Prospects for Measuring the Branching Ratio of the Rare $B_s^0 \to \mu^+\mu^-$ Decay with the ATLAS Experiment

DISSERTATION
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften

vorgelegt von
M.Sc. Valentin Sipica
geb. am 03.06.1979 in Bukarest

eingereicht bei der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen

September 2011

Gutachter der Dissertation: Prof. Dr. P. Buchholz
Prof. Dr. I. Fleck

Abstract

The Large Hadron Collider (LHC) located at the CERN laboratory in Geneva provides $p-p$ collisions at a centre-of-mass energy of $\sqrt{s} = 7$ TeV. The study of the rare $B_s^0 \to \mu^+\mu^-$ decay is among the research topics of ATLAS, one of the main experiments at the LHC. This decay is highly suppressed in the Standard Model of particle physics and may give an indirect evidence for New Physics models.

This PhD thesis investigates prospects for measuring the branching ratio of the $B_s^0 \to \mu^+\mu^-$ decay with the ATLAS experiment. The analysis is based on Monte Carlo data, with $p-p$ collisions generated at a centre-of-mass energy of $\sqrt{s} = 10$ TeV.

The strategy employed is to calculate the $B_s^0 \to \mu^+\mu^-$ branching ratio relative to the branching ratio of the $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decay. The dominant background channel is the $b\bar{b} \to \mu^+\mu^- X$ combinatorial background. True $B_s^0 \to \mu^+\mu^-$ decay candidates are separated from the much larger amount of combinatorial background events using several discriminating quantities.

Upper limits on the $B_s^0 \to \mu^+\mu^-$ branching ratio are computed using a Bayesian and a frequentist method. The expected precision of the branching ratio measurement is estimated for different values of the integrated luminosity. An expected upper limit on the branching ratio is computed to $BR(B_s^0 \to \mu^+\mu^-) < 3.8 \times 10^{-8}$ at a 95% confidence level for 1 fb^{-1}. The precision of the ATLAS measurement of the branching ratio will reach a level compatible with the best current measurements with about $2-5$ fb^{-1} of data.

Zusammenfassung

Der Large Hadron Collider (LHC) am CERN in Genf produziert $p-p$ Kollisionen mit einer Schwerpunktsenergie von $\sqrt{s} = 7$ TeV. Das Studium des seltenen Zerfalls $B_s^0 \to \mu^+\mu^-$ ist eines der Forschungsthemen bei ATLAS, einem der Hauptexperimente am LHC. Dieser Zerfall ist stark im Standardmodell der Teilchenphysik unterdrückt und kann auf Neue-Physik-Modelle hinweisen.

Diese Dissertation erforscht die Möglichkeiten, die Zerfallsrate $B_s^0 \to \mu^+\mu^-$ mit dem ATLAS-Experiment zu messen. Die Analyse ist auf Monte-Carlo-Daten basiert, mit einer Schwerpunktsenergie der $p-p$ Kollisionen von $\sqrt{s} = 10$ TeV.

Die verwendete Strategie ist, die $B_s^0 \to \mu^+\mu^-$ Zerfallsrate relativ zu der Zerfallsrate $B^+ \to J/\psi(\mu^+\mu^-)K^+$ zu messen. Der Hauptuntergrund ist $b\bar{b} \to \mu^+\mu^- X$. Die wahren $B_s^0 \to \mu^+\mu^-$ Zerfälle sind von dem viel grösseren Beitrag des kombinatorischen Untergrunds mittels mehrerer Unterscheidungsvariablen zu trennen.

Obergrenzen der $B_s^0 \to \mu^+\mu^-$ Zerfallsrate werden mit einer Bayesischen und einer frequentistischen Methode berechnet. Die erwartete Genauigkeit der Messung der Zerfallsrate wird für verschiedene Werte der integrierten Luminosität bestimmt. Die erwartete Obergrenze der Zerfallsrate wird zu $BR(B_s^0 \to \mu^+\mu^-) < 3.8 \times 10^{-8}$ (95% Konfidenzintervall) für eine Luminosität von 1 fb^{-1} bestimmt. Die Genauigkeit der ATLAS-Messung der Zerfallsrate wird mit den besten aktuellen Messungen bei einer Datenmenge entspechend einer Luminosität von $2-5$ fb^{-1} kompatibel werden.

Contents

Chapter 1

Introduction

Understanding the properties of elementary particles and their interactions has been at the boundary of human knowledge for about the last hundred years and is the key to understand our universe in detail. The most successful theory employed to describe these phenomena is the Standard Model (SM) of particle physics (introduced in Chapter 2). The SM describes three of the observed fundamental interactions (the strong, the weak and the electromagnetic interaction) by a single internally consistent theory. Its success stems from the experimental verification of most of its predictions. The notable exception is the Higgs boson, which is yet to be observed. Nevertheless, the SM is not a complete model, as it does not incorporate the fourth fundamental interaction, gravity. Other observations describe phenomena, which cannot be explained in the SM. For example, cosmological observations report the existence of dark matter (i.e., matter which does not interact with electromagnetic radiation, and hence is not visible, but interacts gravitationally with the surrounding visible matter). Such problems motivate the need to formulate and test new theories, which are able to explain New Physics (NP) phenomena.

One popular theory extending the SM, with the potential to solve the above mentioned issues, is supersymmetry (SUSY). SUSY postulates, that each elementary SM particle has a corresponding partner particle (i.e., a superpartner). Since these superpartners have not been observed in previous experiments, if they exist, they have to be massive. In order to observe them, high energy experiments have to be performed. The Large Hadron Collider (LHC) (introduced in Chapter 3) is the accelerator producing collisions with the highest centre-of-mass energy worldwide (currently $\sqrt{s} = 7$ TeV, planned to be increased to $\sqrt{s} = 14$ TeV). Several experiments located at the LHC are currently investigating SUSY effects. However, a direct observation of SUSY superpartners will not be possible, if their mass scale is above the LHC centre-of-mass energy.

An alternative, indirect method to probe NP phenomena is to investigate processes, which are sensitive to extensions of the SM. Such processes are the di-muon decays $B_d^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ (introduced in Chapter 2). These decays are highly suppressed in the SM. The SM predictions of their branching ratios are $(1.0 \pm 0.1) \times 10^{-10}$ and $(3.2 \pm 0.2) \times 10^{-9}$ [Bur10], respectively. However, the branching ratios of these decays are sensitive to several extensions of the SM. An

increase of several orders or magnitude is predicted by some NP models, or a significant decrease is expected in other models. Therefore, any deviation of the observed branching ratios from the values expected in the SM will give an indirect evidence of NP effects. Moreover, the branching ratio values can be used to constrain the parameter space, which describes the NP models.

The study of di-muon decays of B_d^0 and B_s^0 mesons is one of the main research topics of the ATLAS experiment. ATLAS (detailed in Chapter 3) is one of the main experiments operating at the LHC, at the CERN laboratory. The expected $b\bar{b}$ production cross-section is about 500 μb and, therefore, it is expected that enough B_d^0 andB_s^0 mesons will be produced, such that the measured branching ratios of the $B_d^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays are compared to the SM expectation with statistical significance.

This thesis is focused on the study of the $B_s^0 \to \mu^+\mu^-$ decay using Monte Carlo (MC) simulated data. The aim is to prepare in detail the analysis for this measurement and to prospect the ATLAS sensitivity to the $B_s^0 \to \mu^+\mu^-$ branching ratio. An overview of the strategy employed for this measurement is given in Chapter 4. The $B_s^0 \to \mu^+\mu^-$ branching ratio is computed with respect to the branching ratio of another decay channel ($B^+ \to J/\psi(\mu^+\mu^-)K^+$), which is measured with an increased statistical accuracy. This approach requires a correction of the observed number of $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decay candidates with their corresponding acceptances, trigger efficiencies and selection efficiencies.

Several MC data sets are used in this analysis (described in Chapter 5). Independent data samples are produced corresponding to the signal channel ($B_s^0 \to \mu^+\mu^-$), to the normalisation channel ($B^+ \to J/\psi K^+$) and to the dominating background channel ($b\bar{b} \to \mu^+\mu^- X$). All samples are fully reconstructed with the ATLAS reconstruction software, so they have a format identical to that of recorded data. Therefore, the analysis presented in this thesis can also be performed, with slight modifications, with recorded data.

Chapter 6 presents a method to compute the acceptances of the signal and normalisation channels based on the production cross-sections computed by software tools used in data generation. The effect of the trigger selection on the signal and normalisation channels is analysed in Chapter 7 and the trigger efficiencies are estimated.

The offline analysis is presented in detail in Chapter 8. There are three topics of interest. Firstly, the candidates for the $B_s^0 \to \mu^+\mu^-$ and the $B^+ \to J/\psi K^+$ decays have to be reconstructed from the available detector data. This reconstruction is implemented in two separate offline algorithms (one for each channel). Secondly, true $B_s^0 \to \mu^+\mu^-$ decays have to be separated from the much larger (by several orders of magnitude) amount of background events. This is achieved by applying cuts on several selection quantities designed for a high rejection of background events and a high retention of true signal decays. After the optimisation of these selection criteria (see Chapter 9) the expected number of background events is diminished to a level compatible to the number of signal events. Lastly, the selection efficiencies for the signal and normalisation channels are computed.

The last step of this analysis is to estimate upper limits on the $B_s^0 \to \mu^+\mu^-$ bran-

ching ratio, described in Chapter 9. Two techniques are used in the upper limit calculation, a Bayesian approach and a frequentist approach. The expected ATLAS sensitivity to the $B_s^0 \to \mu^+\mu^-$ branching ratio is evaluated as a function of the integrated luminosity. A comparison with the current experimental measurements is given. Finally, concluding remarks are given in Chapter 10.

Chapter 2

Rare Decays of B-Mesons in Theory and Experiment

This chapter motivates the study of rare decays of B-mesons from a theoretical and an experimental point of view. Section 2.1 introduces the Standard Model of particle physics. Section 2.2 describes these decays as predicted by the Standard Model. Several other theoretical models, which extend the SM, are mentioned in Section 2.3. Such models are expected to affect the branching ratios of these decays, making them candidates for searches of New Physics phenomena. Finally, the current experimental observations are presented in Section 2.4.

2.1 Standard Model of Particle Physics

The Standard Model (SM) of particle physics is a quantum field theory described by the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where the $SU(3)_C$ group describes the colour charge, the $SU(2)_L$ group describes the weak isospin and the $U(1)_Y$ describes the weak hypercharge.

The building blocks of matter are represented in the SM by 12 elementary particles with spin 1/2 (fermions) and their corresponding antiparticles (particles with the same mass and opposite electric charge). There are six quarks: up (u), down (d), charm (c), strange (s), bottom (b) and top (t). These are grouped in three families, such that each family contains a quark with charge $2/3\,e$ (u, c, t) and one quark with charge $-1/3\,e$ (d, s, b), with e being the elementary charge. There are six leptons: $e, \nu_e, \mu, \nu_\mu, \tau, \nu_\tau$. These are also grouped in three families, such that each family contains a lepton of charge $-1\,e$ (e, μ, τ) and a lepton of charge 0 (ν_e, ν_μ, ν_τ).

In the SM, the interactions between fermions are mediated by field-carrying spin 1 particles (bosons). The strong interaction is mediated by gluons and manifests itself only between quarks. It is described by quantum chromodynamics (QCD) [Gre07], a non-Abelian gauge theory with a symmetry group $SU(3)$. It introduces three colour fields, defined by three corresponding quantum states (i.e, colours). Each gluon is defined as a singlet of two color states, in total 8 types of gluons. The potential between quarks increases linearly with the distance between

the quarks, due to the nature of the colour charge field carried by gluons. Therefore, quarks are bound to each other forming colour neutral hadrons. Hadrons are composite particles formed by either two valence quarks (mesons) or three valence quarks (baryons). The valence quarks define the quantum numbers of hadrons. Virtual $q\bar{q}$ pairs are also contained in hadrons (the quark sea), but they do not contribute to the hadrons' quantum numbers.

The electromagnetic force manifests itself between particles with an electric charge. Interactions between charged particles are described by quantum electrodynamics (QED) [Gre02], an Abelian gauge theory with a symmetry group $U(1)$. The field carrier is the photon, which is chargeless and massless.

Weak interactions appear between all fermions (leptons and quarks). It is mediated by three bosons, $W^{\pm}$ and Z^0. Unlike gluons and photons, these bosons are massive, with masses of $m_W \simeq 80$ GeV and $m_Z \simeq 91$ GeV. Fermions can be either left-handed, or right-handed (V-A theory). The $W^{\pm}$ couple only to left-handed particles and right-handed antiparticles and have charges of $+1\,e$ and $-1\,e$. The Z^0 couples to left-handed particles as well as left-handed antiparticles and is chargeless.

The theory of weak interactions introduces the weak isospin (I), which creates a symmetry group $SU(2)$. The terminology usually denotes its third component (I_3) as the weak isospin. It relates the electric charge (q) to the weak hypercharge (Y_W):

$$Q = I_3 + \frac{Y_W}{2}. \tag{2.1}$$

The electromagnetic and weak interactions are thus united into the electroweak interaction, described by a $SU(2)_L \otimes U(1)_Y$ symmetry group.

The weak isospin (I_3) is $+1/2$ for right-handed fermions, $-1/2$ for left-handed fermions and +1 (-1) for W^+ (W^-). I_3 is conserved by weak interactions. Therefore, in order to describe charged weak interactions (e.g., $l^- \to W^-\nu_l$ or $q \to Wq'$, with q any of the up-type quarks and q' any of the down-type quarks), fermions are grouped in doublets with opposite weak isospins (either $+1/2$ or -1/2). This is particularly important for describing weak interactions between quarks. Up-type quarks (u, c, t) always decay weakly into down-type quarks (d, s, b), and vice-versa (i.e., weak decay is only possible in quarks of opposite weak isospin, but forbidden to quarks of the same isospin). This was first formulated by Cabibbo [Cab63] in order to describe the different amplitudes observed in $s \to uW^-$ (strangeness violating) and $d \to uW^-$ (strangeness conserving) transitions by assigning factors of $\sin\theta_C$ and $\cos\theta_C$ (with θ_C the Cabibbo angle) to the amplitude. This picture was later extended by Glashow, Iliopoulos and Maiani (GIM) [Gla70], when studying the flavour changing neutral current (FCNC) $K^0 \to \mu^+\mu^-$ decay. The GIM mechanism suggested the existence of the c-quark (only the u, d, s-quarks were known at the time). The decay is forbidden at tree level in the SM and it is only possible in loop diagrams. The transitions between different flavours of quarks are described by the orthogonal weak eigenstates d' and s':

$$d' = d\cos\theta_C + s\sin\theta_C, s' = -d\sin\theta_C + s\cos\theta_C. \tag{2.2}$$

This was later extended by Kobayashi and Maskawa [Kob73] by adding a third family of quarks (t, b). The transitions between all flavours of quarks are described by the Cabibbo, Kobayashi, Maskawa (CKM) matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.3}$$

The probability that a quark of flavour i decays weakly into a quark of flavour j is proportional to $|V_{ij}|^2$. The CKM matrix has to be unitary.

An additional scalar massive boson (Higgs) is included in the SM. It is the force carrier boson of the Higgs field postulated in the Higgs mechanism [Hig64], [Eng64], [Gur64]. The electroweak gauge symmetry is spontaneously broken in the Higgs mechanism. This mechanism explains how the SM particles acquire mass by interacting with the Higgs field. The Higgs boson has yet to be discovered experimentally and, therefore, its mass is not known. The search for the Higgs boson is one of the main research topics at two of the main LHC experiments.

Although most SM predictions are proven correct by experiments, the SM is not a complete model. One of the most important items missing is gravitation as an interaction, which is not included in the theory. Some observations cannot be explained by the Standard Model. For example, the observed difference between matter and antimatter is larger than what the SM predicts. The hierarchy problem (different masses of particles and different strengths of the interaction forces) is also not predicted. Therefore, several theories are proposed, that extend the SM in order to incorporate New Physics (NP) phenomena. Some of them are being investigated either directly or indirectly at the LHC experiments. An indirect method is the study of the rare decays, e.g., the $B^0_s \to \mu^+\mu^-$ decay, which are expected to be sensitive to most of these models.

2.2 $B^0_{s(d)} \to \mu^+\mu^-$ in the Standard Model

The $B^0_{s(d)} \to l^+l^-$ decays [Ber], [Art08] represent a particular class of B^0_s and B^0_d decays. These decays are highly suppressed in the SM. At the vertex level, $b \to s(d)$ transitions are FCNCs, which are forbidden as tree decays by the GIM mechanism. An additional internal quark of the up-type is required in the transition. The contributions of transitions involving a c- or a u-quark are suppressed by the small values of the corresponding V_{cb} and V_{ub} CKM matrix elements. These decays are possible in loop diagrams, which are electroweak-loop suppressed. The simplest loop diagrams are the penguin and the box diagrams shown in Fig. 2.1. The contribution of the box diagram is suppressed with respect to the penguin diagram by an additional factor of m_W^2/m_t^2, with m_W, the mass of the $W^\pm$ boson and m_t, the mass of the t-quark. In addition, the decays shown in both diagrams are helicity suppressed by a factor of m_l^2/m_B^2, with m_l the final state lepton mass and m_B the B meson mass.

The amplitude of these decays is described in the perturbation theory using the effective Hamiltonian in the form of the Wilson expansion [Buc96]:

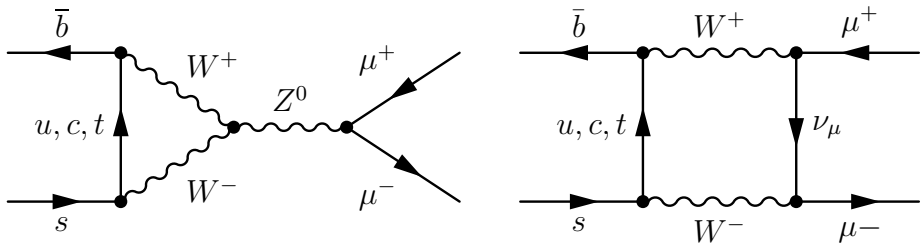


Figure 2.1: Penguin (left) and box (right) diagrams of the $B_s^0 \to \mu^+\mu^-$ decay.

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{(2)}} V_{tb}^* V_{tq} \sum_i C_i(\mu) O_i(\mu), \tag{2.4}$$

where G_F is the weak coupling constant, V_{tb} and V_{tq} are the corresponding CKM matrix elements with q being the s- or the d-quark for B_s^0 and B_d^0, respectively. C_i are the Wilson coefficients, O_i are local operators and μ is the renormalisation scale, which is approximately equal to the b-quark mass ($\mu \simeq 5$ GeV). In the simplest approach, there are 10 Wilson coefficients. The dominating operator in the $B_s(d)^0 \to l^+l^-$ diagrams is the semi-leptonic Q_{10} operator (the C_{10} Wilson coefficient). In addition, contributions of the hypothetical Higgs boson (C_S coefficient) or Goldstone boson (C_P coefficient) can appear. These two contributions are, however, suppressed by a factor m_b^2/m_W^2 and are negligible in the SM, but may be significant in other models. Thus, (2.4) becomes:

$$\begin{aligned} \mathcal{H}_{eff} &\simeq -\frac{4G_F}{\sqrt{2}} V_{tb}^* V_{tq} [C_{10}O_{10} + C_S O_S + C_P O_P] \\ &\simeq -\frac{4G_F}{\sqrt{2}} V_{tb}^* V_{tq} C_{10} O_{10}. \end{aligned} \tag{2.5}$$

The branching ratio of $B_q^0 \to l^+l^-$ decays is derived to:

$$BR(B_q \to l^+l^-) = \frac{G_F^2 \alpha^2 \tau_{B_q} f_{B_q}^2}{64\pi^3} |V_{tb}^* V_{tq}|^2 4m_l^2 \sqrt{m_{B_q}^2 - 4m_l^2} C_{10}^2, \tag{2.6}$$

where α is the fine structure constant, τ_{B_q} is the B_q meson decay life and f_{B_q} is the B_q meson decay constant. The C_{10} Wilson coefficient is estimated to -4.2 in the SM [Ber]. For two muons in the final state, the branching ratios of the two decays are computed as [Bur10]:

$$BR(B_s \to \mu^+\mu^-) = (3.2 \pm 0.2) \times 10^{-9} \tag{2.7}$$

and

$$BR(B_d \to \mu^+\mu^-) = (1.0 \pm 0.1) \times 10^{-10}. \tag{2.8}$$

2.3 $B_s^0 \to \mu^+\mu^-$ in New Physics Models

Extensions of the SM may create additional contributions to the effective Hamiltonian (e.g., the scalar and pseudoscalar field contributions corresponding to the C_S and C_P coefficients in (2.5)). Several New Physics (NP) models affect the $B_s^0 \to \mu^+\mu^-$ branching ratio. Therefore, studying the $B_s^0 \to \mu^+\mu^-$ decay can give an indirect evidence of new phenomena. If the observed branching ratio is significantly different from the value expected in the SM, it could be used to constrain the parameter space of NP models.

Multiple Higgs Doublets Models

Most NP models extend the Higgs sector. In the simplest approach, a model with two Higgs doublets is considered (2HDM) [Ber], [Art08]. Such a model would allow FCNCs at tree level, with a significant enhancement of $B - \bar{B}$ mixing. Since this is not observed experimentally, such a model has to exclude transitions at tree level. This is avoided by imposing a discrete symmetry between the Higgs and the fermion fields. The symmetry is achieved by requiring that up-type quarks couple only to one Higgs doublet, H_u, and down-type quarks only to the other Higgs doublet, H_d (type II 2HDM). A specific case is to allow coupling at the vertex level as defined in type II 2HDM, but allow the coupling of any Higgs doublet to any quark type at the one-loop level (type III 2HDM).

The Minimal Supersymmetric Standard Model (MSSM) is the minimal extension (i.e., with the smallest number of new particles) of the SM, which predicts that each SM particle has a heavier superpartner. The Higgs sector contains two doublets, described by the type III 2HDM formalism. The Higgs field contains five Higgs eigenstates: one pseudoscalar CP-odd neutral boson, A^0, two CP-even neutral bosons, h_0 and H_0 and two charged bosons, $H^\pm$. The masses of the five Higgs bosons are described in terms of $\tan\beta$, the ratio of the vacuum expectation values of the two Higgs doublets, and m_{A^0}, the mass of the CP-odd Higgs boson. The scalar and pseudoscalar Wilson coefficients C_S and C_P in (2.5) are proportional to $\tan^3\beta$ and are non-negligible. They are larger by a factor of $(m_b^2/m_W^2)\tan^2\beta$ than the C_{10} coefficient and at large values of $\tan\beta$ their contribution will be dominant. This effectively results in:

$$BR(B_q \to l^+l^-) \propto \frac{m_b^2 m_l^2}{m_{A^0}^2}\tan^6\beta, \tag{2.9}$$

which produces an enhancement of up to three orders of magnitude with respect to the value expected in the SM. The most recent experimental measurements (see Section 2.4) show that the observed value of the branching ratio does not exceed the SM expectation by a factor of more than about three. This observation confines already the parameter space in MSSM.

In type II 2HDM, the $W^\pm$ bosons can be exchanged with $H^\pm$ bosons in the penguin or box diagrams of the decays (e.g., in Fig. 2.1). This is reflected in additional contributions to the C_S and C_P Wilson coefficients, which are computed as:

$$C_S = C_P = \frac{m_l}{2m_W^2} \tan^2\beta \frac{\ln r}{r-1}, \tag{2.10}$$

where $r = m_{H^\pm}^2/m_t^2$. This results in a branching ratio $BR(B_q \to l^+l^-)$ formula depending on $\tan^4\beta$ and $m_{H^\pm}$. This dependence may shift the value of the branching ratio to a higher value than that expected in the SM (if $\tan\beta/m_{H^\pm}$ is very large) or even to a lower value (e.g., for $\tan\beta \simeq 60$ and $m_{H^\pm} \simeq 500$ GeV a decrease of about 50% is observed).

mSUGRA

Another theoretical model relevant for the study of the $B_s^0 \to \mu^+\mu^-$ decay is mSUGRA, the minimum scenario of the supergravity (SUGRA) theory [Ber], [Art08]. This theory includes general relativity by defining gravity as a field, whose force carrier boson is the graviton.

mSUGRA offers an explanation for the measured value of the anomalous muon magnetic moment $a_\mu = (g-2)/2$, which deviates from the value expected in the SM with a statistical significance of $3.4\,\sigma$ [Hag07]. This difference depends on $\tan\beta$ and on the universal scalar and gaugino masses M_0 and $M_{1/2}$:

$$\delta a_\mu = a_\mu^{exp} - a_\mu^{SM} \propto tan\beta \frac{f(M_0)}{M_{1/2}}. \tag{2.11}$$

A similar behaviour (increasing with $\tan\beta$ and decreasing with increasing $M_{1/2}$) is expected for the $B_s^0 \to \mu^+\mu^-$ branching ratio:

$$BR(B_s^0 \to \mu^+\mu-) \propto tan^6\beta \frac{M_{1/2}^2}{(M_{1/2}^2 + M_0^2)^3}. \tag{2.12}$$

Therefore, a large enhancement is expected (up to one order of magnitude) for large values of $\tan\beta$.

2.4 Experimental Evidence

The $B_s^0 \to \mu^+\mu^-$decay is currently investigated by several experiments. The statistically most significant results are obtained by the experiments located at the Tevatron and the LHC colliders. Tevatron produces $p-\bar{p}$ collisions at a centre-of-mass energy of $\sqrt{s} = 1.96$ TeV. The LHC produces $p-p$ collisions at $\sqrt{s} = 7$ TeV. All experiments based at these colliders benefit from a large amount of events, due to the large $b\bar{b}$ production cross-section.

The most significant result from the Tevatron is obtained by the CDF experiment. An excess of events is observed in $7\,\text{fb}^{-1}$ of recorded data [Aal11]. The probability that this excess is produced only by background events is estimated to 0.27%. The branching ratio is determined to $BR(B_s^0 \to \mu^+\mu-) = (1.8^{+1.1}_{-0.9}) \times 10^{-8}$ and an upper limit is computed to $BR(B_s^0 \to \mu^+\mu-) < 4.0 \times 10^{-8}$ at a confidence level of 95%. However, the excess observed at CDF is not confirmed by LHC experiments

with an amount of data of slightly higher statistical significance. DØ, the other Tevatron experiment, reports an upper limit of $BR(B_s^0 \to \mu^+\mu-) < 5.1 \times 10^{-8}$ at a 95% confidence level with 6.1 fb^{-1} recorded data [Aba10]. DØ does not observe the excess of events as announced by CDF.

The current best measurements are reported by the LHC experiments. An upper limit on the $B_s^0 \to \mu^+\mu^-$ branching ratio is computed at the LHCb experiment to $BR(B_s^0 \to \mu^+\mu-) < 1.3 \times 10^{-8}$ at a 95% confidence level with 300 pb^{-1} recorded data [Ser11]. The upper limit estimated by the CMS experiment is $BR(B_s^0 \to \mu^+\mu-) < 1.9 \times 10^{-8}$ at a 95% confidence level with 1.14 fb^{-1} [Cha11]. A combined analysis using the data collected at both the LHCb and the CMS experiments results in an improved upper limit of $BR(B_s^0 \to \mu^+\mu-) < 1.08 \times 10^{-8}$ at a 95% confidence level [LHC11]. The excess of events observed by CDF does not appear in either the LHCb or the CMS recorded data.

ATLAS, one of the other major experiments at the LHC, is also investigating the $B_s^0 \to \mu^+\mu^-$ decay and the analysis of the data recorded so far is under preparation. This thesis describes the analysis of this decay and gives prospects for measuring the $B_s^0 \to \mu^+\mu^-$ branching ratio, based on Monte Carlo data with a slightly higher centre-of-mass energy ($\sqrt{s} = 10$ TeV) than that of current LHC collisions. The estimates of the branching ratio are given in Chapter 9 and the expected ATLAS sensitivity is compared with the current measurements.

Chapter 3

The ATLAS Experiment

This chapter gives an overview of the ATLAS (A Toroidal LHC ApparatuS) experiment [Aad08], [Aad09]. In Section 3.1, the Large Hadron Collider (LHC) is introduced. The ATLAS detector is described in detail in Section 3.2 and the luminosity determination in Section 3.3. The ATLAS trigger system is explained in Section 3.4. Finally, the computing infrastructure is described in Section 3.5.

3.1 The Large Hadron Collider

The ATLAS detector is one of the main experiments located at the LHC [Eva08], at the CERN laboratory in Geneva, Switzerland. The particle accelerator LHC is designed as an almost circular storage ring with a circumference of about 26.7 km. It is composed of eight straight sections, where charged particles are accelerated in a system of superconducting cavities, and eight circular regions, where the particle beams are forced on a circular trajectory by superconducting dipole magnets. The LHC contains two rings, each accelerating charged particle beams in opposite directions. The two beams are then collided almost frontally at four different points along the LHC trajectory, where four detectors are located. The accelerated particles are either protons or lead ions, each being used in separate fills. Most physics studies at ATLAS are performed with proton collision fills (including the study of the $B_s^0 \to \mu^+\mu^-$ decay). As of 2010, protons are accelerated to an energy of 3.5 GeV each, such that each collision occurs at a centre-of-mass energy of $\sqrt{s} =$ 7 GeV. This represents half of the design energy. The beams are split in trains of bunches (currently 1380 bunches), each containing about 10^{11} protons, with a total of about 10^{14} protons per beam. The time gap between consecutive bunches is at least 50 ns. The intensity of the beams has been gradually increased since the first collisions by increasing the number of bunches in the LHC (equivalent to decreasing the time gap between consecutive bunches). The current instantaneous luminosity is of the order of 10^{33} cm^{-2}s^{-1}. According to plans, the instantaneous luminosity will be further increased up to the design value of 10^{34} cm^{-2}s^{-1}. This will be achieved by decreasing the shortest time gap between bunches to 25 ns and by increasing the number of protons per bunch.

Other large experiments located at the LHC are CMS [Ado08], a general

purpose detector with a similar physics programme as ATLAS, LHCb [Alv08], a forward spectrometer designed with the aim of studying processes involving B hadrons, and ALICE [Aam08], a detector designed to investigate the quark-gluon plasma produced in ion-ion collisions.

3.2 The ATLAS Detector

The ATLAS detector [Aad08] is the largest of the four main detectors installed at the LHC, with a length of about 44 m and a height of about 25 m. The overall design is characterised by a barrel like shape in the central region (parallel to the beam axis) and a disk-like shape at each side of the barrel region (perpendicular to the beam axis), such that collisions take place at its centre. The detector has a multi-layer design, with different sub-components installed at increasing distances from the center, each performing a specific task. Three main components are distinguished. The Inner Detector (ID) is the innermost component, whose most important role is detecting hit positions of charged particles traversing its volume. These hits are then used in software algorithms to reconstruct the trajectories of charged particles (i.e., tracks). The momenta of these particles are also determined by analysing their bending curvatures in the magnetic field (see Section 3.2.2 and Section 3.2.3). Another major component is the calorimeter system. Its main role is the energy measurement of photons, electrons/positrons and hadrons (detailed in Section 3.2.4). The third important component is the Muon Spectrometer (MS), designed to detect hits produced by muons. These hits are used to reconstruct muon tracks and muon momenta (see Section 3.2.5). Muon hits are also used by the trigger system (see Section 3.4). The granularities of the detector components decrease with the distance from the interaction point. The detector covers the full range of the azimuth angle. The pseudorapidity (η) coverage is limited to $|\eta| < 2.5$ for the ID, to $|\eta| < 4.9$ for the calorimeters and to $|\eta| < 2.7$ for the MS. A schematic view of the detector, with its most important sub-detectors, is shown in Fig. 3.1.

3.2.1 ATLAS Conventions

Several conventions are used within the ATLAS collaboration to define the coordinate system, some specific variables and several units of measurement. These conventions are also used throughout this thesis.

The ATLAS coordinate system is centred at the nominal $p-p$ interaction point, which also represents the centre of the detector. The x-axis is defined as pointing to the centre of the LHC ring, the y-axis points upwards and the z-axis is parallel to the beam direction at the collision point. The azimuth angle $\phi = 0$ corresponds to the positive x direction and is defined in the $[-\pi, \pi]$ range. The polar angle θ is computed from the positive z direction and is bound to the $[-\pi, \pi]$ interval. The pseudorapidity η is defined as:

$$\eta = -\ln(\tan\frac{\theta}{2}), \tag{3.1}$$

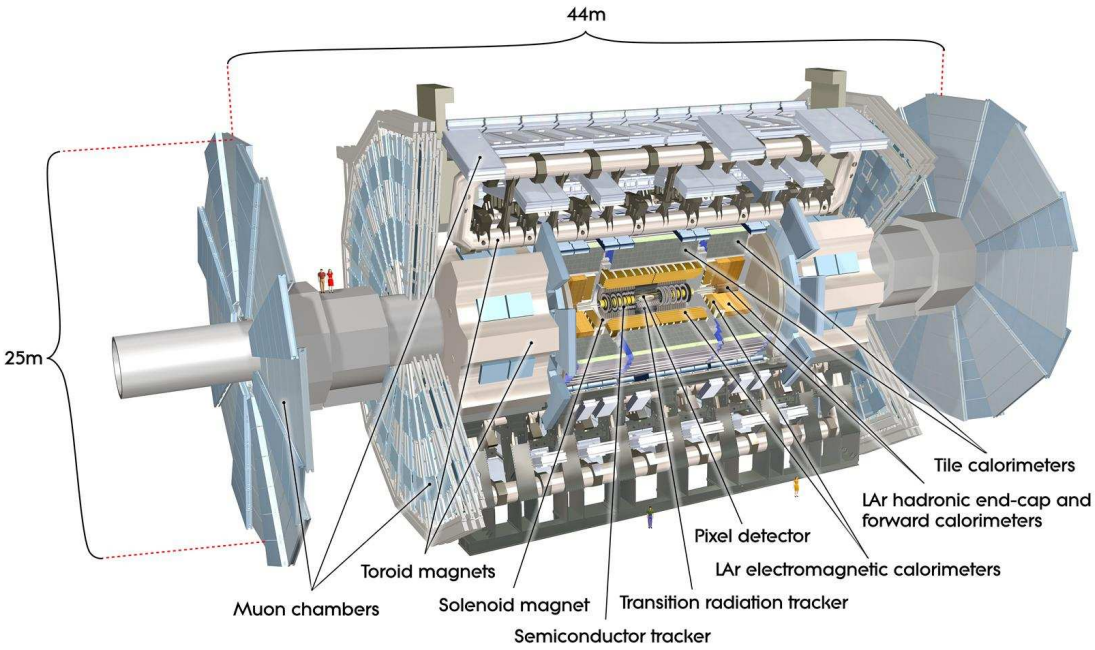


Figure 3.1: Overview of the ATLAS detector [Aad08].

with $\eta = 0$ corresponding to the transverse $(x - y)$ plane. Differences in pseudorapidities are invariant under Lorentz transformation. The distance ΔR in the $\phi - \eta$ space is defined as:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{3.2}$$

The transverse momentum p_T is defined as the projection of the momentum on the $x - y$ (transverse) plane.

Natural units (using $c = 1$, with c being the speed of light) are used to quantify energy, momentum and mass, i.e., all are given in MeV or GeV.

3.2.2 Magnet System

Four large superconducting magnet systems are deployed at ATLAS [Aad08]. Any charged particle (usually with a high momentum) traversing the detector is deflected in the magnetic field produced by the magnet system. The purpose is to use the known strength of the magnetic field in order to determine the charge and transverse momentum of particles (the ATLAS track reconstruction provides q/p_T). Measuring the transverse momentum with a high accuracy (see Section 3.2.3) is crucial for most physics analyses.

A solenoid magnet, the central solenoid, is the innermost magnet. It has an inner diameter of 2.46 m and an axial length of 5.8 m. It encompasses the entire volume of the Inner Detector and provides a 2 T axial magnetic field at its centre. Charged particles traversing the ID volume are bent in the transverse plane. A

second large magnet, the barrel toroid, is placed at the exterior of the detector, with an inner (outer) diameter of 9.4 m (20.1 m) and a length of 25.3 m. It consists of 8 coils and provides a 0.5 T toroidal magnetic field. Two smaller toroid magnets (the end-cap toroids) are placed at each side of the central region. Each has a length of 5 m and an inner (outer) diameter of 1.65 m (10.7 m) and is composed of 8 coils. The magnetic field strength is about 1 T. Charged particles traversing the inner volume of the toroid magnets (barrel and end-cap) are curved in the longitudinal plane.

3.2.3 Inner Detector

The most important function of the Inner Detector (ID) [Aad08] is to detect hit points of charged particles along their trajectories, which are used in a separate step to reconstruct their tracks (i.e., trajectories). By analysing the curvatures of the tracks in the solenoid magnetic field (see Section 3.2.2), particle momenta are determined. Particle tracks are also used to reconstruct primary and secondary vertices. The ID occupies the space closest to $p-p$ interaction point (defined by a radius of about 1.15 m and a length of about 5.4 m) and is completely immersed in the magnetic field of the central solenoid magnet. The ID is composed of three different sub-detectors, disposed in a multi-layer pattern. These are (from the innermost), the Pixel Detector, the Semiconductor Tracker and the Transition Radiation Tracker. Fig. 3.2 shows a schematic view of the ID, where all three sub-detectors are seen. The granularity of a sub-detector is the highest for the innermost component. This is necessary in order to achieve a good separation of charged particles in the reconstruction process (since the density of produced particles is the highest closest to the interaction point).

The material budget of the ID has to be as small as possible, in order to minimise the energy loss of particles traversing its volume. The overall weight is equivalent to less than 2 radiation lengths (X_0) or 0.6 interaction lengths (λ).

On average, about 47 hits are recorded in the ID for each charged particle traversing its volume (excluding the region not covered by the TRT, defined by $|\eta| > 2$). These hits have a well defined position, given by the granularity of the detector sensors in each sub-component (described for each sub-component in the next subsections). The precision of detected hits affects the precision of reconstructed particle tracks and of particle momenta. Performance studies result in a relative transverse momentum resolution less than 0.1% (0.2%) in the barrel (end-cap) region [Atl10b]. Tracks can be extrapolated to the region of $p-p$ interactions. Using the information of all reconstructed tracks, the position of the primary vertex (the point where the $p-p$ collision occurred) is also reconstructed. The precision of the position calculation increases with the number of reconstructed tracks and approaches 30 μm ($x-y$) and 50 μm (z) [Atl10d]. Vertices of secondary decays are reconstructed using the track information of the decay products. On average, the spatial resolution of 2-prong secondary decay vertices is 130 μm ($R-\phi$) and 0.7 mm (z) [Aad09].

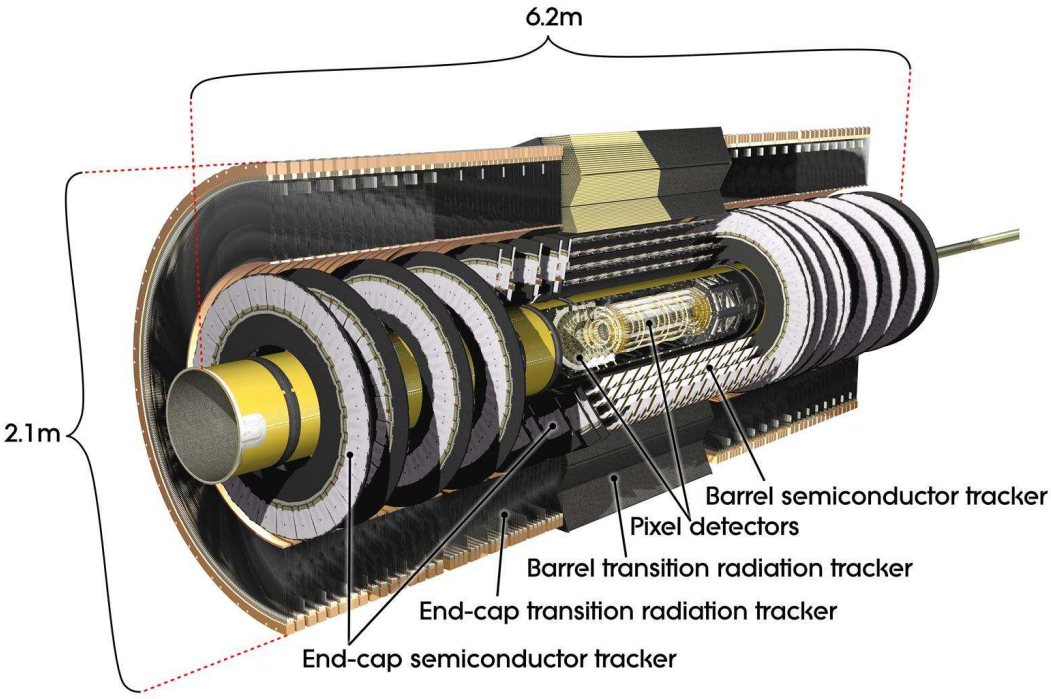


Figure 3.2: Overview of the Inner Detector [Aad08].

