PYTHIA, the ratio is estimated to be $(5.0 \pm 0.5)\%$. However, the constraint has not been applied in order to avoid biases. The fit value is nevertheless in good agreement with the MC estimate.

5.5.2.1 Fit Quality

The quality of the fit is determined from the residual distributions shown in Figures 5.17(c) and 5.17(d). They show the difference between the binned data and the fit function. A χ^{2}/NDF

(a) Fit in dimension $m_{J/\psi K^\pm}$

(b) Fit in dimension $\delta m_{J/\psi K^\pm}$

(c)

(d)
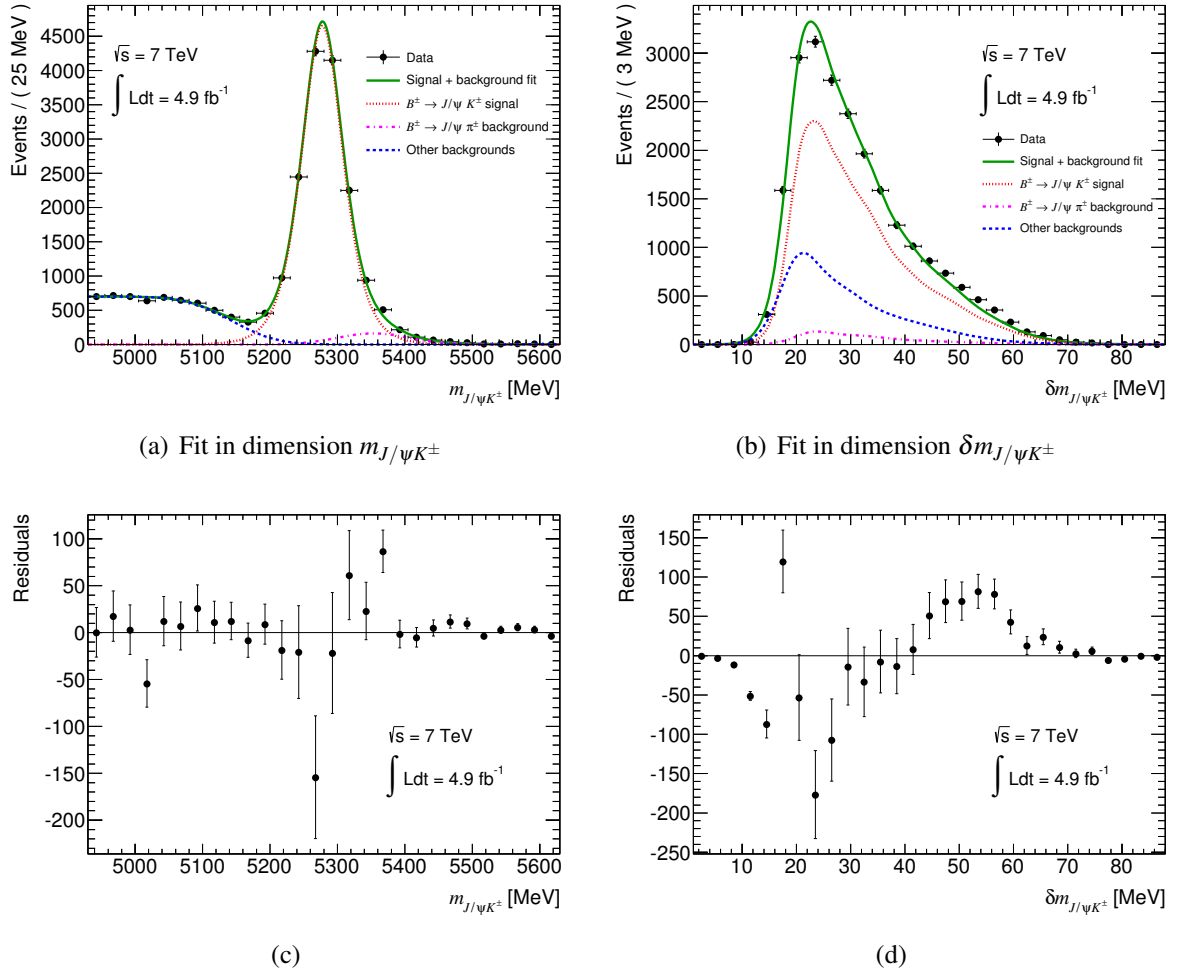
Figure 5.17: The fit to the $B^\pm$ invariant mass spectrum (single-bin case). The plots in the top row show the projection of the fit function (solid green curves) overlaid on the data histogram (solid black dots). The other curves show the contributions from the various model components. The difference between the data histogram and the fit function is shown in the lower plots.

Parameter	Value
$m_{B^\pm}$	$(5277.523 \pm 0.309)\,\mathrm{MeV}$
s	1.1846 ± 0.0104
λ	$(0.00552 \pm 0.00563)\,\mathrm{MeV}^{-1}$
μ_S	$(5142.46 \pm 2.57)\,\mathrm{MeV}$
σ_S	$(75.36 \pm 5.06)\,\mathrm{MeV}$
N_sig	15538 ± 142
N_C	35 ± 23
N_S	5951 ± 86
N_P	915 ± 64

Table 5.8: Final values of the fit parameters (single-bin case).
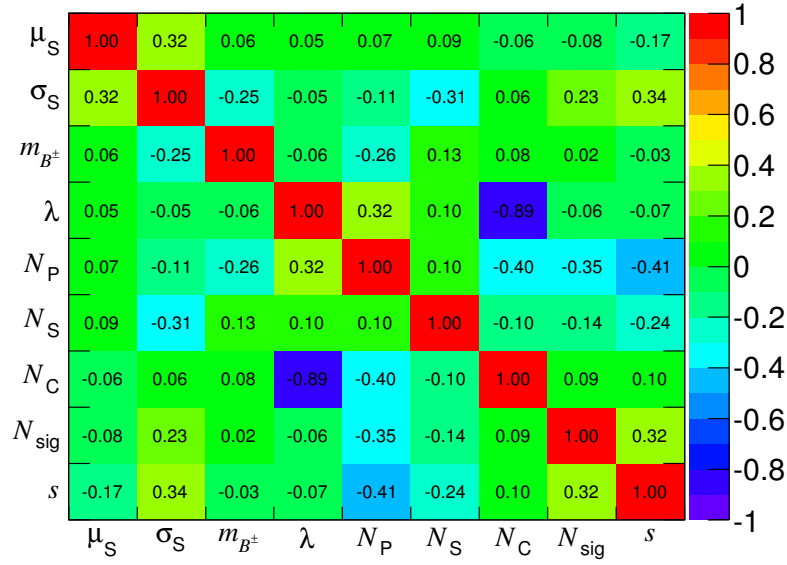
Figure 5.18: Correlation matrix of the fit parameters.

value of 2.2 is computed from the residual distribution in the $m_{B^\pm}$ dimension.

The minimization of $-\log\mathscr{L}$ determines a local minimum in the multi-dimensional parameter space. Therefore, it is important to check if the algorithm did not miss the global minimum. This is checked for each parameter in a wide range of the parameter by recomputing the log-likelihood (as described above) with the final parameter values. The log-likelihood as a function of the parameters at the global minimum is shown in Figure 5.19.

5.5.2.2 Fit Consistency

It is important to ensure that the fit results are not biased due to inconsistencies in the fitting procedure, if any. The procedure can be tested by repeating it on a large number of statistically independent datasets. Since it is impossible to perform the same in collision data due to limited event statistics, the datasets for the test are produced using toy MC simulations.

The model in Equation 5.6 is used to generate a MC dataset. The number of events generated in the dataset is varied around the number of events in the collision data following a Poisson distribution. The MC datasets are then fitted and the fit results are saved. The same fit configuration was used in each fit as in the main fit. A total of 4 000 test fits were performed with 22 439 mean number of events per dataset.

A variable pull is computed for each fit parameter from the test fit results using the expression:

$$\text{pull} = \frac{p_i - \overline{p_i}}{\delta_{p_i}}, \tag{5.17}$$

where, p_i and δ_{p_i} are the parameter and its uncertainty from the test fit, and $\overline{p_i}$ is the mean value of the parameter from the main fit. The pull distributions of the signal model fit parameters are shown in Figure 5.20. The distributions are well described by a Gaussian PDF (solid green curves) of width 'one'. Table 5.9 summarizes the widths and the means of the pull distributions.
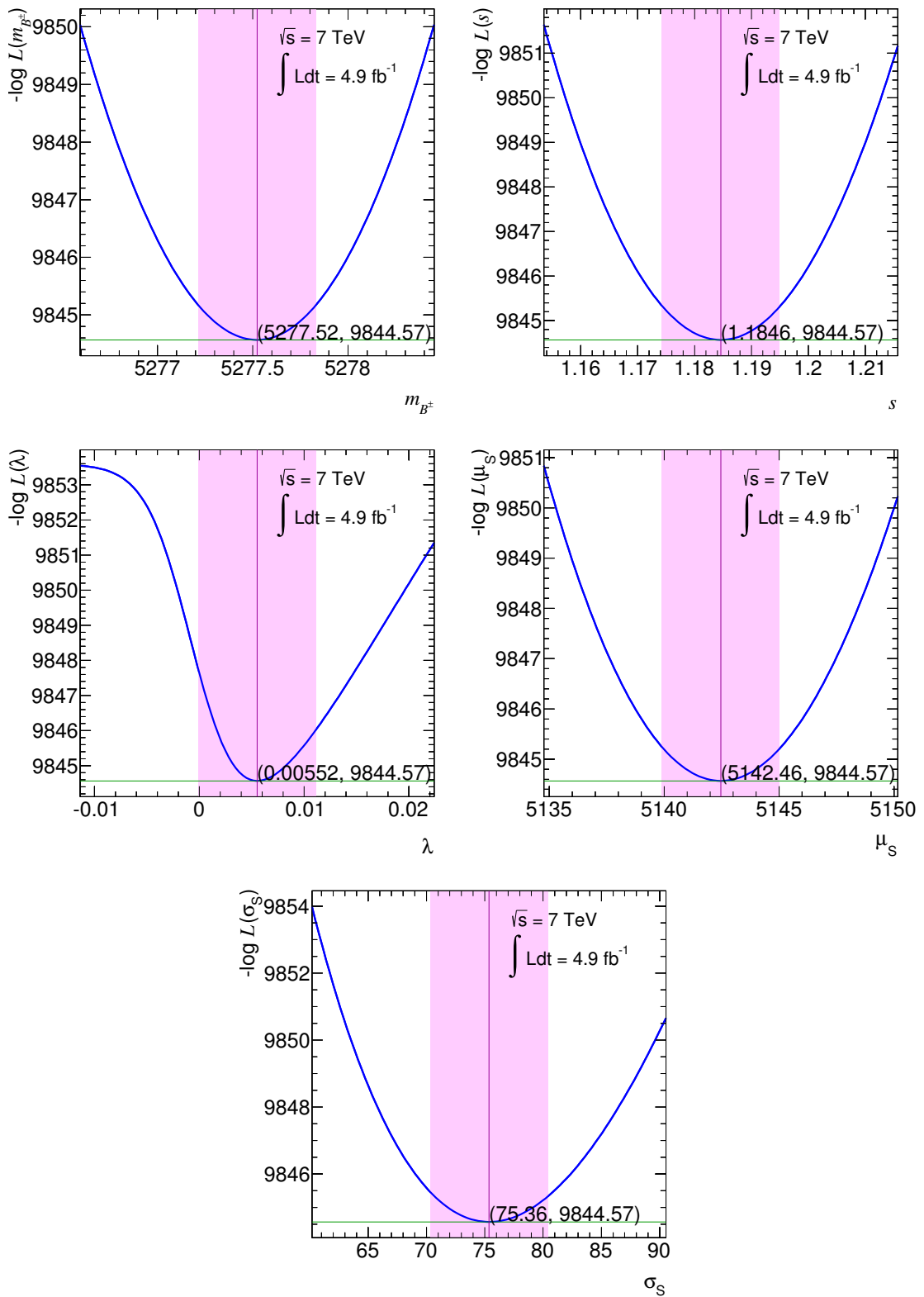
Figure 5.19: Log-likelihood as a function of the fit parameters. The plots are zoomed to reveal the shape of the distribution near the global minimum. In each case the solid vertical line represents the final value of the parameter, and the vertical band depicts the estimated (symmetrized) uncertainty of the parameter.