Pixel Detector

The Pixel Detector [Aad08] is the innermost component of the ID. It is composed of three cylindrical layers in the central region, with a length of 0.8 m each and radii of 50.5 mm, 88.5 mm and 122.5 mm. At each end-cap it consists of three disks positioned at a distances of 0.5 m, 0.58 m and 0.65 m from the centre of the detector. A total of 1744 identical pixel modules are mounted on the support structure of each layer/disk. Each module is composed of a silicon pixel sensor and the corresponding front end read-out electronics. The pixel sensor has a thickness of 250 μm and consists of 47232 individual pixel cells. About 90% of all pixels have a size of 50×400 μm, the remaining having a size of 50×600 μm. Therefore, each pixel hit provides a precision (in the μm range) hit measurement in 3 coordinates. In total, the detector has about 80 million read-out channels.

The Pixel Detector is designed to measure precisely at least three points of the particle trajectory, which is confirmed with performance studies on recorded data [Atl10f], [Atl10g]. The small sizes of the pixels offer a very good space resolution of recorded hits. The intrinsic space resolution is 10 μm $(R - \phi)$ and 115 μm (z) [Aad08]. This is crucial for reconstructing tracks close to the interaction point and for separating primary vertices from secondary decay vertices.

Semiconductor Tracker

The Semiconductor Tracker (SCT) [Aad08] is the second component of the ID. In general, its design is similar to that of the Pixel Detector. Four cylindrical layers are present in the central region, with a length of 1.5 m and radii of 299 mm, 371 mm, 443 mm and 514 mm. Nine disks are present at each end-cap, the first (last) at a distance of 0.85 m (2.7 m) from the centre of the detector. Each barrel layer/end-cap disk is tiled by a total of 4088 SCT modules. Each module contains two pairs of single sided silicon 80 μm micro-strip sensors, mounted back to back with a 40 mrad stereo angle. Each sensor has a thickness of 295 μm. The barrel SCT modules are rectangular in shape and contain 786 pairs of active strips. Each pair of strips is daisy-chained to a total length of 12.6 cm. The end-cap modules have a trapezoidal shape. Three designs are found, with different lengths of the active sensitive strips: outer (with a length of 11.9 mm), middle (with a length of 11.5 mm) and inner (with a length of 5.9 mm). The total number of read-out channels is about 6.2 million.

Due to the length of the SCT strips, each SCT hit has a high precision (in the μm scale) in only 2 coordinates. By combining two hits in both sensors found on the same module, a stereo hit is obtained, with a high precision in 3 coordinates. The SCT design aims to provide at least 4 stereo hit position measurements, which complement the 3 hit positions provided by the Pixel Detector. The intrinsic accuracy is 17 μm $(R-\phi)$ and 580 μm (z) [Aad08].

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) [Aad08] occupies the space between the SCT and the solenoid magnet. It extends the hit measurement capabilities of the Pixel Detector and the SCT with about 36 two-dimensional hit points (less in the transition region between the barrel and the end-caps, around $|\eta| \simeq 1$) and provides electron identification. It consists of 52544 axial drift tubes (straws) in the barrel region and 245760 straws in the end-cap regions. All straws have the same design and differ only in length, the longest being the ones in the barrel region (about 1.4 m). Each straw is 4 mm in diameter and is filled with a $Xe/CO_2/O_2$ gas mixture. The anode is represented by a 31 μm tungsten wire plated with $0.5-0.7$ μm gold. The cathodes are maintained at -1530 V. The signals produced by ionisation induced by charged particles traversing the straws provide hit points with a spatial resolution of about 130 μm $(R-\phi)$. The space between the straws is filled with interleaved layers of materials with different dielectric properties. Ultra-relativistic particles traversing these layers produce transition radiation, with an energy proportional to the Lorentz factor γ of the particle [Gru]. The emission of transition radiation is more pronounced for light particles, such as electrons, which produce a characteristic signal. This is used for electron identification, complementary to the electron identification performed in the calorimeters.

3.2.4 Calorimeters

The energies of particles produced in $p - p$ collisions and in subsequent decays are measured with the ATLAS calorimeter system [Aad08]. Particles entering the calorimeter initiate particle showers [Gru]. For high energetic photons or electrons/positrons the particle shower is produced by electromagnetic interactions (mainly by electron-positron pair production for photons and bremsstrahlung for electrons/positrons). Photons and electrons/positrons are stopped inside the calorimeter and, therefore, their entire energy is absorbed and measured. Hadrons initiate showers by strong (nuclear) interactions inside the calorimeter. Some of the resulting decay products (e.g, neutral pions decaying into two photons) may initiate an electromagnetic cascade as described above, others lose energy by ionisation. However, a fraction of the total initial energy cannot be detected (e.g., that transferred to long lived neutral particles, which leave the detector volume). Therefore, the energy measurement of hadrons is generally less precise. Muons loose much less energy than other particles (mainly by ionisation) and are usually not absorbed in the calorimeters. Therefore their energies are not measured completely with the calorimeter system.

All ATLAS calorimeters are sampling calorimeters. They are composed of layers of active materials (such as liquid argon or plastic scintillators) interleaved with layers of absorbing material (usually high density metals). This enables a compact design of the calorimeters. The shape of the particle shower is sampled at several stages of the shower evolution (at each layer), from which the total energy is determined. Two main groups of calorimeters are distinguished in ATLAS: the electromagnetic calorimeters (for measuring the energy deposited by photons and electrons/positrons) and the hadronic calorimeters (for measuring the energy deposited by hadrons). These are shown in Fig. 3.3.

Unlike the Inner Detector, the material depth of the calorimeter system (expressed either in radiation lengths of interaction lengths) has to be as high as possible. This is needed in order to increase the probability that a particle's energy is completely absorbed. This is detailed in the next section for each sub-component.

Electromagnetic Calorimeters

The electromagnetic calorimeters [Aad08] are designed primarily to provide measurements of the energies of photons and electrons/positrons. All have liquid argon (LAr) as the active material, with thin copper read-out plates, and lead plates as the absorbing material. They are positioned closest to the interaction point (i.e., right after the solenoid magnet). Three main detectors can be distinguished. The electromagnetic barrel calorimeter (LAr EM Barrel) occupies the central region, with a length of 3.2 m and an inner (outer) radius of 2.8 m (4 m). The pseudorapidity coverage is $|\eta| < 1.475$. An electromagnetic end-cap calorimeter (EMEC) is placed on each side of the barrel calorimeter, covering the pseudorapidity range of $1.375 < |\eta| < 3.2$.

The absorber lead plates and the copper read-out plates are folded into an accordion geometry for all calorimeter components. This ensures full ϕ coverage

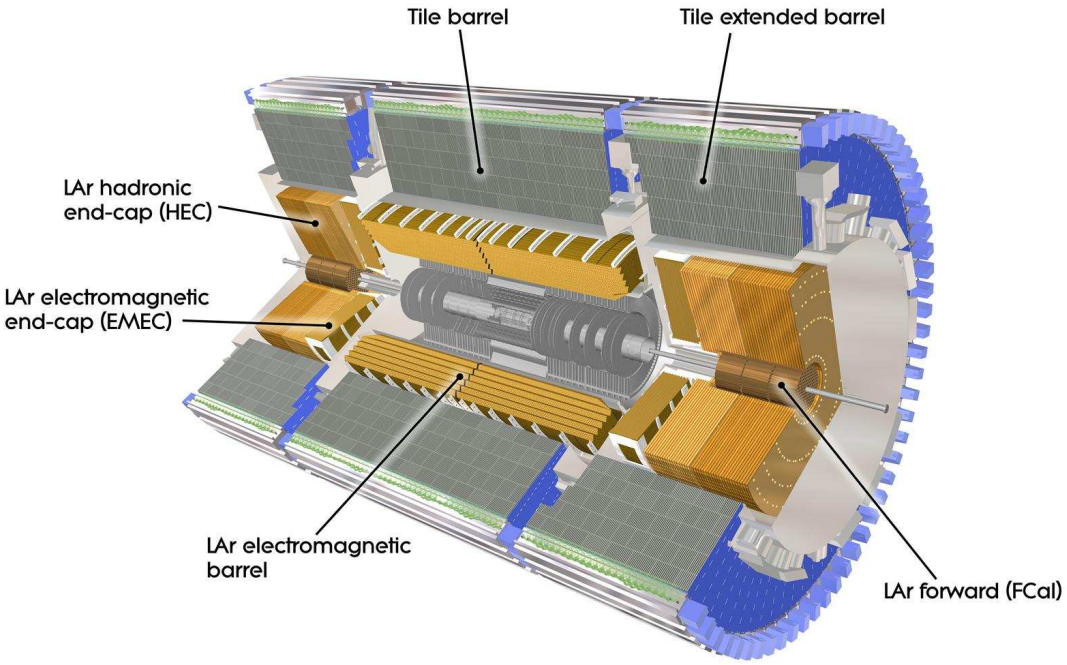


Figure 3.3: Overview of the calorimeter systems [Aad08].

and a fast signal extraction. The accordion waves are axial (in the barrel calorimeter) and radial (in the end-cap calorimeters). The folding angles of the waves are designed such, that a constant LAr gap between the layers is kept. The thickness of the absorber varies in different η regions such, that the total thickness of the barrel calorimeter is about 22 radiation lengths (X_0) for $\eta \simeq 0$ increasing to about $33\,X_0$ for $|\eta| \simeq 1.3$. Similarly, the thickness of each EMEC varies between 24 and $38\,X_0$. The material width represents on average about 2 interaction lengths (λ), which means that most hadrons traverse the electromagnetic calorimeter volume without being absorbed.

Each calorimeter component is further segmented in cells arranged in three layers. Cells in each layer have a different granularity of the read-out components (decreasing towards the exterior). Performance studies result in a relative energy resolution of about 0.5-0.7% [Aad08].

Hadronic Calorimeters

The hadronic calorimeters [Aad08] are designed for the energy measurement of hadrons, which, unlike photons or electrons, usually traverse the volume of the electromagnetic calorimeters without being absorbed. The hadronic calorimeters are therefore positioned on the exterior of the electromagnetic calorimeters. The total material width represents on average about $11\,\lambda$. Several components can be distinguished.

In the central region, the Tile calorimeter is found, with a length of 5.8 m, and the two extended barrel calorimeters on each side, each with a length of 2.6 m. A total pseudorapidity range of $|\eta| < 1.7$ is covered. These calorimeters use steel as the absorber material and plastic scintillators as the active material (which is the most economic design). The detector is composed of modules, each module containing tiles of scintillator material sandwiched in between tiles of absorber in an almost periodical structure. The orientation of the tiles is perpendicular to the beam axis. Wavelength-shifting fibers collect the light produced in the scintillators and guide it to photomultipliers mounted at the rear of the entire structure.

In the end-cap regions ($1.5 < |\eta| < 3.2$) the hadronic end-cap calorimeters (HEC) are mounted. These are built using LAr as the active material and copper plates as absorbers. This design offers the advantage to withstand operation at higher levels of radiation than the design of the Tile calorimeter. The forward regions ($3.1 < |\eta| < 4.9$) are covered by the forward hadronic calorimeters (FCal). These use LAr as the active material and copper or tungsten plates as absorbers. Their design offers a better material budget (the same interaction length with a shorter length of the material) and ability to operate in environments with very high radiation levels expected close to the beam axis.

The relative energy resolution provided by the tile calorimeter varies between 6% and 14%, with a strong η dependence (worse at high η). A similar energy resolution is observed for the HEC and the FCal [Aad08].

3.2.5 Muon Spectrometer

The Muon Spectrometer (MS) [Aad08] is the outermost and the largest detector component. Its role is the detection of hit positions along the trajectories of charged particles, that escape the calorimeters' volume. These hits are used to identify muons, to reconstruct their tracks and to compute their momenta. They also provide a fast muon detection used by the trigger system (explained in detail in Section 3.4). The pseudorapidity coverage is $|\eta| < 2.7$. The MS is composed of many smaller sub-detectors disposed in three cylindrical layers in the barrel region, and in four disks in the end-cap regions. Two main categories can be distinguished. The precision-tracking chambers are designed for accurate momentum measurement (the monitored Drift Tube (MDT) chambers and the Cathode Strip Chambers (CSC)). The trigger chambers are designed with the aim of a short reaction time and less for precision measurements (the resistive plate chambers (RPC) and the Thin Gap Chambers (TGC)). All these components are shown in Fig. 3.4.

Muon identification stems from the fact that, unlike all other charged particles entering the calorimeter volume, muons lose only a small amount of energy, mainly by ionisation, and are not absorbed in the calorimeters. This results in a high probability (above 99.9%), that a particle detected in the MS is a muon.

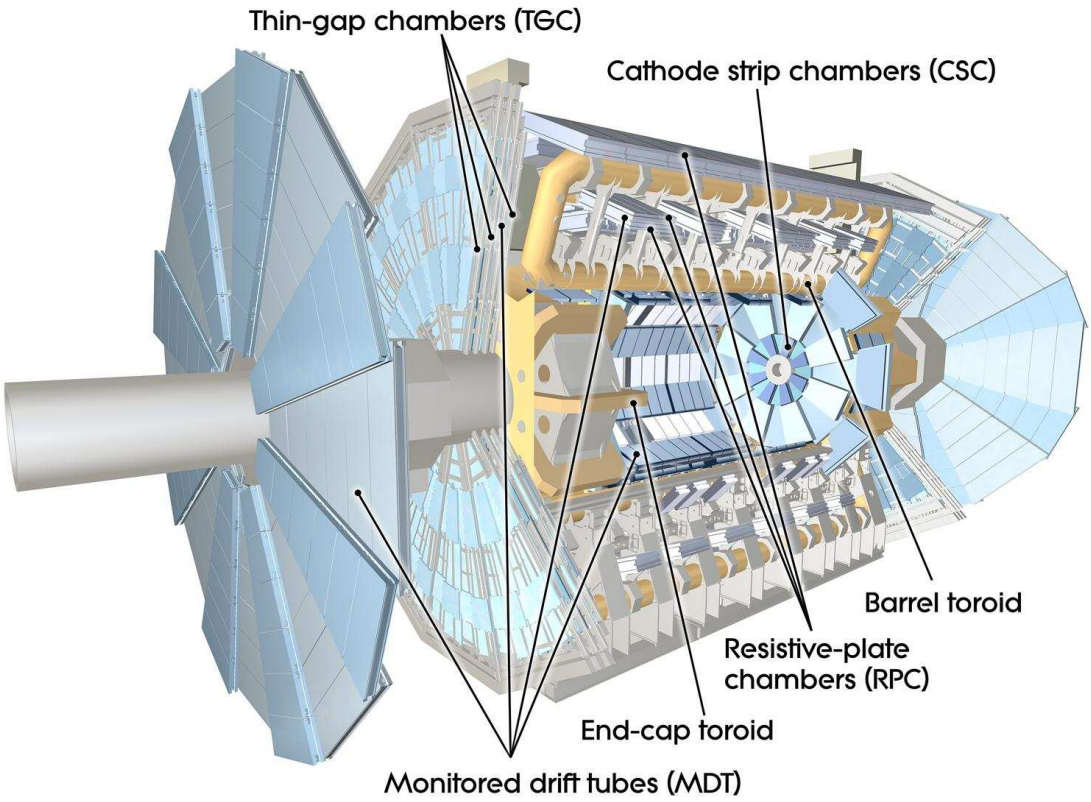


Figure 3.4: Overview of the Muon Spectrometer [Aad08].

Precision Tracking Chambers

The MDT system is assembled in three cylindrical layers in the barrel region. The three layers are placed before, in the interior of and exterior to the toroid barrel magnets, at radii of 5 m, 7.5 m and 10 m. In each end-cap region, they are assembled in four disks centred on the beam axis, at distances of 7.4 m, 10.8 m, 14 m and 21.5 m from the centre of the detector. They cover almost the entire ϕ range, with small gaps in the regions occupied by the detector's support feet (around $\phi \simeq -1.2$ and -2). The pseudorapidity coverage is within the range $|\eta| < 2.7$, with a small gap between chambers found around $\eta \simeq 0$. In the forward direction, in the $2 < |\eta| < 2.7$ range, the MDT chambers of the first disk are replaced with CSCs, which are more suited to the high particle rate in that region.

All 1088 MDT chambers are multi-wire proportional chambers, consisting of 3 to 8 layers of monitored tubes (the number of layers increases from the innermost to the outermost chambers). Each tube is 29.97 mm in diameter, with a 50 μm diameter tungsten-rhenium wire in the middle and filled with an Ar/CO_2 gas mixture. Chambers are rectangular in the barrel region and trapezoidal in the end-cap regions, with tubes directed tangential to the toroid magnetic field. They provide a space resolution of about 80 μm (z) per tube or about 35 μm (z) per chamber.

The CSC system consists of two disks, one at each end-cap, with each disk com-

prising of 8 trapezoidal chambers. The CSCs are multi-wire proportional chambers, with the wires oriented in the radial direction. The two cathodes found on each side of the wire's plane are segmented, one with strips parallel to the wires and the other with strips perpendicular to the wires (which provide also the transverse coordinate). The space resolution observed is about 40 μm (R).

Trigger Chambers

The role of the trigger chambers is to provide a fast muon detection capability and a rough estimate of the muon trajectory and momentum. This information is used by the trigger system (see Section 3.4). With a bunch spacing of 25 ns expected for the nominal LHC operation, a detector response within a few nanoseconds is essential to tag the beam-crossings.

The RPC system consists of three layers of detectors positioned in the barrel region ($|\eta| < 1.05$). Chambers of the first two layers are mounted on either side of the central MDT chambers, while the chambers of the third layer are mounted on top of the exterior MDT chambers. Each chamber consists of two independent detector layers. Each layer contains two grids of electrode strips orthogonal to each other. Therefore, each chamber provides two $\phi - \eta$ measurements of a muon trajectory. The spatial resolution of a hit is about 10 mm (z) and the time resolution of the read-out is about 1.5 ns.

The TGC system is disposed in four disks in each end-cap region ($1.05 < |\eta| < 2.4$). The chambers of the first TGC disk are mounted on the interior of the innermost MDT chambers, the chambers of the second and third TGC disks sandwich the chambers of the third MDT disk, while chambers of the last TGC disk are mounted in the immediate vicinity of the third TGC disk. TGCs are multi-wire proportional chambers with a small wire-to-cathode distance (1.4 mm). Wires are tangential to the toroidal magnetic field. The chambers contain also radial read-out strips (orthogonal to the wires). The space resolution of a muon hit is about $2 - 6$ mm (R) and the time resolution of the read-out is about 4 ns. This design is preferred to the RPC design as it provides a better hit resolution, with a slight decrease of the time resolution.

3.3 Luminosity Determination at ATLAS

Several devices are installed at ATLAS for measuring the instantaneous luminosity delivered by the LHC. The integrated luminosity (i.e., instantaneous luminosity integrated over time) is a measure of the total recorded data. An accurate estimate of the integrated luminosity is important for many measurements at ATLAS (e.g., cross-section measurements).

An absolute luminosity measurement (depending only on the number of particles detected for each $p - p$ interaction) is performed with the ALPHA detector [Aad08]. It consists of two detector systems placed at ± 240 m from the interaction point, around the beam line. It is designed to detect protons deflected in low angle (< 3 μrad) elastic collisions, using scintillator fibers. Due to the hard

radiation conditions close to the beam pipe, it is only used in special LHC runs at low instantaneous luminosities ($< 10^{31}$ cm^{-2}s^{-1}). The LUCID (LUminosity measurement using Cherenkov Integrating Detector) detector [Aad08] provides an absolute luminosity measurement in the forward direction by detecting charged particles originating in inelastic $p-p$ collisions. It consists of 20 Cherenkov tubes placed around the beam pipe at about 17 m at each side of the interaction point. LUCID is designed to operate at lower instantaneous luminosities than the design value (up to 10^{33} cm^{-2}s^{-1}).

Several independent relative luminosity measurements (depending on the number of observed events and additional beam parameters) are achieved with different detectors. The LUCID detector is employed (it is also used in an absolute measurement). The Minimum Bias Trigger Scintillators (MBTS) [Aad08], [Aad11] are scintillators mounted in the gap between the barrel and the end-cap calorimeters ($|z| = 356$ cm). They provide a fast trigger response for minimum-bias events and are used in luminosity calculations. They are designed to operate at low instantaneous luminosities ($< 10^{31}$ cm^{-2}s^{-1}). The Beam Conditions Monitor (BCM) system [Aad08], [Aad11] consists of detectors located at each end-cap of the ID ($|z| = 356$ cm). Each detector is composed of transistor sensors capable to measure the ionising dose and the neutron fluence. Besides its role in the luminosity measurement, it also has the task to send a fast (about 1 ns) warning signal in the event that the LHC operation might result in damage to the detectors (e.g., the proton beam hits the collimators in front of the detector). The BCM is designed to operate under high particle fluxes and, therefore, is used throughout the entire experiment lifetime (up to the design luminosity of 10^{34} cm^{-2}s^{-1}). The Zero-Degree Calorimeters (ZDM) [Aad08] are also used in luminosity determination by observing neutrons and photons produced in the forward direction ($|\eta| > 8.6$). It consists of an electromagnetic calorimeter module (with a width of about $29\,X_0$) and three hadronic calorimeter modules (each with a width of about $1.14\,\lambda$). The ZDM will operate at instantaneous luminosities up to 10^{33} cm^{-2}s^{-1}.

Several relative luminosity measurements have been performed using the data collected by these detectors [Aad11], [Atl11b]. The current luminosity estimate has an uncertainty of about 3.4% [Atl11b].

3.4 Trigger System

The bunch crossing rate expected for the design LHC operation is 40 MHz (corresponding to a time gap between consecutive bunches of 25 ns). On average, the data size of stored event information is about 1.3 MB [Aad08]. It is obvious that recording all collision data produced would require a huge data storing capacity, which is not available with the current technology. The maximum rate of stored events is 200 Hz.

On the other hand, not all events containing inelastic $p-p$ interactions are relevant for the ATLAS physics programme. Most physics analyses study interactions with high transverse energy/momentum transfers. The cross-sections of such processes represent a small fraction of the total $p-p$ cross-section. In addition,

most studies search for events with specific final state particles (e.g., muons, electrons or photons) or other characteristic properties (e.g., hadronic jets, missing transverse energy). The ATLAS trigger system [Aad08], [Aad09] has the role to select potentially interesting events, therefore reducing the input rate.

The trigger system is implemented in three consecutive stages:

- First Level Trigger (L1),
- Second Level Trigger (L2),
- Event Filter (EF).

Each trigger stage is seeded by the previous trigger level. The L1 stage is implemented in hardware and provides a fast, but less precise decision using only a small fraction of the entire detector information. The L2 and EF stages are collectively named as the High Level Trigger (HLT). They are implemented in software executed on computer farms. At each superior stage, the time constrains are more relaxed, due to the bandwidth reduction in the previous trigger level. Therefore, more complex algorithms are used to improve the event selection. In addition, with each increasing level, more detector information is available to the computation. Therefore, each next stage provides an additional precision in the selection.

The different trigger stages are explained in detail in the following paragraphs, with a focus on the muon trigger. Due to the significant probability that a B-meson contains a muon in its decay chain ($> 10\%$ [Nak10]), detecting a muon signature is an indicator for an event containing $b\bar{b}$ production. This is why the muon trigger is used by many B Physics studies at ATLAS. A muon trigger is also used in the study of the $B_s^0 \to \mu^+\mu^-$ decay (see Chapter 4), which is the topic of this thesis.

3.4.1 First Level Trigger

The First Level Trigger (L1) [Aad08], [Aad09] has the role to reduce the input rate from a maximum of 40 MHz to a maximum of 75 kHz (upgradeable to 100 kHz). The L1 trigger searches for muon signatures (using the fast trigger chambers of the MS), energy deposited in the calorimeters (either photons/electrons or hadronic jets), or transverse missing energy. In addition to detecting these signatures, a rough estimate of transverse momentum/energy is performed and threshold cuts on these quantities are applied. Two general L1 trigger categories exist: the muon and the calorimeter L1 trigger.

The trigger decision has to be fast (within 2.5 μs after each bunch crossing). Therefore, the L1 trigger is implemented using custom designed electronics. Furthermore, only the detector information from the relevant sub-components is available (i.e., either from the MS or from the calorimeters). This limits the accuracy of the momentum/energy estimates. The entire event information is buffered in

pipe-line memories located on the front-end electronics of each detector component. This information is passed to the L2 trigger only if the event is accepted by the L1 trigger.

The L1 trigger decision is made by the Central Trigger Processor (CTP). A total of 256 different trigger menu entries (i.e., configurations) are defined, which offer a high degree of flexibility. The output rate of certain menu entries are further reduced by applying a scaling factor. The output of the L1 trigger is represented by Regions of Interest (RoI). These are search regions used by the L2 trigger.

The L1 muon trigger uses the information provided by the dedicated fast trigger chambers of the MS (the RPCs and the TGCs). The barrel region ($|\eta| < 1.05$) is covered by the RPCs. The algorithm searches for a hit in one of the RPCs of the central layer (the pivot plane), shown schematically in Fig. 3.5. In the case of a low p_T threshold (up to 10 GeV) menu entry, another hit is searched for in one of the RPCs of the innermost layer along the direction of a straight line pointing towards the centre of the detector. The region examined is defined in a predetermined look-up table and depends on the p_T threshold (a larger region for a lower p_T threshold). A hit coincidence in both RPCs is defined as a 3-out-of-4 hit coincidence found in the four RPC read-out strips (each RPC contains 2 layers of orthogonal read-out strips). The muon momentum is determined from the deviation of its trajectory from a straight line. For a high p_T threshold (above 10 GeV) menu entry, the algorithm searches for an additional coincidence hit in an RPC of the outermost layer (besides a hit coincidence in RPCs of the first two layers). The search region is similarly determined by a predefined look-up table.

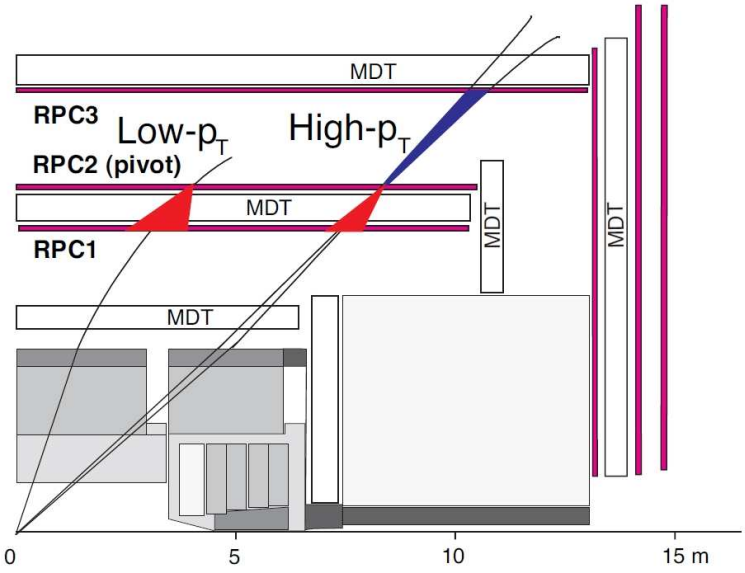


Figure 3.5: Overview of the L1 Muon trigger chambers. Coincidence search regions are highlighted for chambers in the barrel region (RPCs) [Aad08].

In the end-cap region ($1.05 < |\eta| < 2.4$) the L1 trigger information is provided by the TGCs. The hit coincidence search algorithm is similar to that applied for

the search in the barrel region. Notable differences are that the pivot plane (from which the search starts) is the plane of the outermost TGCs and the search is only performed inwards.

Besides different trigger p_T threshold configurations, two main options are used: the single muon trigger (where only one muon signature is required) or the di-muon trigger (where two different muon signatures are required). A RoI is constructed for each signature found (i.e., a coincidence in either the RPCs or the TGCs). The RoI is defined along the estimated muon direction with a size of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the barrel region or 0.3×0.3 in the end-cap region.

3.4.2 Second Level Trigger

The Second Level Trigger (L2) has the aim to further reduce the input rate to a maximum of 3.5 kHz. It is implemented in software algorithms running on a computer farm of about 500 quad-core CPUs. The time allowed for the L2 trigger decision is, on average, 40 ms, i.e., much larger than for the L1 trigger.

The L2 trigger is seeded by the RoIs obtained at the L1 trigger. It refines the L1 trigger decision by accessing the data provided by all high granularity detector components (i.e., the precision chambers of the MS, the high granularity calorimeter cells and the ID sub-components). Usually, only the data from sub-detectors contained in the RoI volume are investigated, which represent about 1-2% of the entire data. Only if the LHC operates at low instantaneous luminosities (around 10^{31} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$) can data from the entire detector volume be used. Using only the data within the RoIs decreases the computing time significantly, which is essential when operating at high instantaneous luminosities.

For the L2 muon trigger, the muon track reconstruction is performed using the hits recorded in the MDTs and CSCs within the RoI volume. A pattern recognition approach is used to reconstruct track segments in each chamber. The muon track is approximated as a series of track segments. Reconstructed muon tracks are then extrapolated to the Inner Detector and statistically combined with reconstructed ID track segments. This improves the muon p_T resolution. A single-muon or a di-muon L2 trigger is used (described in Chapter 4).

3.4.3 Event Filter

The Event Filter (EF) reduces the input rate further to a maximum of 200 Hz. The EF is implemented in software algorithms running on a computer farm of about 1800 dual quad-core CPUs. The average processing time is approximately 4 s per event.

The EF refines the L2 trigger decision by applying reconstruction algorithms similar to those used in the offline ATLAS reconstruction (described in Chapter 5). For example, the alignment and calibration information improves the accuracy of the reconstruction algorithms. These algorithms perform more advanced tasks, such as reconstructing vertices of particle decays. In addition, cuts are applied on specific quantities, e.g., on the invariant mass of reconstructed particles. The

event data corresponding to events accepted by the EF are passed further to offline reconstruction and permanent storage.

3.5 ATLAS Computing Infrastructure

The ATLAS computing infrastructure has to perform two main tasks: storing the data recorded by the detector and processing these data. The expected input rate at the design LHC operation is about 260 MB/s (for a maximum rate of 200 Hz and a data size of 1.3 MB per event [Aad08]). This results in an amount of about 15 PB of data, that needs to be stored yearly [Duc].

Since this cannot be achieved using only the resources available at CERN, a distributed computing model is employed. This model is implemented in the LHC Computing Grid (LCG) [Duc], which distributes the tasks of data storage and processing to a large number of computing facilities at different locations around the world. The LCG interface facilitates access to these data from virtually any location via secure connections over the Internet. The LCG has a hierarchical structure, implemented in three main tiers performing specific functions. All LHC experiments have access to the LCG infrastructure.

The first tier is Tier-0, located at CERN. Amongst its main roles is to temporarily copy the raw data produced by the detector and to distribute them for permanent storage to Tier-1 locations. Tier-0 also contains a computing farm for processing raw data using offline reconstruction algorithms (described in Chapter 5). Data formats, which are more suitable for physics analyses, are thus produced (ESD/AOD). These data are transferred to Tier-1.

Tier-1 contains several storage/computing facilities at various locations. Its main roles are permanent data storage and computing power required by data reprocessing.

Tier-2 sites provide primarily computing facilities for Monte Carlo (MC) event data simulations and end-user analysis. Since the distributed analysis model implies that an analysis job is executed at the same location as where the data is stored, a large amount of analysis-ready data (mostly AOD) is stored at Tier-2.

Chapter 4

The Strategy for the Study of Rare Leptonic B Decays

This chapter presents an overview of the general strategy for detecting the $B_s^0 \to \mu^+\mu^-$ decay. The focus is set on determining its branching ratio or an upper limit on the branching ratio value. Computations of limits based on simulated data follow a different approach than computations based on real data, and these differences are also discussed. The decay channels required for the measurement are introduced together with the relevant background channels. Finally, the strategy for the trigger is presented.

4.1 General Overview

The study of leptonic decays of B mesons ($B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$) is part of a wider B Physics programme at the ATLAS experiment. It is of importance since it might give an indirect evidence of New Physics beyond the Standard Model (SM). The observables to be investigated are the branching ratios of these decays. The branching ratios predicted by the Standard Model are $(3.2 \pm 0.2) \times 10^{-9}$ and $(1.0 \pm 0.1) \times 10^{-10}$ [Bur10], respectively. However, there are several extensions of the SM, which propose higher values of the branching ratios (see Chapter 2 for more details). Hence, an experimental measurement of these branching ratios and their comparison with theoretical predictions will either confirm one of the theoretical models (the SM hypothesis, or its extension), or reject them, in which case a new theory can be considered. It is important to note that, to this date, no clear experimental signature of these decays could be recorded by any of the experiments involved in this search. The most notable results are coming from the experiments located at the Tevatron. Measuring the branching ratio is currently limited by low statistics and the current best experimental evidence is an upper limit from the LHCb experiment: $BR(B_s^0 \to \mu^+\mu^-) < 1.3 \times 10^{-8}$ at a confidence level of 95% [Ser11]. This is still a factor 4 higher than the SM prediction. The Large Hadron Collider (LHC) will provide collisions with a higher center of mass energy, which directly influences the production rate of B mesons. Therefore, it is expected that the experiments at the LHC will detect this decay.

The object of interest in the scope of this thesis is the $B_s^0 \to \mu^+\mu^-$ decay. It has a simple signature, with two muons in the final state. The Muon Spectrometer of ATLAS has good capabilities for identifying muons and, together with the Inner Detector, muon momenta are precisely measured and the particles are accurately tracked to the point of their origin. Consequently, the properties of the B_s^0 mesons, from which the two muons originate, can be well measured (good spatial resolution of the decay vertex and precise momentum measurement). However, as already mentioned, the SM predicts a very low branching ratio. This creates the challenge of separating the signal events from other events with a similar signature, i.e., events with two muons in the final state. There are many such possible decays, with a branching ratio several orders of magnitude higher than that of $B_s^0 \to \mu^+\mu^-$. These are presented in Section 4.3 of this chapter. An efficient trigger strategy (presented in Section 4.4) and a careful offline event selection (described in detail in Chapter 8), are required to ensure a very high rejection level of background events, whilst accepting as many signal events as possible.

Once the background events are suppressed, the computation of the branching ratio is performed (even in the case that no signal events are observed). A common practice is to calibrate the number of the observed signal events with respect to another decay channel (the reference channel), whose measurement benefits from larger statistics and is therefore more precise. The strategy employed at ATLAS (and the other experiments located at the Tevatron or the LHC) is to measure the branching ratio of $B_s^0 \to \mu^+\mu^-$ relative to the branching ratio of the decay $B^+ \to J/\psi K^+$, with the subdecay $J/\psi \to \mu^+\mu^-$. A more detailed description of these channels is given in Section 4.2. The formula to calculate the branching ratio is given by:

$$BR(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s^0}}{N_{B^+}} \frac{\alpha_{B^+}}{\alpha_{B_s^0}} \frac{\varepsilon_{B^+}}{\varepsilon_{B_s^0}} \frac{f_u}{f_s} BR(B^+ \to J/\psi K^+) BR(J/\psi \to \mu^+\mu^-), \tag{4.1}$$

where:

- $N_{B_s^0}$ and N_{B^+} are the numbers of observed events after event selection for the signal and the reference channels, including charge conjugates, $\bar{B}_s^0$ and B^-,
- $\alpha_{B_s^0}$ and α_{B^+} are the kinematic and geometric acceptances corresponding to the two channels,
- $\varepsilon_{B_s^0}$ and ε_{B^+} are the total event selection efficiencies for the signal and reference channels,
- f_s and f_u are the b-quark fragmentation probabilities corresponding to the $b \to s$ and the $b \to u$ scenarios; the ratio $\frac{f_u}{f_s}$ is measured independently from this study,
- the decay rates $BR(B^+ \to J/\psi K^+)$ and $BR(J/\psi \to \mu^+\mu^-)$ are determined from an independent study at the same experiment.

The number of events for the signal and the reference channels are obtained from the same data sample. It is important to note that charge conjugates, $\bar{B}_s^0$ and B^- are also included in $N_{B_s^0}$ and N_{B^+}, respectively. This is necessary due to the difficulty of separating between the $B_s^0 \to \mu^+\mu^-$ and $\bar{B}_s^0 \to \mu^+\mu^-$ decays. The numbers of observed events, $N_{B_s^0}$ and N_{B^+}, need then to be adjusted with their corresponding geometric and kinematic acceptances, the selection efficiencies and the b-quark fragmentation probabilities. In practice, it is helpful to use the ratios of these quantities, as in the above formula, since most of the systematic uncertainties of these quantities are expected to cancel each other. The accurate determination of these quantities and their errors is nonetheless of utmost importance.

In the event, that sufficient statistics is available, a direct measurement of the branching ratio can be performed. If this is not the case, an upper limit will be determined using probabilistic methods, based on observation and expectation. There are several approaches to determine upper limits, based on either Bayesian or frequentist methods, or a combination of the two, and they are discussed in detail in Chapter 9. For now, it is sufficient to mention that all approaches give a limit value on the expected number of signal events. This limit describes, within a certain degree of belief, the probability that the number of observed events in the signal region is statistically compatible with the expectation for the number of signal and background events. Therefore, the following parameters are considered in any attempt to compute the upper limit:

- n, the total number of observed events in the signal region,
- s, the expectation for the number of signal events,
- ε and $\Delta\varepsilon$, the efficiency for the signal and its error,
- b and Δb, the expectation for the number of background events and its uncertainty.

For recorded data, b is estimated by extrapolating the number of background events from sidebands into the signal region. n is the total number of observed events in the signal region, i.e., before sideband subtraction. s is the expectation for the signal and is computed by incorporating various assumptions, e.g., the Standard Model expectation, or that given by a different theoretical model.

For Monte Carlo (MC) data, which is the source of the results presented in this thesis, some differences have to be considered. Firstly, the expectation for the number of background events, b, cannot be determined by extrapolating the number of events from the sidebands into the signal region, due to the limited statistics available. As an alternative, the cuts used in the selection are factorised and an overall selection efficiency is estimated from these factors. The expectation for the number of background events results from the overall efficiency and the cross-section values of the background channels considered. Secondly, the expected number of signal events, s, is unknown for MC data, since it depends on the theoretical model considered. Any assumption on the model predicting the value of s will prevent an accurate result. Therefore, the computations of the limit will

assume the null signal hypothesis, i.e., $s = 0$. n is, as in the case of recorded data, the total number of observed events in the signal region before subtracting the expected number of background events, and will be denoted with $N_{B_s^0}$ in the following. Different hypotheses for $N_{B_s^0}$ are considered and checked, if they are compatible with the expectation for the number of background events:

$$N_{B_s^0} = \sum_i N_{B_s^0,i} \cdot P(N_{B_s^0,i}, b), \tag{4.2}$$

where $P(N_{B_s^0,i}, b)$ is the Poisson probability of observing $N_{B_s^0,i}$ events given the expectation value b for the number of background events in the absence of signal events, as suggested by [Lun10]. A limit is computed for each term $N_{B_s^0,i}$. Then the final result is calculated as the weighted average of these limits:

$$N_{B_s^0}^{(UL)} = \sum_i N_{B_s^0,i}^{(UL)} \cdot P(N_{B_s^0,i}, b). \tag{4.3}$$

The limit on $N_{B_s^0}$ is then used to determine the limit on the branching ratio using (4.1).