The distributions are centered around zero. The pull distributions show that there is no bias due
to the fitting procedure.

(a) N_{sig} pull

(b) $m_{B^\pm}$ pull

(c) Pull of the mass resolution scale parameter s
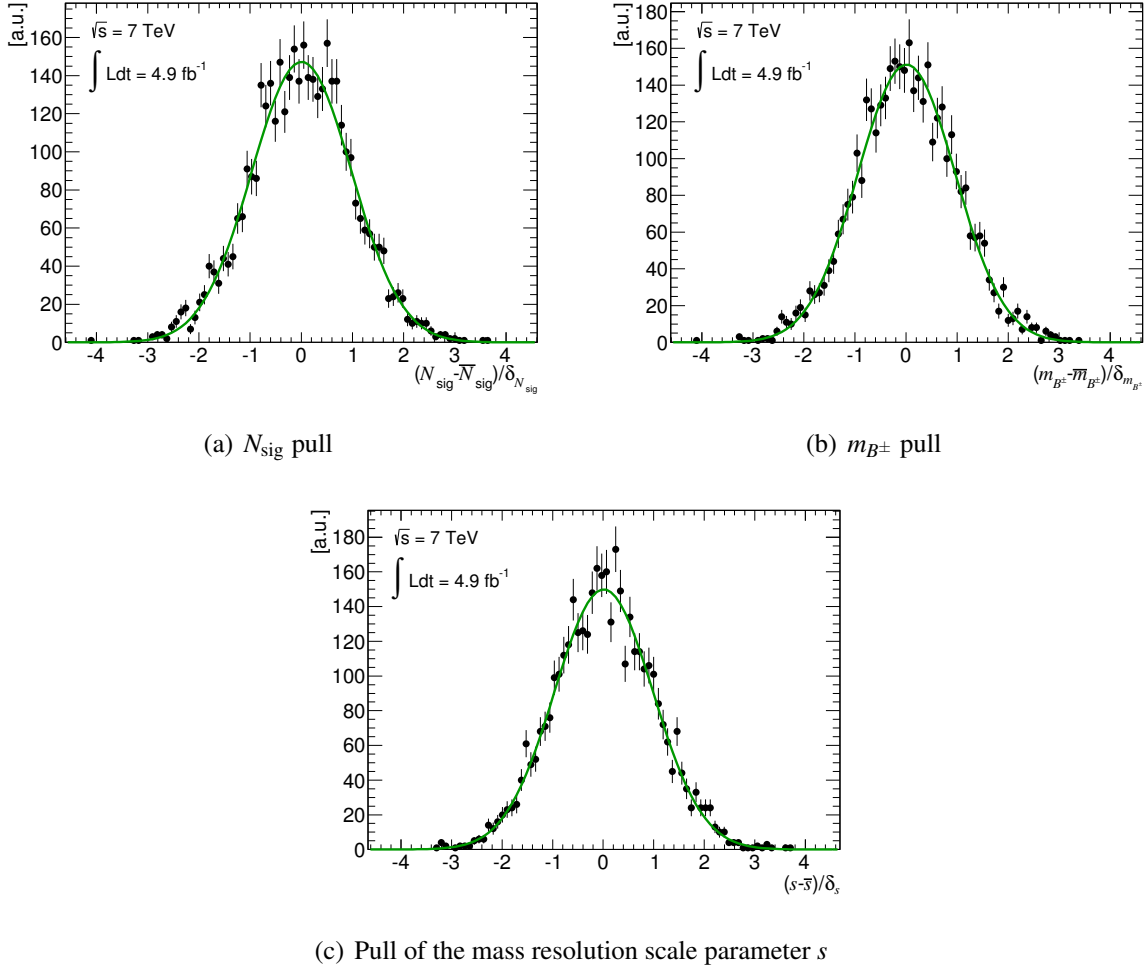
Figure 5.20: Pull distributions of the signal model fit parameters from the toy MC closure test
(single-bin case).

Parameter	Pull width	Pull mean
N_{sig}	0.978 ± 0.012	0.008 ± 0.016
$m_{B^\pm}$	0.955 ± 0.011	0.010 ± 0.015
s	0.975 ± 0.011	0.011 ± 0.016

Table 5.9: The widths and the means of the pull distributions in Figure 5.20 obtained by fitting
them with a Gaussian PDF.

5.5.3   Computation of $B^{\pm}$ Yield

The $B^{\pm}$ yield is derived from the area of the signal Gaussian. The tails of the Gaussian PDF are influenced by the parametrization of the background components. Therefore, they are cut out before the yield is computed. The invariant mass range $m_{J/\psi K^{\pm}} \in [5180, 5380]$MeV, symmetric around the mean $B^{\pm}$ mass, is chosen as the signal band. The yield is computed in the band by integrating the signal PDF (Equation 5.7):

$$N_{B^{\pm}} = N_{\text{sig}} \int_{\text{a}}^{\text{b}} \int_{0}^{\infty} F_{\text{sig}}(m_{J/\psi K^{\pm}}, \delta m_{J/\psi K^{\pm}}; \vec{\mu}) \, dm_{J/\psi K^{\pm}} \, d\delta m_{J/\psi K^{\pm}}. \tag{5.18}$$

In the above expression $\vec{\mu}$ represents the one-dimensional vector (of size n) of the fit parameters $(n = 2 : \{m_{B^{\pm}}, s\})$ from the fit results, and the range [a, b] is the band specified above. The yields for the single-bin case and the three mass resolution categories are summarized in Table 5.11.

    The computation of the statistical uncertainty on the above integral is described in the next section. An estimate of the systematic uncertainty is presented in Section 5.5.3.2.

5.5.3.1   Statistical Uncertainty

The propagation of the uncertainties of the fit parameters to the final result ($N_{B^{\pm}}$) is performed using a toy MC sampling method.

    First a multidimensional Gaussian, $\mathscr{G}(\vec{x}; \vec{\mu}, \Sigma)$, is constructed using the fit results. The dimension of the Gaussian is equal to the number of free parameters n. Here $\vec{\mu}$ is the complete set (of size n) of free parameters used in the model M, Σ is the $n \times n$ covariance matrix of those parameters, and $\vec{x}$ is a random variable vector of dimension n.

    The Gaussian, $\mathscr{G}$, is used as a toy MC model to produce a sample of events of size n_{samp}. If each toy MC event is a parameter set, $\vec{P}_i$, distributed around the mean $\vec{\mu}$, then the distribution of $\vec{P}_i$ $(i = 1, 2, ..., n_{\text{samp}})$ is described by $\mathscr{G}$.

    The integral in Equation 5.18 is evaluated for each $\vec{P}_i$, using the new $m_{B^{\pm}}$ and s values, to obtain a distribution of $N_{B^{\pm}}$. These values are ordered and plotted (Figure 5.21). The uncertainty in $N_{B^{\pm}}$ is read directly from this plot. A sample number k_{samp} is calculated such that it represents a band on the abscissa that marks 68% (1σ) deviation in $N_{B^{\pm}}$. The statistical uncertainties $-\delta N_{B^{\pm}}^{\text{stat}}$, and $+\delta N_{B^{\pm}}^{\text{stat}}$ (indicated by the horizontal red lines in Figure 5.21) correspond to abscissa points k_{samp}, and $n_{\text{samp}} - k_{\text{samp}}$ (indicated by the vertical green lines in Figure 5.21). The value of k_{samp} is computed from the p-value ($\mathscr{Q}$-function[9]) and n_{samp}:

$$k_{\text{samp}} = n_{\text{samp}} \cdot \mathscr{Q}(1). \tag{5.19}$$

This method returns asymmetric uncertainties in $N_{B^{\pm}}$, $-\delta N_{B^{\pm}}^{\text{stat}}$ and $+\delta N_{B^{\pm}}^{\text{stat}}$ (Table 5.11), that take into account the fit result uncertainties.

5.5.3.2   Systematic Uncertainty

The systematic uncertainty is estimated by varying the background models. Two alternate models are used:

---

    [9]The $\mathscr{Q}$-function is a partial integral of a Gaussian, $\mathscr{Q}(z) = \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{-t^2/2} \, dt$
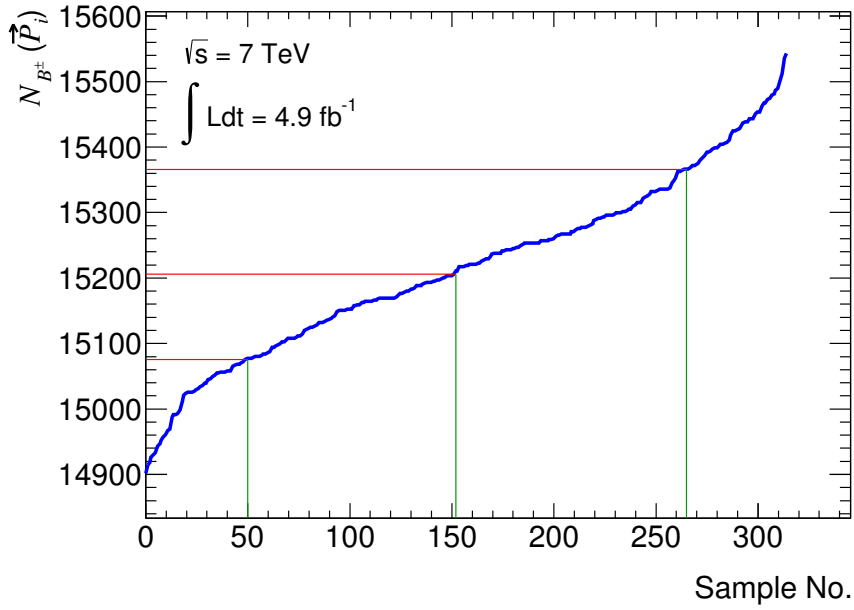
Figure 5.21: The variation of $N_{B^\pm}$ with the fit parameters (parameter set $\vec{P}_i$ of the sample no. *i*). The fit parameters are varied around their original values ($\vec{\mu}$) distributed in an *n*-dimensional Gaussian distribution ($\mathscr{G}$), and the yield is re-computed. The plot is used for the computation of the statistical uncertainties of $N_{B^\pm}$. The central vertical line is the median of the ordered set of yield values, and the left and right vertical lines mark the k_{samp} and $n_{\text{samp}} - k_{\text{samp}}$ value (indicating a 1σ band). The top horizontal line gives $+\delta N_{B^\pm}^{\text{stat}}$, the line in the middle is the $N_{B^\pm}$ value itself from the parameter set $\vec{\mu}$ (for comparison), and the bottom line gives $-\delta N_{B^\pm}^{\text{stat}}$.

- Alternate model I: the model estimates the change in the $B^\pm$ yield if the $B^\pm \to J/\psi \pi^\pm$ background is not modeled. The term F_P and the corresponding parameter N_P are dropped from the likelihood model in Equation 5.6. The combinatorial background and the partially reconstructed decays are modeled by exponential and complimentary error functions as in the main fit.

- Alternate model II: the model estimates the change in the $B^\pm$ yield (w.r.t. alternate model I) if a linear function is used to describe the combinatorial background. The function in Equation 5.14 is replaced by:

$$C(m_{J/\psi K^\pm}) = 1 + s_C \, m_{J/\psi K^\pm}, \tag{5.20}$$

where s_C is a slope parameter. The partially reconstructed decays are modeled by a complimentary error function as in the main fit.

The differences of the yields obtained from the two models w.r.t. the main result are listed in Table 5.10. The largest deviation from the main result is taken as an estimate of the systematic uncertainty. In terms of the fit χ^2/NDF (computed as explained in Section 5.5.2.1) both models perform worse compared to the main fit model.

	Yield	$\delta N_{B^\pm}^{\mathrm{syst}}$	$\frac{\chi^2}{\mathrm{NDF}}$
Main fit	15 210		2.2
Alt. model I	15 577	+367	7.1
Alt. model II	–	–	> 10

Table 5.10: Estimation of the systematic uncertainty of $N_{B^\pm}$. The first row summarizes the main fit result. The yield is computed using an alternate model (second row), and the deviation of this result ($\delta N_{B^\pm}^{\mathrm{syst}}$) from the main result is taken as an estimate of the systematic uncertainty. In the single-bin case the result of the alternate model II fit was not used, because the fit quality was poor. The estimation of the multi-bin systematics is provided in the appendix Table C.2.

5.6 Summary of Results

The final result of the $B^\pm$ yield measurement is summarized in Table 5.11. The table also lists, for comparison, the results in the three mass resolution categories (detailed in Appendix C). The single-bin result is more precise than the multi-bin result due to the higher event statistics.

The results have been checked for biases due to the fitting procedure by means of a closure test. No biases were found. The per-event uncertainties on the $B^\pm$ invariant mass, obtained from the vertex fit, are propagated to the final $B^\pm$ yield. It includes precise estimates of the statistical uncertainties obtained using a toy MC method. The analysis avoids dependence on the MC as much as possible. Any MC samples used have been kinematically re-weighted to match the distributions of the collision data.

Category	$N_{B^\pm}$	$+\delta N_{B^\pm}^{\text{stat}}$	$-\delta N_{B^\pm}^{\text{stat}}$	$\pm\delta N_{B^\pm}^{\text{stat}}$ (%)	$+\delta N_{B^\pm}^{\text{syst}}$	$-\delta N_{B^\pm}^{\text{syst}}$	$\pm\delta N_{B^\pm}^{\text{syst}}$ (%)	$\frac{\chi^2}{\text{NDF}}$
Single bin	15 210	160	130	1.0	367	0	2.4	2.2
BB	9 089	121	95	1.2	169	0	1.9	2.7
BT	3 407	72	71	2.1	157	9	4.6	0.9
AE	3 038	71	65	2.2	236	0	7.8	1.1

Table 5.11: $B^\pm$ yield and its uncertainties in different mass resolution categories.

The result of the single-bin analysis is used in the $B_s^0 \to \mu^+\mu^-$ branching fraction calculation. Like the signal channel the $B^\pm$ yield is computed on even numbered events in collision data. Estimates of the selection efficiencies in the two channels are provided in Chapter 6.

Chapter 6

Summary of the $B_s^0 \to \mu^+\mu^-$ Analysis

The strategy of the $B_s^0 \to \mu^+\mu^-$ analysis has been discussed in detail in Chapter 4. Here the signal channel analysis is summarized. The selection criteria of the B_s^0 candidates is very similar to that of the reference channel (Section 5.3.1). The few minor differences are listed below.

All selection cuts on the two muons in the two channels are the same. The di-muon invariant mass window of the topological trigger in this case is $[4.0, 8.5]$GeV. The B_s^0 candidates are selected with a tighter cut on the vertex fit quality. Those with vertex fit $\chi^2/\text{NDF} < 2$, in a more constrained invariant mass range $m_{\mu^+\mu^-} \in [4.766, 5.966]$GeV, are retained for further analysis. Finally, the BDT cut of $Q > 0.1177$, for the single-bin case, is applied on the data sample.

In Chapter 5 the measurement of the reference channel yield has been described. In this thesis the yield has been determined to be $15\,210 \pm 1.0\%(\text{stat}) \pm 2.4\%(\text{syst})$. An independent measurement of the $B^\pm$ yield has been performed by means of MC-driven maximum likelihood fits [2, 68]. This approach constrains the main fit parameters by simultaneously fitting [73] the corresponding distributions in the MC samples ($B^+ \to J/\psi K^+$, $B^+ \to J/\psi \pi^+$ and $b\bar{b} \to \mu\mu X$). It offers automatic propagation of the uncertainties in the MC fits to the final fit results. The two methods provide consistent results, however, the latter depends more on the MC. The $B^\pm$ yield used in the $B_s^0 \to \mu^+\mu^-$ branching fraction calculation is a weighted average of the two results [2]: $N_{B^\pm} = 15\,214 \pm 1.1\%(\text{stat}) \pm 2.4\%(\text{syst})$.

The other ingredients required for the computation of the Single Event Sensitivity (SES) (Equations 4.6 and 4.7) are the ratios of acceptances and efficiencies [2]:

$$R_{A\varepsilon} = \frac{A_{B^\pm}\varepsilon_{B^\pm}}{A_{B_s^0}\varepsilon_{B_s^0}} = \frac{1.317 \pm 0.008\%(\text{stat})}{4.929 \pm 0.084\%(\text{stat})} = 0.267 \pm 1.8\%(\text{stat}) \pm 6.9\%(\text{syst}). \quad (6.1)$$

They are calculated from the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ MC samples. The products of acceptance and efficiency are computed by normalizing the number of events surviving the selection cuts to the events generated in the fiducial volume, $p_T^B > 8.0$GeV and $|\eta^B| < 2.5$. The systematic uncertainty of 6.9% on $R_{A\varepsilon}$ covers the residual discrepancies between the collision data and the MC after the MC re-weighting (Section 5.3.3). The various sources of systematic uncertainty on the SES are listed in Table 6.1. The largest uncertainty originates from the ratio of the hadronization probabilities, f_u/f_s (8%). Table 6.2 summarizes the inputs for the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ limit extraction including the obtained SES.

The upper limit on $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is computed using the CL$_\text{s}$ method [79]. The following

Description	Syst. uncertainty
PDG branching fractions and f_s/f_d	8.5%
$K^\pm$ tracking efficiency	5%
Vertexing efficiency	2%
$K^\pm$ charge asymmetry in $B^\pm \rightarrow J/\psi K^\pm$	1%
$N_{B^\pm}$	2.4%
$R_{A\varepsilon}$	6.9%
Total (comb. in quadrature)	12.5%

Table 6.1: Systematic uncertainties on the Single Event Sensitivity [2].

likelihood expression is used [2]:

$$
\begin{aligned}
\mathscr{L} =& \mathrm{Poisson}(N_{\mu^+\mu^-}^{\mathrm{obs}} | \varepsilon\mathscr{B} + N_{\mathrm{bkg}} + N_{B_{s,d}^0 \rightarrow h^+h'^-}) \mathrm{Poisson}(N_{\mathrm{bkg,SB}}^{\mathrm{obs}} | R_{\mathrm{bkg}} N_{\mathrm{bkg}}) \times \\
& \mathrm{Gauss}(\varepsilon^{\mathrm{obs}} | \varepsilon, \sigma_\varepsilon) \mathrm{Gauss}(R_{\mathrm{bkg}}^{\mathrm{obs}} | R_{\mathrm{bkg}}, \sigma_{R_{\mathrm{bkg}}}),
\end{aligned}
\tag{6.2}
$$

where,

- $N_{\mu^+\mu^-}^{\mathrm{obs}}$ is the observed number of (even and odd numbered) events in the signal search window, $\pm\Delta m_{\mu^+\mu^-} = \pm 121\,\mathrm{MeV}$, centered around the nominal B_s^0 mass, $m_{B_s^0} = 5366.33\,\mathrm{MeV}$, after un-blinding the collision data;

- $N_{\mathrm{bkg,SB}}^{\mathrm{obs}}$ is the observed number of (even numbered) events in the two sideband;

- N_{bkg} is the estimated number of continuum background events in the signal search window obtained by the interpolation of the sidebands;

- $N_{B_{s,d}^0 \rightarrow h^+h'^-}$ is the peaking background events in the signal search window estimated using MC samples [2];

- the product $\varepsilon \times \mathscr{B}$ is the product of $\varepsilon = 1/\mathrm{SES}$ and the branching fraction of interest (after rearranging Equation 4.7); and

- $R_{\mathrm{bkg}}^{\mathrm{obs}} = \frac{1}{2} \times \frac{300\,\mathrm{MeV}}{\Delta m_{\mu^+\mu^-}}$ is a scaling factor computed from the widths of the sidebands (300 MeV) and the width of the signal search window [2].

The parameters ε and R_{bkg} are nuisance parameters constrained to the observed values $\varepsilon^{\mathrm{obs}}$ and $R_{\mathrm{bkg}}^{\mathrm{obs}}$.

An expected upper limit of 1.6×10^{-8} on $\mathscr{B}(B_s^0 \rightarrow \mu^+\mu^-)$ is obtained. The collision data in the blinded region is analyzed, and the signal yield (Table 6.2) is measured in the optimized search window (Figure 6.1(a)). The observed upper limit is 1.5 (1.2) $\times 10^{-8}$ at 95% (90%) confidence level (CL). The observed limit is comparable with the expected limit (Figure 6.1(b)). Reference [2] provides further details of the analysis.

The very low branching fraction of $B_s^0 \rightarrow \mu^+\mu^-$ made this decay, until recently, inaccessible to the experiments. Thanks to the high luminosity reach of the LHC, the LHC experiments have
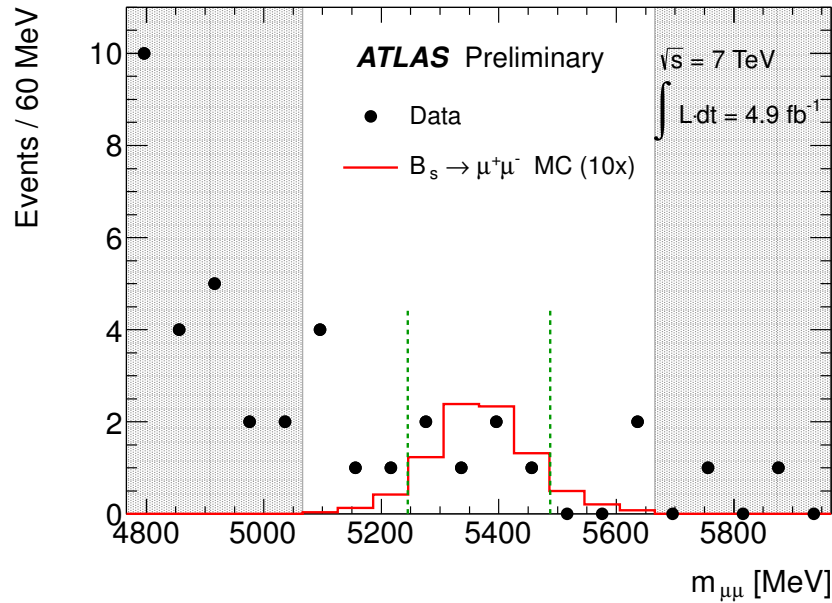
(a) The B_s^0 invariant mass spectrum after un-blinding the collision data (solid dots). The search window is shown by dashed vertical green lines. The shaded gray regions indicate the sidebands used in the analysis. The solid red curve shows the predicted $B_s^0 \to \mu^+\mu^-$ distribution scaled by a factor of 10 (assuming $\mathscr{B}(B_s^0 \to \mu^+\mu^-) = 3.5 \times 10^{-9}$).
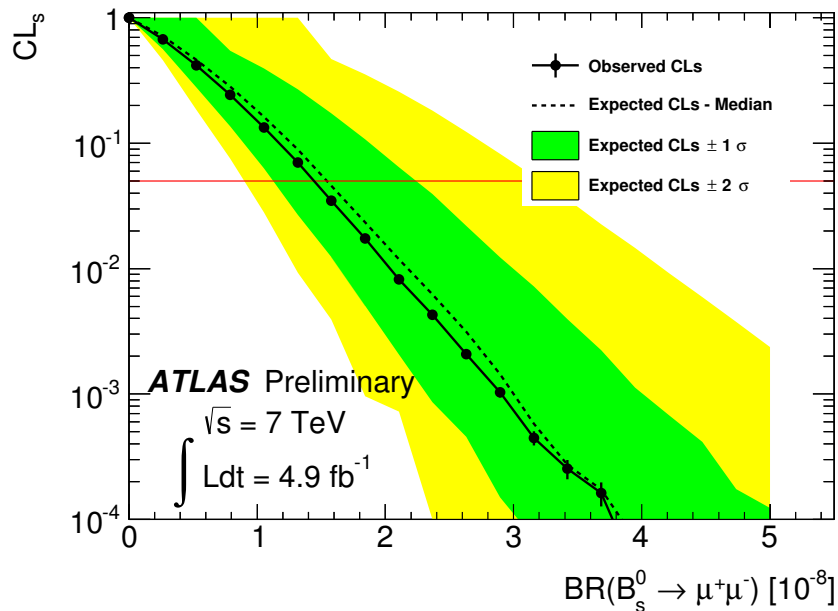
(b) The observed CL_s (solid dots) as a function of the $B_s^0 \to \mu^+\mu^-$ branching fraction. The upper limit is read at the intersection of the red line with the observed CL_s line, which corresponds to 95% CL. The green and yellow bands correspond to the $\pm 1\sigma$ and $\pm 2\sigma$ ranges of the background-only pseudo-experiments with the median of the expected CL_s (dashed line).