4.2 Decay Channels Investigated

Section 4.1 briefly introduced the signal and the reference channels considered in ATLAS for the study of rare leptonic decays of B mesons. This section describes these channels in more detail. Firstly, a few preliminary remarks are given regarding the production of $b\bar{b}$-quark pairs at the LHC. The accelerator will provide a copious source of b-quarks. Two bunches of protons are expected to collide in the center of ATLAS, at a design center-of-mass energy of $\sqrt{s} = 14$ TeV (current collisions take place at $\sqrt{s} = 7$ TeV) and a design instantaneous luminosity of 10^{34} cm^{-2}s^{-1} (currently about 10^{30} cm^{-2}s^{-1}). The prediction for the total $p - p$ cross-section at 14 TeV is 102 mb, to which inelastic processes contribute 79 mb [Aad09]. This yields a rate of about 790 million inelastic $p - p$ collisions every second, with about 1 in 100 collisions producing one $b\bar{b}$ pair. Each b-quark hadronises, producing a B-meson or baryon.

4.2.1 Signal Channel

The channels relevant for the study of rare leptonic decays of B-mesons are: $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$. The latter has a branching ratio an order of magnitude less and is not considered in this thesis. The B_s^0-meson has a mean life time of $\tau = (1.425 \pm 0.041)$ ps, which translates into a proper decay length of $c\tau = 441\ \mu$m [Ams08]. This allows the particle to travel far enough in the detector before decaying, such that its decay vertex is well separated from the vertex of origin. This is a characteristic of most particles containing a b-quark and is an important ingredient in algorithms for distinguishing beauty flavour hadrons from other types of particles, the so-called b-tagging algorithms. For the search of

B_s^0 particles, b-tagging is not used directly, but information regarding the decay length is used as a discriminating variable for background suppression.

The experimental signature of this decay is the presence of two oppositely charged muons in the final state. This implies several considerations regarding the strategy for the trigger and for the offline analysis. The trigger strategy involves decision taking based on the signals left by one or two muons, the single-muon or di-muon triggers. Additionally, a threshold can be imposed on the minimum transverse momenta of these muons, p_T, to reduce the input rate. Several options are available, which are discussed in more detail in Section 4.4.

The offline analysis is based on the selection of B_s^0 candidates decaying into two muons with opposite charges. The format of the data stored by the detector contains detailed information for particle tracks, e.g., their direction (azimuthal angle, ϕ, and pseudorapidity, η), transverse momentum, charge and impact parameter. These variables are obtained in the offline reconstruction, based on information provided by the Inner Detector, and they are used to uniquely identify each track. By combining the track information from the Inner Detector with that given by the Muon Spectrometer, tracks of muons are identified among tracks of other particles. They are subsequently used in a vertex finding algorithm in order to reconstruct candidates of B_s^0 particles. Further details concerning the reconstruction of the B_s^0 candidates and their ensuing selection are given in Chapter 8.

It is important to note that the offline analysis also includes the charge conjugates $\bar{B}_s^0$ in the search. There are techniques to determine the flavour of the B-candidate, i.e., B_s^0 or $\bar{B}_s^0$, based on the flavour of the the other B-meson coming from the same $b\bar{b}$ pair, but this introduces an additional uncertainty in the computation. As an alternative, charge conjugates are included in the search for both, the signal and the reference channels, so that the term $\frac{N_{B_s^0}}{N_{B^+}}$ in (4.1) remains constant.

4.2.2 Reference Channel

The channel chosen to normalise the measurement of the branching ratio is $B^+ \to J/\psi K^+$, with the subdecay $J/\psi \to \mu^+\mu^-$. Its branching ratio is measured to $(3.4 \pm 1.8) \times 10^{-4}$ for the B^+ decay, while the J/ψ decays into two muons with a branching ratio of 5.93% [Ams08]. This is about three orders of magnitude higher than the expected one for the signal channel, and therefore, it is expected that enough statistics will be recorded for a precise measurement.

As in the case of the B_s^0, the B^+ has a relatively large decay time, $\tau = (1.638 \pm 0.011)$ ps, and proper decay length, $c\tau = 491.1\ \mu$m [Ams08]. This allows a good separation of the B^+ decay vertex from its vertex of origin and a cut on the decay length is used to suppress background.

The characteristic of the decay is similar to that of the signal channel. The J/ψ particle decays almost instantaneously, leaving two muons of opposite charges in the final state. The existence of the muon pair implies a similar strategy regarding the trigger: a single-muon or di-muon trigger is used. The presence of the kaon leads to differences regarding the offline analysis. Firstly, the geometric and

kinematic acceptance corresponding to this channel is different from the one of the signal, which affects the term $\frac{\alpha_{B^+}}{\alpha_{B_s^0}}$ in (4.1). This aspect is looked at in detail in Chapter 5. Secondly, the offline algorithm used in the search for B^+ candidates has to take the additional particle (kaon) into account. The search starts with the tracks corresponding to the two muons. Similar to the signal channel, the vertex finding algorithm combines these tracks and returns one or more J/ψ candidates. In a second step, each J/ψ is combined with another track, assumed to be the kaon, by subsequently applying the vertex finder tool. The final result is a set of candidates of B^+ particles. The final offline selection of B^+ candidates is performed exactly as in the case of the signal channel, i.e., identical cuts are applied on the same selection variables.

4.3 Background Channels Investigated

There are several types of decays which can be mistakenly reconstructed as a $B_s^0 \to \mu^+\mu^-$ signal decay. They are grouped into two categories: exclusive and inclusive decays. The first category is represented by individual channels, which are typically contained by the inclusive backgrounds. For a Monte Carlo based analysis it is useful to enhance the statistics of exclusive channels in a controlled manner, in order to study their potential contribution in the signal region.

4.3.1 Exclusive Channels

The exclusive channels include decays with a similar topology to that of the signal channel, such as two-body or three-body decays of B-mesons, with one or two hadrons in the final state. These hadrons are typically pions or kaons and may be misidentified as muons in the reconstruction process. A list of the most important exclusive decays falling into this category is found in Table 4.1.

Decay Channel	Branching ratio	Ref.
$B_d^0 \to K^+\pi^-$	$(1.94 \pm 0.06) \times 10^{-5}$	[Ams08]
$B_d^0 \to \pi^+\pi^-$	$(5.13 \pm 0.24) \times 10^{-6}$	[Ams08]
$B_d^0 \to K^+K^-$	$< 4.1 \times 10^{-7}$ @ 90% CL	[Ams08]
$B_s^0 \to \pi^+\pi^-$	$< 1.7 \times 10^{-6}$ @ 90% CL	[Ams08]
$B_s^0 \to K^-\pi^+$	$< 5.6 \times 10^{-6}$ @ 90% CL	[Ams08]
$B_s^0 \to K^+K^-$	$(3.3 \pm 0.9) \times 10^{-5}$	[Ams08]
$B_s^0 \to K^-\mu^+\nu_\mu$	$\sim 1.36 \times 10^{-4}$	[Aad09]
$B_s^0 \to \pi^-\mu^+\nu_\mu$	$(1.36 \pm 0.15) \times 10^{-4}$	[Aad09]

Table 4.1: Two-body or three-body exclusive background channels and their branching ratios.

Two of these channels, namely $B_s^0 \to K^-\mu^+\nu_\mu$ and $B_s^0 \to K^-\pi^+$, have been investigated in [Aad09], and their contribution is shown to be several orders of magnitude smaller than the contribution of the inclusive background. In addition,

the contributions of the other channels mentioned in Table 4.1 are expected to be similar to or smaller than these two exclusive channels.

4.3.2 Combinatorial Background

The other main source of background, the inclusive decays, is given by events containing random combinations of muon tracks, which may be misidentified as originating in a $B_s^0 \to \mu^+\mu^-$ decay. These decays are generically named combinatorial background. Several inclusive decays are the source of these events:

- $b\bar{b} \to \mu^+\mu^- X$,
- $c\bar{c} \to \mu^+\mu^- X$,
- Drell-Yan processes involving two muons.

$b\bar{b}$ and $c\bar{c}$ events, with muons produced through the decay chain of the beauty or charm quarks, produce a continuum of muon pairs in the di-muon invariant mass spectrum. Drell-Yan processes may also produce muon pairs by annihilation of a quark and an anti-quark contained in the two incoming protons. As already mentioned, these types of background give a much larger contribution than exclusive decays. They are the only source of background considered in the offline analysis in this thesis (Chapter 8).

4.4 Trigger Strategy

The ATLAS trigger system is described in detail in Chapter 3. The main task is to reduce the event data bandwidth to a level compatible to the data storage capability. The event rate has to be reduced from a maximum of 40 MHz to a maximum of 200 Hz. About 10-15% of the total bandwidth is available for B Physics studies. The trigger strategy employed in the study of the $B_s^0 \to \mu^+\mu^-$ decay relies on detecting muon signatures. Several options exist: using a single-muon trigger or a di-muon trigger and imposing different threshold cuts on the muon transverse momentum. The trigger configuration depends mainly on the instantaneous luminosity, which affects the trigger rate. In order to preserve a low bandwidth, the rate of a certain trigger item can be decreased by applying a prescale factor (i.e., an additional selection of events with a frequency determined by the prescale factor). As an alternative to applying prescale factors, the bandwidth is preserved by raising the trigger selection requirements, e.g., using a di-muon trigger instead of single-muon trigger and increasing the muon p_T threshold cuts. Since the study of the $B_s^0 \to \mu^+\mu^-$ decay is expected to be performed throughout the LHC operation period (including the design instantaneous luminosity), the trigger strategy is based on a di-muon trigger. Therefore, all trigger scenarios considered in this thesis are di-muon triggers. The relevant trigger items are summarised in the following paragraphs. A study of the trigger performance is described in Chapter 7.

4.4.1 First Level Trigger

The First Level Trigger (L1) di-muon trigger requires two muons signatures. It is generically labeled `L1_2MUx`, when the same p_T threshold cut is applied for both muons, or `L1_2MUx_MUy`, when different threshold cuts are applied for each muon. The following L1 trigger items are included in the B Physics trigger menu: `L1_2MU0`, `L1_2MU4` (used only during 2011 detector operation) and `L1_2MU6`. The `L1_2MU0` trigger entry requires two muon signatures without applying a p_T threshold cut. For the studies presented in this thesis, an additional trigger item is considered: `L1_2MU4_MU6`. It is used only for simulated data. A trigger performance study [Atl10e] performed on collision data shows that the `L1_2MU4` and `L1_2MU6` trigger items can be used (without a prescale factor) up to an instantaneous luminosity of about 10^{33} cm^{2}s^{-1}.

4.4.2 Second Level Trigger

There are two main di-muon L2 trigger algorithms. The first algorithm is seeded by one L1 muon and a search of a second muon is performed inside an enlarged Region of Interest (RoI) with a size of $\Delta\eta \times \Delta\phi \simeq 0.75 \times 0.75$. This algorithm is employed in conjunction with a single-muon L1 trigger and is therefore only used at instantaneous luminosities below 10^{33} cm^{2}s^{-1}. The second algorithm (the topological di-muon trigger) is seeded by two L1 trigger muons. Each muon is reconstructed in the RoI provided by the L1 trigger. Since this algorithm is seeded by a di-muon L1 trigger, it is used at high instantaneous luminosities (up to the design value). The two approaches are shown schematically in Fig. 4.1.

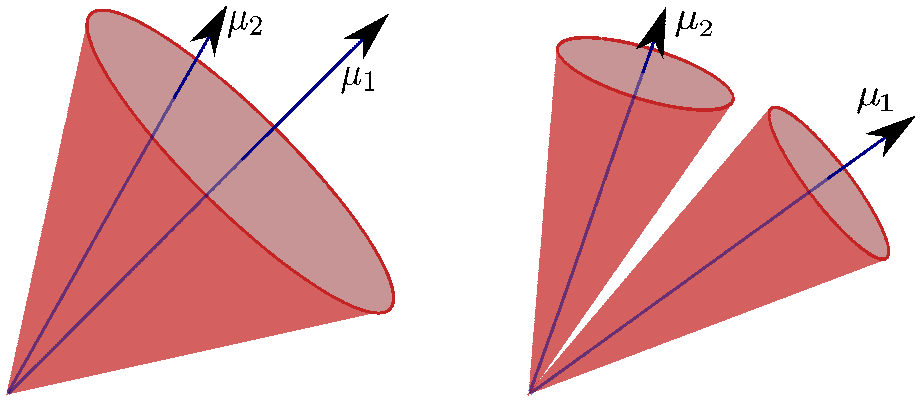


Figure 4.1: Schematic view of the di-muon L2 trigger algorithms seeded by one (left) or two (right) L1 trigger muons.

The L2 di-muon trigger entries included in the B Physics menu are: `L2_2mu4` and `L2_2mu6`. The lower case letters in the names are used to signify that a confirmation of the L1 muons is performed. `L2_mu4_mu6` is another trigger item considered for studies based on simulated data. All these L2 trigger items can be used without a prescale factor up to instantaneous luminosities of about 10^{33} cm^{2}s^{-1}.

4.4.3 Event Filter

The di-muon Event Filter (EF) triggers considered are similar to the di-muon L2 trigger items: `EF_2mu4`, `EF_2mu6` and `EF_mu4_mu6`.

An additional requirement imposed at the EF is that the two muons originate in a common vertex. A vertex finding algorithm similar to that from the offline reconstruction software (explained in Chapter 8) is used to verify the common vertex assumption. `EF_2mu4_DiMu` is such an EF trigger item and it has the advantage of a much lower output rate.

An even more advanced EF trigger item is `EF_2mu4_Bmumu`. This trigger requires that the two muons originate in a common vertex and imposes a cut on the reconstructed di-muon invariant mass. The invariant mass window of accepted events is $[4\ \text{GeV}, 7\ \text{GeV}]$. This cut rejects significant contributions coming from $c\bar{c}$ and $b\bar{b}$ resonances. This trigger item is specifically tuned for the study of the $B_s^0 \to \mu^+\mu^-$ decay.

4.4.4 Trigger Chains

Trigger chains are sequences of trigger items from each trigger stage, such that all trigger items in the chain have similar selection criteria. The trigger chains considered in this thesis are:

- `L1_2MU4`, `L2_2mu4`, `EF_2mu4`,
- `L1_2MU4`, `L2_2mu4`, `EF_2mu4_DiMu`,
- `L1_2MU4`, `L2_2mu4`, `EF_2mu4_Bmumu`,
- `L1_2MU4_MU6`, `L2_mu4_mu6`, `EF_mu4_mu6`,
- `L1_2MU6`, `L2_2mu6`, `EF_2mu6`.

Chapter 5

Data Samples

This chapter reviews the data samples needed for the analysis presented in this thesis. All data sets are Monte Carlo (MC) generated data. The tools used in the production of these data sets are described. Different settings are used in the generation process of each data sample. These differences and their effects on kinematic distributions and production cross-sections are discussed.

5.1 Monte Carlo Production

The production of Monte Carlo events typically involves three steps: *generation* of events containing the physics processes to be studied, *simulation* of interaction processes between the generated particles and the detector and *digitisation* of the detector response. A scheme showing these steps is presented in Fig. 5.1, where boxes represent the stages in data production and ellipses represent data formats.

At generation level, hard scattering processes in p-p collisions are simulated. Parton interactions with initial and final state radiation, parton hadronisation and particle decays are also simulated. Decays take place with certain probabilities given by theoretical models. A selection of events containing only specific processes is possible. The particles generated from these interations are described by their four momenta and are passed on to the next stage. Here, physical interactions between particles and the detector material are simulated, including interactions with sensitive detector materials. The output is represented by particle hits, i.e., energy deposits in the detector volumes. The digitisation involves simulating the detector response, including processing the hits, and aquisition of data in digital format. The output data has the same format as real data recorded by the detector. These stages are detailed in [Aad10]. The tools specific to this analysis are described in more detail in the following sections.

An additional step is the event *reconstruction*. This employs the same software tools used in the processing of real data. The output data are stored in the form of objects that are more suitable for physics analysis, e.g., particle tracks and their properties, decay vertices and others.

The detector introduces uncertainties in measuring true parameters due to the limited granularity of the sensitive components. This is simulated in the simula-

tion and digitisation steps. The use of stochastic methods in the reconstruction step introduces additional uncertainties. Therefore, distributions of reconstructed parameters are smeared around the true values with a function which describes the imperfections of the detector and of the reconstruction methods. In the following, the notions of reconstructed and truth values of particle parameters will be used. The following analysis is based on reconstructed parameters, unless mentioned otherwise.

Producing large statistics of simulated data using the entire software chain is prohibited by the long computing time, especially in the simulation step. This process can be shortened by smearing certain truth values with the known resolution of the detector, as an alternative to simulating all interactions of particles in the detector. This process is named fast simulation. Several approaches exist, in which the fast simulation is performed in all detector components, or only in specific subdetectors, while full simulation is still performed in the remaining subdetectors.

The final step shown in the diagram is the *analysis* of produced MC events. The aim of this step is e.g., to find candidates for the investigated decays. This is independent of the event production, but is an essential part of the analysis chain. In depth details on the analysis procedure are given in Chapter 8.

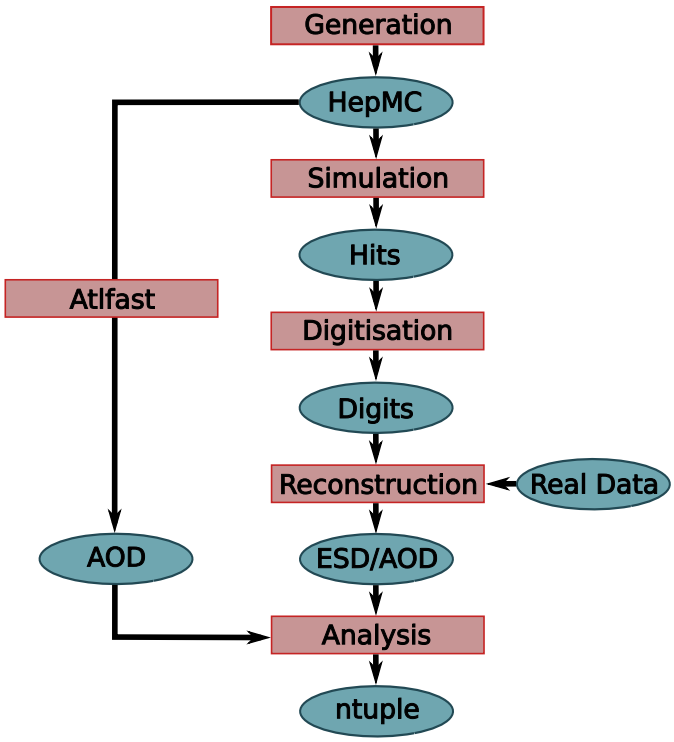


Figure 5.1: Schematic view of Monte Carlo data simulation and reconstruction.

ATLAS Software Framework

The software framework used to produce and process MC events in ATLAS is called Athena [Aad10]. It is mostly implemented in the PYTHON [vR09], C++ [Str97] and FORTRAN [Met85] programming languages and it has a modular structure. PYTHON scripts offer a high degree of flexibility in configuring production jobs. They interface a large number of indivual algorithms performing specific tasks. These are typically implemented in the C++ or FORTRAN programming languages, which are optimised for speed of computing.

Athena evolves rapidly in time, new algorithms being constantly added, or existing ones being improved. In order to keep track of all constituent packages, a versioning system is used. The data sets were produced with the latest stable version of Athena available at the time of production. For event generation, Athena version 14.2.0 was used. For simulation, digitisation and reconstruction, Athena version 14.2.25 was employed. The $B_s^0 \to \mu^+\mu^-$analysis was implemented and run with Athena version 15.3.0.

5.2 Review of Data Samples

Chapter 4 describes the strategy used for the study of $B_s^0 \to \mu^+\mu^-$decays. The main channels of interest have been introduced there: the signal channel, the reference channel and several inclusive background channels. Data sets corresponding to each channel have been produced in order to prepare the analysis chain and are summarised in Table 5.1. These are fully simulated and reconstructed Monte Carlo samples. Specific details regarding the generation process are given in Section 5.3. Additionally, several other data sets have been generated (but not reconstructed) for studying differences at MC truth level. These are described in detail in Chapter 6.

Decay channel	Label	Number of reconstructed events	$p_T(\mu_1) >$ [GeV]	$p_T(\mu_2) >$ [GeV]
$B_s^0 \to \mu^+\mu^-$	`Bs_mu6mu4`	50 000	6	4
$B_s^0 \to \mu^+\mu^-$	`Bs_mu2p5mu2p5`	150 000	2.5	2.5
$B^+ \to J/\psi(\mu^+\mu^-)K^+$	`B+_mu6mu4K+`	150 000	6	4
$b\bar{b} \to \mu^+\mu^- X$	`bb_mu6mu4X`	1 000 000	6	4
$b\bar{b} \to \mu^+\mu^- X$	`bb_mu2p5mu2p5X`	1 000 000	2.5	2.5
$c\bar{c} \to \mu^+\mu^- X$	`cc_mu2p5mu2p5X`	1 000 000	2.5	2.5
$DY \to \mu^+\mu^- X$	`DY_mu2p5mu2p5X`	1 000 000	2.5	2.5

Table 5.1: Fully simulated data sets used in the analysis.

All data sets are generated such that each event contains at least one of the corresponding decays. The table also shows the number of available reconstructed events in each sample. Finally, the thresholds applied on the muons' transverse momenta at generation level are shown.

The $B_s^0 \to \mu^+\mu^-$ data sets are used to tune the analysis for the proper selection of signal events. The muon p_T thresholds of 6 and 4 GeV are used to increase the effective MC statistics available, assuming the same thresholds are used in the trigger and in the offline selection. This data set is used in the analysis proper, while the sample with thresholds of 2.5 GeV on both muons is used to investigate effects of the muon trigger.

The $B^+ \to J/\psi(\mu^+\mu^-)K^+$ sample is used to investigate how the analysis affects the reference channel. The efficiency of selecting these events is important in the the computation of the $B_s^0 \to \mu^+\mu^-$ branching ratio.

The $b\bar{b} \to \mu^+\mu^- X$ data samples are used in order to optimise the analysis for background suppression. The set with muon p_T thresholds of 6 and 4 GeV is employed in the analysis proper, while the set with muon p_T thresholds of 2.5 GeV is used in trigger studies.

The $c\bar{c} \to \mu^+\mu^- X$ and $DY \to \mu^+\mu^- X$ samples are also studied in order to determine the influence of additional sources of background, i.e., events containing $c\bar{c}$ pairs decaying into two muons and Drell-Yan processes.

All data sets have been centrally produced by the collaboration for preparing the analysis tools before data taking started. Events are generated assuming collisions at a centre-of-mass energy of $\sqrt{s} = 10$ TeV. This value is lower than the design centre-of-mass energy of the LHC ($\sqrt{s} = 14$ TeV) and higher than the value used during first data taking period ($\sqrt{s} = 7$ TeV). Preliminary studies using the design centre-of-mass energy have been performed earlier [Aad09]. The decision to perform similar studies using $\sqrt{s} = 10$ TeV is a consequence of the idea to run LHC in a first phase at a lower centre-of-mass energy than the initially planned value. These studies commenced before the decision to start data taking with collisions at $\sqrt{s} = 7$ TeV. The studies presented in this thesis are meant to verify and complete the $B_s^0 \to \mu^+\mu^-$ analysis presented in [Aad09].

5.3 Event Generation

Several tools are available for particle generation. The tool used to generate the events used in this thesis is Pythia [Sjo06]. This a general purpose event generator, which has been extensively tested and validated with data from previous experiments. Version 6.415 of this software tool was used in the generation process. Pythia is integrated into the Athena framework by means of the PythiaB package [Smi05]. This is an ATLAS specific modification of the original Pythia program, tuned for the production of events used in B Physics analysis. It increases the efficiency of generating b- and c-quarks in QCD showers. It also facilitates several parameters used in event selection.

Deep inelastic scattering between the two colliding protons is simulated at $\sqrt{s} = 10$ TeV. The hard scattering processes between the partons contained by the colliding protons are simulated by low order QCD calculations. These are typically $2 \to 2$ scattering processes between coloured partons: $q_1q_2 \to q_1q_2$, $q_1\bar{q_1} \to q_2\bar{q_2}$, $q\bar{q} \to gg$, $qg \to qg$, $gg \to q\bar{q}$ and $gg \to gg$. Additionally, parton showers are produced by initial-state and final-state QCD and QED radiation. In the lowest

approximation, the shower can be described by a sequence of $1 \to 2$ processes: $q \to qg$, $g \to gg$, $g \to q\bar{q}$, $q \to q\gamma$ and $l \to l\gamma$.

The parton shower evolves further up to the hadronisation of existing quarks. In this process, coloured partons are joined to form colourless hadrons. Each meson or baryon produced in this step is described by its measured properties, e.g., mass, lifetime, decay modes and branching ratios. The values for these parameters are usually taken from [Nak10], but it is also possible to adjust these values. The produced hadrons or leptons are subject to further decays with certain probabilities, according to their decay modes and branching ratios. If the decay occurs before the particle reaches the inner-most detector component, the particle is considered unstable. The decay process is then repeated for daughter particles. Particles that reach the inner-most detector component are considered stable, for example p, n, π, K, e, μ, ν, γ.

Before the event is stored, the program checks the list of produced particles (stable and unstable) for the required decay. The event is only stored if a match is found. The format of the output follows the HepMC format [Dob01]. This has a tree structure, where each node represents a decay vertex and each branch represents a particle. Therefore, each event contains a list of particles and their properties (e.g., particle code, mass, charge), their kinematic variables (e.g., transverse momentum p_T, pseudorapidity η, azimuthal angle ϕ) and the indices of their mother and daughter particles. Additionally, a list of vertices is present, with their polar coordinates and the indices of their mother and daughter particles on the particle list.

5.3.1 MC Truth Event Selection

It is expected that events used in data analysis contain final state particles within a limited phase space. Muons can only be observed within a pseudorapidity range of $[-2.5, 2.5]$ in the Inner Detector, due to the limited geometrical acceptance. In addition, the p_T spectra are limited by the steep rising curves produced by trigger thresholds on the transverse momenta. This is also reflected in the kinematic properties of the B mesons from which the muons originate. It is desired that the simulated data are as similar as possible to those expected to be recorded by the detector. It is important to notice that, due to the random nature of event generation, the production of samples consisting only of events containing specific decays can be very time consuming. In order to efficiently produce events with the desired kinematic properties, Pythia allows to tune the generation process by applying thresholds on several variables.

One possibilty is to impose cuts on the kinematic variables of the particles produced in the $2 \to 2$ hard scattering interactions. These are the allowed ranges for the rapidity of the decay products with the highest (lowest) rapidity in their centre-of-mass, `CKIN(9)-(10)` (`CKIN(11)-(12)`). The allowed range in both cases is $[-3.5, 3.5]$.

Another variable available for tuning is `CKIN(3)`. It is defined as the momentum of outgoing particles perpendicular to the direction of the incoming particles in the centre-of-mass system of the hard scattering process, as schematically shown in

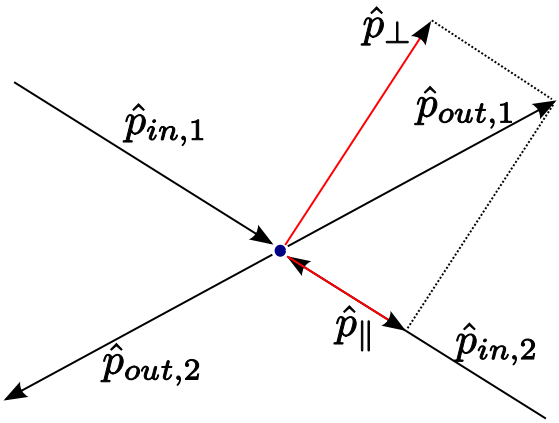


Figure 5.2: Schematic view of $\hat{p}_\perp$ in the rest frame of the hard scattering process.

Fig. 5.2. This variable is denoted $\hat{p}_\perp$. The thresholds imposed on the $\hat{p}_\perp$ values in the generation of each data sample are shown in Table 5.2. $\hat{p}_\perp$ is highly correlated with the transverse momentum of the B mesons produced after hadronisation, as shown in Fig. 5.3. Since $\hat{p}_\perp$ and $p_T(B)$ apply to different particles and different systems of reference, they are not identical. Therefore, it is not ensured that a threshold applied on $\hat{p}_\perp$ will affect also the transverse momentum of the final B meson. However, this cut is applied at a very early stage in the event generation. A careful tuning of this parameter has the biggest impact on optimising the efficiency of event generation.

Data set	[CKIN(9), CKIN(10)] [CKIN(11), CKIN(12)]	$\hat{p}_\perp >$ [GeV]	$p_T(\bar{b}) >$ [GeV]	$\eta(\bar{b}) \in$
Bs_mu6mu4	[-3.5, 3.5]	10	10	[-2.5, 2.5]
Bs_mu2p5mu2p5	[-3.5, 3.5]	5	2.5	[-2.5, 2.5]
B+_mu6mu4K+	[-3.5, 3.5]	15	10	[-2.5, 2.5]
bb_mu6mu4X	[-3.5, 3.5]	10	7	[-4.5, 4.5]
bb_mu2p5mu2p5X	[-3.5, 3.5]	8	7	[-4.5, 4.5]
cc_mu2p5mu2p5X	[-3.5, 3.5]	6.5	4	[-4.5, 4.5]
DY_mu2p5mu2p5X	-	-	-	-

Table 5.2: Selection cuts used in event selection at generation level. For the bb_mu6mu4X and bb_mu2p5mu2p5X data samples the values in the columns under $p_T(\bar{b})$ and $\eta(\bar{b})$ refer to either the b or $\bar{b}$-quarks. For the cc_mu2p5mu2p5X, these values refer to either the c or $\bar{c}$-quarks. No such kinematic selection variables were used for DY_mu2p5mu2p5X.

Further optimisation is possible by tuning the range of the kinematic variables of the b- or c-quarks produced. These are the ranges of the pseudorapidity and the transverse momentum. In the case of signal samples, these cuts are applied only on the $\bar{b}$-quark, since the $\bar{b}$-quark is the one required to hadronise into a B_s^0 or

B^+ meson. In the case of combinatorial background, it is required that either a b- or $\bar{b}$-quark fulfill the kinematic selection criteria. For the $c\bar{c}$ background, the constraints are applied on either quark of the $c\bar{c}$ pair. The values of these cuts are shown in Table 5.2.

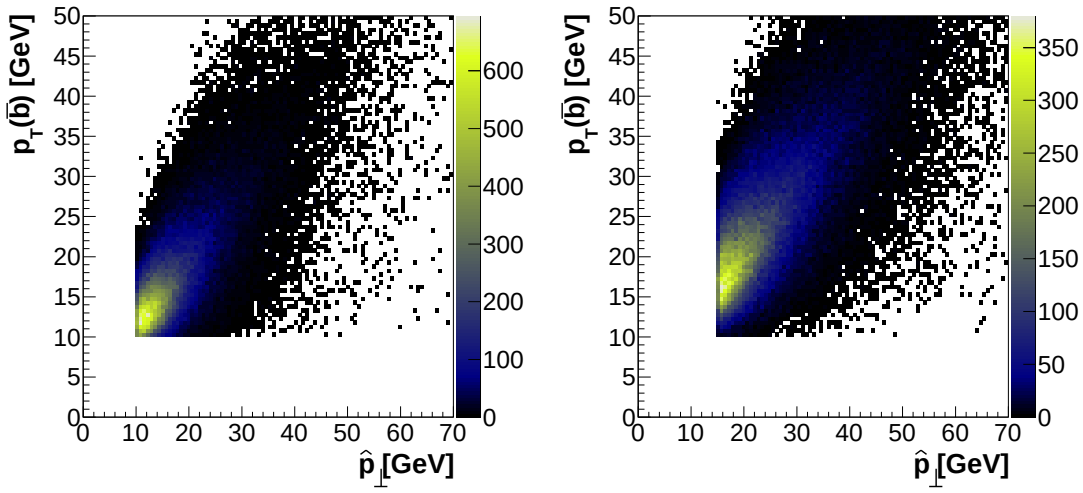


Figure 5.3: Correlations of $\hat{p}_\perp$ with the transverse momentum of the hardest $\bar{b}$-quark for the generated $B_s^0 \to \mu^+\mu^-$(left) and $B^+ \to J/\psi(\mu^+\mu^-)K^+$(right) MC data samples.

Finally, the cuts on the transverse momenta of final state muons are applied as detailed in Table 5.1. This is to ensure that the MC data samples have similar thresholds as the nominal trigger items that are to be used.

5.3.2 MC Truth Level Kinematic Variables

The kinematic variables of the generated samples have been analysed. The aim is to verify whether these variables behave as expected and to identify differences between the signal and reference channels.

Firstly, the multiplicity of muons is investigated. Fig. 5.4 shows the distribution of the total number of muons observed in each event for the signal (left) and reference (right) channels. Both positive and negative muons are considered. On average, 2.3 muons are generated in each event, in both cases and no event contains less than two muons. Besides the decay of the generated channels, muons can be produced as well in other decays, such as decays of pions or kaons and J/ψ resonances.

In a next step, the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decay chains are extracted based on truth information in each data set. The p_T spectra of the particles produced in these decays are of particular interest and are shown in Fig. 5.5. The cuts imposed on the two muons at 4 and 6 GeV are clearly visible. Very few muons fall below the threshold. This is possible in events with a higher muon multiplicity, when the hardest p_T muons do not originate in the decay of

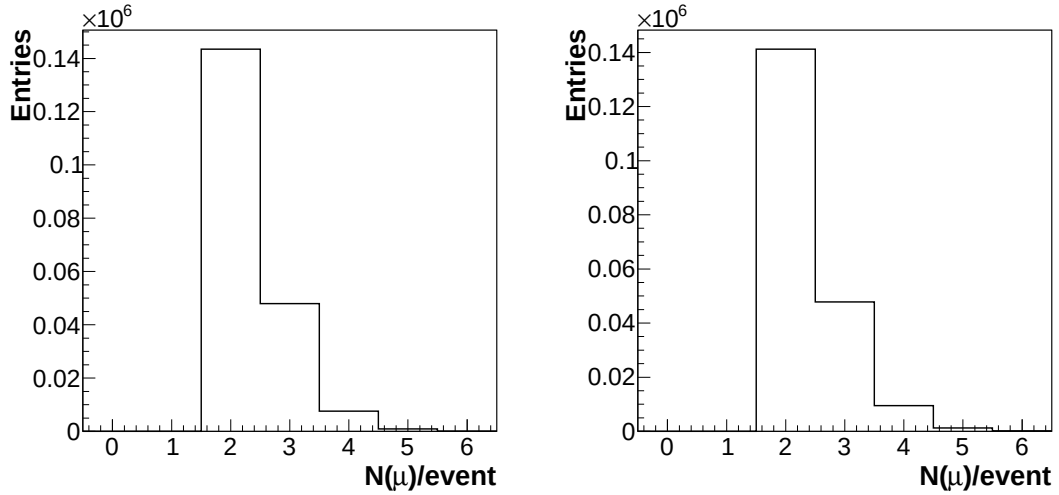


Figure 5.4: Muon multiplicity in the generated $B_s^0 \to \mu^+\mu^-$ (left) and $B^+ \to J/\psi(\mu^+\mu^-)K^+$ (right) samples.

the B meson. In the case of the reference channel, an additional particle is found in the decay products of the B^+ meson, the kaon. Its p_T spectrum shows the cut imposed at 0.5 GeV.

Finally, the p_T, η and ϕ distributions for the produced B_s^0 and B^+ mesons are analysed, as displayed in Fig. 5.6. A clear difference is seen in the p_T distributions. The B^+ spectrum is harder. This is due to two reasons. Firstly, the B meson p_T is highly correlated with the transverse momenta of its decay products. For B_s^0, the decay products have a p_T cut at 4 and 6 GeV, which explains the sharp rising edge around $p_T(B_s^0) = 10\,\text{GeV}$. For B^+, the rising edge is shifted to higher values due to the existence of an additional particle in the decay chain. Secondly, the B meson p_T spectrum is highly correlated with the $\hat{p}_\perp$ and $p_T(\bar{b})$ selection variables. For B^+, a higher threshold imposed on $\hat{p}_\perp$ determines a harder p_T spectrum.

A slight difference is seen in the η distributions, the peak around $\eta = 0$ being more pronounced in the case of the B^+ mesons. This is explained by the differences observed in the p_T spectra, due to the higher probability that high p_T particles are oriented around $\eta = 0$. No obvious differences are present in the ϕ distributions. Several other kinematic variables are discussed in more detail in Appendix A.

5.3.3 Cross-Section Calculations

Knowing the production cross-section for each channel is important for determining the expected number of events for a given amount of data. This is computed by:

$$N_k^{expected} = \sigma_k^{prod} \cdot \mathcal{L}_{int}, \tag{5.1}$$

where σ_k^{prod} is the production cross-section of channel k and $\mathcal{L}_{int}$ is the integrated luminosity.

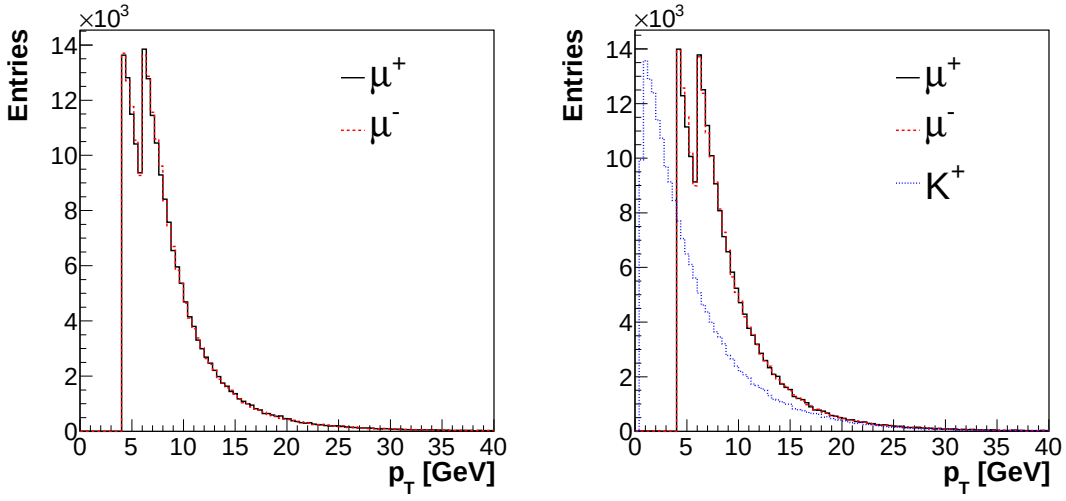


Figure 5.5: p_T spectra of the B_s^0 (left) and B^+ (right) decay products.

Pythia provides several output variables that estimate the production cross-section at different stages of the generation process. The first is σ_k^{PX}, which represents the cross-section of channel k after the hard scattering process and before applying the selection cuts on the b or c-quarks. This cross-section depends strongly on the choice of the `CKIN` variables, but does not depend on the cuts applied to the b or c-quark kinematic variables. Another variable is σ_k^{BX}, which represents the cross-section calculated after applying cuts on the kinematic variables of the b and c-quarks and of the final state muons.

Each data set is generated in smaller subsamples containing 1 000 events. The σ_k^{PX} and σ_k^{BX} variables are available for each subsample. An average cross-section has been computed using 200 subsamples (in total 200 000 generated events). Fig. 5.7 shows the distribution for σ_k^{PX} observed for the signal and reference channels. Similarly, Fig. 5.8 depicts the σ_k^{BX} distribution for the two channels. The average for each variable is obtained by the mean of a Gauss function used to fit the shown distributions. The resulting values are shown in Table 5.3.

In the case of an exclusive decay channel, Pythia forces this decay by setting its branching ratio to 1. Therefore, the cross-section returned by Pythia has to be corrected by the branching ratio of the respective decay. An additional correction of a factor of 2 has to be applied, in order to account for the charge conjugate decays, which are not explicitly simulated. The final production cross-section is given by:

$$\sigma_k^{prod} = 2 \cdot \mathcal{B}_k \cdot \sigma_k^{BX}. \tag{5.2}$$

For inclusive decay channels, where no specific decays are forced, these corrections are not necessary.