Figure 6.1: The latest result from the ATLAS experiment [2].

Quantity	Value
$N_{B^\pm}$	$15\,214 \pm 1.10\% \pm 2.39\%$
$R_{A\varepsilon}$	$0.267 \pm 1.8\% \pm 6.9\%$
SES	$(2.07 \pm 0.26) \times 10^{-9}$
$R_{\text{bkg}}^{\text{obs}}$	1.240 ± 0.050
$N_{\mu^+\mu^-}^{\text{exp}} \mid N_{\mu^+\mu^-}^{\text{obs}}$	$6.75 \mid 6$
$N_{\text{bkg,SB}}^{\text{obs}}$	8
$N_{B_{s,d}^0 \rightarrow h^+ h'^-}$	0.30

Table 6.2: Inputs for the extraction of the upper limit [2]. [The first two rows summarize the inputs for the SES calculation.]

now the sensitivity to observe this decay. An access of events in the B_s^0 invariant mass spectrum was first observed by the LHCb experiment [44,45] (Figure 6.2). Their result has a significance of 4.0σ. Table 6.3 compares the latest results from the three LHC experiments involved in the studies of $B_{s,d}^0 \rightarrow \mu^+\mu^-$. The latest ($4.3\sigma$) result from the CMS experiment estimates $\mathscr{B}(B_s^0 \rightarrow \mu^+\mu^-)$ to be $(3.0_{-0.9}^{+1.0}) \times 10^{-9}$ [48], which is very close to the SM expectation (Equation 3.14). The LHC experiments are working towards improving this measurement.

Experiment	$B_s^0 \rightarrow \mu^+\mu^-$	$B_d^0 \rightarrow \mu^+\mu^-$	$\int L dt$	
ATLAS	$< 1.5 \times 10^{-8}$	$-$	$4.9\,\text{fb}^{-1}$	[2]
LHCb	$(2.9_{-1.0}^{+1.1}) \times 10^{-9}$	$< 7.4 \times 10^{-10}$	$1+2\,\text{fb}^{-1}$	[45]
CMS	$(3.0_{-0.9}^{+1.0}) \times 10^{-9}$	$< 1.1 \times 10^{-10}$	$5+20\,\text{fb}^{-1}$	[48]

Table 6.3: $B_{s,d}^0 \rightarrow \mu^+\mu^-$ results (at 95% CL) from the various LHC experiments. The last column compares the amount of data analyzed (integrated luminosity).
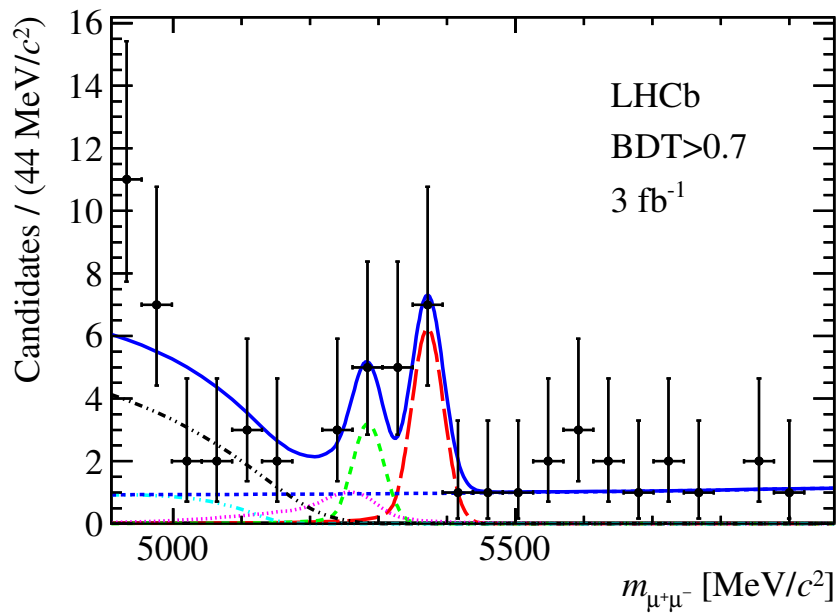
Figure 6.2: The latest result from the LHCb experiment showing the $B^0_{s,d}$ invariant mass spectrum [45]. The solid black dots show the data distribution. The solid blue curve shows an unbinned maximum-likelihood fit to the data. The red long dashed line shows the $B^0_s \to \mu^+\mu^-$ signal, and the green medium dashed line shows the $B^0_d \to \mu^+\mu^-$ signal contribution. The other curves show the various contributing background components: combinatorial background (blue medium dashed), $B^0_s \to h^+h'^-$ (magenta dotted), $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ (light blue dot-dashed), $B^0_d \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ (black dot-dashed).