Taking into account the effects produced by different selection criteria on the production cross-section is essential to correctly estimate the expected number of events. This is a crucial factor when comparing the number of observed events in

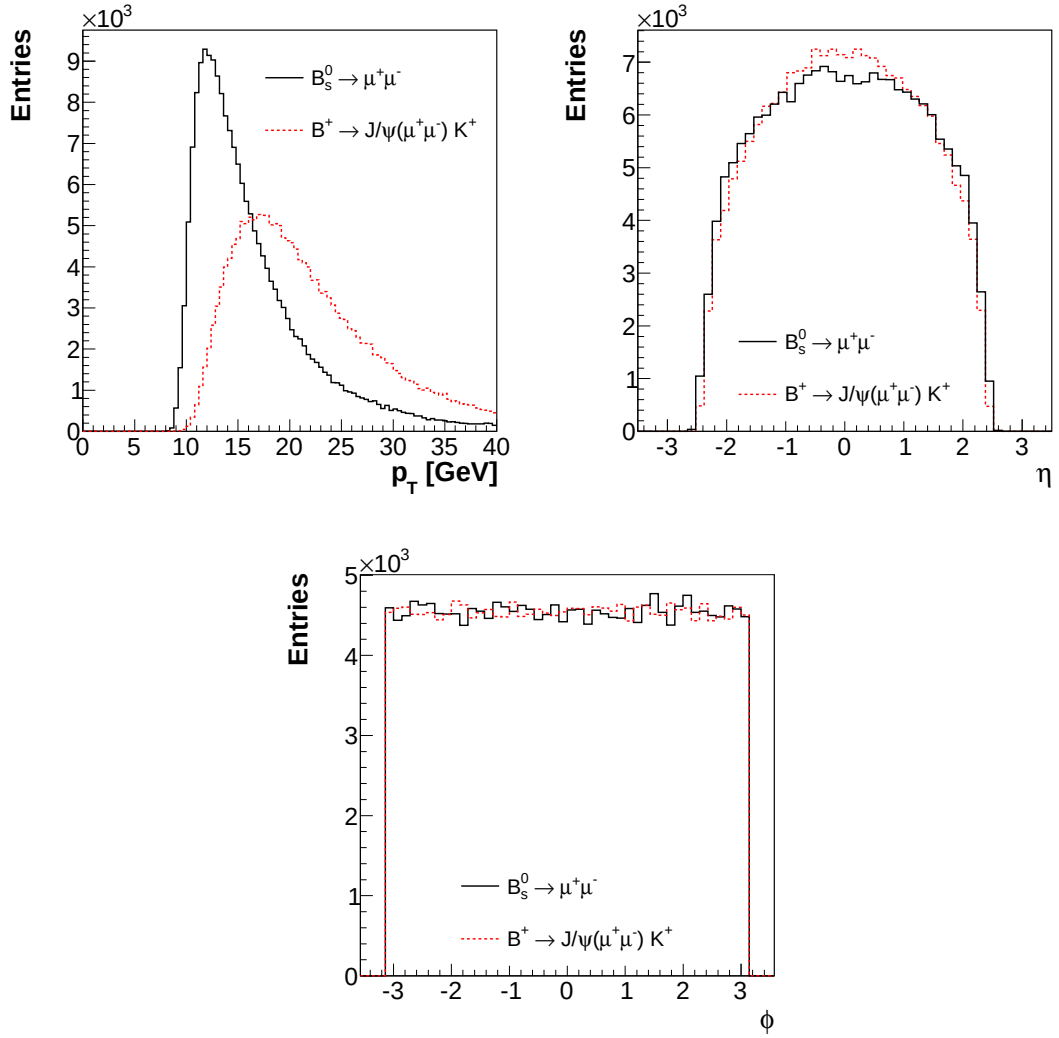


Figure 5.6: p_T, η and ϕ spectra of the generated B_s^0 and B^+ mesons.

the signal and reference channels, as shown in (4.1). In addition, the ratio of the geometric and kinematic acceptances for the two channels is determined from MC truth data. This is discussed in Chapter 6 taking the effects of the different tuning parameters into account.

5.4 Event Simulation

The simulation of particle interactions with the detector material is achieved with the GEANT4 software toolkit [Ago03], [All06]. It is implemented in C++ and the Athena framework provides a PYTHON wrapper for easy manipulation. Details of ATLAS specific implementations are given in [Aad10]. GEANT4 provides several key functions: interpreting detector geometry, modelling particle propagation in the detector material and simulating physical processes that particles undergo

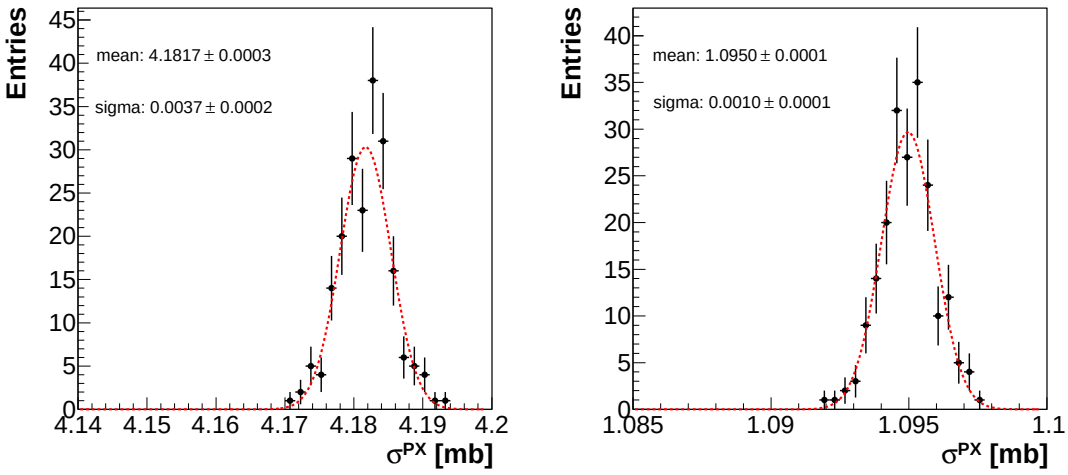


Figure 5.7: Distribution of the σ_k^{PX} variable for the $B_s^0 \to \mu^+\mu^-$ (left) and $B^+ \to J/\psi(\mu^+\mu^-)K^+$ (right) channels. The Gauss function resulting from the fit is shown as a red dotted line.

Decay Channel	σ^{PX} [mb]	σ^{BX} [μb]	correction factor	σ^{prod} [μb]
Bs_mu6mu4	4.1817 ± 0.0003	1.585 ± 0.004	$(6.4 \pm 0.4) \times 10^{-9}$	$(10.1 \pm 0.1) \times 10^{-9}$
Bs_mu2p5mu2p5	33.305 ± 0.004	14.05 ± 0.04	$(6.4 \pm 0.4) \times 10^{-9}$	$(89.9 \pm 0.8) \times 10^{-9}$
B+_mu6mu4K+	1.0950 ± 0.0001	1.696 ± 0.004	$(1.20 \pm 0.04) \times 10^{-4}$	$(2.04 \pm 0.07) \times 10^{-4}$
bb_mu6mu4X	-	0.0654	1	0.0654
bb_mu2p5mu2p5X	-	0.6	1	0.6
cc_mu2p5mu2p5X	-	0.261	1	0.261
DY_mu2p5mu2p5X	-	0.307	1	0.307

Table 5.3: Cross-sections of generated data samples.

within the detector.

The geometry of the ATLAS detector is represented by the parametrized objects physical volumes, described by material, shape and position in the detector. Due to the high granularity of various subcomponents, a large number of such volumes are required for an accurate description of the detector. More than 316 000 different physical volumes are required. The geometrical shapes that are used have to be compatible with the shapes recognised by GEANT4. These are stored in the ATLAS geometry database, which is available during the simulation process. In addition, the ATLAS conditions database stores various data, such as calibration and misalignment parameters, as well as lists of dead channels.

The physics processes considered by GEANT4 fall into three main categories. In flight decays are calculated based on the particle mean life time and decay modes. Electromagnetic processes include ionisation produced by charged particles in the detector material, bremsstrahlung, multiple scattering, electronuclear and photonuclear reactions. The computation of energy loss for these processes is also

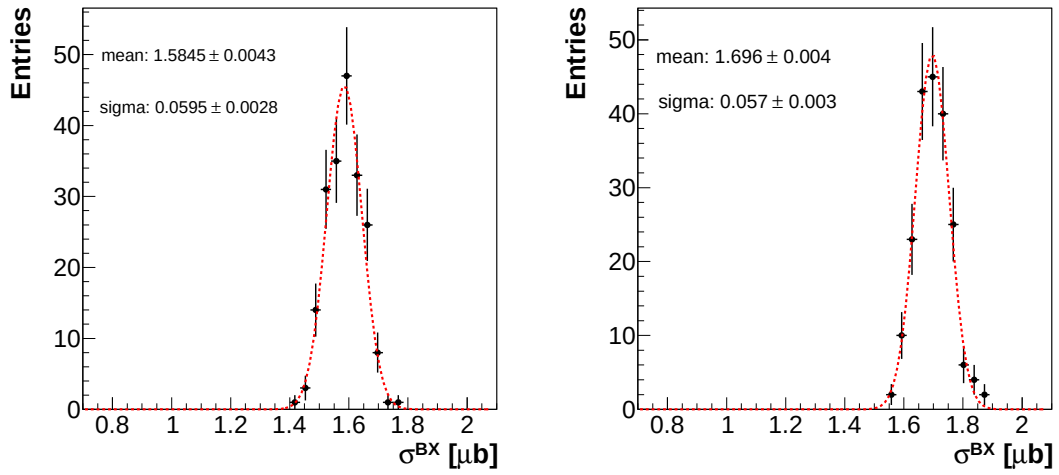


Figure 5.8: Distribution of the σ_k^{BX} variable for the $B_s^0 \to \mu^+\mu^-$ (left) and $B^+ \to J/\psi(\mu^+\mu^-)K^+$ (right) channels. The Gauss function resulting from the fit is shown as a red dotted line.

performed. Finally, hadronic interactions include nuclear reactions of hadrons with the detector material.

The output is represented by hits, which are a representation of the total number of charged particles accumulated in a sensitive detector volume. Truth information for the generated particles used as input by GEANT4 is also available in the output.

5.5 Digitisation

The digitisation step simulates the detector response to GEANT4 hits, which is represented by digits. A digit is typically produced if the resulting voltage or current rises above a pre-defined threshold. Thresholds are specific to each subdetector. The digit usually stores the threshold that has been exceeded, but signal shapes are also stored in the case of some subdetectors.

The digitisation tool also simulates cross-talk between readout channels, electronic noise and channel variations within a subdetector. All these parameters are specific to the subdetector under consideration.

An additional functionality is simulating pile-up events, i.e., additional events produced by multiple interactions in the same bunch crossing or in neighbouring bunch crossings. This is achieved by overlaying hits from several events as input to the digitisation step. Pile-up events should be differentiated from underlying events, which represent multiple interactions produced in the same p-p collision. The latter can be produced in the generation step, using PYTHIA.

The output is stored as Raw Data Objects (RDOs), whose format is identical to that of real data. The typical size for one event is about 2.5 MB. This may

increase depending on the amount of pile-up events included.

5.6 Event Reconstruction

The ATLAS reconstruction software aims at rebuilding the event information based on signals recorded in the detector. It has a modular structure, with a large number of algorithms performing specific tasks [Aad09]. Several categories can be distinguished.

One of the most important categories is represented by *track reconstruction* algorithms. These employ several fitting techniques to reconstruct trajectories of charged particles in the Inner Detector (ID) from space points. The high granularity of the Pixel Detector and the SCT provides precise space points of particle hits. These points are used to reconstruct particle trajectories. TRT drift circle information is also used to complete the particle trajectory in the fit. Track segments are propagated from the first pixel layer to the TRT. A complementary algorithm extrapolates unused TRT segments towards the first pixel layer. The exact knowledge of the magnetic field inside the ID is required for an accurate track extrapolation. Track parameters are also computed. The charge and transverse momentum (q/p_T) are determined based on the curvature of the track in the known magnetic field. The direction of the track (given by the azimuthal ϕ and polar θ angles) is determined after extrapolating it to the point of closest approach to the centre of the detector. The transverse (d_0) and longitudinal (z_0) impact parameters are computed as the distances of closest approach to the centre of the detector. Succesfully reconstructed tracks are stored in a collection.

Another set of algorithms falls in the category of *primary vertex reconstruction*. These algorithms extrapolate reconstructed tracks to the p-p interaction region and fit them to a common vertex. One or more vertices may be found. Typically, the vertex associated to the tracks with the highest total p_T corresponds to the hard scattering process and is labeled the primary vertex. The other vertices are labeled pile-up vertices. The simulation of the samples used in this analysis does not include multiple p-p collisions, so it is not expected to have any pile-up vertices besides the primary vertex. The set of vertices is stored in a collection, together with several parameters, such as the spatial coordinates of the vertex position and the list of tracks associated with the vertex.

Muon track reconstruction is another set of algorithms for reconstructing muon tracks. There are three main approaches. Standalone muon tracks are built based on the track segments provided by the Muon Spectrometer (MS) and then extrapolated to the beam spot. Combined muon tracks are reconstructed by matching tracks based on MS information with Inner Detector (ID) tracks. The first approach has the advantage of a higher pseudorapidity coverage (up to $|\eta| < 2.7$), but presents a lower efficiency around $\eta = 0$ and $|\eta| = 1.2$, regions which are not completely covered by the detector. Combined muon tracks have a lower pseudorapidity coverage (up to $|\eta| < 2.5$), due to the limited acceptance of the ID. But they have the advantage of a better rejection of fake muons, due to the additional ID information. The default muon track collection used in the scope of this analysis is

the combined muon track collection. Finally, a third approach is available, which propagates ID tracks to the MS and matches them with a muon track segment.

Other algorithms important for many types of analyses, but not relevant for the $B_s^0 \to \mu^+\mu^-$ analysis are: *reconstruction of electrons and photons*, *jet reconstruction*, *missing transverse energy reconstruction* and *b-tagging* algorithms.

Chapter 6

Study of the Acceptance Ratio

This chapter presents a Monte Carlo (MC) based method to determine the ratio of the values of the geometric and kinematic acceptances corresponding to the $B_s^0 \to \mu^+\mu^-$ ($\alpha_{B_s^0}$) and $B^+ \to J/\psi(\mu^+\mu^-)K^+$ (α_{B^+}) channels. Section 6.1 explains the procedure adopted for this study. There are two parameters needed for this calculation. The first is the ratio of the b-quark fragmentation probabilities into an u or s-quark, which is determined in Section 6.2. The second parameter is the ratio between the production cross-sections of the two B decay channels. Differences at the generator level and results are discussed in Section 6.3.

6.1 Procedure for Determining the Acceptance Ratio

In order to determine the ratio of the acceptances of the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi(\mu^+\mu^-)K^+$ channels, it is necessary to calculate and compare the corresponding cross-sections. Several computed production cross-sections are available, as explained in Chapter 5. One of them is returned by the Pythia program, σ^{PX}, the cross-section obtained after the hard scattering process and before b-quark hadronisation. This includes the threshold cuts on the `CKIN` variables associated to the hard scattering process. Another is returned by the PythiaB package, σ^{BX}, the cross-section computed after the simulation of the B hadron decay. The decay is forced in Pythia by setting its branching ratio to 1. The σ^{BX} cross-section also includes selection of events based on cuts on kinematic and geometric quantities of final state particles.

The production cross-section of a specific channel k is the product of the corresponding σ_k^{BX}, its branching ratio $\mathcal{B}_k$ and a correction factor 2. In case, subdecays are also forced, the branching ratio is given by the product of all branching ratios, $\prod_i \mathcal{B}_{k,i}$. The factor 2 accounts for charge conjugates, which are not specifically generated. Altogether:

$$\sigma_k^{prod} = 2 \cdot \sigma_k^{BX} \cdot \prod_i \mathcal{B}_{k,i}. \tag{6.1}$$

At the same time, the production cross-section can be defined as σ^{PX} corrected

by the probability $f_{b\bar{b}}$, that a $b\bar{b}$ pair is produced in the hard scattering process, by the probability f_k^{Py}, that the b-quark hadronises into the B meson corresponding to channel k, by the product $\prod_i \mathcal{B}_{k,i}$ of the branching ratios of the forced decays, by a factor α_k corresponding to the selection of final state particles and a factor 2 for charge conjugates:

$$\sigma_k^{prod} = 2 \cdot \sigma^{PX} \cdot f_{b\bar{b}} \cdot f_k^{Py} \cdot \prod_i \mathcal{B}_{k,i} \cdot \alpha_k. \tag{6.2}$$

The cuts on the final state particles limit the allowed η and p_T ranges. Muons coming from either the $B_s^0 \to \mu^+\mu^-$ or the $B^+ \to J/\psi K^+$ decay are only allowed in an η range of [-2.5, 2.5]. This corresponds to the geometrical acceptance of the Inner Detector. Kaons produced in the $B^+ \to J/\psi K^+$ decay have a similar allowed η range. Threshold cuts on muon p_T are set at 4 and 6 GeV, similar to the threshold cuts used in the di-muon trigger selection. Kaons are only accepted with a p_T higher than 0.5 GeV, which is the approximate threshold for reconstructing charged tracks in ATLAS. Plots of these distributions are found in Chapter 5 and in Appendix A. Taking into account, that these geometric and kinematic properties of generated data approximately correspond to those expected in recorded data, it is reasonable to assume the correction factor α_k is a good estimate of the acceptance expected for the particular channel.

(6.1) and (6.2) lead to:

$$\alpha_k = \frac{\sigma_k^{BX}}{\sigma^{PX} \cdot f_{b\bar{b}} \cdot f_k^{Py}}, \tag{6.3}$$

and the ratio of acceptances corresponding to the signal and reference channels becomes:

$$\frac{\alpha_{B^+}}{\alpha_{B_s^0}} = \frac{f_s^{Py}}{f_u^{Py}} \cdot \frac{\sigma_{B^+}^{BX}}{\sigma_{B_s^0}^{BX}}. \tag{6.4}$$

It is assumed, that the cross-section of the hard scattering process, σ^{PX}, does not depend on the decay channel. Therefore, this term cancels out in (6.4). However, this quantity is affected by cuts on the `CKIN` variables applied in the generation. A discussion of the effect of the different cuts applied on $\hat{p}_\perp$ in the generation of the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ data samples is given in Section 6.3.

This method has been proposed and discussed in [Wal09]. The results presented in this Chapter have been independently computed and compared with the original results.

6.2 b-quark Fragmentation Ratio

This section presents a method to compute the f_s^{Py}/f_u^{Py} factor in (6.4). It is important to note, that these variables are specific to the internal structure of the Pythia generator. Therefore, they are different from those in the f_u/f_s factor (4.1), which are determined experimentally. In order to determine this ratio several

inclusive samples $b\bar{b} \to X$ are generated using PYTHIA. Different settings for $\hat{p}_\perp$ and $p_T(\bar{b})$ are used for each sample in order to investigate whether these variables have an effect on the final result. Samples are generically labeled `bbXi`, where `i` represents a specific setting of selection variables at generator level, as shown in Table 6.1. Each data set consists of 200 000 events.

Data set	$\hat{p}_\perp >$ [GeV]	$p_T(\bar{b}) >$ [GeV]	σ_i^{PX} [mb]	σ_i^{BX} [μb]
`bbX1`	10	10	4.1812 ± 0.0005	54.20 ± 0.15
`bbX2`	10	2.5	4.1817 ± 0.0008	146.1 ± 0.4
`bbX3`	5	10	33.3024 ± 0.0015	103.3 ± 0.2
`bbX4`	5	2.5	33.309 ± 0.003	537.9 ± 1.3
`bbX5`	15	10	1.0950 ± 0.0017	25.45 ± 0.06
`bbX6`	15	2.5	1.0951 ± 0.0003	50.30 ± 0.12

Table 6.1: Data sets used to determine the f_s^{Py}/f_u^{Py} ratio, the corresponding selection cuts at generator level and average cross-sections returned by PYTHIA and PythiaB. The data sets are provided by [Wal09].

The cross-section variables, σ_i^{PX} and σ_i^{BX}, are calculated by the PYTHIA generator using subsamples of 1000 events. The average values are shown in the last two columns of Table 6.1. As expected, there is no statistical difference between σ_i^{PX} values for samples with the same threshold cut applied on $\hat{p}_\perp$. A looser threshold cut on $\hat{p}_\perp$ ($p_T(\bar{b})$) results in a higher value of σ_i^{PX} (σ_i^{BX}).

The number of generated $\bar{b}$-quarks, $N_i^{\bar{b}}$, is calculated for each sample. The decay chain produced after the hadronisation of $\bar{b}$-quarks is analysed and the number of generated B_s^0 (B^+), $N_i^{B_s^0}$ ($N_i^{B^+}$), particles is computed. Then the fragmentation probability ratios are calculated:

$$f_{s,i}^{Py} = \frac{N_i^{B_s^0}}{N_i^{\bar{b}}} \tag{6.5}$$

and

$$f_{u,i}^{Py} = \frac{N_i^{B^+}}{N_i^{\bar{b}}}. \tag{6.6}$$

These variables are computed for each data set and their ratios, $f_{s,i}^{Py}/f_{u,i}^{Py}$, are given in Table 6.2. The same procedure is then repeated for charge conjugates and the ratios $f_{\bar{s},i}^{Py}/f_{\bar{u},i}^{Py}$ are shown in the last column of Table 6.2. All computed values are statistically consistent with each other. The average value of f_s^{Py}/f_u^{Py} is calculated to 0.2123 ± 0.0010.

In a next step, the uniformity of the fragmentation ratio with respect to various kinematic variables is investigated. Fig. 6.1 shows the dependence of the

Data set	$f_{s,i}^{Py}/f_{u,i}^{Py}$	$f_{\bar{s},i}^{Py}/f_{\bar{u},i}^{Py}$
bbX1	0.2121 ± 0.0018	0.2118 ± 0.0018
bbX2	0.2133 ± 0.0018	0.2137 ± 0.0018
bbX3	0.2136 ± 0.0018	0.2129 ± 0.0018
bbX4	0.2109 ± 0.0018	0.2107 ± 0.0018
bbX5	0.2131 ± 0.0018	0.2108 ± 0.0018
bbX6	0.2120 ± 0.0018	0.2128 ± 0.0018

Table 6.2: Computed b-quark fragmentation probabilities into an s- or u-quark. $f_{s,i}^{Py}/f_{u,i}^{Py}$ represents the ratio of probabilities that a $\bar{b}$-quark fragments into an s- or a u-quark. $f_{\bar{s},i}^{Py}/f_{\bar{u},i}^{Py}$ represents the ratio of probabilities that a b-quark fragments into an $\bar{s}$- or a $\bar{u}$-quark.

f_s^{Py}/f_u^{Py} ratio on the B-meson transverse momentum determined from the data sets where a threshold cut of 10 (left) or 2.5 GeV (right) applied on $p_T(\bar{b})$. The drop below 10 GeV observed in the left plot is induced by the threshold cut applied on $p_T(\bar{b})$. In those events, where $p_T(\bar{b})$ is just above threshold, the energy available for the B meson produced in the hadronisation is limited, and therefore the probability of a hadronisation into the lighter B^+ is higher. A similar behaviour is observed for the data sets with a $p_T(\bar{b})$ threshold cut of 2.5 GeV.

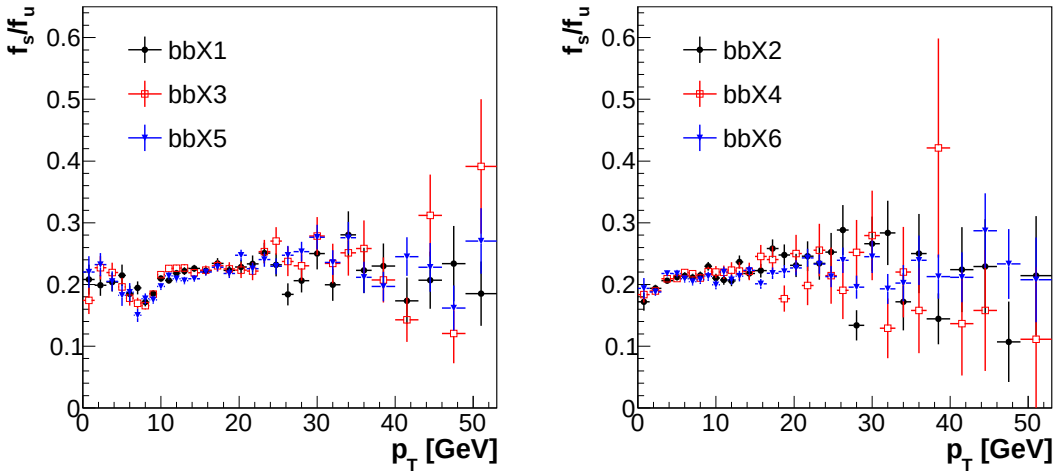


Figure 6.1: Dependence of f_s^{Py}/f_u^{Py} on the B meson transverse momentum determined on the data samples with a threshold cut on $p_T(\bar{b})$ of 10 (left) or 2.5 GeV (right).

Fig. 6.2 shows a similar dependence of $f_{\bar{s}}^{Py}/f_{\bar{u}}^{Py}$ on the $\bar{B}$ meson transverse momentum. As expected, the drop observed in the previous case is not present here, as there is no threshold cut imposed on $p_T(b)$. The dependences on other kine-

matic variables (η and ϕ) are studied in Appendix B, but no significant differences are observed.

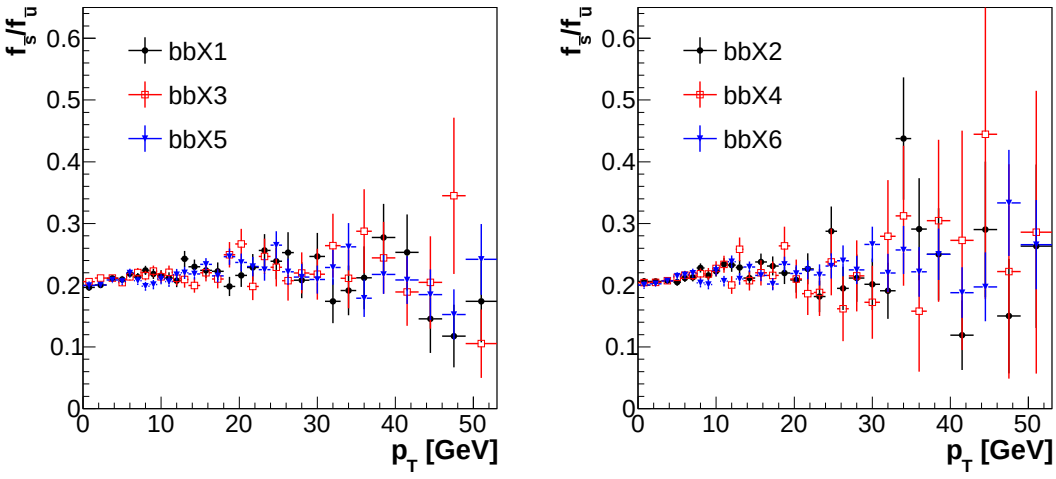


Figure 6.2: Dependence of $f_{\bar{s}}^{Py}/f_{\bar{u}}^{Py}$ on the B meson transverse momentum determined on the data samples with a threshold cut on $p_T(\bar{b})$ of 10 (left) or 2.5 GeV (right).

6.3 Determining the Cross Section Ratio

The next factor in the ratio of acceptances (6.4) to be investigated is the ratio of the PYTHIA cross-sections of the generated channels. In the ideal case, the two samples should be produced with the same set of parameters describing the hard scattering process. This ensures that the σ^{PX} factor cancels out in (6.4). However, in the central production of the fully simulated MC data sets, two different threshold cuts were imposed on the $\hat{p}_\perp$ variable, which affects the production mechanism of each decay channel. This is an undesired feature, since it is expected, that events recorded by the detector are produced under similar conditions of p-p collisions. The $b\bar{b}$ production is the same regardless of the type of B mesons produced in the subsequent hadronisation process and of the decay chain of the B mesons. This is an artifficial effect introduced by the way the fully simulated MC samples are produced. Differences in the acceptance corresponding to each channel are determined by detector acceptances to observe their final state particles and must not depend on the $b\bar{b}$ production mechanisms.

In order to ensure similar conditions for the $b\bar{b}$ production in the generation of the two decay channels, several other data sets are produced and analysed, complementary to those presented in Table 5.2. These are summarised in Table 6.3. The `Bs_1` (`B+_1`) data set is the same as the `Bs_mu6mu4` (`B+_mu6mu4K+`) data set presented in Table 5.2. The `Bs_2` data set contains at least one $B_s^0 \to \mu^+\mu^-$ decay in each event and has the same generator level cuts as the `B+_1` data set. Similarly,

the `B+_2` data set contains at least one $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decay per event and has the same generator level cuts as the `Bs_1` data set. All data sets contain 200 000 generated events each. The values of the σ^{PX} and σ^{BX} cross-sections have been computed by averaging the values returned by the PYTHIA generator and PythiaB package for smaller subsamples of each data set. As expected, it is observed, that the values of the σ^{PX} variable are identical for data sets with the same threshold cut on $\hat{p}_\perp$.

Data set	$\hat{p}_\perp >$ [GeV]	$p_T(\bar{b}) >$ [GeV]	σ^{PX} [mb]	σ^{BX} [μb]
`Bs_1`	10	10	4.1817 ± 0.0003	1.585 ± 0.004
`Bs_2`	15	10	1.0948 ± 0.0001	0.919 ± 0.003
`B+_1`	15	10	1.0950 ± 0.0001	1.696 ± 0.004
`B+_2`	10	10	4.1814 ± 0.0001	2.296 ± 0.006

Table 6.3: Data sets used in acceptance studies, the cuts used at the generator level and the corresponding PYTHIA and PythiaB cross-sections. The data sets are provided by [Wal09].

The situation with similar $\hat{p}_\perp$ threshold cuts is investigated. Two cases are distinguished: the threshold cut on $\hat{p}_\perp$ in both data sets is set to 10 GeV (the `Bs_1` and `B+_2` data sets), or 15 GeV (the `Bs_2` and `B+_1` data sets). The results are summarised in Table 6.4. The value given for $f^{Py}_{B^0_s}/f^{Py}_{B^+}$ is the average computed in the previous section. The ratio of the cross-sections is given in the fourth column of the table.

Data sets	$\hat{p}_\perp >$ [GeV]	$f^{Py}_{B^0_s}/f^{Py}_{B^+}$	$\sigma^{BX}_{B^+}/\sigma^{BX}_{B^0_s}$	$\alpha^{BX}_{B^+}/\alpha^{BX}_{B^0_s}$
`Bs_1` and `B+_2`	10	0.2123 ± 0.0010	1.449 ± 0.005	0.3078 ± 0.0019
`Bs_2` and `B+_1`	15		1.845 ± 0.005	0.391 ± 0.002

Table 6.4: Ratio of the PythiaB cross-sections and of the computed acceptances for the data sets with threshold cuts of 10 and 15 GeV on $\hat{p}_\perp$.

The last column of the table shows the results for the acceptance ratio in the two cases. The final result is computed as the average of the two values. The differences in the result, caused by different settings at generator level, are assigned to the systematic uncertainty. The relative systematic uncertainty is about 12% and dominates the relative statistical uncertainty of about 0.6%.

$$\frac{\alpha_{B^+}}{\alpha_{B^0_s}} = 0.349 \pm 0.002(stat) \pm 0.042(syst). \tag{6.7}$$

Chapter 7

Trigger Studies

In this chapter the investigation of the event selection performed by the trigger system is described. The ATLAS trigger system was explained in detail in Chapter 3 and the trigger strategy was introduced in Chapter 4. The search for $B_s^0 \to \mu^+\mu^-$ decays is based on a di-muon trigger chain. Here the effects of several di-muon trigger chains are studied on dedicated Monte Carlo (MC) data samples corresponding to the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays. The aim is to estimate the trigger selection efficiencies for these particular data sets. The ratios of these efficiencies are among the inputs employed in the branching ratio calculation, as given in (4.1).

Section 7.1 describes the method used to determine the trigger efficiency. An estimate of the efficiencies of the di-muon trigger with threshold cuts of 6 and 4 GeV are also given for the different MC data samples. The behaviour of the trigger efficiency as a function of various kinematic variables is studied in Section 7.2. Alternative trigger scenarios are briefly discussed in Section 7.3.

7.1 Trigger Efficiency Determination

The trigger efficiency can be determined from recorded data using the tag and probe method [Has10], [Aad09]. Data unbiased by the trigger selection under investigation is analysed and events containing muon pairs originating in a common vertex are isolated. Candidates for $J/\psi \to \mu^+\mu^-$ or $\Upsilon \to \mu^+\mu^-$ decays are ideal for trigger performance studies in the low muon p_T regime, due to the high statistics available. Stringent reconstruction and identification criteria are imposed on one of the decay muons. This muon is labeled tag muon. The second muon is labeled probe muon. The method investigates whether the probe muon passed the trigger decision that is being investigated. The single muon efficiency is determined as a function of the kinematic and geometric properties (p_T, η, ϕ) of the probe muon. Di-muon efficiency maps are then determined based on the single muon efficiencies, as a function of the p_T, η and ϕ properties of each muon. Determining the trigger efficiency for a specific decay involves the convolution of the efficiency map with a function describing the corresponding muon p_T, η and ϕ spectrum.

In the case of MC data, a simpler method is used. A trigger simulation software

is executed during the event reconstruction process. This software tool simulates the muon reconstruction and trigger decision processes in the three trigger stages (Level 1, Level 2 and Event Filter). Each event is then given a set of flags, each flag corresponding to the decision of a certain trigger item. The efficiency can then be statistically estimated based on the total number of events and the number of events that pass a certain trigger.

The efficiency is computed as:

$$\varepsilon^{trigger} = \frac{N_{pass}}{N_{total}}. \tag{7.1}$$

It is determined based on a Bayesian approach to construct the probability distribution function $P(\varepsilon|N_{pass}, N_{total})$ of the true efficiency ε given N_{pass}, the number of events, that pass the selection criteria, and N_{total}, the total number or events [Ull08], [Cas09]. The mode of such a probability distribution function is the classical interpretation of efficiency: $\frac{N_{pass}}{N_{total}}$. The expectation value is computed to $\frac{N_{pass}+1}{N_{total}+2}$. In this interpretation, the expectation value of the efficiency is biased in the case of small N_{total}, but approaches the value of the mode when N_{total} is reasonably large. This is the case in most instances of efficiency calculations presented in this thesis. The Bayesian approach offers the advantage of a correct treatment of statistical uncertainties in the limiting cases, when N_{pass} approaches 0 or N_{total}. In such situations, an incorrect treatment of the probability distribution function of the true efficiency (e.g., a binomial distribution) results in an unphysical statistical uncertainty of 0. Using the approach mentioned above, the statistical uncertainty can be computed as a 1-σ interval centred around the expectation value, which yields a non-zero value of the uncertainty in the limiting cases. This corresponds to an asymmetric error interval around the mode. This method is implemented in the `TGraphAsymmErrors` class of the ROOT [Ant09] analysis package, which is used throughout the analysis in this thesis.

The `Bs_mu6mu4`, `B+_mu6mu4K+` and `bb_mu6mu4X` data samples are analysed and the di-muon trigger chain with muon p_T threshold cuts of 6 and 4 GeV is investigated. All data samples are generated with similar p_T threshold cuts imposed on muon transverse momenta. Thus, at truth level, each event fulfills the requirements to pass this trigger. The trigger decision is investigated for each event and the efficiency is determined from (7.1) using the total number of events of each sample and the number of events passing the trigger selection. Efficiencies are determined at each stage of the trigger chain (Level 1, Level 2 and Event Filter) and are summarised in Table 7.1. The overall efficiency is about 34% (35%) for the `Bs_mu6mu4` (`B+_mu6mu4K+`) data sample. This results in a ratio of:

$$\frac{\varepsilon^{trigger}_{B^+}}{\varepsilon^{trigger}_{B^0_s}} = 1.042 \pm 0.007\,(stat), \tag{7.2}$$

which is later used in the calculation of the $B^0_s \to \mu^+\mu^-$ branching ratio.

There are two main sources for the efficiency loss. One is the geometry of the ATLAS detector, which contains several regions not fully covered by detector components. These are investigated in Section 7.2. Another source is the limited

Trigger signature	Bs_mu6mu4	B+_mu6mu4K+	bb_mu6mu4X
L1_2MU4_MU6	0.630 ± 0.002	0.6394 ± 0.0012	0.5844 ± 0.0005
L2_mu4_mu6 (w.r.t. L1)	0.812 ± 0.002	0.7906 ± 0.0013	0.6912 ± 0.0006
L2_mu4_mu6	0.512 ± 0.002	0.5056 ± 0.0013	0.3965 ± 0.0005
EF_mu4_mu6 (w.r.t. L2)	0.666 ± 0.003	0.7028 ± 0.0012	0.6427 ± 0.0008
EF_mu4_mu6	0.341 ± 0.002	0.3554 ± 0.0012	0.2548 ± 0.0004

Table 7.1: Trigger efficiencies computed for the Bs_mu6mu4, B+_mu6mu4K+ and bb_mu6mu4X data sets, for the trigger chain with thresholds of 6 and 4 GeV on muon p_T.

precision in reconstructing the muons' transverse momenta. This is typically seen as a turn-on curve of the efficiency as a function of muon p_T in the region of the threshold value. Muons with truth p_T above the threshold value may have a lower reconstructed p_T, thus failing the trigger decision.

Another effect is produced by muons, with truth p_T below the threshold values, which are accepted by the trigger (due to misreconstructed p_T). Since all data samples are generated with the same muon p_T threshold cuts as in the EF_mu4_mu6 trigger item, this effect has to be studied on samples generated with lower muon p_T threshold cuts. For this purpose, the Bs_mu2p5mu2p5 and bb_mu2p5mu2p5X samples are analysed (where the p_T of both muons at generator level is required to be above 2.5GeV). The Bs_mu2p5mu2p5 sample contains similarly generated decays as the Bs_mu6mu4 sample (i.e., $B_s^0 \to \mu^+\mu^-$). Since no data sample corresponding to B+_mu6mu4K+ exists in the official production, the bb_mu2p5mu2p5X sample can be used as a replacement by selecting only the subset of events containing true $B^+ \to J/\psi K^+$ decays.

The total number of events N_{truth} containing two muons with truth transverse momenta above 6 and 4 GeV are given in Table 7.2 for both data sets. The number of events N_{trig} with both muons above the 6 and 4 GeV threshold cuts and fulfilling the trigger decision is also shown. The efficiency computed as $\varepsilon = N_{trig}/N_{truth}$ corresponds in rough terms to the efficiency computed above. The number of events N_{below}, with at least one muon truth p_T below the threshold cuts, but accepted by the trigger, is also given. The fraction of these events with respect to N_{trig} is computed to 0.055 ± 0.002 (0.14 ± 0.04) for the Bs_mu2p5mu2p5 (bb_mu2p5mu2p5X) data sets. This represents a systematic uncertainty of the trigger decision due to the momentum uncertainty in the threshold region. This uncertainty is then propagated to the ratio of trigger efficiencies $\varepsilon_{B^+}^{trigger}/\varepsilon_{B_s^0}^{trigger}$ resulting in a systematic uncertainty of about 15%:

$$\frac{\varepsilon_{B^+}^{trigger}}{\varepsilon_{B_s^0}^{trigger}} = 1.042 \pm 0.007\,(stat) \pm 0.156\,(syst). \tag{7.3}$$

Differences between the fraction of events N_{below}/N_{trig} observed for the Bs_mu2p5mu2p5 and the bb_mu2p5mu2p5X data sets stem from slightly different cuts applied at the

	Bs_mu2p5mu2p5	bb_mu2p5mu2p5X
N_{truth}	30 216	197
N_{trig}	10 056	79
$\varepsilon = N_{trig}/N_{truth}$	0.333 ± 0.003	0.40 ± 0.03
N_{below}	557	11
$f = N_{below}/N_{trig}$	0.055 ± 0.002	0.14 ± 0.04

Table 7.2: Number of truth decays N_{truth} with muons above the threshold cuts of 6 and 4 GeV. Number of events N_{trig} with both muons above the threshold cuts, which are accepted by the trigger. Number of events N_{below} with at least one muon below the threshold cuts, which are accepted by the trigger.

generator level. Slightly higher threshold cuts applied on the $\hat{p}_\perp$ and the $\bar{b}$-quark p_T (see Chapter 5) produce a harder p_T spectrum of the decay muons for the bb_mu2p5mu2p5X sample. The p_T spectra of the hardest and second hardest muons are shown in Fig. 7.1.

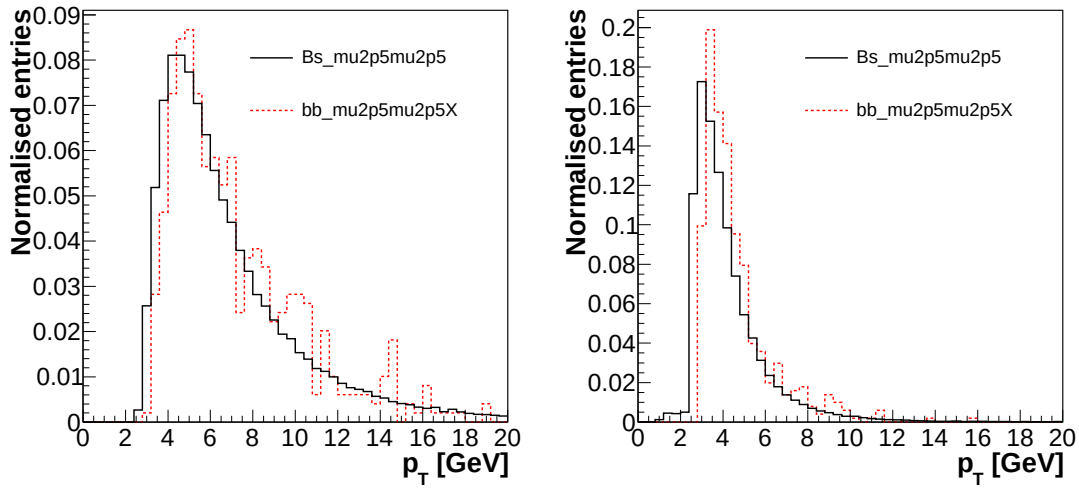


Figure 7.1: Reconstructed p_T spectrum of the hardest (left) and second hardest (right) muons observed in the Bs_mu2p5mu2p5 (black solid) and bb_mu2p5mu2p5X (red dashed) data sets. Each sample is normalised to unity.

The overall trigger efficiency of about 25% observed for the bb_mu6mu4X data sample is considerably smaller than for the other two data sets (see Table 7.1). This is also due to different selection cuts used in the generation (see Chapter 5). A lower threshold cut used for the accepted $\bar{b}$-quark transverse momentum produces a softer muon p_T spectrum for this data set. This is shown in Fig 7.2, where the p_T distributions of the hardest and second hardest muons in each data sample are displayed. The trigger efficiency for the bb_mu6mu4X sample plays no role in the

computation of the branching ratio, since the same data set is used to estimate the number of $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays, and is only given for reference.

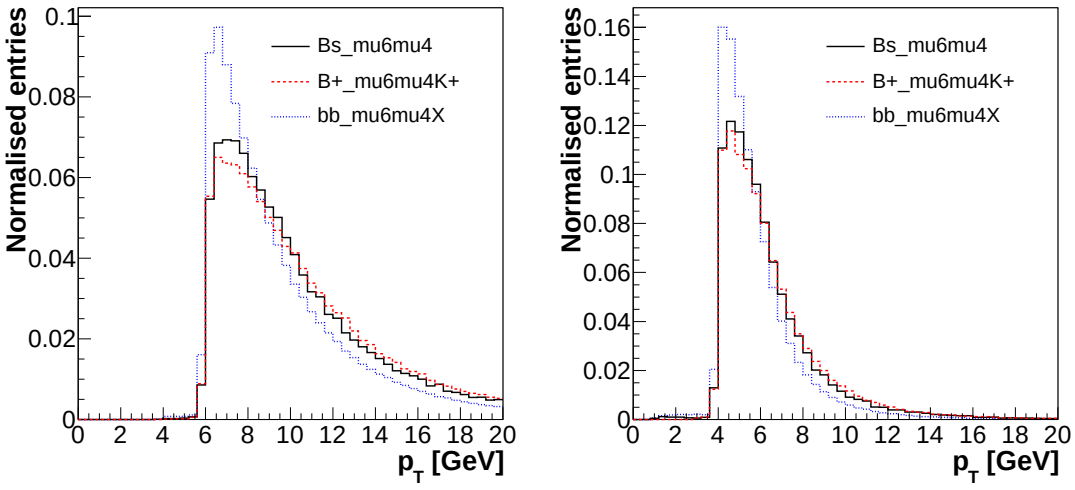


Figure 7.2: Reconstructed p_T spectrum of the hardest (left) and second hardest (right) muons observed in the `Bs_mu6mu4` (black solid), `B+_mu6mu4K+` (red dashed) and `bb_mu6mu4X` (blue dotted) data sets. Each sample is normalised to unity.

The slight differences observed in the muon p_T spectra for the `Bs_mu6mu4` and the `B+_mu6mu4K+` samples are also due to a higher cut applied on the $\hat{p}_\perp$ variable at generator level for the `B+_mu6mu4K+` sample. This produces a slightly harder muon p_T spectrum (visible in the tail).

7.2 Trigger Efficiency as a Function of Kinematic Variables

The trigger efficiency as a function of the true B_s^0 (B^+) p_T, η and ϕ is studied for the `Bs_mu6mu4` (`B+_mu6mu4K+`) data samples. Each trigger stage is investigated separately. Fig. 7.3, Fig. 7.4 and Fig. 7.5 show the corresponding distributions for the Level 1, Level 2 and Event Filter, respectively.

Several efficiency drops are observed at Level 1 in the efficiency as a function of η and ϕ. The drop around $|\eta| \sim 0$ is due to passive material. Differences in the shape are expected due to different opening angles between muons originating in B_s^0 or B^+ decays. Since B_s^0 particles have on average lower transverse momenta (due to generator cuts), the muon opening angles are larger and the drop around $|\eta| \sim 0$ becomes wider. Other efficiency losses are observed around $|\eta| \sim 1.05$, which is the region of transition from barrel to endcap, where the probability of a coincidence to be found is smaller. Similar drops are found in the efficiency as a function of ϕ around $\phi \sim -1.2$ and -2, which are also regions of passive

material (ATLAS support structures). Similar effects are observed in dedicated trigger studies, e.g. [Aad09].

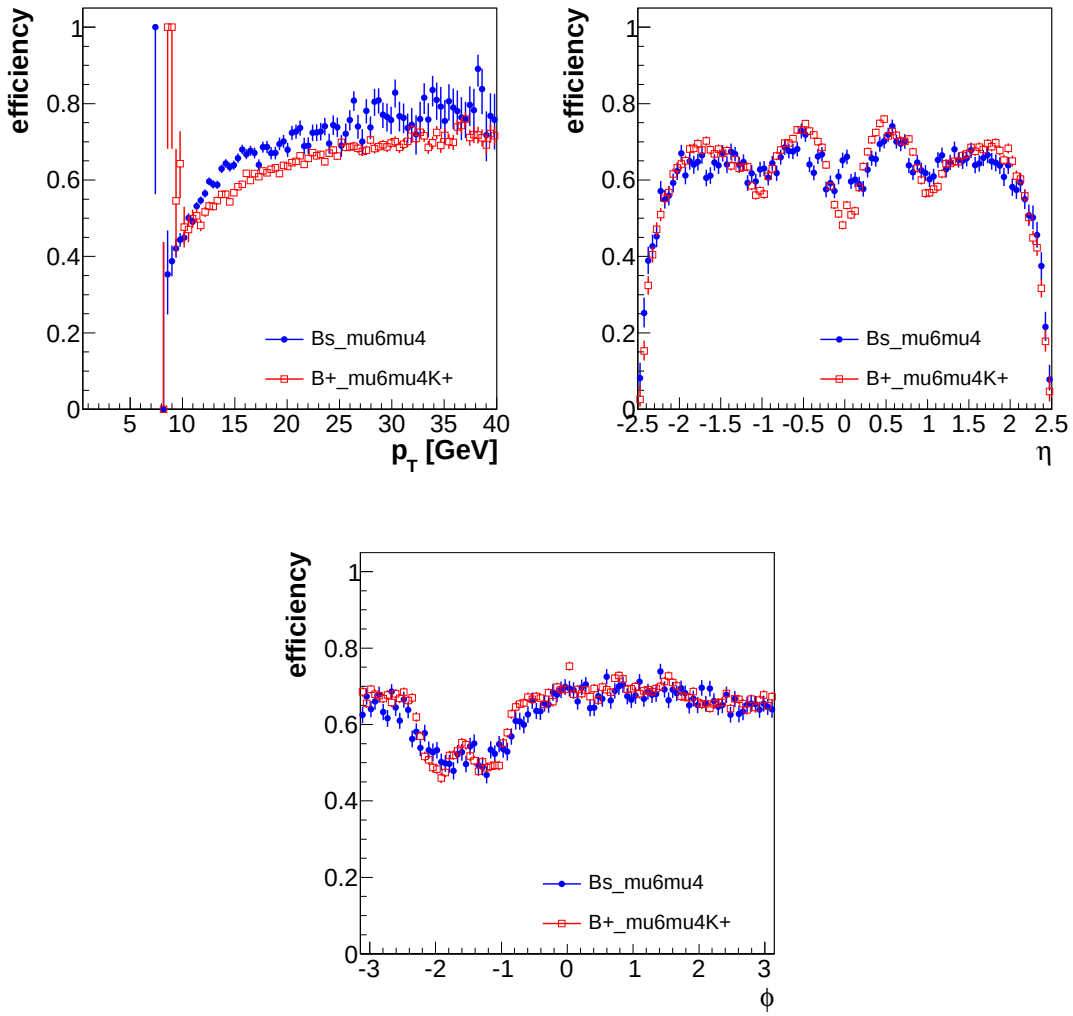


Figure 7.3: `L1_2MU4_MU6` trigger efficiency as a function of the true B_s^0 (blue circles) and B^+ (red squares) p_T, η and ϕ.

These efficiency losses are propagated to higher trigger stages, since only events that pass the Level 1 decision are accepted at the next stages. The only noticeable effect is that the shape of the efficiency drops (at $\eta = 0$, etc.) become wider due to the smearing effects inherent to the reconstruction processes at Level 2 and Event Filter.

Another difference is observed in the efficiency as a function of B_s^0 and B^+ transverse momenta. It is caused by the generation process, which forces higher B^+ transverse momenta for events generated with the same muon p_T threshold cuts (due to the momentum of the additional particle produced in the decay). This shifts the efficiency to lower values when plotted against B-meson p_T. The effect is seen

at all trigger stages.

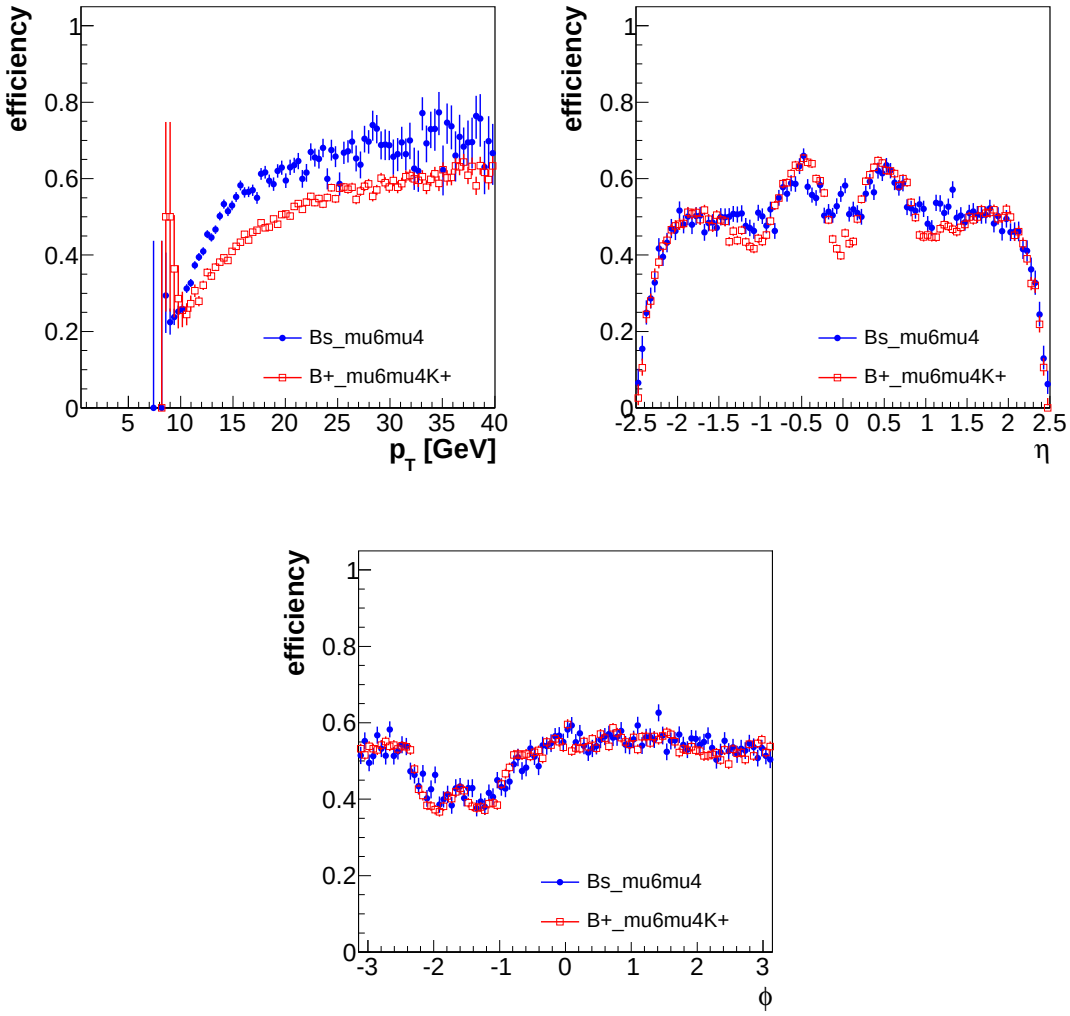


Figure 7.4: Total `L2_mu4_mu6` trigger efficiency as a function of the true B_s^0 (blue circles) and B^+ (red squares) p_T, η and ϕ.

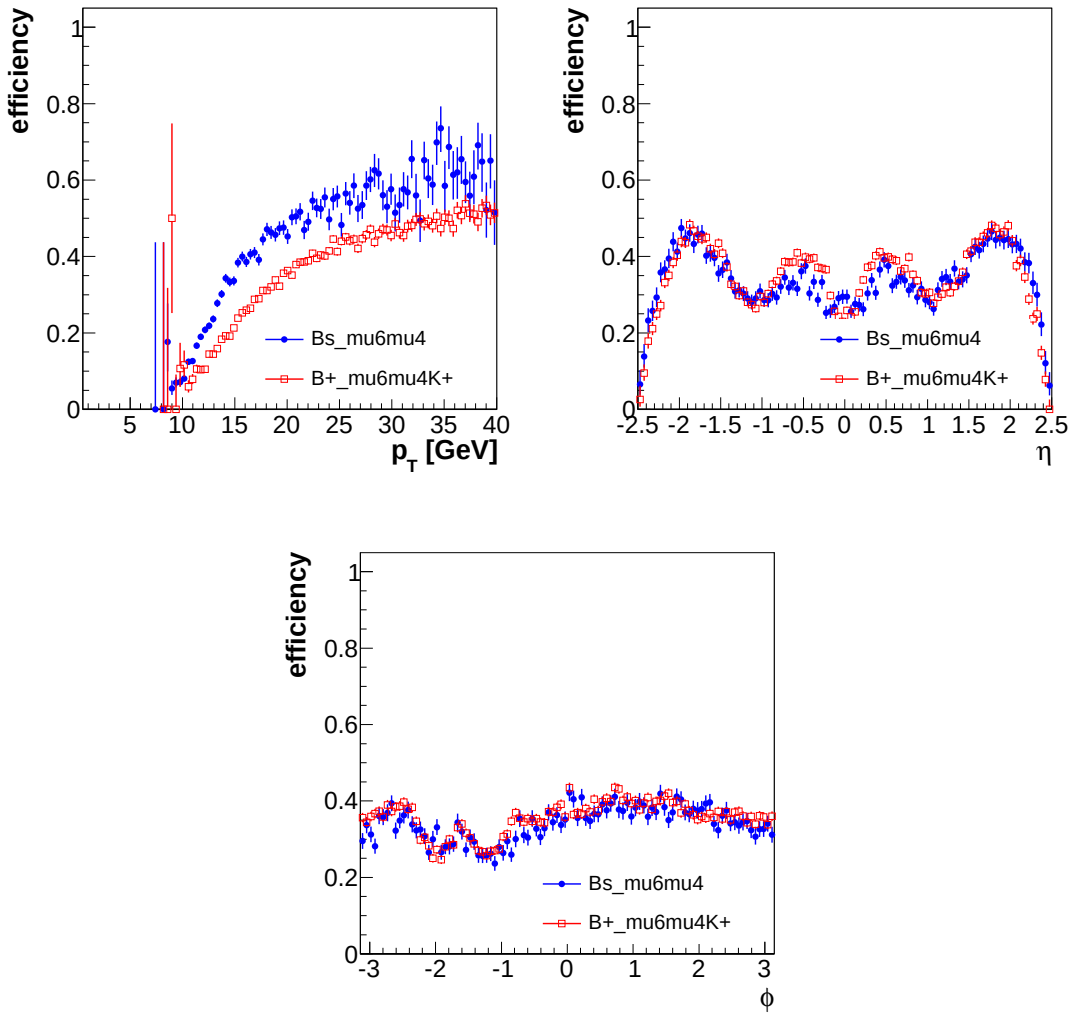


Figure 7.5: Total `EF_mu4_mu6` trigger efficiency as a function of the true B_s^0 (blue circles) and B^+ (red squares) p_T, η and ϕ.

7.3 Alternative Trigger Scenarios

Several other trigger scenarios are available for data taking. Simple di-muon triggers with threshold cuts of 4 (`EF_2mu4`) or 6 GeV (`EF_2mu6`) on both muons exist in the ATLAS trigger menu. Another trigger chain is based on two 4 GeV muons, with an additional requirement in the Event Filter that both muons originate in a common vertex (`EF_2mu4_DiMu`). An even more complex trigger is that where, in addition to the common vertex requirement, an invariant mass cut is imposed on the di-muon vertex (`EF_2mu4_Bmumu`). The invariant mass range accepted for the di-muon vertex is [4 GeV, 7 GeV]. All these scenarios are summarised in Table 7.3. First studies of trigger performance with $\sqrt{s} = 7$ TeV data [Atl10e] have shown that the `EF_2mu4` trigger rate is small enough to be used at lower instantaneous

luminosities (up to about $10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$). For higher instantaneous luminosities additional requirements are necessary (e.g., higher threshold values or the common vertex requirements).

Trigger signature	Bs_mu6mu4	B+_mu6mu4K+
EF_2mu4	0.356 ± 0.002	0.3757 ± 0.0013
EF_2mu6	0.2039 ± 0.0018	0.2049 ± 0.0010
EF_2mu4_DiMu	0.312 ± 0.002	0.2722 ± 0.0012
EF_2mu4_Bmumu	0.705 ± 0.002	0.00379 ± 0.00016

Table 7.3: Trigger efficiency for the EF_2mu4, EF_2mu6, EF_2mu4_DiMu, EF_2mu4_Bmumu trigger items observed in the Bs_mu6mu4 and B+_mu6mu4K+ data sets.

For the analysis presented in this thesis, however, these scenarios are not suitable. The trigger signatures with two 4 GeV muons are below the threshold cuts used in the generation, and their effects cannot be accurately estimated. The trigger signature with two 6 GeV muons is also not ideal as it would drastically limit the available statistics of the already limited MC data sets. Therefore, the EF_mu4_mu6 trigger signature, which is similar to other trigger signatures used in the first data taking periods, will be used in further analysis.

Chapter 8

Offline Event Selection

This chapter presents the event selection procedure based on reconstructed Monte Carlo (MC) events. The aim is to extract the number of candidates for the $B_s^0 \to \mu^+\mu^-$ (the signal channel) and $B^+ \to J/\psi(\mu^+\mu^-)K^+$ (the reference channel) decays from the same data sample. The combinatorial background sample is used for this purpose, as discussed in Chapter 4. The number of observed candidates serves as input to determine an expected upper limit on the $B_s^0 \to \mu^+\mu^-$ branching ratio, as explained in Chapter 4. Other inputs for the branching ratio calculation are the candidate reconstruction and selection efficiencies. For this purpose, dedicated data samples containing the B_s^0 and B^+ decay channels are analysed with similar selection criteria as the background sample.

Two algorithms are implemented for reconstructing B_s^0 or B^+ candidates. These are presented in Section 8.1. The selection of candidates requires the candidates to pass trigger selection criteria, summarised in Section 8.2, and a series of baseline quality cuts, described in Section 8.3. In a next step, a cut-based analysis is employed. The aim is to suppress background events as much as possible, while achieving a high retention rate of true signal events. Several discriminating variables are introduced and their rejection power is studied on the background and dedicated signal data sets. These variables are discussed in Section 8.4 and their selection efficiencies for the dedicated signal samples are described in Section 8.6. Other efficiencies inherent to the reconstruction process are also discussed. The background sample is analysed in Section 8.7 and the number of B_s^0 and B^+ candidates is estimated.

8.1 Reconstruction of B_s^0 and B^+ Candidates

The search for B_s^0 or B^+ candidates is performed on reconstructed data. A general overview of the ATLAS reconstruction is given in Chapter 5. The MC data considered here are stored in AOD format, which is identical for real events. Therefore, the B_s^0 (B^+) finding algorithms can also be executed with minor adjustments on recorded data. The algorithms are implemented and executed within the Athena software framework [Aad10]. Two distinct algorithms are implemented for the search of B_s^0 and B^+ candidates. Most differences stem from the fact that B_s^0 can-

didates are represented by 2-prong vertices ($\mu\mu$), while B^+ candidates are represented by 3-prong vertices ($\mu\mu K$).

8.1.1 B_s^0 Finding Algorithm

The B_s^0 finding algorithm, named `Bs2mumu`, starts by searching for two muon candidates. Only muons reconstructed by combining Inner Detector (ID) and Muon Spectrometer (MS) information are considered. The geometrical acceptance of these muons is defined by a range in pseudorapidity of $|\eta| < 2.5$, roughly corresponding to the detector's acceptance region. Combined muon tracks are then matched against ID tracks based on their kinematic properties (p_T, η, ϕ). In the following, muon track candidates are defined as ID tracks successfully matched with combined muon tracks. In addition, a track quality selection is performed, based on the number of hits the reconstructed track leaves in the Inner Detector, the impact parameter with respect to the reconstructed primary vertex and the minimum reconstructed transverse momentum. Each track is required to have at least:

- 1 hit in the B-layer,
- 2 hits in the Pixel Detector (except the B-layer),
- 7 hits in the silicon-based detectors (Pixel Detector and SCT).

The cuts on the transverse (longitudinal) impact parameter, d_0 (z_0), are:

- $|d_0| < 2$ mm,
- $|z_0| < 10$ mm.

In addition, only tracks with a reconstructed p_T above 0.5 GeV are accepted.

The track selection criteria are based on previous Inner Detector performance studies [Aad09]. The requirement of a minimum number of hits (in the Pixel Detector and SCT) improves the quality of tracks, resulting in a better precision in subsequent calculations. It also rejects incompletely reconstructed TRT track segments. The cuts on the track impact parameter reject tracks from material interactions, such as photon conversions, which are not desired in this analysis. These selection criteria slightly affect the selection efficiency, which is later discussed in Section 8.6.

If an event contains at least two successfully reconstructed muon tracks, a vertex finder will be employed. The `VKalVrt` vertex finder [Kos03] is based on a Kalman filter method. This algorithm extrapolates input particle tracks towards the center of the detector and checks whether they originate in a common vertex. If the fit is successful, the covariance matrix of the fit is computed based on the errors of the reconstructed tracks. The invariant mass of the decaying particle is also calculated based on the momenta and masses of the decay products. The particle track reconstruction gives information regarding the momenta and charges, but no information regarding the masses of particles. Therefore, particle types

considered in the vertex fit are generally unknown and a mass hypothesis has to be assigned for the invariant mass calculation. For example, in the case of a di-muon vertex fit, the known muon mass is assigned to the particle tracks. Successfully reconstructed vertices may include true particles different from those assumed in the mass hypothesis (i.e., wrongly assumed particle masses). This causes a certain amount of combinatorial background. Muon tracks make use of additional information in comparison with ID-only tracks. By design, most non-muon particles are stopped in the calorimeters. Therefore, reconstructed muon tracks are mostly produced by true muon particles, with a low rate of fake muon tracks (around 0.1% in the low p_T regime [Aad09]). Thus, the mass hypothesis is expected to be more accurate, when muon tracks are used in comparison to when ID only tracks are employed. The mass value for muon candidates considered in the fit is 105.6 MeV.

No additional selection criteria are used at this point in the `Bs2mumu` algorithm. Several kinematic variables, used later for background suppression, are computed for each successfully reconstructed candidate. These are stored into a ROOT [Ant09] ntuple for further analysis (described in Section 8.4).

8.1.2 B^+ Finding Algorithm

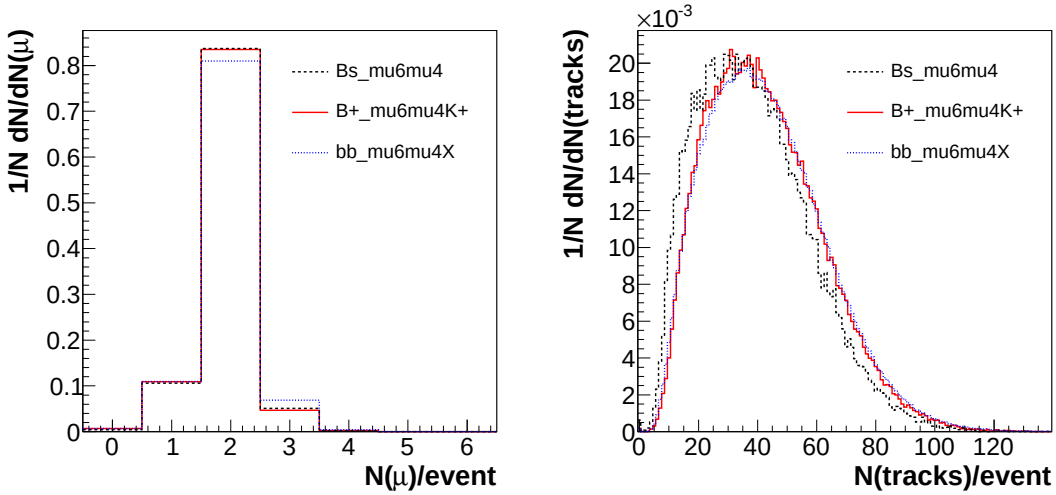
The B^+ finding algorithm, named `Bplus2JpsiKplus`, uses identical selection criteria for muon and ID tracks as the B_s^0 finding algorithm. A different approach is needed due to the different topology of the decay (practically a 3-prong vertex). Therefore, in a first step, the two muon candidates are fitted to a common vertex, corresponding to the $J/\psi \to \mu^+\mu^-$ subdecay, in a similar manner to the di-muon vertex fit used in the `Bs2mumu` algorithm. However, in this case an additional constraint is used in the fit, requiring that the resulting vertex invariant mass is within a 150 MeV window around the known J/ψ mass (3097 MeV [Nak10]). A good quality of the di-muon vertex and a low rate of fake vertices due to combinatorics is desired at this point. In a next step, the J/ψ candidate is used in conjunction with another ID track in a second vertex fit. The ID track represents the kaon candidate and the mass assumed for this particle is 493.6 MeV [Nak10]. Each successfully reconstructed $\mu\mu K$ vertex represents a B^+ candidate. As with the `Bs2mumu` algorithm, no additional selection criteria are used at this point and all candidates are stored into a ROOT ntuple.

8.1.3 Candidate Yield

The `Bs2mumu` algorithm is executed separately on the `Bs_mu6mu4` and `bb_mu6mu4X` data sets, corresponding to the signal and the $b\bar{b}$ combinatorial background channels. Similarly, the `Bplus2JpsiKplus` algorithm is run on the `B+_mu6mu4K+` and `bb_mu6mu4X` data sets, corresponding to the reference channel and the same $b\bar{b}$ combinatorial background channel.

The combined muon track and ID track multiplicities are analysed in each sample. These do not depend on whether the `Bs2mumu` or the `Bplus2JpsiKplus` algorithm has been used. Plots showing these multiplicities are shown in Fig. 8.1.

The average multiplicities in each of the `Bs_mu6mu4`, `B+_mu6mu4K+` and `bb_mu6mu4X` data samples are given in Table 8.1. On average, about 1.9 combined muons are observed in each event in all data samples, with small differences explained by the different kinematic requirements used in the event generation. In Chapter 5 it has been shown that the muon multiplicity is about 2.3 at the generation stage. The efficiency loss occurs during the reconstruction process and is discussed in Section 8.6.



(a) **(b)**

Figure 8.1: Combined muon (a) and ID track (b) multiplicities in the `Bs_mu6mu4` (black, dashed), `B+_mu6mu4K+` (red, solid) and `bb_mu6mu4X` (blue, dotted) data sets.

Data set	`Bs_mu6mu4`	`B+_mu6mu4K+`	`bb_mu6mu4X`
N(μ)/event	1.941 ± 0.002	1.923 ± 0.001	1.9551 ± 0.0004

Table 8.1: Average muon multiplicity in the `Bs_mu6mu4`, `B+_mu6mu4K+` and `bb_mu6mu4X` data samples.

The average number of ID tracks observed in each event is about 38.58 ± 0.08, 42.40 ± 0.04 and 42.92 ± 0.02 in the `Bs_mu6mu4`, `B+_mu6mu4K+` and `bb_mu6mu4X` data samples, respectively. The slightly higher multiplicity observed in the `bb_mu6mu4X` data sample is expected, since both the b- and $\bar{b}$-quarks produced in the hard scattering process are open to all possible decays in the simulation with PYTHIA [Sjo06]. In the `Bs_mu6mu4` and `B+_mu6mu4K+` data samples only the b-quark is open to all decays, while the $\bar{b}$-quark is forced to decay into the simulated channel, with only 2 or 3 particles in the final state. In addition, the higher multiplicity observed in the `B+_mu6mu4K+` data set in comparison with the `Bs_mu6mu4` sample is explained by the higher cut on the $\hat{p}_\perp$ variable (see Chapter 5) applied in the generation of

the hard scattering process (i.e., 15 GeV instead of 10 GeV). This implies, that, on average, the b-quark produced has a higher energy, which leads to an increased number of particles in the final state.

The number of events containing at least one reconstructed $\mu\mu$ vertex is summarised in Table 8.2 for each data set analysed. The number of successfully reconstructed B_s^0 or B^+ candidates with the above mentioned selection criteria is also shown. A total of 46 934 and 838 600 B_s^0 (i.e., $\mu\mu$) candidates are found in the `Bs_mu6mu4` and `bb_mu6mu4X` samples respectively (with the `Bs2mumu` algorithm). The loss of events seen in the candidate reconstruction is mostly due to the efficiency loss in the muon reconstruction. Similarly, 125 332 and 82 147 J/ψ candidates are recorded in the `B+_mu6mu4K+` and `bb_mu6mu4X` data samples (with the `Bplus2JpsiKplus` algorithm). In this case, the window imposed on the reconstructed di-muon vertex mass restricts the number of candidates to those close to the true J/ψ mass value. This explains the low number of reconstructed J/ψ candidates with respect to the total number of events in the $b\bar{b}$ combinatorial background sample. Finally, the number of B^+candidates reconstructed in the `B+_mu6mu4K+` and `bb_mu6mu4X` samples is 989 769 and 604 013, respectively. The much higher multiplicity observed in B^+ (i.e., $\mu\mu K$) candidates is due to the high number of combinations found for the di-muon vertex and another ID track.

Data set	`Bs_mu6mu4`	`bb_mu6mu4X`	`B+_mu6mu4K+`	`bb_mu6mu4X`
Algorithm	`Bs2mumu`		`Bplus2JpsiKplus`	
Total events	50 000	1 000 000	150 000	1 000 000
Reconstructed $\mu\mu$ events	44 129	760 125	125 148	82 113
$\mu\mu$ candidates	46 934	838 600	125 332	82 147
$\mu\mu K$ candidates	-	-	989 769	604 013

Table 8.2: Number of events containing at least one reconstructed $\mu\mu$ vertex and number of successfully reconstructed B_s^0, J/ψ and B^+ candidates.

The multiplicities of the reconstructed $\mu\mu$ and $\mu\mu K$ candidates are displayed in Fig. 8.2. Only events with at least one reconstructed $\mu\mu$ candidate are considered here. On average, 1.063 ± 0.001 (1.1014 ± 0.0004) B_s^0 candidates are observed in the `Bs_mu6mu4` (`bb_mu6mu4X`) samples (with the `Bs2mumu` algorithm). Similarly, 1.0015 ± 0.0001 (1.00041 ± 0.00007) J/ψ candidates are reconstructed in each event from the `B+_mu6mu4K+` and `bb_mu6mu4X` samples (with the `Bplus2JpsiKplus` algorithm). A much higher average multiplicity of B^+ candidates of 7.47 ± 0.01 (6.92 ± 0.01) is found in the `B+_mu6mu4K+` (`bb_mu6mu4X`) data set. This is due to the increased combinatorics resulting from using many tracks as kaon candidates in the vertex fit.

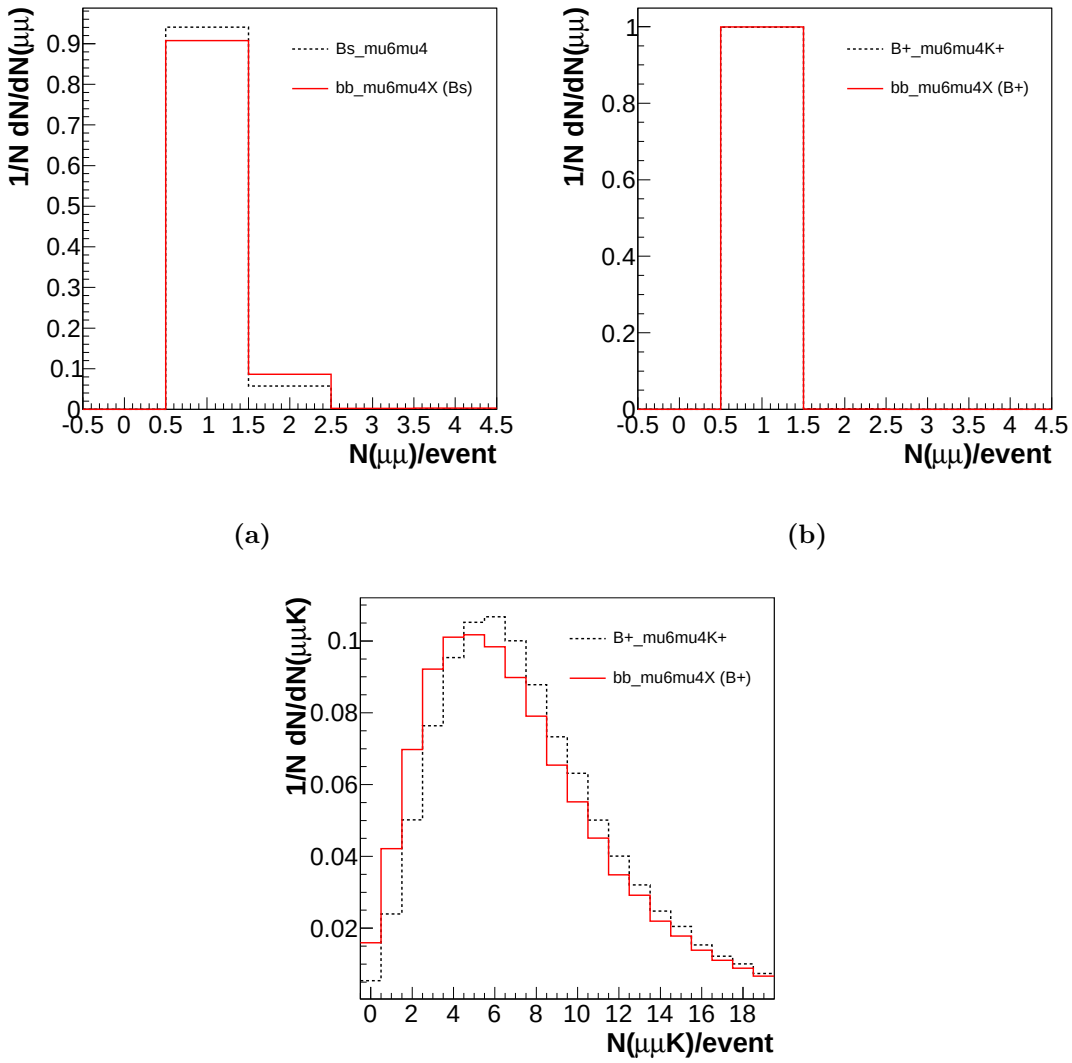


Figure 8.2: Multiplicities of reconstructed $B_s^0 \to \mu^+\mu^-$ (a), $J/\psi \to \mu^+\mu^-$ (b) and $B^+ \to J/\psi K^+$ (c) candidates in dedicated signal (black, dashed) or background (red, solid) data samples.

8.2 Trigger Selection

The next step in the selection verifies whether events containing successfully reconstructed B_s^0 or B^+ candidates fulfil the trigger requirements. This is important since the trigger requirements may alter the kinematic distributions of reconstructed candidates. For this purpose the trigger simulation software is executed within the Athena framework independent of the `Bs2mumu` and the `Bplus2JpsiKplus` algorithms. This is detailed in Chapter 7. The trigger chain chosen is based on di-

muon events, with turn-on curves corresponding to cuts of 6 and 4 GeV on muon transverse momenta. The number of events that pass the EF_mu4_mu6 requirement is summarised in Table 8.3. B_s^0, J/ψ and B^+ multiplicities are mostly unaffected by the trigger selection.

Data set	Bs_mu6mu4	bb_mu6mu4X	B+_mu6mu4K+	bb_mu6mu4X
Algorithm	Bs2mumu		Bplus2JpsiKplus	
Reconstructed $\mu\mu$ events	44 129	760 125	125 148	82 113
Pass EF_mu4_mu6 events	16 289	213 508	48 558	25 882
$\mu\mu$ candidates	17 552	239 332	48 679	25 901
$\mu\mu K$ candidates	-	-	400 481	206 647

Table 8.3: Number of events satisfying the EF_mu4_mu6 trigger requirement and number of successfully reconstructed B_s^0, J/ψ and B^+ candidates passing the EF_mu4_mu6 trigger.

8.3 Baseline Selection

Following the trigger selection, a series of cuts are imposed on the $\mu\mu$ and $\mu\mu K$ candidates. These cuts affect some of the kinematic variables of the candidates or their decay products and the quality of the vertex fits. In the following, these are collectively named baseline selection cuts.

8.3.1 Selection of B_s^0 Candidates

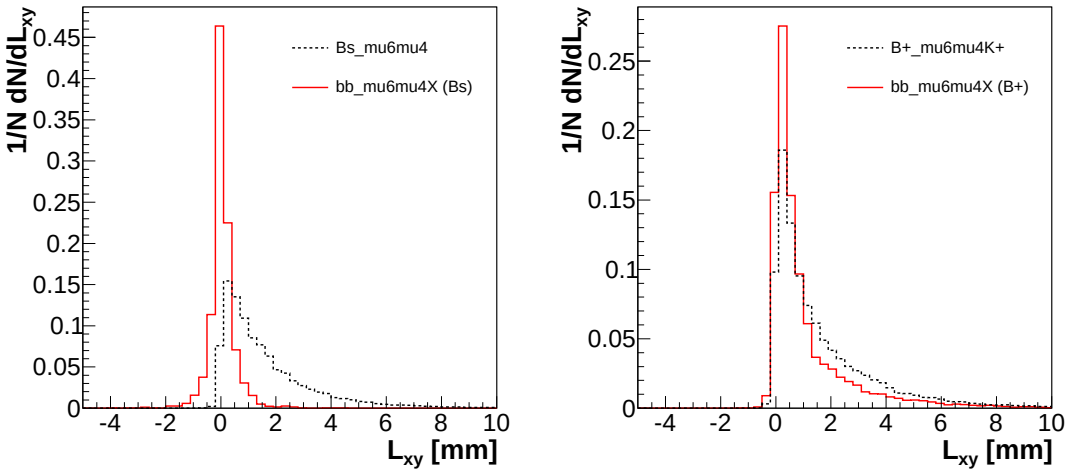
For the selection of B_s^0 candidates, the two muons originating in the $\mu\mu$ vertex are required to have opposite charges. The minimum transverse momenta have to be 6 and 4 GeV. In addition, their pseudorapidity η is restricted to the $[-2.5, 2.5]$ interval.