Appendix A

Discriminating Variables

There are 13 analysis variables used for the computation of the BDT classifier. The variables are discussed in Section 5.3.2. Figure 5.3 shows the distributions of $I_{0.7}$ and $|\alpha_{2D}|$. Figures A.1 through A.11 show the distributions of the other 11 variables. The shaded green histograms shows the $B_s^0 \rightarrow \mu^+\mu^-$ MC distribution after the re-weighting, the solid red dots shows the combinatorial background distribution from the $b\bar{b} \rightarrow \mu\mu X$ MC sample, and the solid black dots show the sideband distribution in collision data. The Data/MC plots show the agreement between the two background distributions.
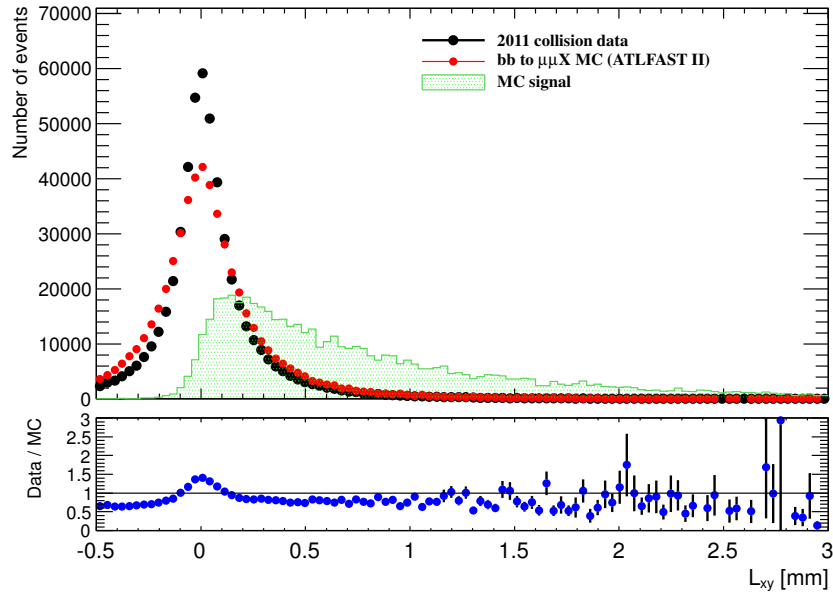
Figure A.1: L_{xy} distribution [68].
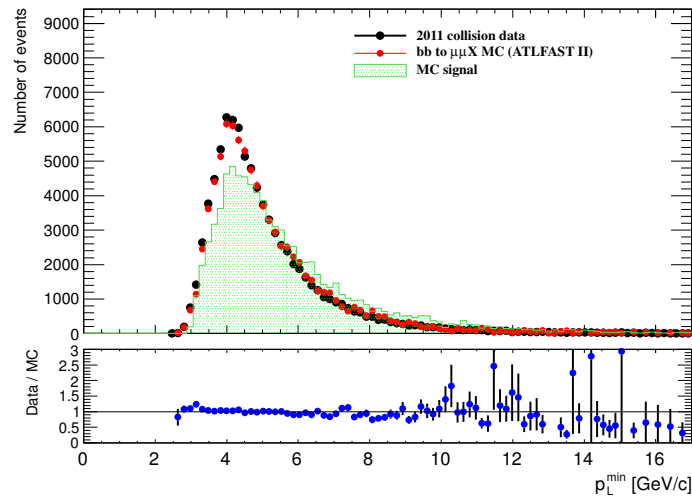
Figure A.2: $p_{\mathrm{L}}^{\mathrm{min}}$ distribution of events with $L_{\mathrm{xy}} > 0.2\,\mathrm{mm}$ [68].
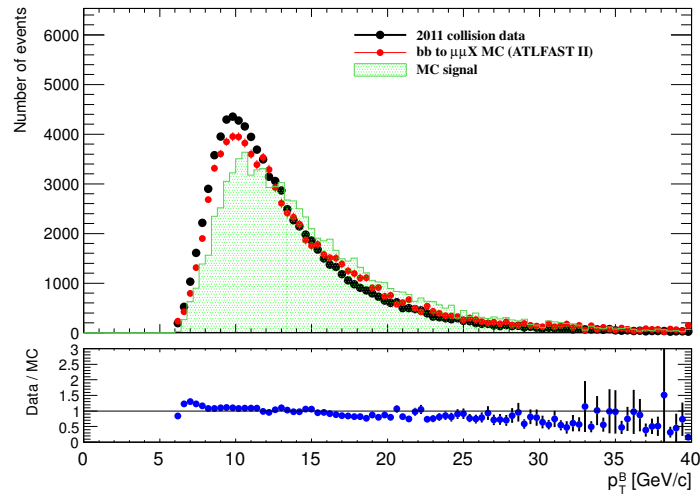
Figure A.3: p_{T}^{B} distribution of events with $L_{\mathrm{xy}} > 0.2\,\mathrm{mm}$ [68].
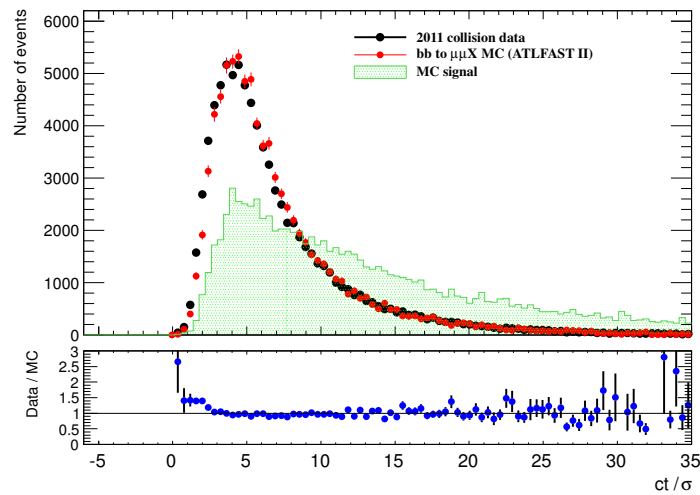
Figure A.4: Proper decay length significance (ct/σ) with $L_{\mathrm{xy}} > 0.2\,\mathrm{mm}$ [68].
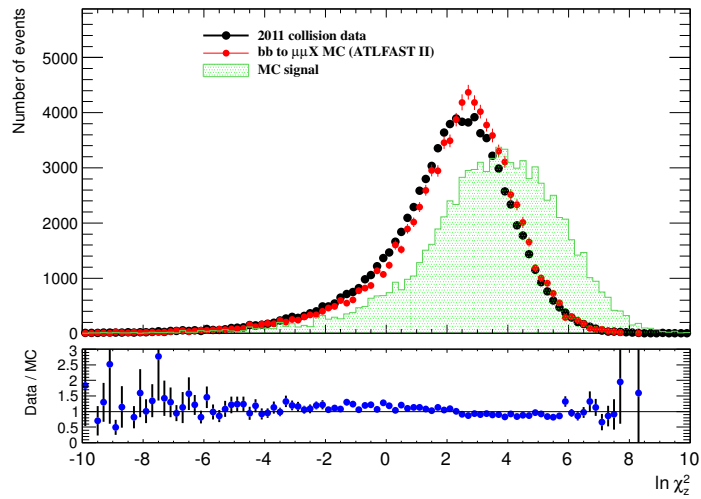
Figure A.5: $\ln \chi_z^2$ distribution of events with $L_{xy} > 0.2\,\text{mm}$ [68].
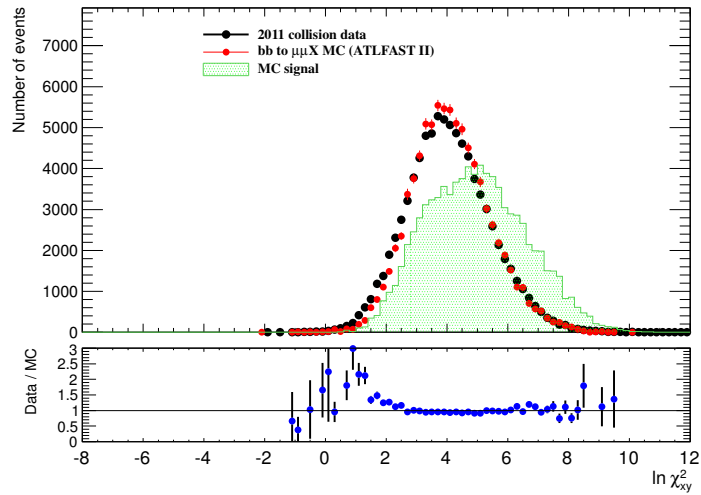
Figure A.6: $\ln \chi_{xy}^2$ distribution of events with $L_{xy} > 0.2\,\text{mm}$ [68].
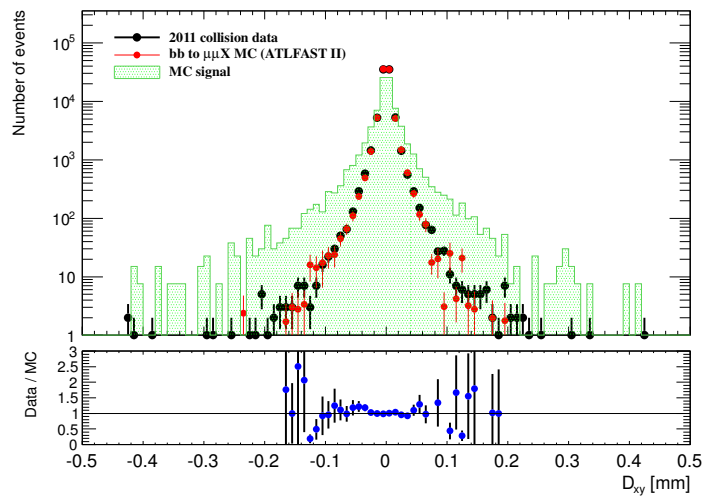
Figure A.7: D_{xy} distribution of events with $L_{xy} > 0.2\,\text{mm}$ [68].
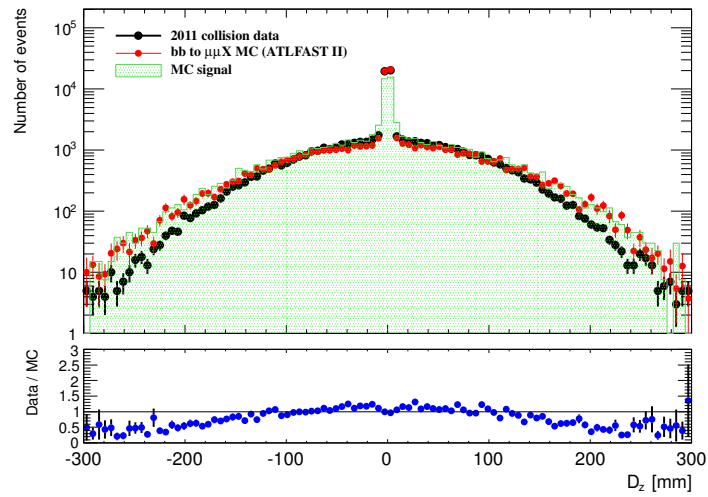
Figure A.8: D_z distribution of events with $L_{xy} > 0.2\,\text{mm}$ [68].
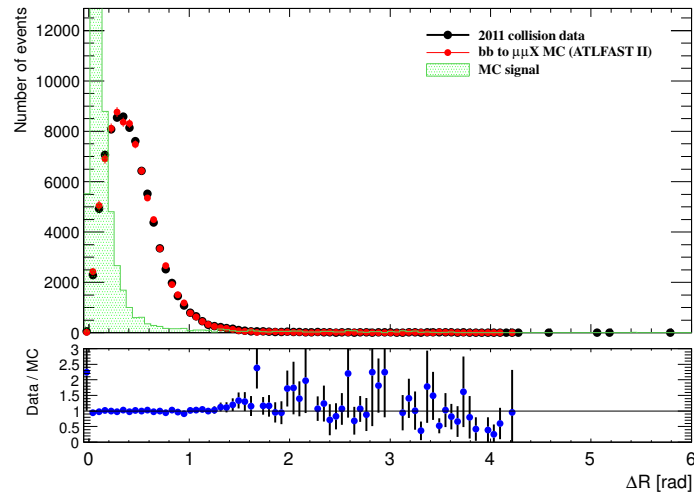
Figure A.9: ΔR distribution of events with $L_{xy} > 0.2\,\text{mm}$ [68].
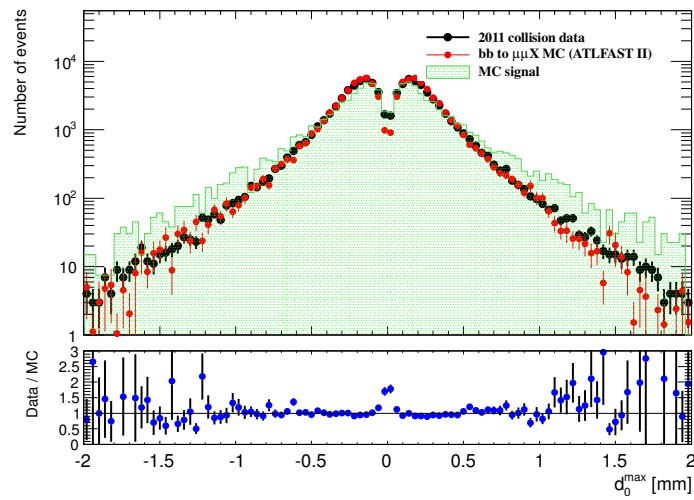
Figure A.10: d_0^{max} distribution of events with $L_{xy} > 0.2\,\text{mm}$ [68].
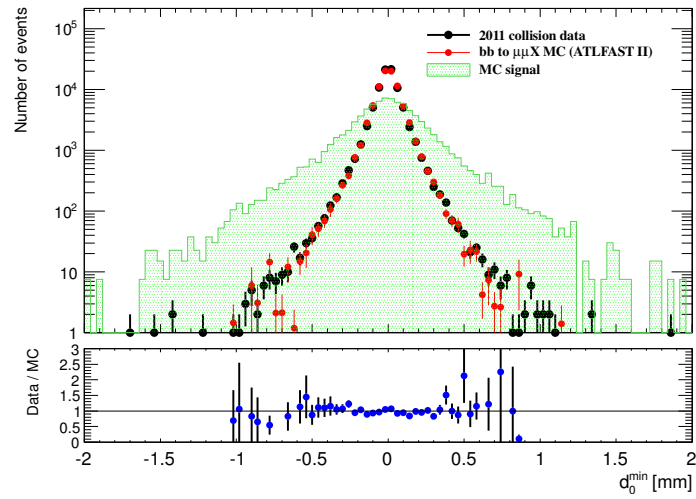
Figure A.11: $d_0^{\min}$ distribution of events with $L_{xy} > 0.2\,\text{mm}$ [68].

Appendix B

Variation of Mass Resolution

The signal model used for the $B^\pm$ yield extraction (Equation 5.11) uses a scale factor s to allow its width to be influenced by the per-event mass uncertainties. Ideally, s would have a value of '1'. Any other value of s indicates that the per-event uncertainties are either underestimated or overestimated. Figure B.1 shows the so called 'pull distributions' computed using the reconstructed $B^\pm$ invariant mass, $m_{J/\psi K^\pm}$, its uncertainty, $\delta m_{J/\psi K^\pm}$, and the truth mass, $m_{J/\psi K^\pm}^{\mathrm{truth}}$ (the values are taken from the $B^+ \to J/\psi K^+$ MC sample). The pull, $\frac{m_{J/\psi K^\pm} - m_{J/\psi K^\pm}^{\mathrm{truth}}}{\delta m_{J/\psi K^\pm}}$, is plotted on the abscissa. Gaussian fits to the pull distributions show $\sigma > 1$ in all cases, indicating that the mass uncertainties from the vertex fit are generally underestimated. This is in agreement with the observed values of s listed in Tables 5.8 and C.1.

In order to check if the mass resolution scales linearly with $\delta m_{J/\psi K^\pm}$, the computed pull values are split into bins of $\delta m_{J/\psi K^\pm}$. The pull distributions are fitted with a Gaussian model, and the width of the pull distribution in each bin is plotted as a function of the median $\delta m_{J/\psi K^\pm}$. The plots are shown in Figure B.2. The bins are of varying widths to have enough event statistics in each bin. The linear dependence of the mass resolution on the mass uncertainty can be seen in these plots. This supports the choice of a linear term for the mass resolution in Equation 5.11.

(a) Single-bin

(b) BB

(c) BT

(d) AE
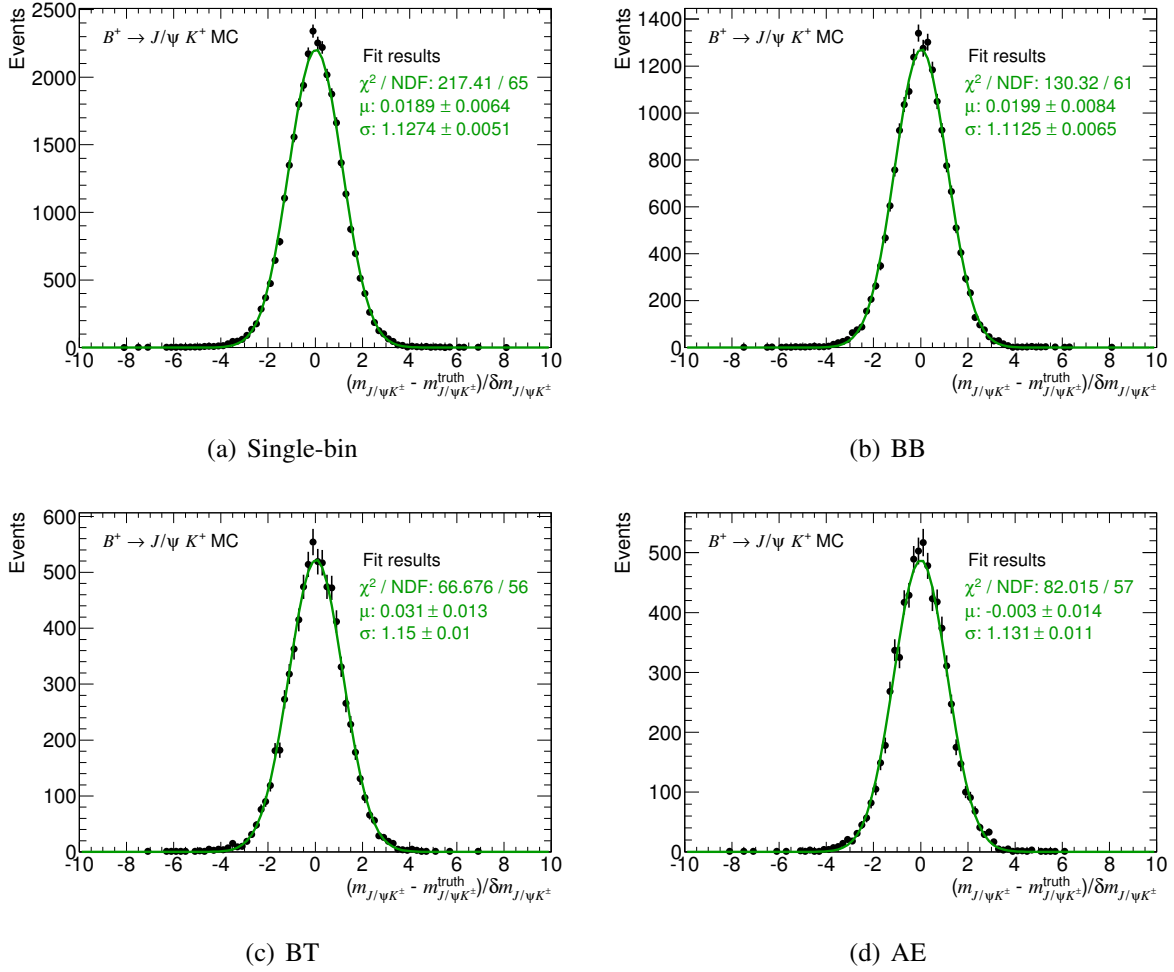
Figure B.1: Mass resolution pull distributions $(\frac{m_{J/\psi K^{\pm}} - m_{J/\psi K^{\pm}}^{\text{truth}}}{\delta m_{J/\psi K^{\pm}}})$ computed using the $B^+ \rightarrow J/\psi K^+$ MC for the four categories.

(a) Single-bin

(b) BB

(c) BT

(d) AE
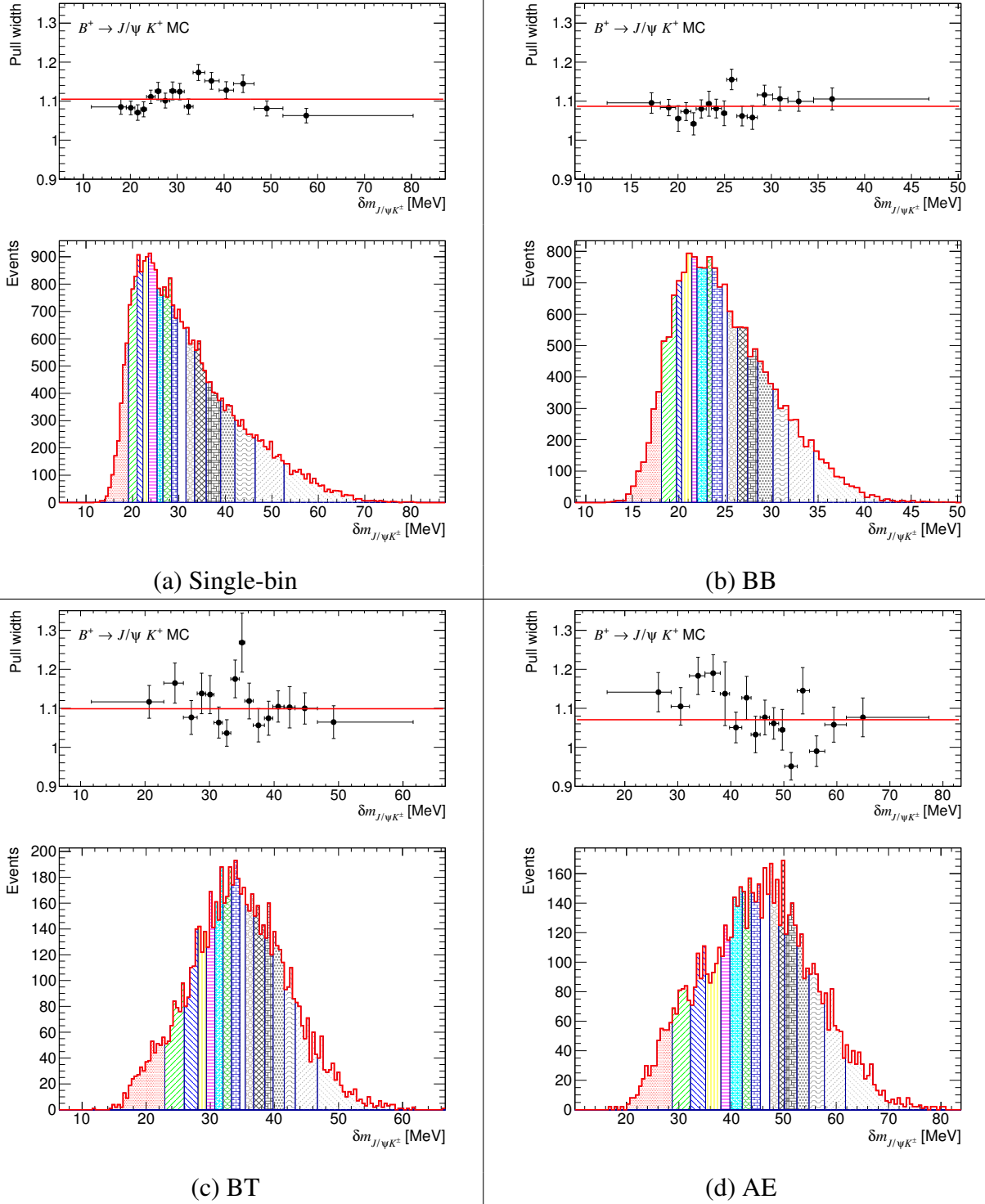
Figure B.2: The dependence of the $B^{\pm}$ mass resolution on the mass uncertainties. The four sets of plots correspond to the four categories. In each case the upper plot shows the pull widths as a function of median $\delta m_{J/\psi K^{\pm}}$ (the horizontal red line depicts the average), and the lower plot shows the $\delta m_{J/\psi K^{\pm}}$ bins corresponding to the upper plot.

Appendix C

Multi-bin Results

This chapter lists the results obtain in the three mass resolution categories (BB, BT and AE). The final values of the fit parameters are listed in Table C.1. The fit plots are shown in Figures C.1, C.2 and C.3.

Table C.2 corresponds to Table 5.10. It lists the yields obtained from the various models used for the estimation of the systematic uncertainties on the main fit result.

Parameter	BB	BT	AE
$m_{B^\pm}$ (MeV)	5277.791 ± 0.334	5276.617 ± 0.803	5273.72 ± 1.24
s	1.1838 ± 0.0126	1.1788 ± 0.0229	1.1431 ± 0.0287
λ (MeV^{-1})	0.00325 ± 0.00923	-0.0143 ± 0.0143	0.0090 ± 0.0129
μ_S (MeV)	5139.1 ± 3.4	5151.45 ± 7.16	5144.43 ± 5.31
σ_S (MeV)	82.71 ± 6.58	64.8 ± 13.3	53.39 ± 9.09
N_sig	9118 ± 103	3482 ± 71	3239 ± 85
N_C	18 ± 28	73 ± 115	5 ± 6
N_S	3578 ± 66	1270 ± 120	1245 ± 42
N_P	467 ± 39	320 ± 35	396 ± 57

Table C.1: Final values of the fit parameters (multi-bin case).
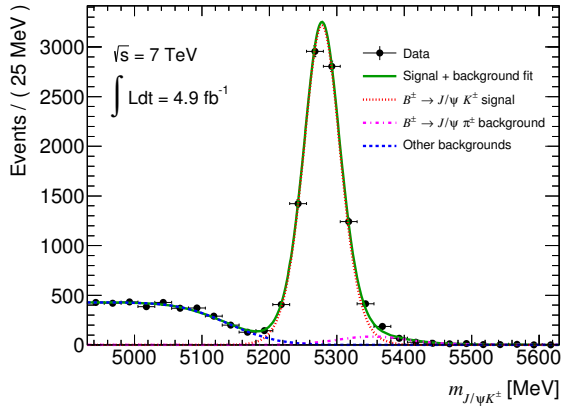
(a) Fit in dimension $m_{J/\psi K^{\pm}}$
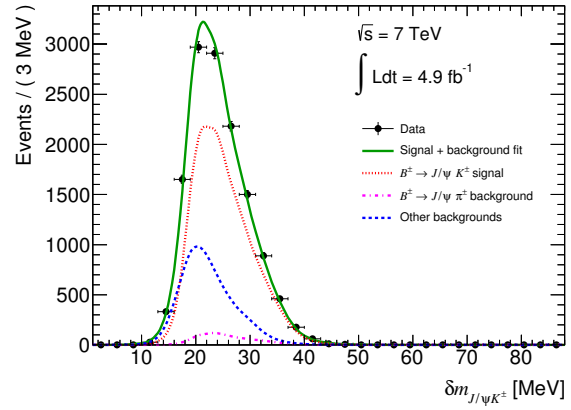
(b) Fit in dimension $\delta m_{J/\psi K^{\pm}}$
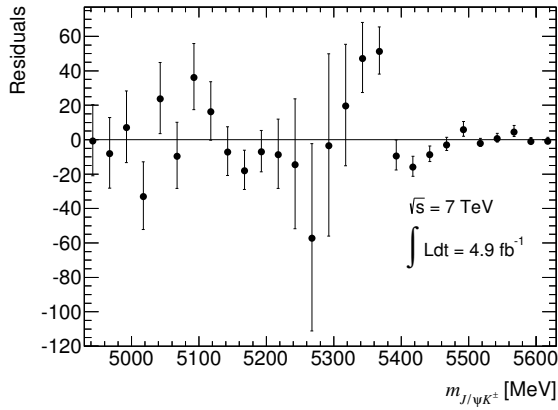
(c)
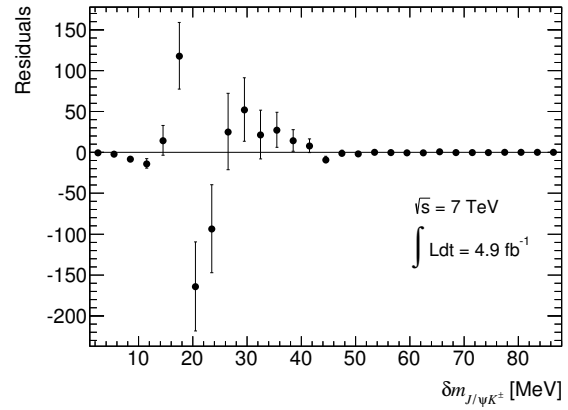
(d)