The transverse decay length L_{xy} is defined as the distance between the point of origin of the B_s^0 candidate (always assumed to be the primary vertex) and the point of decay (the position of the $\mu\mu$ vertex), projected onto the direction of the B_s^0 transverse momentum:

$$L_{xy} = \vec{L} \cdot \frac{\vec{p_T}}{|\vec{p_T}|}, \tag{8.1}$$

where $\vec{L}$ is the distance vector connecting the primary and secondary vertices. Since B-mesons are expected to have a high average life time ($\sim$ 1.5 ps), this variable is an efficient discriminator against $\mu\mu$ candidates produced by prompt J/ψ decays or Drell-Yan processes. This is described in more detail in Section 8.4. At this point, only a quality cut against wrongly reconstructed vertices is applied. Vertex positions of candidates reconstructed very close to the primary vertex may

produce negative values of L_{xy}, due to the finite space resolution for reconstructing particle tracks. These wrongly reconstructed vertices are rejected by the requirement $L_{xy} > 0$ mm. The L_{xy} spectra of all candidates fulfilling the baseline requirements, except the cut on L_{xy}, are shown in Fig. 8.3.



(a) **(b)**

Figure 8.3: L_{xy} distributions for B_s^0 (a) and B^+ (b) candidates after baseline selection, except the $L_{xy} > 0$ mm requirement.

A cut on the vertex fit quality is also employed in order to suppress wrongly reconstructed candidates. Therefore, a $\chi^2/ndf < 10$ is required for $\mu\mu$ vertices. If the event contains several candidates, the one with the lowest χ^2/ndf is chosen. The probability, that an event contains more than one real $B_s^0 \to \mu^+\mu^-$ decay, is very small and events containing multiple B_s^0 candidates usually occur when more than two muons are reconstructed. Contributions from fake candidates are thus minimised. Finally, the invariant mass of $\mu\mu$ candidates is required in the [4.3 GeV, 7 GeV] region. This is in order to avoid regions with major contributions from J/ψ or Υ decays with two muons in the final state.

The purity of the baseline selection is determined as the fraction of accepted candidates, that can be matched to true decays from the total number of accepted candidates. Truth matching is achieved by comparing the flight directions of reconstructed and truth B_s^0 decay products (given by η and ϕ) and $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is used as a discriminant. B_s^0 candidates are considered to be matching true $B_s^0 \to \mu^+\mu^-$ decays if $\Delta R < 0.01$ for both decay particles. About 98% of all accepted events are successfully matched to true decays.

8.3.2 Selection of B^+ Candidates

The selection of B^+ candidates starts by applying similar cuts on the $\mu\mu$ vertex. The two muons originating in the vertex are required to have opposite charges, their

p_T values are required to be above 6 and 4 GeV, respectively, and their pseudorapidity η has to fit in the $[-2.5, 2.5]$ interval. Only candidates with $\chi^2/ndf < 10$ are accepted. In addition, the di-muon invariant mass is required to fit inside a window of ± 120 MeV around the J/ψ mass value of 3096.9 MeV [Nak10].

The selection of B^+ candidates requires a minimum transverse momentum of the kaon track candidate above 1.5 GeV. In addition, a cut is imposed on the transverse decay length $L_{xy} > 0$ mm. A quality cut $\chi^2/ndf < 10$ is then required on $\mu\mu K$ candidates. Finally, the invariant mass of the $\mu\mu K$ candidate is required to fit in the $[4.3\,\text{GeV}, 7\,\text{GeV}]$ interval. If an event contains several candidates fulfilling all selection criteria, the candidate with the smallest χ^2/ndf will be chosen.

The purity of the event selection is determined, as in the case of the B_s^0 selection, by matching reconstructed B^+ decay products to truth particles. A B^+ candidate matches a true $B^+ \to J/\psi K^+$ decay if all three final state particles fulfil the $\Delta R < 0.01$ requirement. About 90% of all accepted events are successfully matched to true decays. The lower purity, than that found for the B_s^0 selection (about 98%), is due to the higher number of combinations specific to the B^+ reconstruction. It could be improved by requiring harder cuts on the χ^2/ndf of the $\mu\mu K$ vertex, or on the p_T of the kaon candidate. However, this would reduce the selection efficiency, which is not desired in this analysis. Fake events will be suppressed by statistically subtracting sideband contributions in the invariant mass signal region (see Section 8.7).

8.3.3 Baseline Selection Yield

The numbers of successfully reconstructed candidates, that pass all baseline selection cuts, are summarised in Table 8.4. A total of 15 390 (2 049) B_s^0 candidates are observed in the `Bs_mu6mu4` (`bb_mu6mu4X`) data set, if the event reconstruction is performed with the `Bs2mumu` algorithm. In addition, 35 862 (6 731) B^+ candidates are observed in the `B+_mu6mu4K+` (`bb_mu6mu4X`) data set, when the event reconstruction is performed with the `Bplus2JpsiKplus` algorithm.

Data set	`Bs_mu6mu4`	`bb_mu6mu4X`	`B+_mu6mu4K+`	`bb_mu6mu4X`
Algorithm	`Bs2mumu`		`Bplus2JpsiKplus`	
Candidates (baseline selection)	15 390	2 049	35 862	6 731

Table 8.4: Number of successfully reconstructed B_s^0, J/ψ and B^+ candidates fulfilling the baseline selection cuts. Here, the number of candidates is equivalent to the number of events containing at least one candidate.

8.4 Cut Based Selection

Following the trigger and the baseline selections, the number of B_s^0 candidates in the signal (`Bs_mu6mu4`) MC sample is compared to the number of candidates in the background (`bb_mu6mu4X`) sample (using the `Bs2mumu` algorithm). Fig. 8.4 shows the invariant mass spectra of all candidates observed in the two samples. The number of events in both cases is scaled to the same integrated luminosity of 1 fb^{-1}. It is seen, that, at this stage, it is impossible to distinguish true signal decays from the much larger number of candidates observed in the combinatorial background sample. In order to suppress the large amount of background events, while keeping a high retention rate of true signal decays, several selection variables are investigated.

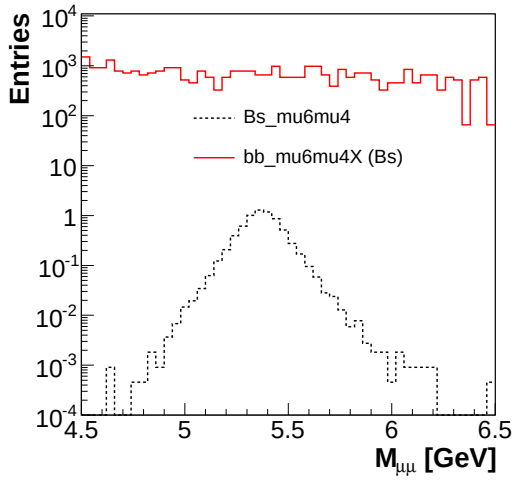


Figure 8.4: Invariant mass spectra of B_s^0 candidates after the trigger and baseline selections observed in the `Bs_mu6mu4` and `bb_mu6mu4X` data samples. Entries are scaled to an integrated luminosity of 1 fb^{-1}.

This section presents the selection variables and their distributions, which are compared for the signal and background samples. The analysis is tuned for the selection of $B_s^0 \to \mu^+\mu^-$ decays and all candidates are reconstructed with the `Bs2mumu` algorithm. At a later stage, in Section 8.5, the same selection criteria are applied on B^+ candidates. However, in this case the purpose is not background suppression, but applying the same selection criteria for B_s^0 and B^+ candidates, which may result in the partial cancellation of the corresponding systematic uncertainties.

8.4.1 Transverse Decay Length

The mean lifetime τ and the proper decay length $c\tau$ are intrinsic properties of each particle type. For B_s^0 particles the measured value is $c\tau = 441$ μm [Nak10]. Other B-mesons have slightly different measured proper decay lengths, e.g. 457 μm for

B_d^0 and 491 μm for B^+ [Nak10]. The differences between the proper decay lengths of b-flavoured and c-flavoured mesons are much larger (e.g., $c\tau = 122$ μm for D^0 while the J/ψ proper decay length is several orders of magnitude lower than that of B-mesons). The distribution of the measured decay length usually follows an exponential function, with a constant factor in the exponent characteristic for each particle type. Large differences are usually observed in such distributions corresponding to particles with different mean lifetimes. Therefore, applying a cut on the decay length offers a good suppression of $\mu\mu$ pairs originating in decays of prompt J/ψ particles or c-flavoured mesons. A similar effect is expected for suppressing $\mu\mu$ pairs produced in Drell-Yan processes, which also take place very close to the interaction point.

The decay length is determined as the distance between the primary vertex (the point where the B-meson is produced) and the secondary vertex (the point of the B-meson decay). Due to the better point resolution in the $x-y$ plane than in the z direction, using only the $2D$ (i.e., $x-y$) information instead of the full $3D$ information leads to a better precision of this variable. In addition, by projecting the decay length on the transverse momentum, additional rejection power may be gained from the fact that true signal candidates, unlike background events, tend to have momenta better aligned to their flight directions. The L_{xy} variable has been introduced in Section 8.3 and has been defined in (8.1). The basic elements used in its calculation are schematically drawn in Fig. 8.5 (a), showing the projections in the transverse plane.

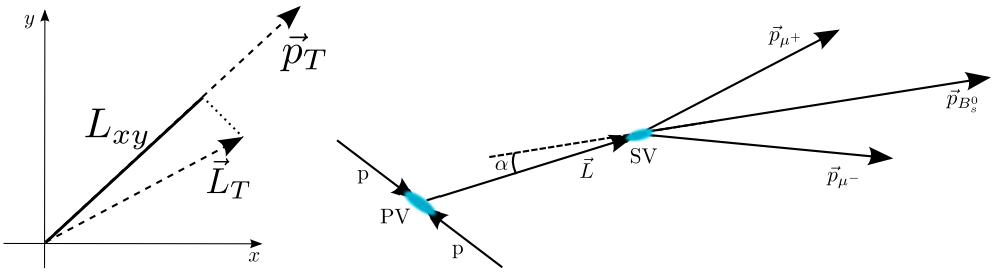


Figure 8.5: (a) Schematic view of the L_{xy} variable in the transverse plane. (b) Schematic view of the pointing angle.

A lower cut on the L_{xy} variable is used to reject short lived particles. The cut suggested in [Aad09] is $L_{xy} > 0.5$ mm. The efficiencies of this cut are discussed in Sections 8.6 and 8.7. The cut will be revised and optimised in Chapter 9, with the aim of obtaining the best expected upper limit on the branching ratio.

The L_{xy} distributions of B_s^0 candidates passing the baseline selection from the `Bs_mu6mu4` and `bb_mu6mu4X` data samples are displayed in Fig. 8.6. The distribution of candidates observed in the signal sample follows approximately an exponential distribution, as expected, due to the fact that most entries come from true $B_s^0 \to \mu^+\mu^-$ decays. In contrast, the spectrum observed in the background sample is dominated by events with very short decay lengths (mostly prompt J/ψ and

Drell-Yan candidates) and differs significantly from an exponential shape, due to contributions of many different particles.

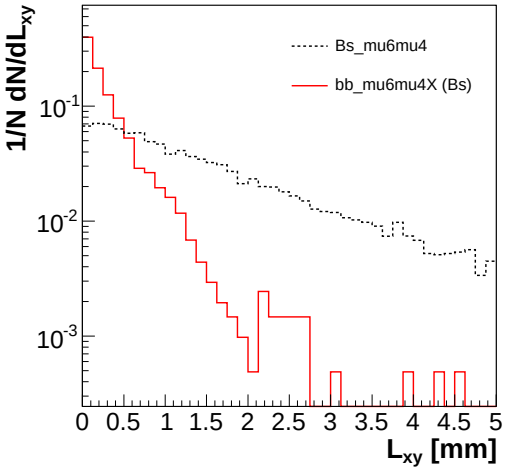


Figure 8.6: Transverse decay length distribution for B_s^0 candidates observed in the `Bs_mu6mu4` (black, dashed) and `bb_mu6mu4X` (red, solid) data samples.

8.4.2 Pointing Angle

Another discriminating variable is the pointing angle α. It is defined as the angle between the flight direction of the B_s^0 candidate and the direction connecting the primary and the secondary vertices:

$$\alpha = \arccos\left(\frac{\vec{L}}{|\vec{L}|} \cdot \frac{\vec{p}}{|\vec{p}|}\right), \tag{8.2}$$

where $\vec{p}$ is the B_s^0 candidate's reconstructed momentum, and $\vec{L}$ is the distance between the primary vertex and the B_s^0 decay vertex. This is also schematically shown in Fig. 8.5 (b).

The pointing angle spectrum is shown in Fig. 8.7 for the `Bs_mu6mu4` and `bb_mu6mu4X` data samples. All candidates considered are required to pass the baseline selection. Differences in the distributions of this variable are expected between signal and background events. In the case of true signal decays, the reconstructed B_s^0 momentum is expected to be well aligned to the particle's flight trajectory and, therefore, small values for α are expected. Fake muon pair candidates originating in a 3-prong decays of B-mesons (e.g., $B_d^0 \to \pi^- l^+ \nu_l$) may be misreconstructed in the B_s^0 mass region. However, in such a case the di-muon vertex is not well aligned to the B_s^0 flight direction. These events may be rejected by accepting only candidates in the low α region. Another situation is that of two B-mesons flying close to each other, each producing one muon in the final state. B-mesons contain a muon in the decay chain with a probability of about

10%, so, if the two muons are fitted to a common vertex, significant background contributions may be produced. In this situation the pointing angle is also a good discriminant against background events.

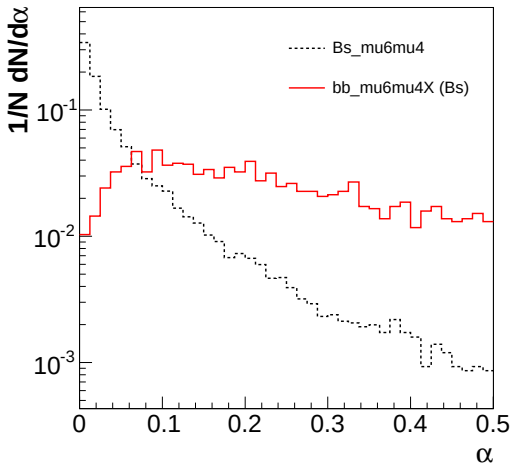


Figure 8.7: Pointing angle distribution for B_s^0 candidates observed in the `Bs_mu6mu4` (black, dashed) and `bb_mu6mu4X` (red, solid) data samples.

As already mentioned, the transverse decay length also contains information regarding the alignment of the reconstructed B_s^0 momentum and flight direction, albeit the pointing angle offers more precision. Therefore, a high degree of correlation is expected between the two variables. This is investigated in the efficiency determination (see Section 8.7).

The upper cut suggested in [Aad09] is $\alpha < 0.017$ rad. The efficiencies of applying this cut on the signal and background samples are analysed in Section 8.6 and 8.7. The cut is re-evaluated and optimised in Chapter 9.

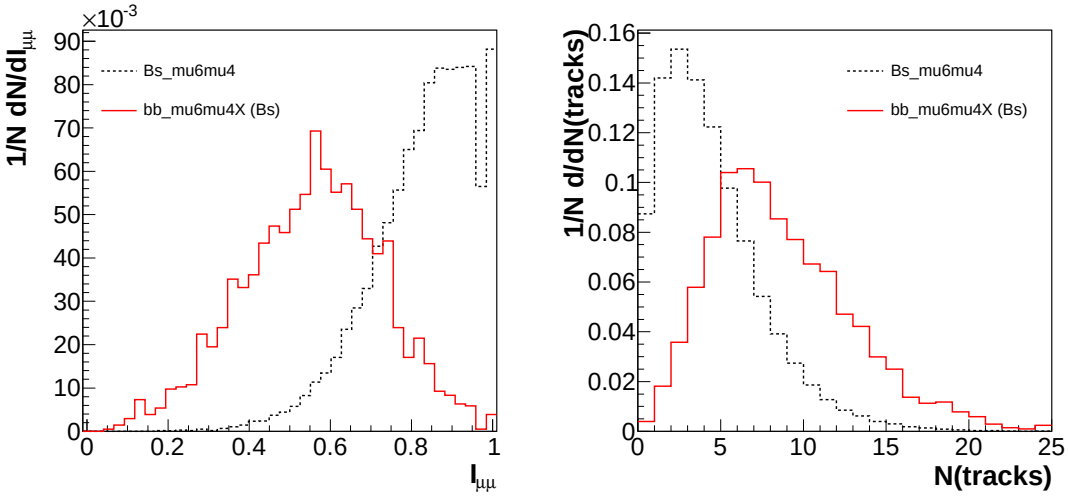
8.4.3 B_s^0 Isolation

Another variable investigated is the B_s^0 isolation, $I_{\mu\mu}$. It is defined as the fraction of the reconstructed B_s^0 candidate transverse momentum in the sum of transverse momenta of all particle tracks around the B_s^0 flight direction:

$$I_{\mu\mu} = \frac{p_T(B_s^0)}{p_T(B_s^0) + \sum_i p_T(track_i)}, \tag{8.3}$$

The two muon tracks resulting from the B_s^0 decay are excluded from the computation. Only tracks within a cone centred around the B candidate flight direction are accepted. The cone width is defined by $\Delta R < 1$, where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

The spectra of the isolation variable are shown in Fig. 8.8 for those candidates that pass the baseline requirements. Events from the `Bs_mu6mu4` and the `bb_mu6mu4X` data sets are shown. The isolation variable is a good discriminant for



(a) (b)

Figure 8.8: Distribution of the isolation variable for B_s^0 candidates observed in the `Bs_mu6mu4` (black, dashed) and `bb_mu6mu4X` (red, solid) data samples (a). Number of tracks in a $\Delta R < 1$ cone around the B_s^0 candidate direction in the `Bs_mu6mu4` (black, dashed) and `bb_mu6mu4X` (red, solid) data samples (b)

true $B_s^0 \to \mu^+\mu^-$ decays, with only few tracks expected inside the cone around the B_s^0 candidate (except the two muons). In this case the ratio in (8.3) tends to have a value close to 1. Background events usually have a considerably larger number of tracks around the B_s^0 candidate, due to more complicated event topologies. These additional tracks shift the peak in the distribution towards 0. Other types of events rejected by this variable are those containing two B-mesons flying in close proximity, each producing one muon in the final state (the situation described in Subsection 8.4.2). These are typically semileptonic decays with $2-3$ hadrons in the decay tree. These additional tracks contribute to a lower value of $I_{\mu\mu}$.

The sharp peak in the $I_{\mu\mu}$ spectrum at $I_{\mu\mu} = 1$ for the `Bs_mu6mu4` sample (i.e., the last bin of the histogram) is produced by those events without any additional reconstructed track in the cone around the B_s^0 direction. The sharp drop observed in the penultimate bin, just before $I_{\mu\mu} = 1$, is an effect of the implicit threshold cut on the track transverse momentum at 0.5 GeV in the Athena reconstruction. The plateau region observed around $I_{\mu\mu} = 0.9$ is produced by high pseudorapidity B_s^0 candidates, where the cone around the B_s^0 flight direction stretches outside the detector acceptance region ($|\eta| < 2.5$). In such cases, a fraction of the tracks around the B_s^0 candidate is missing from the $I_{\mu\mu}$ calculation.

The cut suggested in [Aad09] is $I_{\mu\mu} > 0.9$. The efficiency of this cut is computed in Section 8.6 and 8.7. The cut is then revised in Chapter 9 with the aim of optimising the expected limit on the $B_s^0 \to \mu^+\mu^-$ branching ratio.

8.4.4 Invariant Mass

Another variable used for separating signal from background events is the di-muon vertex invariant mass. This is a powerful discriminant against other types of decays containing di-muon pairs in the final state. For example, $c\bar{c}$ or $b\bar{b}$ resonances manifest themselves in different regions of the invariant mass spectrum (e.g., 3097 MeV for J/ψ and 9460 MeV for Υ [Nak10]). The invariant mass resolution is around 65 MeV for J/ψ and 170 MeV for Υ [Atl10a]. Therefore, they are expected to yield small contributions in the B_s^0 mass region. Other decays, such as hadronic or semi-leptonic decays of B_s^0 or B_d^0 mesons, with pions or kaons in the in the final state being misidentified as muons, are reconstructed below the B_s^0 mass region [Aad09]. This is a consequence of assigning the muon mass to the heavier pions or kaons in the di-muon vertex fit [Aad09].

The invariant mass spectrum of B_s^0 candidates observed in the `Bs_mu6mu4` data sample is displayed in Fig. 8.9 (a). Only the trigger and baseline selections are required. It is seen that the distribution approximately follows a double Gaussian shape. This is due to different point resolutions in the barrel ($|\eta| < 1.05$) and endcap ($|\eta| > 1.05$) regions of the detector. This is explicitly shown in Fig. 8.9 (b) and (c), where only candidates with both muons reconstructed in the barrel (b) and endcap (c) regions are selected. In both cases, the distributions are very close to normal distributions with widths of 66.2 ± 0.7 MeV (barrel) and 133.6 ± 1.4 MeV (endcap) computed from fits with a Gaussian function. The reconstructed mean B_s^0 mass is 5371 ± 1 MeV in both cases, which is consistent with the B_s^0 mass value used in the PYTHIA simulation (5369 MeV). The invariant mass distribution for all candidates is then fitted with a double Gaussian function:

$$G(x) = A_1 \ \exp \frac{-(x-\mu_1)^2}{2\sigma_1^2} + A_2 \ \exp \frac{-(x-\mu_2)^2}{2\sigma_2^2}, \tag{8.4}$$

where the μ_1 and μ_2 parameters are confined to a 20 MeV interval around the B_s^0 mass. The mean value is calculated to 5373 ± 2 MeV as the first moment of the fit function. The resolution is determined as the second central moment of the fit function and its value is computed to 108 ± 3 MeV. The fit functions' parameters (mean and width) obtained after the fits are summarised in Table 8.5.

Parameter	full η range	barrel ($\|\eta\| < 1.05$)	endcap ($\|\eta\| > 1.05$)
μ_1[MeV]	5376 ± 3	5371 ± 1	5372 ± 2
σ_1[MeV]	153 ± 4	66.2 ± 0.7	133.6 ± 1.4
μ_2[MeV]	5371 ± 1	-	-
σ_2[MeV]	71 ± 2	-	-

Table 8.5: Parameters of the fit functions used to describe the invariant mass spectrum. The distribution is fitted with a double Gaussian function, when the full η range is considered, and with a simple Gaussian function, when only events in the barrel or endcap regions are considered.

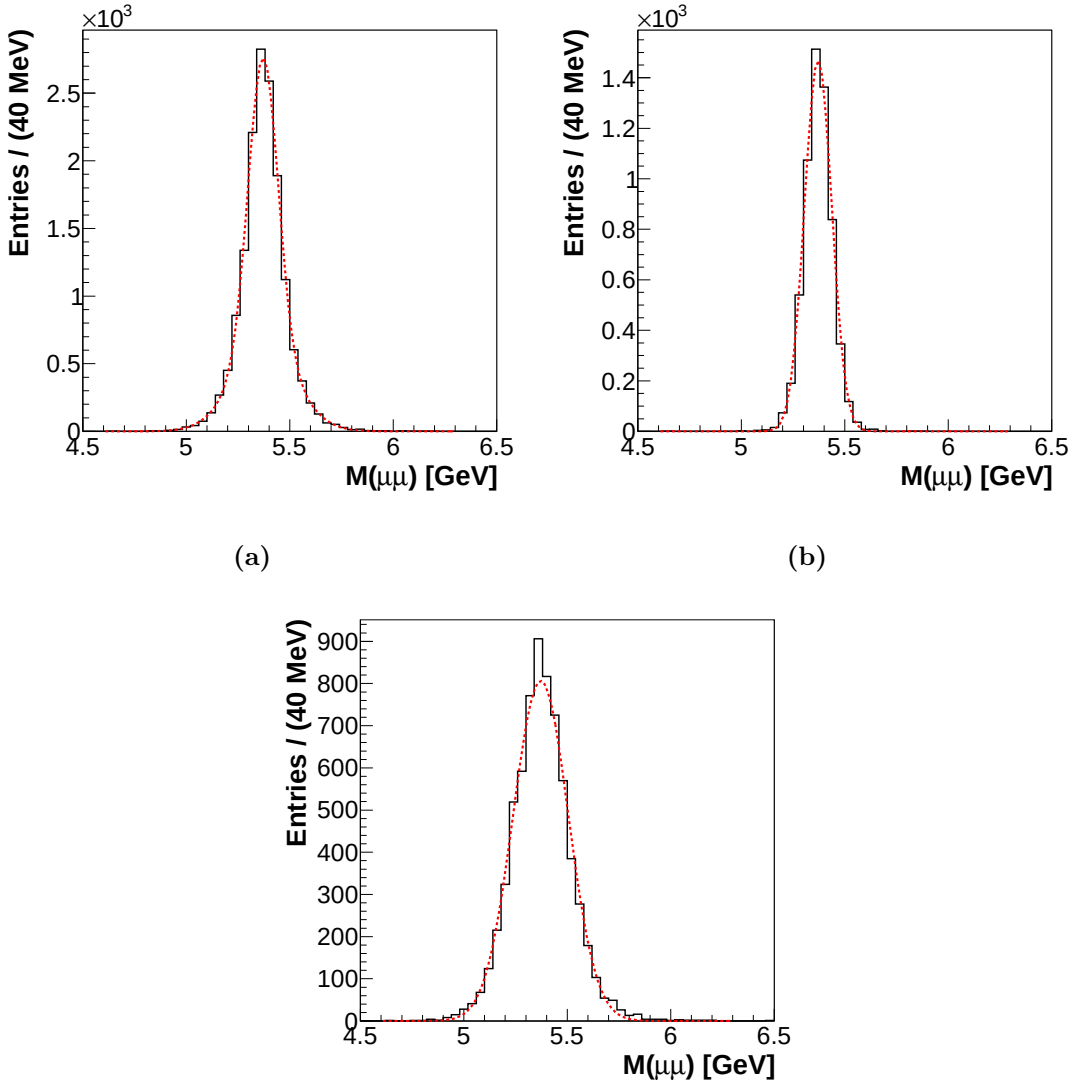
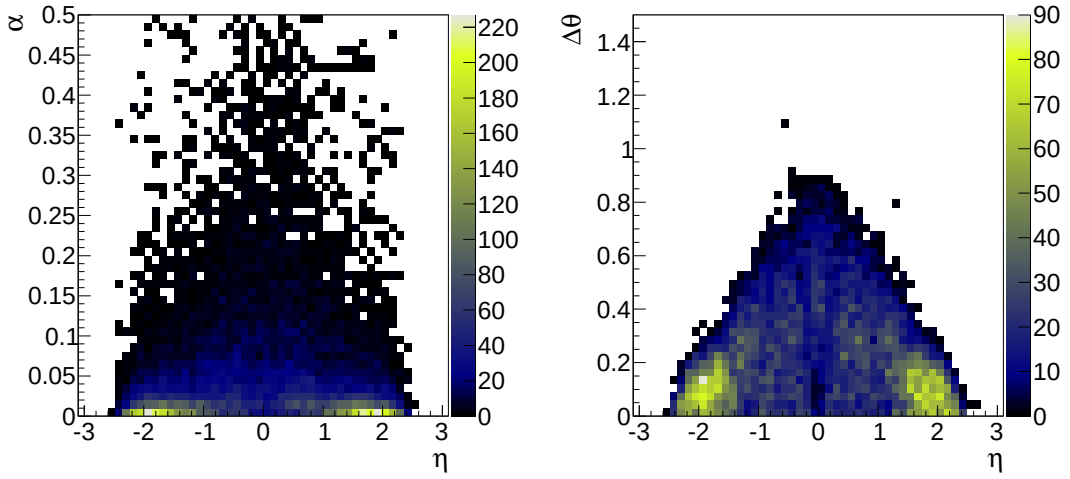


Figure 8.9: Di-muon invariant mass spectrum in the `Bs_mu6mu4` data sample, for all candidates (a) or only for those candidates with reconstructed muons in the barrel (b) or endcap (c) regions. The distributions are fitted by either a double Gauss (a) or by a simple Gauss (b), (c) function, shown with a dotted line.

A signal window is constructed with $M_{\mu\mu} \in [M_{B_s^0} - \sigma, M_{B_s^0} + 2\sigma]$, where $M_{B_s^0}$ is the known B_s^0 mass and σ is the invariant mass resolution, computed as the width of the double Gaussian fit for the entire η range ($\sigma = 108$ MeV). The asymmetric search window is chosen in order to minimise the overlap with the B_d^0 mass region (5279 MeV [Nak10]). This is necessary since $B_d^0 \to \mu^+\mu^-$ decays may also be studied.

The effect on the invariant mass resolution by applying the previous selection cuts is also investigated. Events are selected by applying the $L_{xy} > 0.5$ mm,

$\alpha < 0.017$ and $I_{\mu\mu} > 0.9$ cuts independently. The resolution is calculated in each case as the second central moment of a double Gaussian function which is fitted to the distribution. It is observed, that the cuts on L_{xy} and $I_{\mu\mu}$ do not affect the resolution, the computed values being 108 ± 3 MeV and 110 ± 7 MeV, respectively. A significant effect is produced when selecting events by imposing the cut on α. The calculated resolution is 123 ± 7 MeV in this case. This is an effect of the Lorentz boost in the z-direction. Since a minimum p_T requirement is imposed on the two muons (6 and 4 GeV), a candidate in the high η region has on average a higher total momentum than one in the central region. This results in a smaller opening angle between the two muons. Therefore, the reconstructed B_s^0 momentum is better aligned with the B_s^0 flight direction, inducing a smaller value of α. Fig. 8.10 (a) shows the distribution of the pointing angle versus the pseudorapidity of B_s^0 candidates. It is indeed observed, that candidates in the high η region have typically small values of α. The effect of the Lorentz boost on the di-muon opening angle (in polar coordinates) is clearly seen in Fig. 8.10 (b). Since candidates selected in the low α region are mostly found in the high $|\eta|$ region (where the mass resolution is worse), the overall invariant mass resolution becomes worse. In contrast, the L_{xy} and $I_{\mu\mu}$ distributions are almost constant in all η regions. L_{xy} is computed in the $x - y$ plane (not affected by the boost in the z-direction) and $I_{\mu\mu}$ is calculated based on ΔR differences between particle trajectories, which is also invariant under Lorentz transformation.



(a) **(b)**

Figure 8.10: (a) Pointing angle versus the pseudorapidity of reconstructed B_s^0 candidates. (b) Opening angle between the decay muons in polar coordinates versus the pseudorapidity of reconstructed B_s^0 candidates. In both cases trigger and baseline selection is required.

8.5 Event Selection in the Reference Channel

This section investigates the `B+_mu6mu4K+` and the `bb_mu6mu4X` data sets analysed with the `Bplus2JpsiKplus` algorithm. The aim is to apply a set of selection criteria for B^+ candidates as similar as possible to that used for the B_s^0 candidates. Applying similar cuts is expected to reduce some of the systematic uncertainties. The general approach is described in Chapter 4. Specifically, the use of ratios of selection efficiencies in computing the branching ratio (given in 4.1) may result in the cancellation of some of the systematic uncertainties associated with the efficiencies.

8.5.1 L_{xy}, α and $I_{\mu\mu K}$ Selection

The selection variables are calculated similarly to those used in the B_s^0 selection, with the difference that all quantities are computed with respect to the $\mu\mu K$ vertex. The transverse decay length is defined as in (8.1), where $\vec{p_T}$ is the transverse momentum of the B^+ candidate and $\vec{L}$ is the direction connecting the primary and the $\mu\mu K$ vertices. The pointing angle is defined in (8.2). The isolation variable is explained in (8.3), where the tracks considered in the denominator's sum exclude also the kaon candidate besides the two muons (i.e., all tracks of particles originating in the B^+ decay). The distributions of these three parameters are shown in Fig. 8.11 for those candidates that pass the trigger and baseline requirements in the `B+_mu6mu4K+` and `bb_mu6mu4X` data sets. The `B+_mu6mu4K+` data set is used only for determining the selection efficiencies, with the same cuts as used for B_s^0 candidates. These efficiencies are analysed further in Section 8.6. The `bb_mu6mu4X` sample is used to estimate the number of observed $B^+ \to J/\psi K^+$ events from the same data sample used for the search of the $B_s^0 \to \mu^+\mu^-$ decay. Differences between the shapes of these quantities in the signal and background samples are much smaller than in the case of B_s^0 candidates. The reason is that most B^+ candidates are true events. To the contrary, all B_s^0 candidates observed in the background sample are fake events.

8.5.2 Invariant Mass

The next quantity investigated is the $\mu\mu K$ invariant mass. Fig. 8.12 (a) shows the distribution for B^+ candidates in the `B+_mu6mu4K+` data set that pass the trigger and baseline requirements. The distribution shows a Gaussian structure over-imposed on a linear background. The background is caused by wrongly reconstructed vertices, due to the high number of possible combinations of kaon candidates. Therefore, in order to estimate the invariant mass resolution, a function formed by a first order polynomial plus a double Gaussian function is fitted to the distributions. As in the case of the $\mu\mu$ invariant mass, different resolutions are observed, when the final state particles are detected in different regions of the detector. The distributions for those candidates with both muon tracks detected in the barrel ($|\eta| < 1.05$) and endcap ($|\eta| > 1.05$) regions are shown in Fig. 8.12 (b) and (c), respectively. The shapes are approximated by a first order polynomial

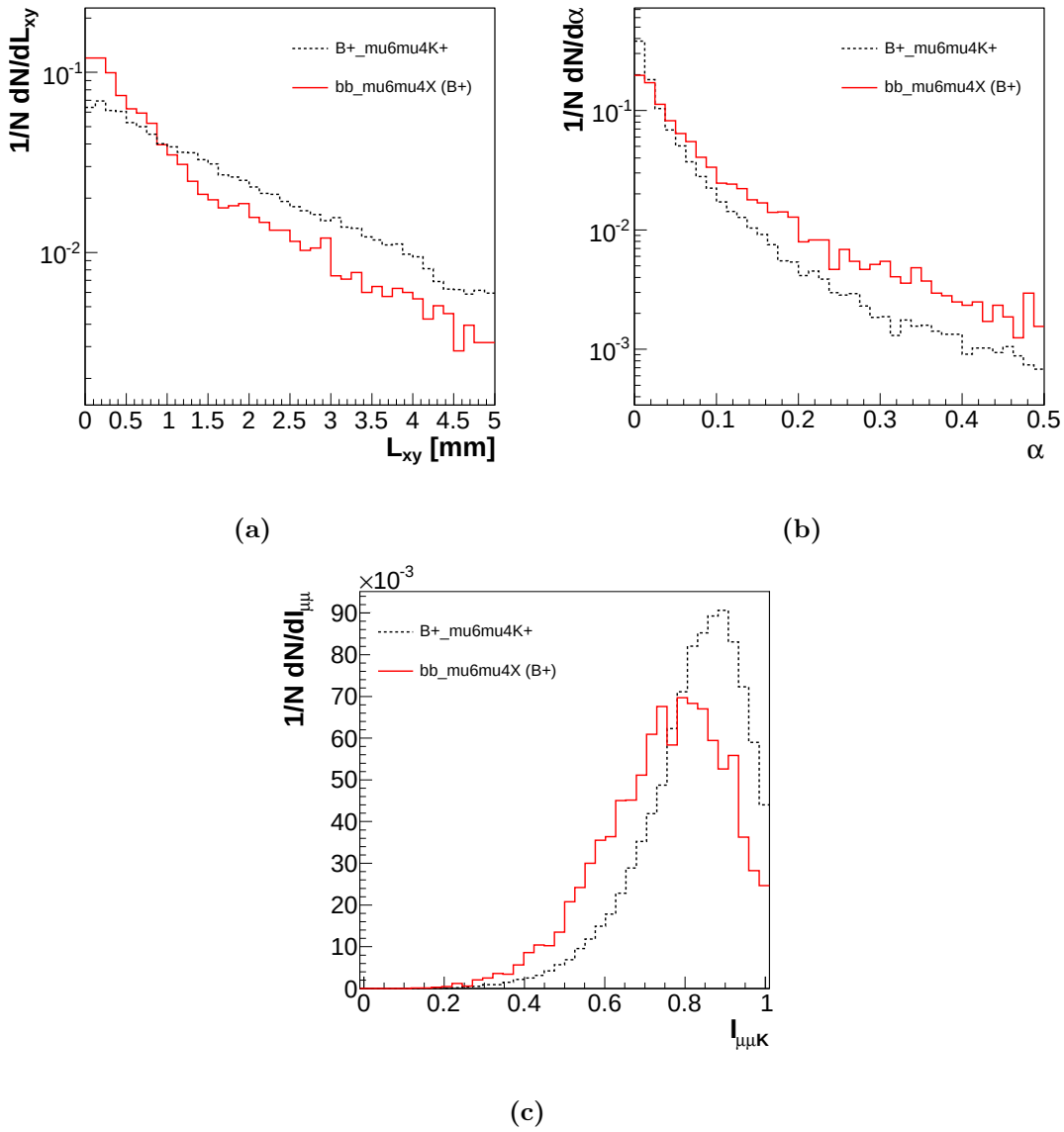


Figure 8.11: Distributions of the transverse decay length (a), pointing angle (b) and isolation (c) variables for B^+ candidates in the `B+_mu6mu4K+` (black) and `bb_mu6mu4X` (red) data sets.

plus a simple Gaussian function and the resolution is calculated as the width of the resulting Gaussian shape. The computed values are 27.9 ± 0.7 MeV in the barrel region and 49.0 ± 1.4 MeV in the endcap region. In order to better approximate the different resolution effects in the two regions, the invariant mass distribution for all candidates is fitted by a first order polynomial plus a double Gaussian function (as defined in (8.4)). The invariant mass resolution is therefore approximated as the second central moment of the resulting double Gaussian function and the calculated value is 43.3 ± 0.1 MeV. The mean mass is computed to 5281 ± 1 MeV, as the first moment of the double Gaussian function. This is consistent with the

B^+ mass of 5279 MeV assumed in the PYTHIA simulation. The parameters of all fit functions are summarised in Table 8.6.