Figure C.1: The fit to the $B^{\pm}$ invariant mass spectrum (BB category).
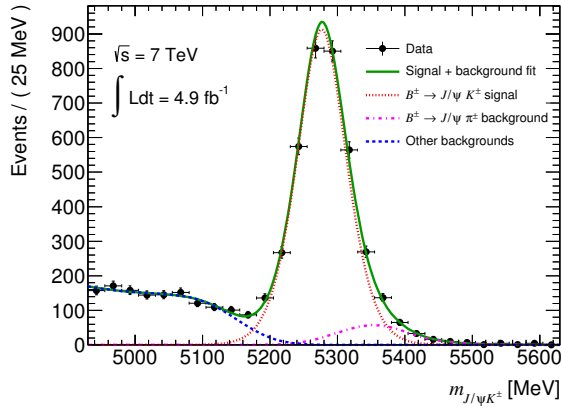
(a) Fit in dimension $m_{J/\psi K^{\pm}}$
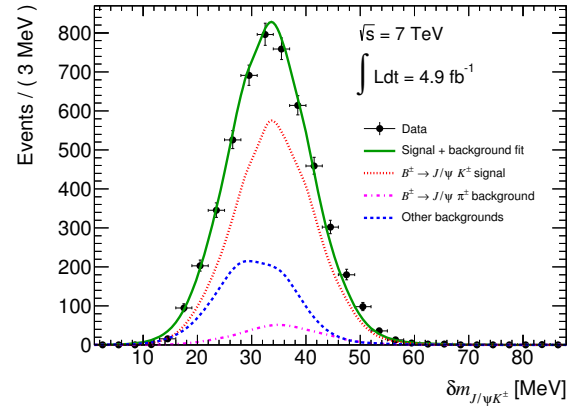
(b) Fit in dimension $\delta m_{J/\psi K^{\pm}}$
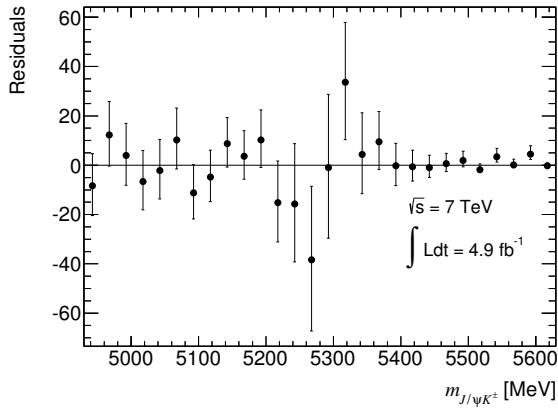
(c)
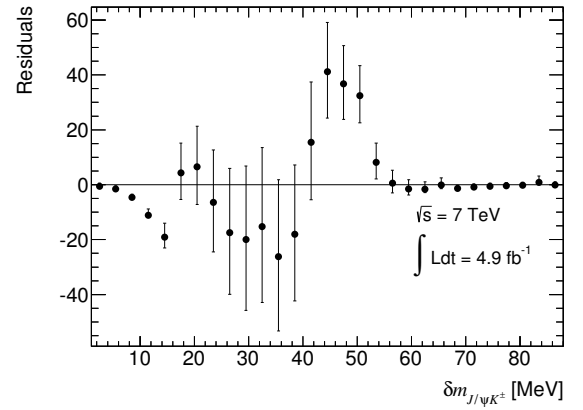
(d)

Figure C.2: The fit to the $B^{\pm}$ invariant mass spectrum (BT category).
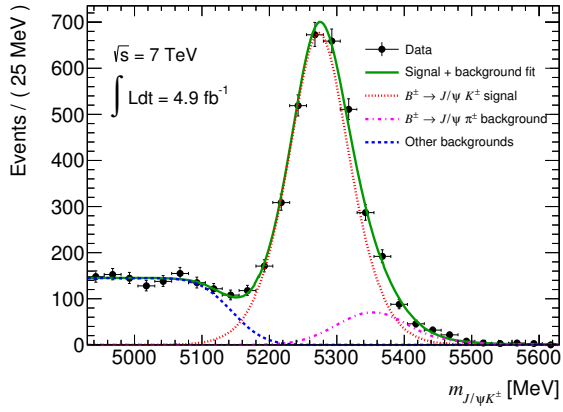
(a) Fit in dimension $m_{J/\psi K^\pm}$
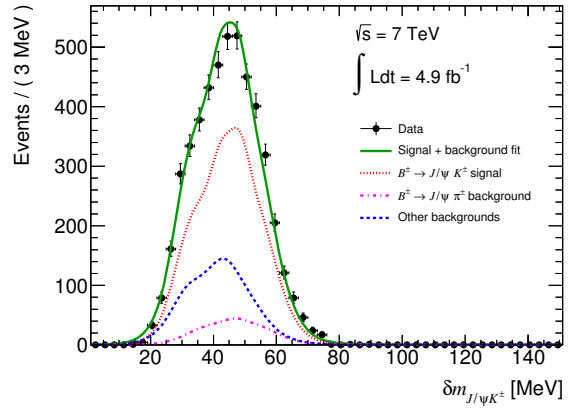
(b) Fit in dimension $\delta m_{J/\psi K^\pm}$
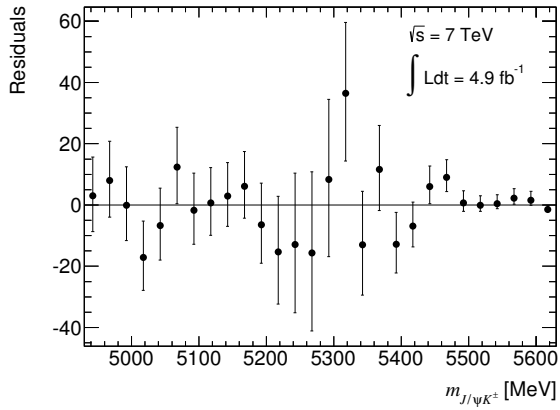
(c)
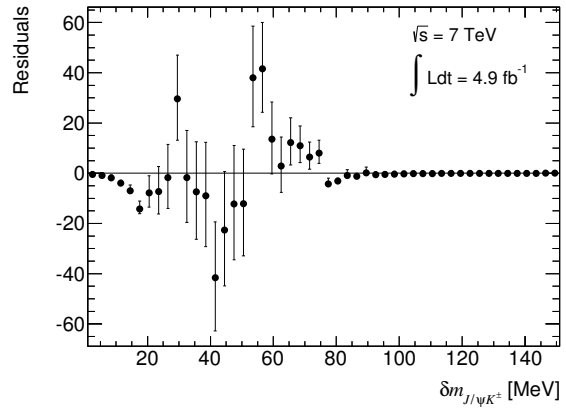
(d)

Figure C.3: The fit to the $B^\pm$ invariant mass spectrum (AE category).

		Yield	$\delta N_{B^\pm}^{\text{syst}}$	$\frac{\chi^2}{\text{NDF}}$
	Main fit	9 089		2.7
BB	Alt. model I	9 258	+169	6.2
	Alt. model II	9 239	+150	4.9
	Main fit	3 407		0.9
BT	Alt. model I	3 398	-9	2.4
	Alt. model II	3 564	+157	1.5
	Main fit	3 038		1.1
AE	Alt. model I	3 273	+236	1.2
	Alt. model II	3 257	+219	1.2

Table C.2: Estimation of the systematic uncertainty of $N_{B^\pm}$ (multi-bin case). Section 5.5.3.2 explains the computation.

List of Abbreviations

List of Figures

List of Tables

Bibliography

[1] ATLAS Collaboration, *Search for the decay $B_s^0 \to \mu^+\mu^-$ with the ATLAS detector*, Phys. Lett., *vol. B713*:387–407, 2012, doi:10.1016/j.physletb.2012.06.013. arXiv:1204.0735 [hep-ex].

[2] ATLAS Collaboration, *Limit on $B_s^0 \to \mu^+\mu^-$ branching fraction based on $4.9\,\mathrm{fb}^{-1}$ of integrated luminosity*, ATLAS-CONF-2013-076, CERN, Geneva, July 2013.

[3] L. R. Evans *et al.*, *LHC machine*, J. Instrum., *vol. 3*:S08 001, August 2008, doi:10.1088/1748-0221/3/08/S08001.

[4] ATLAS Collaboration, *ATLAS detector and physics performance*, *Technical Design Report*, vol. 1, CERN, Geneva, 1999.

[5] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, J. Instrum., *vol. 3*:S08 003, August 2008, doi:10.1088/1748-0221/3/08/S08003.

[6] CERN PhotoLab, *The four main LHC experiments*, June 1999. CERN photo number: CERN-AC-9906026.

[7] CERN PhotoLab, *Computer generated image of the ATLAS detector*, March 2008. CERN photo number: CERN-GE-0803012.

[8] ATLAS Collaboration, *Performance of the ATLAS muon trigger in 2011*, ATLAS-CONF-2012-099, CERN, Geneva, July 2012.

[9] ATLAS Collaboration, *Expected performance of the ATLAS experiment - detector, trigger and physics*, CERN, Geneva, 2008. arXiv:0901.0512 [hep-ex].

[10] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2010*, Eur. Phys. J., *vol. C72*:1849, 2012, doi:10.1140/epjc/s10052-011-1849-1. arXiv:1110.1530 [hep-ex].

[11] ATLAS Collaboration, *ATLAS luminosity public results*.
URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic

[12] O. S. Brüning *et al.*, *LHC luminosity and energy upgrade: a feasibility study*, LHC-Project-Report-626, CERN-LHC-Project-Report-626, CERN, Geneva, December 2002.

[13] R. Assmann *et al.*, *First thoughts on a Higher-Energy LHC*, CERN-ATS-2010-177, CERN, Geneva, August 2010.

[14] P. Vankov, *ATLAS upgrade for the HL-LHC: meeting the challenges of a five-fold increase in collision rate*, in *Proceedings of the Hadron Collider Physics Symposium 2011*, ATL-UPGRADE-PROC-2012-003, 1–3, Paris, 2012. arXiv:1201.5469 [physics.ins-det].

[15] R. Feynman, *Mathematical formulation of the quantum theory of electromagnetic interaction*, Phys. Rev., *vol. 80*:440–457, 1950, doi:10.1103/PhysRev.80.440.

[16] S. Glashow *et al.*, *Weak interactions with lepton-hadron symmetry*, Phys. Rev., *vol. D2*:1285–1292, 1970, doi:10.1103/PhysRevD.2.1285.

[17] S. Weinberg, *A model of leptons*, Phys. Rev. Lett., *vol. 19*:1264–1266, 1967, doi:10.1103/PhysRevLett.19.1264.

[18] A. Salam, *Weak and electromagnetic interactions*, in *Proceedings of the 8th Nobel Symposium 1968*, eConf C680519, 367–377, Stockholm, 1968.

[19] H. Fritzsch *et al.*, *Current algebra: quarks and what else?*, in *Proceedings of the 16th International Conference On High-Energy Physics*, eConf C720906V2, 135–165, Batavia, Illinois, 1972. arXiv:hep-ph/0208010 [hep-ph].

[20] *see, e.g.*, D. Griffiths, *Introduction to elementary particles*, 2nd ed., Wiley, 2008, *and references therein*.

[21] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett., *vol. 13*:508–509, 1964, doi:10.1103/PhysRevLett.13.508.

[22] F. Englert *et al.*, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett., *vol. 13*:321–323, 1964, doi:10.1103/PhysRevLett.13.321.

[23] G. Guralnik *et al.*, *Global conservation laws and massless particles*, Phys. Rev. Lett., *vol. 13*:585–587, 1964, doi:10.1103/PhysRevLett.13.585.

[24] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett., *vol. B716*:1–29, 2012, doi:10.1016/j.physletb.2012.08.020. arXiv:1207.7214 [hep-ex].

[25] CMS Collaboration, *Search for a Standard-Model-like Higgs boson with a mass of up to* 1 TeV *at the LHC*, CERN-PH-EP-2013-050, CERN, Geneva, March 2013. arXiv:1304.0213 [hep-ex].

[26] S. Glashow *et al.*, *Weak interactions with lepton-hadron symmetry*, Phys. Rev., *vol. D2*:1285–1292, 1970, doi:10.1103/PhysRevD.2.1285.

[27] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett., *vol. 10*:531–533, 1963, doi:10.1103/PhysRevLett.10.531.

[28] M. Kobayashi *et al.*, *CP violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys., *vol. 49*:652–657, 1973, doi:10.1143/PTP.49.652.

[29] Particle Data Group, *Review of particle physics*, Phys. Rev. D, *vol. 86*:010 001, July 2012, doi:10.1103/PhysRevD.86.010001.

[30] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa matrix*, Phys. Rev. Lett., *vol. 51*:1945, 1983, doi:10.1103/PhysRevLett.51.1945.

[31] A. Höcker *et al.*, *A new approach to a global fit of the CKM matrix*, Eur. Phys. J., *vol. C21*:225–259, 2001, doi:10.1007/s100520100729. arXiv:hep-ph/0104062 [hep-ph].

[32] CKMfitter Group, *CP violation and the CKM matrix: assessing the impact of the asymmetric B factories*, Eur. Phys. J., *vol. C41*:1–131, 2005, doi:10.1140/epjc/s2005-02169-1. arXiv:hep-ph/0406184 [hep-ph].

[33] A. J. Buras *et al.*, *Probing New Physics with the $B_s^0 \to \mu^+\mu^-$ time-dependent rate*, 2013. arXiv:1303.3820 [hep-ph].

[34] G. Buchalla *et al.*, *Weak decays beyond leading logarithms*, Rev. Mod. Phys., *vol. 68*:1125–1144, 1996, doi:10.1103/RevModPhys.68.1125. arXiv:hep-ph/9512380 [hep-ph].

[35] A. J. Buras *et al.*, *On the Standard Model prediction for $\mathscr{B}(B_{s,d}^0 \to \mu^+\mu^-)$*, Eur. Phys. J., *vol. C72*:2172, 2012, doi:10.1140/epjc/s10052-012-2172-1. arXiv:1208.0934 [hep-ph].

[36] G. Buchalla *et al.*, *The rare decays $K \to \pi\nu\bar{\nu}$, $B \to X\nu\bar{\nu}$ and $B \to l^+l^-$: an update*, Nucl. Phys., *vol. B548*:309–327, 1999, doi:10.1016/S0550-3213(99)00149-2. arXiv:hep-ph/9901288 [hep-ph].

[37] G. Buchalla *et al.*, *QCD corrections to rare K and B decays for arbitrary top quark mass*, Nucl. Phys., *vol. B400*:225–239, 1993, doi:10.1016/0550-3213(93)90405-E.

[38] Heavy Flavor Averaging Group, *Averages of b-hadron, c-hadron, and τ-lepton properties as of early 2012*, SLAC-R-1002, July 2012. arXiv:1207.1158 [hep-ex].

[39] HPQCD Collaboration, *B-meson decay constants from improved lattice NRQCD and physical u, d, s and c sea quarks*, 2013. arXiv:1302.2644 [hep-lat].

[40] LHCb Collaboration, *Precision measurement of the $B_s^0 - \bar{B}_s^0$ oscillation frequency with the decay $B_s^0 \to D_s^- \pi^+$*, CERN-PH-EP-2013-054, LHCB-PAPER-2013-006, CERN, Geneva, April 2013. arXiv:1304.4741 [hep-ex].

[41] G. Raven, *Measurement of the CP violation phase ϕ_s in the B_s^0 system at LHCb*, in *Proceedings of the CKM 2012, the 7th International Workshop on the CKM Unitarity Triangle*, eConf C120928, Cincinnati, 2012. arXiv:1212.4140 [hep-ex].

[42] K. De Bruyn *et al.*, *Branching ratio measurements of B_s^0 decays*, Phys. Rev., *vol. D86*:014 027, 2012, doi:10.1103/PhysRevD.86.014027. arXiv:1204.1735 [hep-ph].

[43] K. De Bruyn *et al.*, *Probing New Physics via the $B_s^0 \to \mu^+\mu^-$ effective lifetime*, Phys. Rev. Lett., *vol. 109*:041 801, 2012, doi:10.1103/PhysRevLett.109.041801. arXiv:1204.1737 [hep-ph].

[44] LHCb Collaboration, *First evidence for the decay $B_s^0 \to \mu^+\mu^-$*, Phys. Rev. Lett., *vol. 110*:021 801, 2013, doi:10.1103/PhysRevLett.110.021801. arXiv:1211.2674 [hep-ex].

[45] LHCb Collaboration, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B_d^0 \to \mu^+\mu^-$ decays at the LHCb experiment*, CERN-PH-EP-2013-128, LHCB-PAPER-2013-046, July 2013. arXiv:1307.5024 [hep-ex].

[46] W. J. Stirling, private communication.

[47] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7\,\mathrm{TeV}$ using the ATLAS detector at the LHC*, CERN-PH-EP-2013-026, CERN, Geneva, February 2013. arXiv:1302.4393 [hep-ex].

[48] CMS Collaboration, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B_d^0 \to \mu^+\mu^-$ with the CMS experiment*, CMS-BPH-13-004, CERN-PH-EP-2013-129, July 2013. arXiv:1307.5025 [hep-ex]. *Submitted to Phys. Rev. Lett.*

[49] LHCb Collaboration, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP, *vol. 1304*:001, 2013, doi:10.1007/JHEP04(2013)001. arXiv:1301.5286 [hep-ex].

[50] H.-J. Yang *et al.*, *Studies of boosted decision trees for MiniBooNE particle identification*, Nucl. Instrum. Meth., *vol. A555*:370–385, 2005, doi:10.1016/j.nima.2005.09.022. arXiv:physics/0508045 [physics].

[51] A. Hoecker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, in *Proceedings of the ACAT 2007, 11th International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, PoS 040, Amsterdam, 2007. arXiv:physics/0703039 [physics].

[52] G. Punzi, *Sensitivity of searches for new signals and its optimization*, in *Proceedings of the PHYSTAT 2003, Statistical problems in Particle Physics, Astrophysics, and Cosmology*, eConf C030908, MODT002, Menlo Park, 2003. arXiv:physics/0308063 [physics].

[53] ATLAS Collaboration, *Flavour tagged time dependent angular analysis of the $B_s^0 \to J/\psi\phi$ decay and extraction of $\Delta\Gamma$ and the weak phase ϕ_s in ATLAS*, ATLAS-CONF-2013-039, CERN, Geneva, April 2013.

[54] ATLAS Collaboration, *Observation of the $B^\pm$ meson in the decay $B^\pm \to J/\psi(\mu^+\mu^-)K^\pm$ in ATLAS*, ATLAS-CONF-2010-098, CERN, Geneva, November 2010.

[55] ATLAS Collaboration, *Measurement of the differential cross-section of B^+ meson production in pp collisions at $\sqrt{s} = 7\,\mathrm{TeV}$ at ATLAS*, CERN-PH-EP-2013-089, CERN, Geneva, June 2013. arXiv:1307.0126 [hep-ex]. *Submitted to JHEP.*

[56] T. Sjostrand *et al.*, Pythia 6.4 *physics and manual*, JHEP, *vol. 0605*:026, 2006, doi:10.1088/1126-6708/2006/05/026. arXiv:hep-ph/0603175 [hep-ph].

[57] ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, ATL-PHYS-PUB-2011-008, CERN, Geneva, April 2011.

[58] ATLAS Collaboration, *ATLAS tunes of* PYTHIA 6 *and* PYTHIA 8 *for MC11*, ATL-PHYS-PUB-2011-009, CERN, Geneva, July 2011.

[59] GEANT4 Collaboration, GEANT4 - *a simulation toolkit*, Nucl. Instrum. Meth., *vol. A506*:250–303, July 2003, doi:10.1016/S0168-9002(03)01368-8.

[60] E. Richter-Was *et al.*, ATLFAST-II *a fast simulation package for ATLAS*, ATL-PHYS-98-131, CERN-ATL-PHYS-98-131, CERN, Geneva, November 1998.

[61] A. Salzburger, *The ATLAS track extrapolation package*, ATL-SOFT-PUB-2007-005, CERN, Geneva, June 2007.

[62] S. Hassani *et al.*, *A muon identification and combined reconstruction procedure for the ATLAS detector at the LHC using the (MUONBOY, STACO, MuTag) reconstruction packages*, Nucl. Instrum. Meth., *vol. A572*:77–79, 2007, doi:10.1016/j.nima.2006.10.340.

[63] T. Cornelissen *et al.*, *Concepts, design and implementation of the ATLAS New Tracking (NEWT)*, ATL-SOFT-PUB-2007-007, CERN, Geneva, March 2007.

[64] V. Kostyukhin, *VKalVrt - package for vertex reconstruction in ATLAS.*, ATL-PHYS-2003-031, CERN-ATL-PHYS-2003-031, CERN, Geneva, August 2003.

[65] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment*, ATLAS-CONF-2010-069, CERN, Geneva, June 2010.

[66] Muon Combined Performance Group - ATLAS Collaboration, private communication.

[67] CMS Collaboration, *Search for $B_s^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow \mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. ett., *vol. 107*:191 802, 2011, doi:10.1103/PhysRevLett.107.191802. arXiv:1107.5834 [hep-ex].

[68] ATLAS Rare *b*-Decays Subgroup, private communication. ATL-COM-PHYS-2013-053.

[69] P. Jussel, private communication.

[70] V. Sipica, private communication.

[71] K. S. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun., *vol. 136*:198–207, 2001, doi:10.1016/S0010-4655(00)00243-5. arXiv:hep-ex/0011057 [hep-ex].

[72] I. Ibragimov, private communication.

[73] W. Verkerke *et al.*, *The RooFit toolkit for data modeling*, in *Proceedings of the CHEP 2003, Conference for Computing in High Energy and Nuclear Physics*, eConf C0303241, MOLT007, La Jolla, 2003. arXiv:physics/0306116 [physics].

[74] F. James *et al.*, *Minuit: a system for function minimization and analysis of the parameter errors and correlations*, Comput. Phys. Commun., *vol. 10*:343–367, July 1975, doi:10.1016/0010-4655(75)90039-9.

[75] M. Oreglia, *A study of the reactions $\psi' \to \gamma\gamma\psi$*, Ph.D. thesis, Stanford Univ., Stanford, December 1980.

[76] J. Gaiser, *Charmonium spectroscopy from radiative decays of the J/ψ and ψ'*, Ph.D. thesis, Stanford Univ., Stanford, August 1982.

[77] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Υ' and Υ resonances*, Ph.D. thesis, DESY, Hamburg, April 1986.

[78] G. Punzi, *Comments on likelihood fits with variable resolution*, in *Proceedings of the PHYSTAT 2003, Statistical problems in Particle Physics, Astrophysics, and Cosmology*, eConf C030908, WELT002, Menlo Park, 2003. arXiv:physics/0401045 [physics].

[79] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys., *vol. G28*:2693–2704, 2002, doi:10.1088/0954-3899/28/10/313.

Acknowledgements

I want to thank my family for their emotional support. They were very supportive and understanding all this time. If I had not received their constant encouragement, it would have been indeed difficult to reach my goals.

I would like to wish my colleagues Oliver Rosenthal and Uwe Fröhlich, who are soon to face their disputations, all the best. They have been very helpful to rescue me from technical issues, and to keep me updated about the procedures concerning dissertation submission.

I owe it to my friends and colleagues who have helped me in many ways. I am grateful especially for their emotional support. I bow in respect to all my teachers without whom I would not have reached here.

It is embarrassing to have missed someone in a word of thanks, such a list is always incomplete, because there are indeed a lot of people behind one's achievement, and I grateful to all of them.