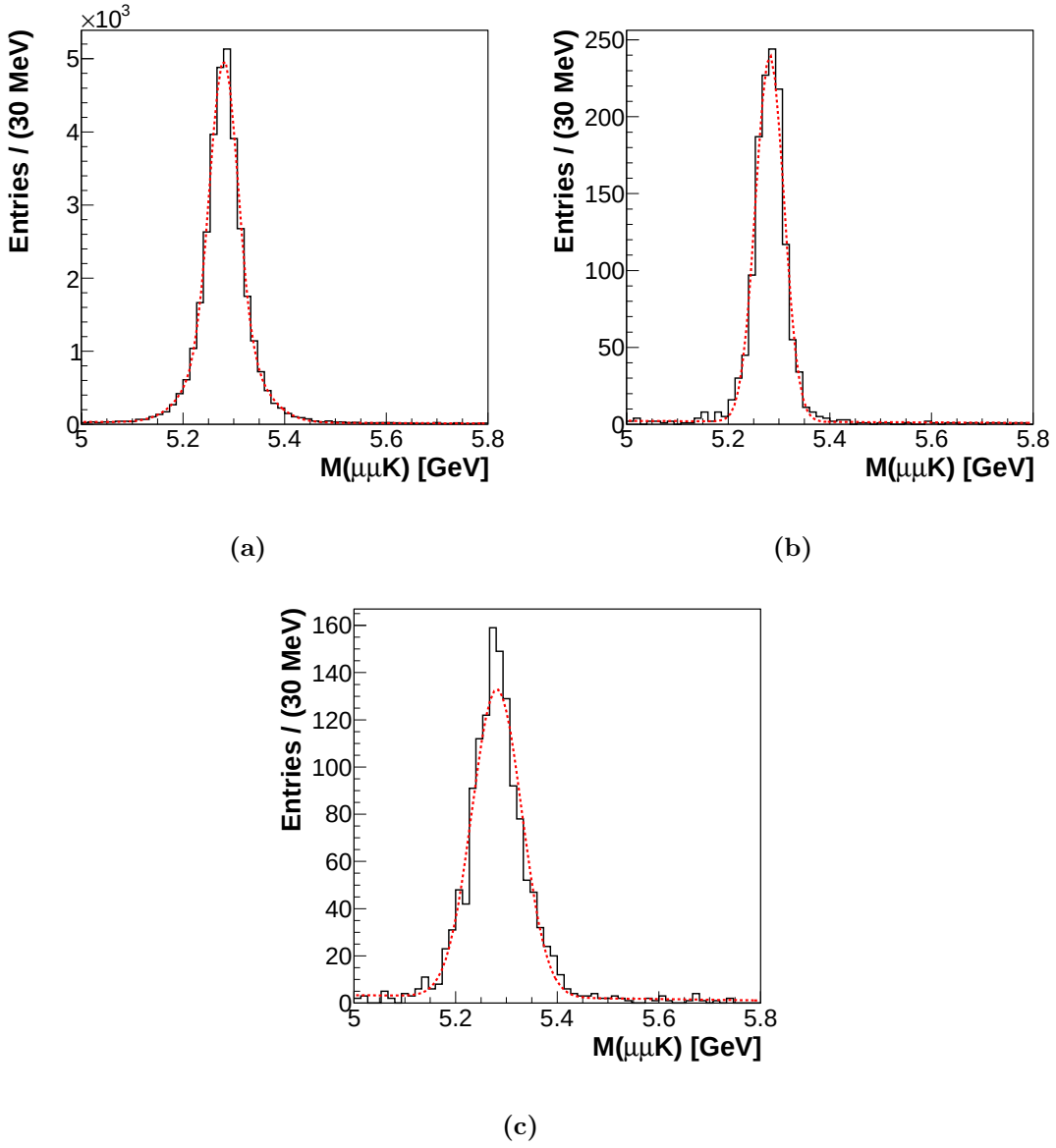


Figure 8.12: $\mu\mu K$ invariant mass spectrum in the `B+_mu6mu4K+` data sample, for all candidates (a) or only for those candidates with reconstructed muons in the barrel (b) or endcap (c) regions. The distributions are fitted by a first order polynomial plus a double Gaussian function (a) or by a first order polynomial plus a simple Gaussian function (b), (c), shown with a dotted line.

The search window region is defined as $\pm 2\sigma$ around the B^+ mass, where σ represents the computed value of the entire $\mu\mu K$ invariant mass spectrum ($\sigma = 43$ MeV). The efficiency of selection based on this mass window is discussed in Section 8.6.

Parameter	full η range	barrel ($\|\eta\| < 1.05$)	endcap ($\|\eta\| > 1.05$)
μ_1 [MeV]	5283 ± 1	5282 ± 1	5281 ± 2
σ_1 [MeV]	61 ± 2	27.9 ± 0.7	49.0 ± 1.4
μ_2 [MeV]	5280 ± 1	-	-
σ_2 [MeV]	27 ± 1	-	-

Table 8.6: Parameters of the fit functions used to describe the $M_{\mu\mu K}$ spectrum. The distribution is fitted by a first order polynomial plus a double Gaussian function, when the full η range is considered, and to a first order polynomial plus a simple Gaussian function, when only events in the barrel or endcap regions are considered.

The effects of applying the $L_{xy} > 0.5$ mm, $\alpha < 0.017$ and $I_{\mu\mu K} > 0.9$ cuts on the $\mu\mu K$ resolution are also investigated. The invariant mass spectrum is analysed independently for each type of selection cut and the mean and resolution are computed as explained above. It is observed that the mean value is kept constant in all cases. The width does not change either if the L_{xy} and $I_{\mu\mu K}$ selections are performed, but it is degraded if the cut on α is applied. The resolution is calculated to 48.8 ± 1.5 MeV in the latter case. This effect is similar to that observed in the analysis of the $\mu\mu$ spectrum and is an effect of the Lorentz boost in the z-direction. Another effect is the decrease of the linear combinatorial background observed before the selection is applied. This is expected, since the selection variables are tuned to reject fake B^+ candidates.

8.6 Determination of Efficiencies

The efficiency of reconstructing and observing B_s^0 and B^+ candidates can be factorised in four main categories: trigger, reconstruction, baseline selection and cut-based selection efficiencies:

$$\frac{\varepsilon_{B^+}}{\varepsilon_{B_s^0}} = \frac{\varepsilon_{B^+}^{trigger}}{\varepsilon_{B_s^0}^{trigger}} \frac{\varepsilon_{B^+}^{rec}}{\varepsilon_{B_s^0}^{rec}} \frac{\varepsilon_{B^+}^{baseline}}{\varepsilon_{B_s^0}^{baseline}} \frac{\varepsilon_{B^+}^{cuts}}{\varepsilon_{B_s^0}^{cuts}}. \tag{8.5}$$

Each of these efficiencies is studied separately in the following subsections. Some of these efficiencies are factorised further in order to investigate independently the distinct steps involved in the reconstruction and selection processes.

The statistical uncertainty of each efficiency is computed using the method introduced in Chapter 7, which is implemented in the `TGraphAsymmErrors` class of the ROOT [Ant09] analysis package. The statistical uncertainty of the efficiency obtained at each selection step is computed taking into account the number of events, that pass the respective selection step, and the number of events available after the previous selection step. In case several selection criteria are applied consecutively, the statistical uncertainty of the overall efficiency is calculated with

respect to the number of events available before the first selection step. For example, the uncertainty of the total efficiency is computed with respect to the total number of events. This takes the correlations between efficiencies of consecutive selection steps into account.

8.6.1 Trigger Efficiency

The efficiency of the trigger chain considered in the analysis (the di-muon trigger based on 6 and 4 GeV p_T muons) has been studied in detail in Chapter 7. The ratio of the overall trigger efficiencies for the $B^+ \to J/\psi K^+$ and $B_s^0 \to \mu^+\mu^-$ data sets has been estimated to (compare to (7.3)):

$$\frac{\varepsilon_{B^+}^{trigger}}{\varepsilon_{B_s^0}^{trigger}} = 1.042 \pm 0.007\,(stat) \pm 0.156\,(syst). \tag{8.6}$$

8.6.2 Reconstruction Efficiency

The reconstruction efficiency has to take two major aspects of the B_s^0 candidate reconstruction into account: the reconstruction of muon candidates and the di-muon vertex fit. The efficiency of each step is investigated. For the B^+ reconstruction two additional ingredients are used: the kaon candidate and the $\mu\mu K$ vertex fit. The efficiencies of these steps are investigated as well. It follows:

$$\frac{\varepsilon_{B^+}^{rec}}{\varepsilon_{B_s^0}^{rec}} = \frac{\varepsilon_{B^+}^{muon}}{\varepsilon_{B_s^0}^{muon}} \frac{\varepsilon_{B^+}^{\mu\mu\ vtx}}{\varepsilon_{B_s^0}^{\mu\mu\ vtx}} \varepsilon_{B^+}^{kaon}\ \varepsilon_{B^+}^{\mu\mu K\ vtx}. \tag{8.7}$$

Muon Reconstruction Efficiency

There are two components which affect the muon reconstruction efficiency (ε^{muon}): the efficiency of muon reconstruction proper ($\varepsilon^{\mu\mu\ rec}$) and the efficiency of the track selection ($\varepsilon^{trk\ sel}$). The former is estimated by comparing the total number of events containing at least two reconstructed combined muons to the total number of events analysed. It is known, that each event contains at least two muons at truth level. The computed efficiencies are:

$$\varepsilon_{B_s^0}^{\mu\mu\ rec} = \varepsilon_{B^+}^{\mu\mu\ rec} = 1, \tag{8.8}$$

with a statistical uncertainty of about 10^{-5}. No efficiency loss is noticed, due to the previously applied trigger requirements. Since the Event Filter performs a similar muon reconstruction as the Athena offline reconstruction, it is expected, that any efficiency loss due to the reconstruction of muons is already included in the trigger efficiency.

However, in order to check whether the efficiency of the muon reconstruction is consistent with other measurements, the efficiency is also studied without the trigger selection. The efficiency computed from the `Bs_mu6mu4` (`B+_mu6mu4K+`) data sets is in this case $0.9190 \pm 0.012\,(stat)$ ($0.9090 \pm 0.0007\,(stat)$). This is consistent

with independent muon performance studies performed with $\sqrt{s} = 14$ TeV MC data [Aad09] and with $\sqrt{s} = 7$ TeV recorded data [Atl10c], [Atl11a]. These studies show, that the efficiency has a strong dependence on the muon p_T. The efficiency rises with p_T in the lower p_T region, reaching a plateau around $p_T > 3\,(5)$ GeV in the endcap (barrel) region of the ATLAS detector. The efficiency at the plateau is above 95%. Several regions with lower efficiencies are observed around $\eta \sim 0$ (a region not covered by active detector elements) and $|\eta| \sim 1.1$ (the transition between the barrel and endcap regions).

A systematic uncertainty comes from misidentifying non-muon particles (typically pions or kaons) as muons. Since combined muons are reconstructed by statistically combining Muon Spectrometer and Inner Detector information, the fraction of misidentified muons is very small (e.g., the fraction of reconstructed muons produced by misidentified pions is $(0.070 \pm 0.012)\%$ [Atl10c]).

Since muon tracks are associated to ID tracks in the B_s^0 and B^+ reconstruction algorithms and a quality selection is imposed on each ID track (see Section 8.1), an additional factor has to be considered. This factor takes the efficiency loss due to these quality cuts into account. It is determined as the ratio of all tracks passing the quality requirements and the total number of tracks with at least 1 hit in the silicon-based detectors (in order to avoid including "TRT only" track segments):

$$\varepsilon_{B_s^0}^{trk\ sel} = 0.8702 \pm 0.0002\,(stat) \tag{8.9}$$

and

$$\varepsilon_{B^+}^{trk\ sel} = 0.86825 \pm 0.00012\,(stat). \tag{8.10}$$

A fraction of this efficiency loss is due to the probability of a charged track to leave a hit when passing through a Pixel or SCT module (which varies between 98% and 99.8% depending on the detectors' sub-components [Lim10]) and the requirements on the impact parameter with respect to the primary vertex.

Finally, the ratio of efficiencies of reconstructing events containing at least two muons becomes:

$$\frac{\varepsilon_{B^+}^{muon}}{\varepsilon_{B_s^0}^{muon}} = \frac{(\varepsilon_{B^+}^{trk\ sel})^2}{(\varepsilon_{B_s^0}^{trk\ sel})^2}\,\frac{\varepsilon_{B^+}^{\mu\mu\ rec}}{\varepsilon_{B_s^0}^{\mu\mu\ rec}} = 0.9942 \pm 0.0004\,(stat) \pm 0.0010\,(syst). \tag{8.11}$$

$\mu\mu$ Vertex Efficiency

The efficiency of reconstructing a di-muon vertex is computed as the ratio of the number of events with at least one successfully reconstructed vertex and the number of events with at least two reconstructed muons (including the ID track requirements). It is computed to:

$$\varepsilon_{B_s^0}^{\mu\mu\ vtx} = \varepsilon_{B^+}^{\mu\mu\ vtx} = 1, \tag{8.12}$$

with a statistical uncertainty of about 10^{-5}. The systematic uncertainties of reconstructed tracks included in the fit are used to compute the χ^2 of each reconstructed

vertex. This is used in a next step to select candidates based on a goodness of fit criterion and the efficiency of this selection is also computed (included in the baseline selection efficiency).

Kaon Reconstruction Efficiency

The efficiency of reconstructing a kaon candidate is equivalent to the efficiency of reconstructing a regular ID track (since no additional requirements are imposed on the kaon candidate in the B^+ reconstruction algorithm). The ID track efficiency is determined from previous MC studies to about 95% [Aad09]. This has to be corrected with the efficiency of the track quality requirements of $\varepsilon_{B^+}^{trk\ sel} = 0.86825 \pm 0.00012\,(stat)$, in a similar manner as in the muon efficiency calculation. A systematic uncertainty is computed from the contributions of fake tracks (about 0.2% of all tracks [Aad09]). Then the kaon reconstruction efficiency becomes:

$$\varepsilon_{B^+}^{kaon} = \varepsilon_{B^+}^{trk\ sel}\ \varepsilon^{ID\ track} = 0.82484 \pm 0.00012\,(stat) \pm 0.0016\,(syst). \tag{8.13}$$

$\mu\mu K$ Vertex Efficiency

The efficiency of reconstructing a $\mu\mu K$ vertex is computed as the ratio of the number of events containing at least one $\mu\mu K$ vertex and the number of events containing at least one $\mu\mu$ vertex:

$$\varepsilon_{B^+}^{\mu\mu K\ vtx} = 1, \tag{8.14}$$

with a statistical uncertainty of about 10^{-5}.

Total Reconstruction Efficiency

The total reconstruction efficiency is then computed by inserting all above mentioned efficiencies into (8.7):

$$\frac{\varepsilon_{B^+}^{rec}}{\varepsilon_{B_s^0}^{rec}} = 0.8201 \pm 0.0006\,(stat) \pm 0.0018\,(syst). \tag{8.15}$$

8.6.3 Baseline Selection Efficiency

The efficiency of the baseline selection is determined as:

$$\frac{\varepsilon_{B^+}^{baseline}}{\varepsilon_{B_s^0}^{baseline}} = \frac{\varepsilon_{B^+}^{\mu\mu\ p_T}}{\varepsilon_{B_s^0}^{\mu\mu\ p_T}}\ \frac{\varepsilon_{B^+}^{\chi^2(\mu\mu)}}{\varepsilon_{B_s^0}^{\chi^2(\mu\mu)}}\ \frac{\varepsilon_{B^+}^{L_{xy}>0}}{\varepsilon_{B_s^0}^{L_{xy}>0}}\ \frac{\varepsilon_{B^+}^{M_{\mu\mu}}}{\varepsilon_{B_s^0}^{M_{\mu\mu}}}\ \varepsilon_{B^+}^{K\ p_T}\ \varepsilon_{B^+}^{M_{\mu\mu K}}\ \varepsilon_{B^+}^{\chi^2(\mu\mu K)}, \tag{8.16}$$

where each factor represents the efficiency of an individual step of the selection (see Section 8.3). All factors are described in the following paragraphs.

The first step in the baseline selection is imposing cuts on the muons' minimum transverse momenta (a cut value of 6 GeV for the hardest muon and a cut value of 4 GeV for the second hardest muon). This yields efficiencies of:

$$\varepsilon_{B_s^0}^{\mu\mu\ p_T} = 0.9956 \pm 0.0005\,(stat) \tag{8.17}$$

and

$$\varepsilon_{B^+}^{\mu\mu\ p_T} = 0.9982 \pm 0.0002\,(stat). \tag{8.18}$$

The finite resolution of the muon p_T introduces a systematic uncertainty, when applying a cut on it. Since the trigger selection employs reconstruction algorithms (in the Event Filter) similar to the offline reconstruction, and the same threshold cuts on muon p_T were used in the baseline and in the trigger selections, this uncertainty is already included in the systematic uncertainty associated to the trigger selection (about 15%).

The next step is the selection of candidates with $\chi^2/ndf < 10$ for each reconstructed $\mu\mu$ vertex, with the efficiencies:

$$\varepsilon_{B_s^0}^{\chi^2(\mu\mu)} = 0.9922 \pm 0.0007\,(stat) \tag{8.19}$$

and

$$\varepsilon_{B^+}^{\chi^2(\mu\mu)} = 0.9924 \pm 0.0004\,(stat). \tag{8.20}$$

The following step is the selection of $\mu\mu$ vertices with $L_{xy} > 0$ mm. This results in the efficiencies:

$$\varepsilon_{B_s^0}^{L_{xy}>0} = 0.9627 \pm 0.0015\,(stat) \tag{8.21}$$

and

$$\varepsilon_{B^+}^{L_{xy}>0} = 0.9934 \pm 0.0004\,(stat). \tag{8.22}$$

The finite resolutions of the positions of the reconstructed vertices (used in the calculation of L_{xy}) produce a systematic uncertainty on the L_{xy} variable and on the efficiency obtained by applying a cut on it. This is discussed in Section 8.6.4, where a tighter cut is applied to suppress background events.

A $\mu\mu$ vertex invariant mass window of $[4300\,\text{MeV}, 7000\ \text{MeV}]$ is required by the B_s^0 selection and a window of $[2977\,\text{MeV}, 3217\,\text{MeV}]$ is required by the B^+ selection. These cuts have efficiencies of:

$$\varepsilon_{B_s^0}^{M_{\mu\mu}} = 0.9935 \pm 0.0006\,(stat) \tag{8.23}$$

and

$$\varepsilon_{B^+}^{M_{\mu\mu}} = 0.9714 \pm 0.0008\,(stat). \tag{8.24}$$

In addition, for the B^+ selection, the cut imposed on the kaon transverse momentum ($p_T > 1.5$ GeV) has an efficiency of:

$$\varepsilon_{B^+}^{K\ p_T} = 0.99974 \pm 0.00008\,(stat) \pm 0.02\,(syst). \tag{8.25}$$

The relative transverse momentum resolution of about 1% (2%) in the $|\eta| < 1.2$ ($|\eta| > 1.2$) region [Atl10b] was used to compute the errors on the kaon p_T. All p_T values are either increased or decreased by their errors resulting in a variation of the selection efficiency of about 2% around the central value (the systematic uncertainty in (8.25)).

The requirement, that the $\mu\mu K$ invariant mass is within the [4300MeV, 7000 MeV] range, has an efficiency of:

$$\varepsilon_{B^+}^{M_{\mu\mu K}} = 0.8201 \pm 0.0018\,(stat), \tag{8.26}$$

and the selection of $\mu\mu K$ vertices with $\chi^2/ndf < 10$ results in an efficiency of:

$$\varepsilon_{B^+}^{\chi^2(\mu\mu K)} = 0.9470 \pm 0.0012\,(stat). \tag{8.27}$$

Finally, the ratio of efficiencies of the baseline selection is computed based on the all above mentioned efficiencies:

$$\begin{aligned}\frac{\varepsilon_{B^+}^{baseline}}{\varepsilon_{B_s^0}^{baseline}} &= \frac{\varepsilon_{B^+}^{\mu\mu\ p_T}}{\varepsilon_{B_s^0}^{\mu\mu\ p_T}} \frac{\varepsilon_{B^+}^{\chi^2(\mu\mu)}}{\varepsilon_{B_s^0}^{\chi^2(\mu\mu)}} \frac{\varepsilon_{B^+}^{L_{xy}>0}}{\varepsilon_{B_s^0}^{L_{xy}>0}} \frac{\varepsilon_{B^+}^{M_{\mu\mu}}}{\varepsilon_{B_s^0}^{M_{\mu\mu}}} \varepsilon_{B^+}^{K\ p_T}\ \varepsilon_{B^+}^{M_{\mu\mu K}}\ \varepsilon_{B^+}^{\chi^2(\mu\mu K)} \\ &= 0.786 \pm 0.003\,(stat) \pm 0.016\,(syst).\end{aligned} \tag{8.28}$$

8.6.4 Cut Selection Efficiency

The efficiency of the event selection using the four discriminating variables is factorised as the product of the efficiency of each cut:

$$\frac{\varepsilon_{B^+}^{cuts}}{\varepsilon_{B_s^0}^{cuts}} = \frac{\varepsilon_{B^+}^{L_{xy}}}{\varepsilon_{B_s^0}^{L_{xy}}} \frac{\varepsilon_{B^+}^{\alpha}}{\varepsilon_{B_s^0}^{\alpha}} \frac{\varepsilon_{B^+}^{I_{\mu\mu K}}}{\varepsilon_{B_s^0}^{I_{\mu\mu}}} \frac{\varepsilon_{B^+}^{M_{\mu\mu K}}}{\varepsilon_{B_s^0}^{M_{\mu\mu}}}. \tag{8.29}$$

The efficiency of each cut is computed with respect to the previous selection step, and, therefore, the total efficiency does not depend on the correlations between the selection variables. The cut imposed on the transverse decay length ($L_{xy} > 0.5$ mm) results in efficiencies of:

$$\varepsilon_{B_s^0}^{L_{xy}} = 0.744 \pm 0.004\,(stat) \pm 0.004\,(syst) \tag{8.30}$$

and

$$\varepsilon_{B^+}^{L_{xy}} = 0.787 \pm 0.002\,(stat) \pm 0.016\,(syst). \tag{8.31}$$

The systematic uncertainty comes from propagating the uncertainties of x and y positions of the reconstructed primary and secondary vertices to the L_{xy} variable. All L_{xy} values are shifted either up or down by their errors and, in each case, the resulting variation of the efficiency around the central value is recorded. The

relative change of the efficiency was computed to about 0.5% for the `Bs_mu6mu4` sample and about 2% for the `B+_mu6mu4K+` sample. The higher relative uncertainty observed for the `B+_mu6mu4K+` sample is explained by a higher ratio of fake candidates (i.e., candidates that cannot be matched to true decays).

The efficiencies of the cut applied on the pointing angle ($\alpha < 0.017$) are:

$$\varepsilon^{\alpha}_{B^0_s} = 0.536 \pm 0.005\,(stat) \tag{8.32}$$

and

$$\varepsilon^{\alpha}_{B^+} = 0.552 \pm 0.003\,(stat). \tag{8.33}$$

The cut on the B^0_s/B^+ isolation ($I_{\mu\mu(K)} > 0.9$) has an efficiency of:

$$\varepsilon^{I_{\mu\mu}}_{B^0_s} = 0.340 \pm 0.006\,(stat) \tag{8.34}$$

and

$$\varepsilon^{I_{\mu\mu K}}_{B^+} = 0.296 \pm 0.004\,(stat). \tag{8.35}$$

As a study of the systematic uncertainty, the error of the isolation variable is computed based on the transverse momenta uncertainties (a relative transverse momentum resolution of about 1% (2%) in the $|\eta| < 1.2$ ($|\eta| > 1.2$) region [Atl10b]). The average computed error is about 0.002, which has no impact on the efficiency. The systematic uncertainty for this efficiency is therefore neglected in the following calculations.

The $\mu\mu$ invariant mass is restricted to the [5261 MeV, 5585 MeV] interval and the $\mu\mu K$ invariant mass is restricted to the [5193 MeV, 5365 MeV] interval. This selection results in an efficiency of:

$$\varepsilon^{M_{\mu\mu}}_{B^0_s} = 0.779 \pm 0.009\,(stat) \pm 0.016\,(syst) \tag{8.36}$$

and

$$\varepsilon^{M_{\mu\mu K}}_{B^+} = 0.896 \pm 0.005\,(stat) \pm 0.018\,(syst). \tag{8.37}$$

The systematic uncertainty of the $\mu\mu$ or $\mu\mu K$ invariant mass selection comes from the fraction of fake events, that could not be matched to truth decays, of about 2%, in both cases. In the case of the $\mu\mu K$ selection, the purity is increased from about 90% (after baseline selection) to about 98%, due to the selection cuts which help to reduce fake events.

Finally, the ratio of the overall cut selection efficiencies is computed from:

$$\begin{aligned} \frac{\varepsilon^{cuts}_{B^+}}{\varepsilon^{cuts}_{B^0_s}} &= \frac{\varepsilon^{L_{xy}}_{B^+}}{\varepsilon^{L_{xy}}_{B^0_s}} \frac{\varepsilon^{\alpha}_{B^+}}{\varepsilon^{\alpha}_{B^0_s}} \frac{\varepsilon^{I_{\mu\mu K}}_{B^+}}{\varepsilon^{I_{\mu\mu}}_{B^0_s}} \frac{\varepsilon^{M_{\mu\mu K}}_{B^+}}{\varepsilon^{M_{\mu\mu}}_{B^0_s}} \\ &= 1.09 \pm 0.03\,(stat) \pm 0.04\,(syst). \end{aligned} \tag{8.38}$$

8.6.5 Summary of Efficiency Calculation

The efficiencies mentioned above are summarised in Table 8.7. The efficiency of each selection step is computed with respect to the previous step. ε^{cuts} represents the product of the four efficiencies of the cut selections ($\varepsilon^{L_{xy}}$, ε^{α}, $\varepsilon^{I_{\mu\mu(K)}}$, $\varepsilon^{M_{\mu\mu(K)}}$). The total efficiency ε^{total} is the product of the trigger, reconstruction, baseline selection and cut selection.

	`Bs_mu6mu4`	`B+_mu6mu4K+`
$\varepsilon^{trigger}$	$0.341 \pm 0.002\,(stat) \pm 0.019\,(syst)$	$0.3554 \pm 0.0012\,(stat) \pm 0.05\,(syst)$
ε^{rec}	$0.7582 \pm 0.0003\,(stat) \pm 0.0005\,(syst)$	$0.6218 \pm 0.0004\,(stat) \pm 0.0013\,(syst)$
$\varepsilon^{baseline}$	$0.9448 \pm 0.0018\,(stat) \pm 0.0000\,(syst))$	$0.742 \pm 0.002\,(stat) \pm 0.015\,(syst)$
$\varepsilon^{L_{xy}}$	$0.744 \pm 0.004\,(stat) \pm 0.004\,(syst)$	$0.787 \pm 0.002\,(stat) \pm 0.016\,(syst)$
ε^{α}	$0.536 \pm 0.005\,(stat) \pm 0.000\,(syst)$	$0.552 \pm 0.003\,(stat) \pm 0.000\,(syst)$
$\varepsilon^{I_{\mu\mu(K)}}$	$0.340 \pm 0.006\,(stat) \pm 0.000\,(syst)$	$0.296 \pm 0.004\,(stat) \pm 0.000\,(syst)$
$\varepsilon^{M_{\mu\mu(K)}}$	$0.779 \pm 0.009\,(stat) \pm 0.016\,(syst)$	$0.896 \pm 0.005\,(stat) \pm 0.018\,(syst)$
ε^{cuts}	$0.106 \pm 0.003\,(stat) \pm 0.002\,(syst)$	$0.1152 \pm 0.0017\,(stat) \pm 0.003\,(syst)$
ε^{total}	$0.0258 \pm 0.0008\,(stat) \pm 0.006\,(syst)$	$0.0189 \pm 0.0004\,(stat) \pm 0.003\,(syst)$

Table 8.7: Summary of reconstruction and selection efficiencies.

8.7 Analysis of the Background Sample

The next step in the analysis chain is extracting the number of observed B_s^0 and B^+ candidates from the same data sample. The combinatorial background sample, `bb_mu6mu4X`, is assumed to describe the most important background contribution in the B_s^0 mass region. Two independent searches are performed to identify event candidates for the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays. The number of resulting candidates is later used to estimate the $N_{B_s^0}/N_{B^+}$ term in (4.1).

8.7.1 Search for $B_s^0 \to \mu^+\mu^-$

Following the trigger and baseline selection 2 049 events are observed, each containing one $B_s^0 \to \mu^+\mu^-$ decay candidate. Fig. 8.4 gives an idea of the large amount of background events, that need to be rejected in order to observe signal events. For this purpose the four discriminating variables introduced in Section 8.4 are employed: L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$. A cut flow approach is investigated. This approach involves applying successive cuts on these quantities and observing the effect produced on the remaining events:

- $L_{xy} > 0.5$ mm,
- $\alpha < 0.017$,

- $I_{\mu\mu} > 0.9$,
- $M_{\mu\mu} \in [5261\ \text{MeV}, 5585\ \text{MeV}]$.

In the ideal case, this is a "cut and count" method. However, the limited statistics of the data sample does not allow this simple approach (i.e., no candidates are left when all cuts are applied). Therefore, a cut factorisation method is investigated. Each cut is applied independently to the total number of candidates left after the baseline selection. The selection efficiency of each cut is then computed, and the total selection efficiency is determined as:

$$\varepsilon^{total} = \varepsilon^{L_{xy}} \cdot \varepsilon^{\alpha} \cdot \varepsilon^{I_{\mu\mu}} \cdot \varepsilon^{M_{\mu\mu}}, \tag{8.39}$$

where $\varepsilon^{L_{xy}}$, ε^{α}, $\varepsilon^{I_{\mu\mu}}$ and $\varepsilon^{M_{\mu\mu}}$ represent the selection efficiencies of applying the corresponding cuts independently. This holds true only if there are no correlations between these variables. The linear correlation coefficients between these variables are shown in Table 8.8

	L_{xy}	α	$I_{\mu\mu}$	$M_{\mu\mu}$
L_{xy}	1	-0.33	-0.08	-0.15
α		1	-0.13	0.17
$I_{\mu\mu}$			1	-0.01
$M_{\mu\mu}$				1

Table 8.8: Linear correlation coefficients between the L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$ variables for B_s^0 candidates passing the baseline selection in the `bb_mu6mu4X` sample. Statistical uncertainties are computed to be 0.02 in each case.

The strongest linear correlation is observed between L_{xy} and α. This is expected due to the way L_{xy} is constructed (i.e., it takes into account how well the B_s^0 flight direction is aligned to the reconstructed momentum in the transverse plane). The other variables are less correlated, but their correlation coefficients are nevertheless not negligible. In order to be able to use the cut factorisation as defined in (8.39), the effect of these correlations has to be considered as a systematic uncertainty on the final result.

A possible approach to take correlations into account is given by the "projection and correlation approximation" (PCA) method [Kar98]. This method takes the covariance matrix and the 1-dimensional probability density functions (p.d.f.) of the discriminating variables into account in order to construct an n-dimensional p.d.f. The method is exact only if the n-dimensional p.d.f. is an n-dimensional Gaussian function:

$$G(\mathbf{y}) = (2\pi)^{-n/2}|V|^{-1/2} \exp(-\frac{1}{2}\mathbf{y}^T V^{-1}\mathbf{y}), \tag{8.40}$$

where $\mathbf{y}$ is the n-dimensional vector of variables and V is its covariance matrix. Since the selection variables resulting in the analysis do not follow a normal distribution, the p.d.f. of each variable can be transformed to a Gaussian distribution by:

$$y_i(x_i) = \sqrt{2}\ \mathrm{erf}^{-1}(2F(x_i) - 1), \tag{8.41}$$

where x_i represents any selection variable, $F(x_i)$ its cumulative distribution function and erf^{-1} the inverse error function. The correlation matrix is preserved under this transformation. In addition, the transformation only ensures that the projections of $G(\mathbf{y})$ follow a normal distribution. The main assumption of PCA is that, if the projections follow a normal distribution, the n-dimensional p.d.f. will be an n-dimensional Gaussian distribution as defined in (8.40).

A toy Monte Carlo tool is implemented based on this approach. The distributions of the four discriminating variables are parametrized by fitting them to different functions describing their shape. The functions obtained from the fit are then used to make the transformation of the initial p.d.f. to a Gaussian distribution according to (8.41). The covariance matrix of the transformed variables is calculated and a 4-dimensional Gaussian p.d.f. is constructed according to (8.40). In a next step, 500 million events are simulated. Each event consists of the 4 selection variables distributed according to the 4-dimensional Gaussian p.d.f. computed in the previous step. The 4 generated normal distributions are then transformed back to the original distributions, using the inverse of the transformation shown in (8.41). The statistics available from the toy MC simulation is large enough to apply the cut based selection without cut factorisation. Efficiencies of the cut selection with and without cut factorisation are then compared. Details regarding the implementation of all steps are given in Appendix C.

It is important to note that, besides the correct parametrization of the discriminating variables' distributions, the linear correlation coefficients must also be preserved in the simulation process. It is observed that the transformation to a Gaussian distribution does not preserve these correlations in all cases. Differences are notably large when large variations of the linear correlation coefficient exist between different space regions. This is the case when analyzing the correlation between L_{xy} and α. A small correlation ($\rho \simeq 0.1$) is observed in the region $0 < \alpha < 0.2$, while in the outer region the correlation is much higher ($\rho \simeq 0.3$). In order to preserve the correlations in the PCA method, events are preselected according to $\alpha < 0.18$. The correlation coefficients computed from preselected events of the `bb_mu6mu4X` sample and from events generated with the toy MC tool are shown in Table 8.9. The choice of the preselection cut is made such that the correlation coefficients of simulated variables are within a one standard deviation interval of the input coefficients. The preselection reduces the amount of events available and, therefore, increases the statistical uncertainty, but offers the advantage of a better estimation of the systematic uncertainty. This systematic uncertainty is determined to be about 13%.

Using the $\alpha < 0.18$ preselection cut, 673 events are available for further analysis. The selection cuts are applied again on this sample and their efficiencies

	L_{xy}	α	$I_{\mu\mu}$	$M_{\mu\mu}$
L_{xy}	1	-0.10	-0.17	-0.16
α		1	-0.09	0.13
$I_{\mu\mu}$			1	0.03
$M_{\mu\mu}$				1

	L_{xy}	α	$I_{\mu\mu}$	$M_{\mu\mu}$
L_{xy}	1	-0.11	-0.17	-0.17
α		1	-0.07	0.12
$I_{\mu\mu}$			1	0.05
$M_{\mu\mu}$				1

Table 8.9: Linear correlation coefficients between the L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$ variables. The coefficients in the left table are computed from the 673 input events obtained after preselection according to $\alpha < 0.18$ with a statistical uncertainty of 0.04. The coefficients in the right table are computed from a simulated sample of 1 million events, attempting to describe the situation when the preselection is employed. The statistical error in this case is 0.0009.

are computed independently. The values of the efficiencies are summarised in Table 8.10. The total selection efficiency is computed from:

$$\begin{aligned} \varepsilon^{total} &= \varepsilon^{pres}\ \varepsilon^{fact} \\ &= \varepsilon^{pres}\ \varepsilon^{L_{xy}}\ \varepsilon^{\alpha}\ \varepsilon^{I_{\mu\mu}}\ \varepsilon^{M_{\mu\mu}}, \end{aligned} \tag{8.42}$$

where $\varepsilon^{L_{xy}}$, ε^{α}, $\varepsilon^{I_{\mu\mu}}$ and $\varepsilon^{M_{\mu\mu}}$ are computed with respect to the preselection according to $\alpha < 0.18$ (i.e., ε^{pres}).

Selection step	Events remaining	Efficiency
baseline selection	2 049	0.00202 ± 0.00004
$\alpha < 0.18$	673	0.328 ± 0.011
$L_{xy} > 0.5$ mm	215	0.319 ± 0.019
$\alpha < 0.017$	25	$0.037\,^{+0.009}_{-0.007}$
$I_{\mu\mu} > 0.9$	13	$0.019\,^{+0.007}_{-0.005}$
$M_{\mu\mu} \in [5261\ \text{MeV}, 5585\ \text{MeV}]$	86	$0.127\,^{+0.014}_{-0.013}$
Cut factorisation	-	$(2.9 \pm 1.1\,(stat) \pm 0.4\,(syst)) \times 10^{-5}$
Total	-	$(9.3 \pm 4\,(stat) \pm 1.3\,(syst)) \times 10^{-6}$

Table 8.10: Summary of efficiencies of selection variables of B_s^0 candidates in the `bb_mu6mu4X` sample.

The expected number of $B_s^0 \to \mu^+\mu^-$ candidates observed in the `bb_mu6mu4X` data sample with the above mentioned selection cuts becomes:

$$\begin{aligned} N_{bgd} &= N^{baseline}\ \varepsilon^{total} \\ &= 0.020 \pm 0.008\,(stat) \pm 0.003\,(syst), \end{aligned} \tag{8.43}$$

which corresponds to about 15 pb^{-1}. This is used to compute an upper limit on the expected $B_s^0 \to \mu^+\mu^-$ branching ratio (in Chapter 9).

Finally, the number of expected background events is extrapolated to an integrated luminosity of 1 fb^{-1} and is shown in Fig. 8.13. The invariant mass spectrum observed in the `Bs_mu6mu4` sample after all selection cuts is also scaled to the same integrated luminosity, assuming the Standard Model cross section. About 0.4 signal events and about 1.3 background events are observed in the signal region. Compared to Fig. 8.4, the expected number of background events is reduced by several orders of magnitude and is compatible with the number of signal events. The number of expected background events in the left and right sidebands has also been estimated using the cut factorisation. The average number of expected background events is about 0.20 events/50 MeV (0.17 events/50 MeV) in the left (right) sideband and it is compatible to the average number of expected background events in the signal region (about 0.23 events/50 MeV).

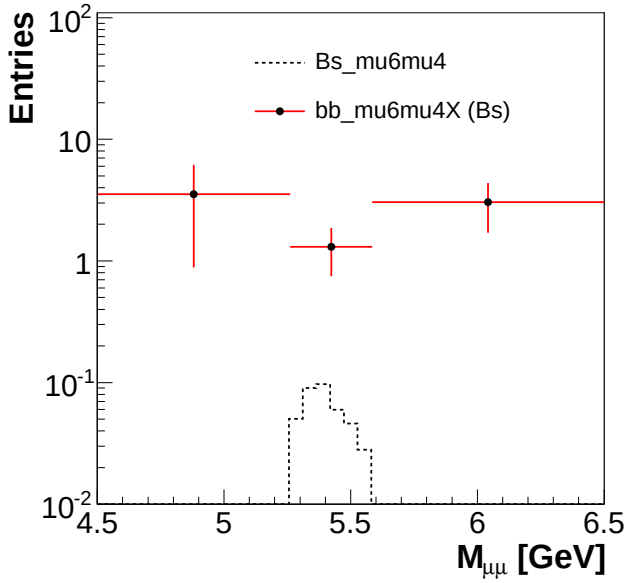


Figure 8.13: Invariant mass spectrum of B_s^0 candidates observed in the `Bs_mu6mu4` sample (line) after the entire selection chain has been applied. The expected number of B_s^0 candidates in the `bb_mu6mu4X` sample (dots) after all the selection cuts (including cut factorisation) have been applied. Entries are scaled to an integrated luminosity of 1 fb^{-1}.

8.7.2 Selection of $B^+ \to J/\psi K^+$

After the trigger and baseline selections, 6 731 events remain in the `bb_mu6mu4X` data set. In the next step, these events are filtered further by applying the same selection cuts as in the analysis performed on the `B+_mu6mu4K+` sample:

- $L_{xy} > 0.5$ mm,
- $\alpha < 0.017$,
- $I_{\mu\mu} > 0.9$.

These cuts result in 363 events remaining in the sample. The invariant mass distribution of all events in the [4.3 GeV, 7 GeV] range observed before and after applying the selection cuts is shown in Fig. 8.14 (a). The selection cuts reduce the amount of combinatorial background on either side of the signal peak, while the shape of the signal peak is still clearly visible.

The next step is the estimation of the number of signal events observed in the sample. Since the combinatorial background is not entirely removed by the cut selection, the contribution of the sideband region has to be statistically subtracted from the signal region. A more detailed view of the invariant mass distribution (in the [5 GeV, 5.8 GeV] range) is shown in Fig. 8.14 (b-d). The aim is to find a function which accurately describes the $M_{\mu\mu K}$ distribution and then to fit this function to the observed distribution. It has been already shown that the shape of the signal peak is well represented by a double Gaussian function (i.e., the sum of two Gaussian functions, each aimed to describe different resolution effects observed in the barrel and endcap regions). The mean of each Gaussian function is kept as a free parameter in the [5275 MeV, 5285 MeV] range during the minimisation process. Due to the small number of events remaining, the sidebands can be approximated by several different functions. Three scenarios are considered, in which the sidebands are parametrized by a first order polynomial, a second order polynomial, or an exponential function. Thus, the functions considered for the fit are:

$$f_1(x) = A_1 \ \exp\frac{-(x-\mu_1)^2}{2\sigma_1^2} + A_2 \ \exp\frac{-(x-\mu_2)^2}{2\sigma_2^2} + a_0 + a_1 x, \tag{8.44}$$

$$f_2(x) = A_1 \ \exp\frac{-(x-\mu_1)^2}{2\sigma_1^2} + A_2 \ \exp\frac{-(x-\mu_2)^2}{2\sigma_2^2} + a_0 + a_1 x + a_2 x^2, \tag{8.45}$$

and

$$f_3(x) = A_1 \ \exp\frac{-(x-\mu_1)^2}{2\sigma_1^2} + A_2 \ \exp\frac{-(x-\mu_2)^2}{2\sigma_2^2} + \exp(a_0 + a_1 x). \tag{8.46}$$

A binned maximum likelihood fit is performed with each of the functions defined above and the fit results are projected onto the distributions in Fig. 8.14 (b-d). The parameters of these functions (means and widths of the Gaussian functions) are shown in Table 8.11. In each case, the mean of the double Gaussian function is computed as the first moment of the double Gaussian function and the width is computed as the second central moment of the double Gaussian function. Their

values are also given in Table 8.11. In all cases, the computed mean is consistent with the B^+ mass value used in the generation (5279 MeV). The width varies between 39 MeV and 56 MeV and is in rough terms similar to that observed in the analysis of the `B+_mu6mu4K+` sample (about 43 MeV).

Parameter	$f_1(x)$	$f_2(x)$	$f_3(x)$
μ_1 [MeV]	5280 ± 5	5281 ± 4	5282 ± 5
σ_1 [MeV]	36 ± 6	34 ± 5	39 ± 6
μ_2 [MeV]	5277 ± 5	5283 ± 6	5282 ± 5
σ_2 [MeV]	100 ± 7	100 ± 7	39 ± 6
a_0 [MeV]	23 ± 6	374 ± 100	22 ± 4
a_1	-4 ± 1	-132 ± 36	-4.0 ± 0.8
a_2 [MeV^{-1}]	-	12 ± 3	-
μ [MeV]	5280 ± 5	5282 ± 5	5282 ± 5
σ [MeV]	39 ± 6	56 ± 5	39 ± 6
N_{sig}	$79 \pm 25\,(stat)$	$83 \pm 23\,(stat)$	$82 \pm 9\,(stat)$

Table 8.11: Resulting parameters of the fit functions used to describe the shape of the signal peak and combinatorial background sidebands: average mean and width of the double Gaussian component (μ, σ), number of signal events N_{sig} estimated after subtracting the background sidebands from the signal region. The statistical errors are computed from the covariance matrix of the fit functions' parameters.

The number of observed B^+ candidates is then estimated as the integral of the double Gaussian function, in the [5193 MeV, 5365 MeV] interval (the same interval used in the analysis of the `B+_mu6mu4K+` data set). The covariance matrix of the parameters obtained for the double Gaussian function is then used to compute the statistical uncertainty. The number of signal events and their statistical errors are also summarised in Table 8.11. In the following, the number of expected B^+ events is estimated based on the exponential parametrization of the background sidebands. The exponential parametrization is the shape which is expected to describe more closely the sidebands in the wider [4.3 GeV, 7 GeV] interval (seen in Fig. 8.14 (a)) and also produces a better result of the fit, reflected in a smaller statistical uncertainty. A systematic uncertainty results from different methods to parametrize the background sidebands. It is computed as the standard deviation from the mean of the number of observed events in each case. The number of observed B^+ candidates in the `bb_mu6mu4X` data set is:

$$N_{sig} = 82 \pm 9\,(stat) \pm 2\,(syst), \tag{8.47}$$

corresponding to an integrated luminosity of about 15 pb^{-1}. This number will later be used in the calculation of an upper limit on the expected branching ratio (in Chapter 9). It can be extrapolated to a certain integrated luminosity by weighting

it with the production cross section computed during the PYTHIA generation of events.

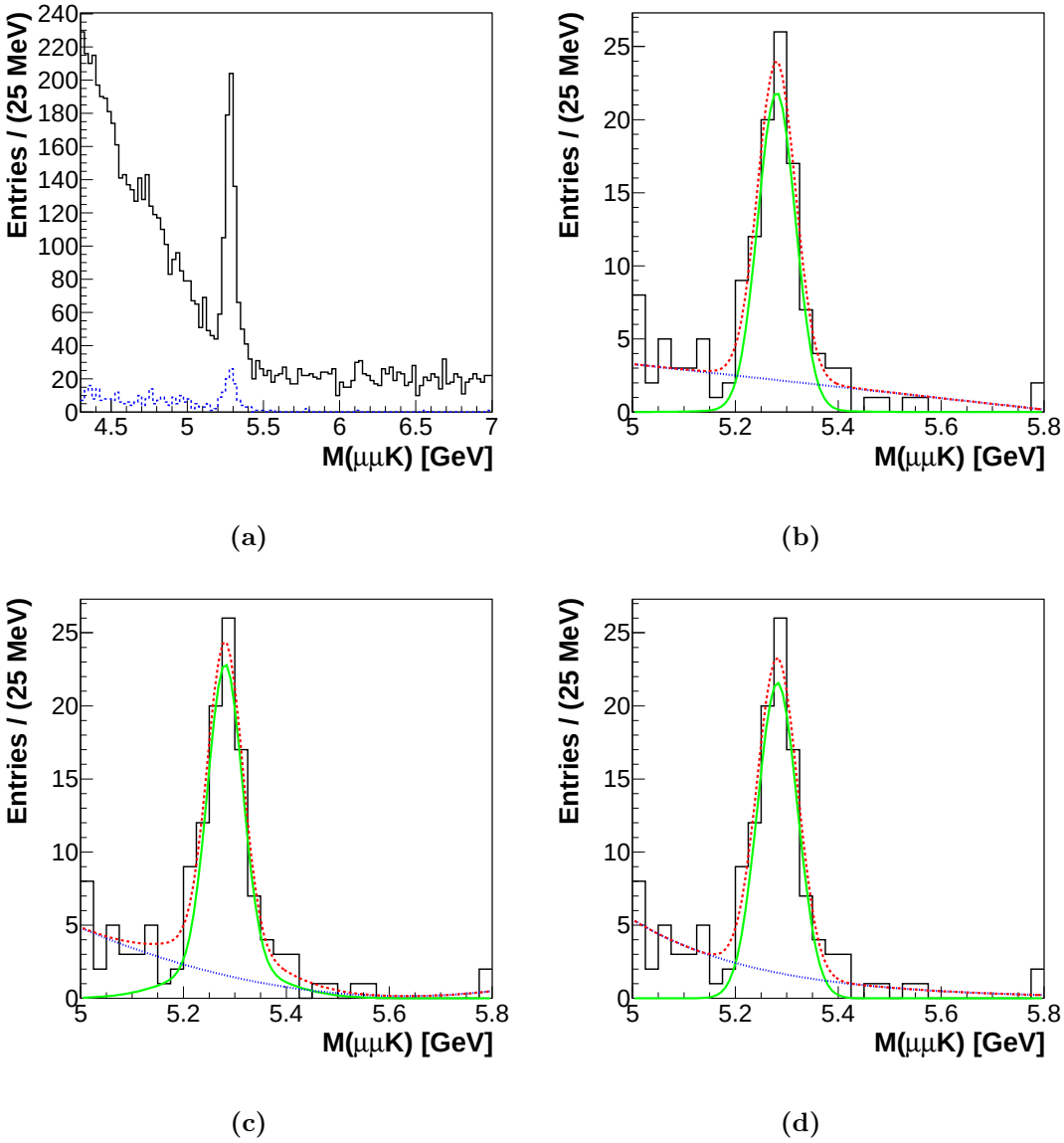


Figure 8.14: The $\mu\mu K$ invariant mass spectrum in the `bb_mu6mu4X` data sample. (a) All candidates in the [4.3 GeV, 7 GeV] interval are considered before (black, solid) or after (blue, dashed) the selection cuts are applied. (b-d) Only candidates in the [5 GeV, 5.8 GeV] window are shown. The results from the fit to the distributions are drawn with a red, dashed line. The fit functions are formed by a double Gaussian function (shown separately as a green, solid shape) plus a first order polynomial (b), a second order polynomial (c) or an exponential function (d). The polynomial and the exponential functions are shown separately as a blue, dotted shape.

Chapter 9

Prospects for Measuring the $B_s^0 \to \mu^+\mu^-$ Branching Ratio

The overall strategy for the study of the $B_s^0 \to \mu^+\mu^-$ decay has been introduced in Chapter 4. The goal of this thesis is to determine an estimate of an upper limit on the $B_s^0 \to \mu^+\mu^-$ branching ratio. Calculations are based on simulated MC data, which require a special treatment of the limit calculation. The master formula used for the branching ratio calculation is:

$$BR(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s^0}}{N_{B^+}} \frac{\alpha_{B^+}}{\alpha_{B_s^0}} \frac{\epsilon_{B^+}}{\epsilon_{B_s^0}} \frac{f_u}{f_s} BR(B^+ \to J/\psi K^+) BR(J/\psi \to \mu^+\mu^-), \tag{9.1}$$

where each term has been described in detail in Chapter 4. The analysis presented in the previous chapters gives estimates for each of the factors in (9.1): the ratio of the acceptances was investigated in Chapter 6, the ratio of the reconstruction and selection efficiencies was studied in Chapters 7 and 8 and the numbers of events corresponding to the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays were estimated in Chapter 8. All other factors are obtained from literature.

In this chapter, several methods to compute upper limits are introduced (a Bayesian approach and a frequentist approach). These methods are then used to determine estimates of the MC limit on the branching ratio. A method to optimise the selection cuts is also introduced, which aims at optimising the expected upper limit. Finally, the ATLAS sensitivity to measure the branching ratio is determined.

9.1 Exclusion Limit Calculation

In many instances, when exclusion limits have to be computed on a specific parameter, the problem is reduced to a counting experiment, where the number of observed events n, representative for the physics study, is recorded. This number is then compared to the number of events μ *expected* to be observed under the same conditions as n. Typically, μ is expressed as:

$$\mu = s\ \epsilon + b, \tag{9.2}$$

where s and b represent the expected number of signal and background events and ϵ is a factor which accounts for the acceptance times the efficiency of observing signal events. An exclusion limit is then determined by a confidence interval of the unknown value of the expected number of signal events s, such that n is consistent with μ within a certain confidence level. Using the probability distribution function (p.d.f.) $P(\mu|n)$ and assuming a one-sided confidence interval:

$$\alpha = \frac{\int_0^{\mu^{UL}} P(\mu|n)d\mu}{\int_0^{+\infty} P(\mu|n)d\mu} = \frac{\int_0^{s^{UL}\epsilon+b} P(\mu|n)d\mu}{\int_0^{+\infty} P(\mu|n)d\mu}, \tag{9.3}$$

where the s^{UL} is the upper limit on the true value of the expected number of signal events and *alpha* represents the degree of belief in the confidence interval (i.e., the confidence level).

9.1.1 A Bayesian Approach

Physics analyses typically give results regarding a certain observable (in this case n), so $P(n|\mu)$ can be determined experimentally, but not $P(\mu|n)$. One solution is to employ the Bayes theorem, which results in:

$$P(\mu|n) = \frac{P(n|\mu)P(\mu)}{\int_0^{+\infty} P(n|\mu)P(\mu)d\mu}, \tag{9.4}$$

and (9.3) therefore becomes:

$$CL = \frac{\int_0^{\mu^{UL}} P(n|\mu)P(\mu)d\mu}{\int_0^{+\infty} P(n|\mu)P(\mu)d\mu}. \tag{9.5}$$

In this interpretation, $P(n|\mu)$ is assumed to be a Poissonian distribution. Several options are available to parametrize the prior $P(\mu)$ for μ, which is further interpreted as a function of the priors for s, ϵ and b [Hei04]. In the following calculations, a flat prior is used for $P(\mu)$. The null signal hypothesis ($s = 0$) is assumed for the expected number of signal events (discussed in Section 9.1.3) and ε and b are considered to be Gaussian distributed around their mean values. The uncertainties of ε and b are taken into account by convoluting the $P(n|\mu)$ probability function with the known Gaussian distributions of ε and b:

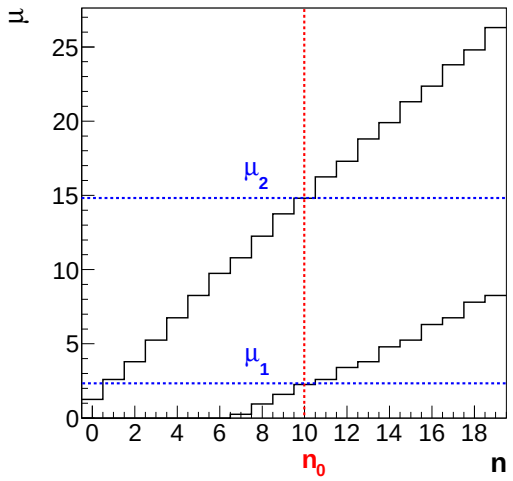
$$P(n|\mu) = P(n|s,\epsilon,\sigma_\varepsilon,b,\sigma_b) = \iint P(n|s(\epsilon+\epsilon')+b+b')g(\epsilon',\sigma_\epsilon)g(b',\sigma_b)d\epsilon' db', \tag{9.6}$$

where $g(\epsilon,\sigma_\epsilon)$ and $g(b,\sigma_b)$ are the Gaussian distributions of ϵ and b. The effect of including uncertainties into the calculation is, that the confidence interval increases and the computed upper limit shifts to a higher value [Con03].

The following calculations of the upper limits using the Bayesian method are performed using an algorithm, which is a C++ reimplementation of the original `bayes.f` algorithm [Con98]. The uncertainties of the efficiency and the number of expected background events are also included in the computations.

9.1.2 The Feldman Cousins Unified Approach

One common problem with the Bayesian approach is the difficult choice of the priors, which must be informative but not biased by one person's expectation. One alternative to this method is the Feldman Cousins interpretation of confidence intervals [Fel98], which determines a confidence interval for the unknown true value of μ without a direct estimation of $P(\mu|n)$. This technique is based on the Neyman construction [Ney37] of a map of probabilities $P(n_i|\mu_j)$ corresponding to each possible value for n_i and μ_j. This map is used to construct a confidence belt $[n_{i_1}, n_{i_2}]$ for each μ_j, such that $P(n_i \in [n_{i_1}, n_{i_2}]|\mu_j) = \alpha$. Using the confidence belt for a given measurement n_0, a confidence interval $[\mu_1, \mu_2]$ is determined, such that $P(\mu \in [\mu_1, \mu_2]) = \alpha$. This is obtained by integrating over all $P(n_0|\mu_j)$ values. An improvement is achieved by reordering the $P(n_0|\mu_j)$ elements considered in the integration. These elements are sorted in decreasing order of $P(n_0|\mu_{best})/P(n_0|\mu_j)$. μ_{best} represents the value of μ, which maximises $P(n_0|\mu)$. The advantage of the re-ordering is that the result is automatically an upper limit or a double sided interval, depending on n_0 and μ. As an example, the confidence belt computed for a 95% confidence interval and for $b = 3$ is shown in Fig. 9.1. The lower (μ_1) and upper (μ_2) limits obtained for a given number of observed events n_0 are also highlighted.



(a)

Figure 9.1: Confidence belt obtained with the Feldman Cousins approach for a 95% confidence level and $b = 3$.

In the following upper limit calculations, a re-implementation of the original algorithm in the `TFeldmanCousins` class from the ROOT package [Ant09] was used. The re-implementation incorporates the uncertainties of ϵ and b into $P(n|\mu)$ in a manner similar to the Bayesian method.

9.1.3 Monte Carlo Limit on the Branching Ratio

Assuming the measured number of $B_s^0 \to \mu^+\mu^-$ decays $N_{B_s^0}$ is small, an upper limit on the number of expected signal events $N_{B_s^0}^{UL}$ can be determined. The computation of the upper limit on the branching ratio is obtained by substituting $N_{B_s^0}$ with $N_{B_s^0}^{UL}$ in (9.1). For $N_{B_s^0} = 1$ in (9.1), the right side expression becomes the single event sensitivity (ses) and:

$$\begin{aligned} BR(B_s^0 \to \mu^+\mu^-)^{UL} &= \frac{N_{B_s^0}^{UL}}{N_{B^+}} \frac{\alpha_{B^+}}{\alpha_{B_s^0}} \frac{\epsilon_{B^+}}{\epsilon_{B_s^0}} \frac{f_u}{f_s} BR(B^+ \to J/\psi K^+) BR(J/\psi \to \mu^+\mu^-) \\ &= N_{B_s^0}^{UL} \cdot ses. \end{aligned} \tag{9.7}$$

The single event sensitivity effectively plays the role of the acceptance times efficiency (ϵ) in the computation of the upper limit on the expected number of signal events. Therefore, the error on ses should also be correctly estimated and included in the calculation.

For limit calculations based on MC samples, the expected number of signal decays is unknown, since it depends on the theoretical model assumed. Thus, in order to have a model independent estimation of the branching ratio, the null signal hypothesis is used, i.e., $s = 0$ in (9.2). Considering the observed number of decays follows a Poissonian distribution $P(N_{B_s^0}|b, \sigma_b)$, $N_{B_s^0}$ is expressed as the sum of all possible outcomes weighted with their probabilities for a given expected number of background events:

$$N_{B_s^0} = \sum_{i=0}^{\infty} N_{B_s^0,i}\ P(N_{B_s^0,i}|b, \sigma_b). \tag{9.8}$$

From this follows:

$$N_{B_s^0}^{UL} = \sum_{i=0}^{\infty} N_{B_s^0,i}^{UL}\ P(N_{B_s^0,i}|b, \sigma_b) \tag{9.9}$$

and

$$BR(B_s^0 \to \mu^+\mu^-)^{UL} = ses\ \sum_{i=0}^{\infty} N_{B_s^0,i}^{UL}\ P(N_{B_s^0,i}|b, \sigma_b), \tag{9.10}$$

the last equation being the one used in the following estimations of upper limits on the branching ratio.

9.2 Upper Limit Estimates on the Branching Ratio

All input quantities needed for the computation of the limit on the branching ratio have been estimated in the previous chapters. The ratio of the acceptances

corresponding to the signal and the reference channels are determined in Chapter 6. The overall reconstruction and selection efficiencies are discussed in detail in Chapter 8. The expected number of $B^+ \to J/\psi K^+$ decays as well as the expected number of background events for the $B_s^0 \to \mu^+\mu^-$ channel are also estimated in Chapter 8. The number of events N_0 (observed or expected) is then scaled to an integrated luminosity of 1 fb^{-1} by:

$$\begin{aligned} N(L) &= \frac{N_0 \ \sigma^{prod} \ L}{N_{sample}} \\ &= N_0 \ k \end{aligned} \tag{9.11}$$

and the statistical and systematic uncertainties are scaled by:

$$\sigma_N^{stat}(L) = \sigma_{N_0}^{stat} \ \sqrt{k} \tag{9.12}$$

and

$$\sigma_N^{syst}(L) = \sigma_{N_0}^{syst} \ k, \tag{9.13}$$

where σ^{prod} is the production cross section of the data set, N_{sample} is the number of generated events in the data sample and L is the integrated luminosity to which the number of events is scaled. A systematic uncertainty, in addition to the errors shown in Table 9.1, is the uncertainty of the b-quark production cross section. This uncertainty is not taken into consideration in the following calculations, since it cancels when using the $N_{B_s^0}/N_{B^+}$ ratio in (9.1). This uncertainty would be important for an independent measurement (i.e., without employing a reference channel). The values for the b-quark fragmentation ratio f_u/f_s and the branching ratio of the reference channel are taken from literature [Yao06], [Nak10]. All quantities are summarised in Table 9.1.

The algorithms based on the Bayesian or the Feldman Cousins approaches to compute upper limits are then executed with the above mentioned input quantities. The computed upper limits on the expected number of events and on the branching ratio are given in Table 9.2. All expected numbers of events are scaled to an integrated luminosity of 1 fb^{-1}. The limits are computed for confidence levels of 90% and 95%. The limit increases with a higher confidence level, due to a larger size of the confidence interval. The Bayesian approach returns an upper limit on a one-sided confidence interval, while the Feldman Cousins approach returns both an upper and a lower limit of a two-sided interval. The Bayesian method always produces larger confidence intervals than the Feldman Cousins approach. This might be caused by an uninformative choice of the priors used in the computation, which results in a more conservative result.

9.3 Cut Optimisation

Since the values of the cuts imposed on the selection variables have been tuned in a previous analysis with data sets simulated at a different center of mass energy,

$\alpha_{B^+}/\alpha_{B_s^0}$	0.35 ± 0.04	
$\varepsilon_{B^+}/\varepsilon_{B_s^0}$	0.73 ± 0.12	
N_{B^+} /events	5350 ± 150	
f_u/f_s	3.86 ± 0.59	[Yao06]
$BR(B^+ \to J/\psi K^+)$	$(1.014 \pm 0.034) \times 10^{-3}$	[Nak10]
$BR(J/\psi \to \mu^+\mu^-)$	$(5.93 \pm 0.06) \times 10^{-2}$	[Nak10]
ses	$(1.0 \pm 0.2) \times 10^{-8}$	
b /events	1.28 ± 0.19	

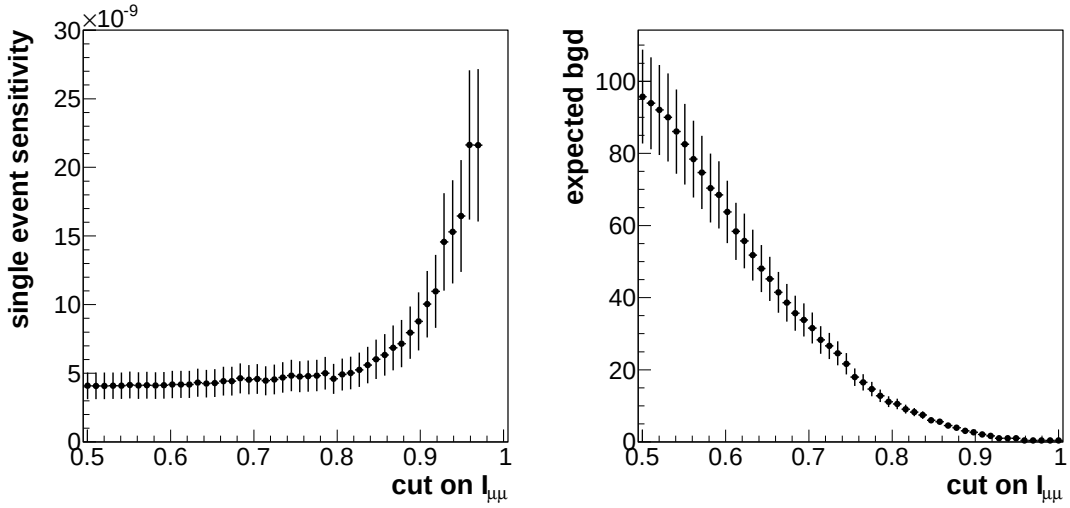
Table 9.1: Input variables used in the upper limit calculation. The number of $B^+ \to J/\psi K^+$ decays N_{B^+} and the number of expected background events b are scaled to an integrated luminosity of 1 fb^{-1}. The corresponding statistical uncertainties are scaled as well.

	Bayesian approach	Feldman Cousins approach
$N_{B_s^0}^{UL@90\%}$	< 4.3	$\in [0.03, 3.7]$
$BR(B_s^0 \to \mu^+\mu^-)^{UL@90\%}$	$< 4.4 \times 10^{-8}$	$\in [2.8 \times 10^{-10}, 3.8 \times 10^{-8}]$
$N_{B_s^0}^{UL@95\%}$	< 5.6	$\in [0.01, 4.7]$
$BR(B_s^0 \to \mu^+\mu^-)^{UL@95\%}$	$< 5.7 \times 10^{-8}$	$\in [1.1 \times 10^{-10}, 4.8 \times 10^{-8}]$

Table 9.2: Expected upper limit on the number of signal events and on the $B_s^0 \to \mu^+\mu^-$ branching ratio for an integrated luminosity of 1 fb^{-1} using the Bayesian and the Feldman Cousins approaches.

an optimisation of the cuts is investigated. The procedure aims to find the optimal set of cuts, which results in the smallest estimated upper limit on the branching ratio. Each cut is varied independently, while keeping the other cuts constant. For each cut value, the analysis chain is executed and the input variables shown in Table 9.1 are re-evaluated. Two opposite effects are observed. For example, if the cut is tightened, the selection efficiency for the signal channel decreases. This causes the single event sensitivity to increase, which results in a higher limit. On the other hand, a tighter cut also decreases the expected number of background events, which results in a lower upper limit. As an example, these anti-correlated effects are shown in Fig. 9.2. A higher value of the cut on the isolation variable (i.e., a tighter cut) produces an increase of the single event sensitivity and a decrease of the expected number of background events.

In order to use independent data samples for the cut optimisation and for the computation of upper limit estimates, all data sets are divided into two roughly equal subsamples. The first, formed by selecting events with odd event number, is used only for tuning the selection cuts. The second, formed by selecting only



(a) **(b)**

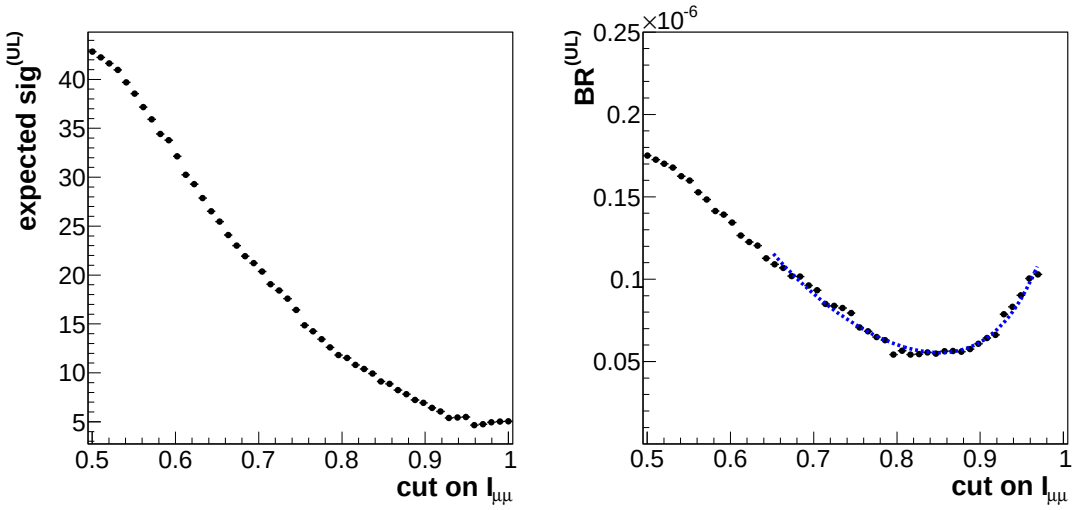
Figure 9.2: Single event sensitivity (a) and expected number of background events (b) as a function of the cut imposed on $I_{\mu\mu}$.

events with even event number, is later used to compute the upper limit with the optimised cut values. Initially, the standard cuts are used:

- $L_{xy} > 0.5\,\text{mm}$,
- $\alpha < 0.017$,
- $I_{\mu\mu} > 0.9$.

The optimisation is performed in several steps such, that only one quantity is tuned in each step. Once the optimum cut value on a quantity is found, it is preserved in any following iterations. This minimises the effects of the correlations between these variables. The expected number of events are extrapolated to an integrated luminosity of $1\,\text{fb}^{-1}$. The upper limits are computed using the Bayesian approach, at a confidence level of 95%.

In a first iteration, only the isolation variable is varied. The upper limits on the number of expected signal events and on the branching ratio are shown in Fig 9.3 as a function of the cut imposed on $I_{\mu\mu}$. A function constructed as the sum of two exponential functions is fitted to the distribution of the upper limit on the branching ratio. The range considered in the fit is between 0.7 and 0.97. In the region below 0.7, the shape of the distribution deviates from that of the fit function. Including this region in the fit would result in a worse approximation in the region of interest (i.e., of the minimum). In the region above 0.97, the cut is too tight and very few events remain after the selection. This causes large fluctuations in the fit on the $\mu\mu K$ invariant mass, which affect the computation of the upper limit. The optimum cut on the isolation variable is then determined as the minimum of the fit function and is computed to be 0.85.



(a) (b)

Figure 9.3: Computed upper limit on the expected number of signal events (a) and on the branching ratio (b) as a function of the cut imposed on $I_{\mu\mu}$. The fit function is shown with a dashed line.

In a second step, the cut on the isolation variable is fixed to the value found previously (i.e., $I_{\mu\mu} > 0.85$) and the procedure is repeated by varying the cut on the pointing angle. Plots showing the upper limit values on the number of expected signal events and on the branching ratio as a function on the cut applied on α are shown in Fig. 9.4 (a), (b). A function formed by the sum of two exponential functions is fitted to the distribution of the upper limit on the branching ratio. The fit range is chosen as 0.006 to 0.075. Below this range, the cut imposed on α is too tight and the branching ratio limit increases, due to the large statistical uncertainty. Above this range the fit function does not accurately describe the distribution. The optimum cut value is determined as the minimum of the fit function and is estimated to be 0.0165.

In a last iteration, the cut values of $I_{\mu\mu}$ and α are fixed to the previously optimised values (i.e., $I_{\mu\mu} > 0.85$ and $\alpha < 0.0165$) and only the cut on the transverse decay length L_{xy} is varied. The upper limit values on the expected number of signal events and on the branching ratio are shown in Fig. 9.4 (c), (d). The flat region observed in the distribution of the upper limit on the expected number of signal events in the region above 1.3 is an effect of the upper limit computation, when $b \simeq 0$. A function formed by the sum of two exponential functions is used to fit the distribution of the upper limit on the branching ratio. The minimum of the fit function is then assumed to the optimum cut on L_{xy}. This is computed to be 1.28 mm.

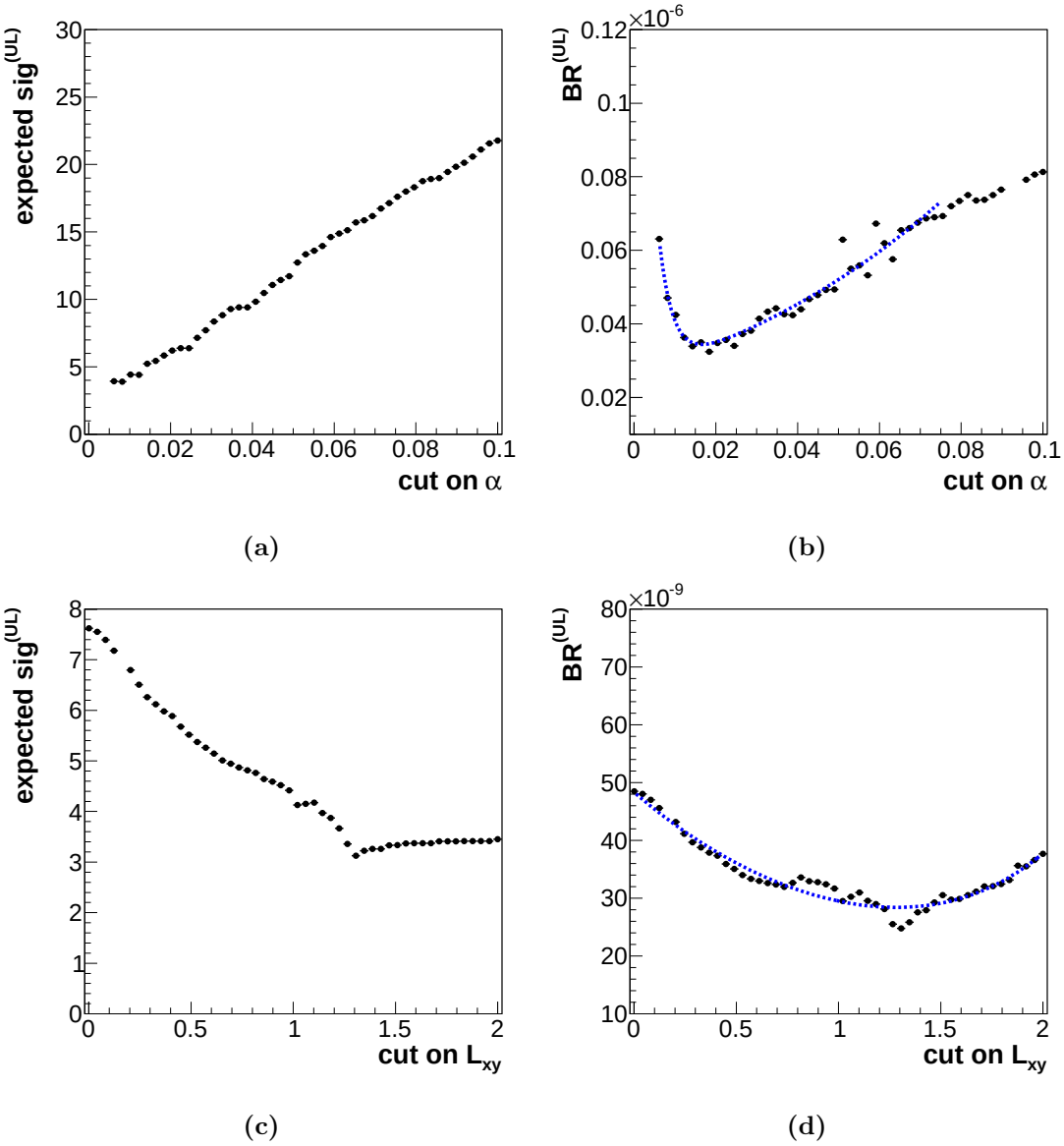


Figure 9.4: Computed upper limit on the expected number of signal events (a),(c) and on the branching ratio (b),(d) as a function of the cut imposed on α (a), (b) or L_{xy} (c), (d). The fit functions are shown with dashed lines.

Therefore, the optimised values of the cuts on the selection variables are:

- $L_{xy} > 1.28\,\text{mm}$,
- $\alpha < 0.0165$,
- $I_{\mu\mu} > 0.85$.

The invariant mass window used in the event selection is determined based on the mass resolution and is not subject to this optimisation procedure.

The optimised cuts are then used on the remaining subsamples, formed by selecting only events with even event numbers. The effect observed on the expected number of signal and background events and on the single event sensitivity are shown in Table 9.3. All values are extrapolated to an integrated luminosity of 1 fb^{-1}. The cut optimisation improves the signal to background ratio by about a factor 3 and the single event sensitivity is also slightly improved.

	Standard cuts	Optimised cuts
s /events	0.282 ± 0.013	0.332 ± 0.014
b /events	2.0 ± 0.3	0.7 ± 0.1
ses	$(1.1 \pm 0.3) \times 10^{-8}$	$(9 \pm 2) \times 10^{-9}$

Table 9.3: Expected number of signal and background events and single event sensitivity obtained with the standard and with the optimised set of cuts.

The upper limits on the branching ratio are then estimated using the quantities in Table 9.3 (i.e., b and ses). These are given in Table 9.4 using the Bayesian and the Feldman Cousins approaches, at a 95% confidence level. In both cases the estimated upper limit improves by about a factor 0.6, when the optimised set of cuts is employed.

	Approach	Standard cuts	Optimised cuts
$BR^{UL@95\%}$	Bayesian	$< 7.0 \times 10^{-8}$	$< 4.4 \times 10^{-8}$
	Feldman Cousins	$\in [2 \times 10^{-10}, 5.9 \times 10^{-8}]$	$\in [6 \times 10^{-11}, 3.8 \times 10^{-8}]$

Table 9.4: Expected upper limits on the branching ratio obtained with the standard and with the optimised set of cuts. Computations are done using the Bayesian and the Feldman Cousins approaches at a 95% confidence level.

Another possible optimisation technique [Pun03] is to maximise the quantity:

$$S = \frac{\varepsilon(t)}{a/2 + \sqrt{b(t)}}, \tag{9.14}$$

where $\varepsilon(t)$ is the signal selection efficiency for a given cut t, $b(t)$ is the corresponding number of expected background events and a is the number of sigmas corresponding the confidence interval considered (i.e., $a \simeq 2$ for a 95% confidence level). The cut optimisation has also been performed using this technique, resulting in a similar set of optimised cuts.

9.4 Sensitivity to the $B_s^0 \to \mu^+\mu^-$ Branching Ratio

The final step of the analysis is to extrapolate the calculations to higher values of the integrated luminosity. The expected number of background events and

the observed number of B^+ events are obtained from the subsample resulting after selecting only even events from the `bb_mu6mu4X` data set (about half the statistics). These are scaled to a certain integrated luminosity using (9.11). The corresponding uncertainties are scaled according to (9.12) and (9.13). The expected number of signal and background events and the single event sensitivity are shown in Fig. 9.5.

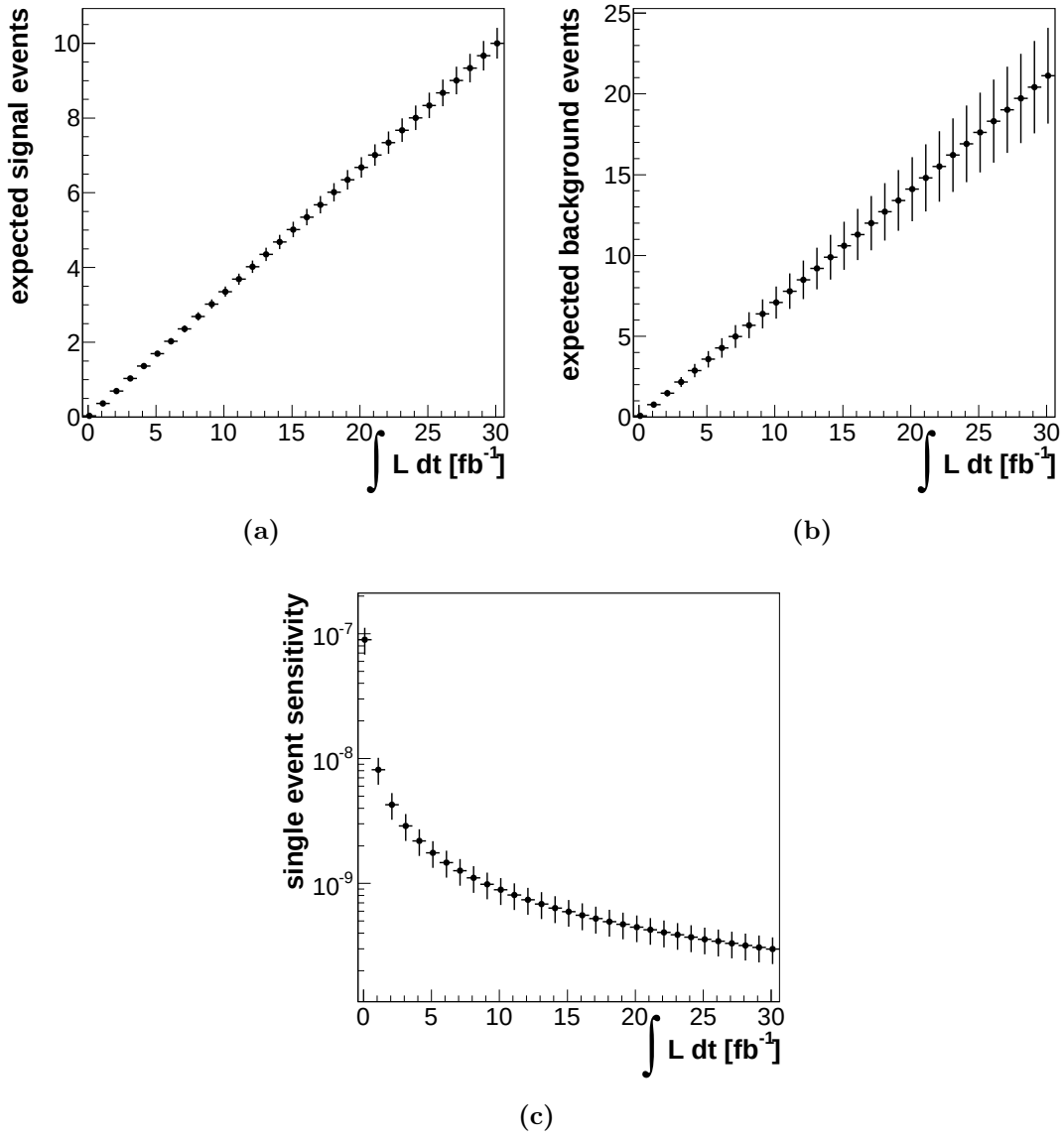


Figure 9.5: Expected number of signal (a) and background (b) events and single event sensitivity (c) as a function of integrated luminosity.

Estimates of the upper limit on the $B_s^0 \to \mu^+\mu^-$ branching ratio are shown in Fig. 9.6 for a 90% and a 95% confidence level. Both the Bayesian and the Feldman Cousins methods are considered. The uncertainties of the input parameters are scaled with the luminosity as well. The extrapolation performed with the Bayesian method is performed up to an integrated luminosity of about $30\,\mathrm{fb}^{-1}$. For the Feld-

man Cousins approach, only the upper limits of the two-sided confidence intervals are shown. The scaling to a large integrated luminosity requires a very large size of the probability map needed in the computations. This may produce an improper convergence of the algorithm. Therefore, the extrapolation is only performed up to about 24 fb^{-1}. As already mentioned, the Feldman Cousins algorithm has the advantage of producing less conservative results than the Bayesian algorithm, but it is less stable at high values of the expected number of background events. All computed upper limit values are given in Appendix D.

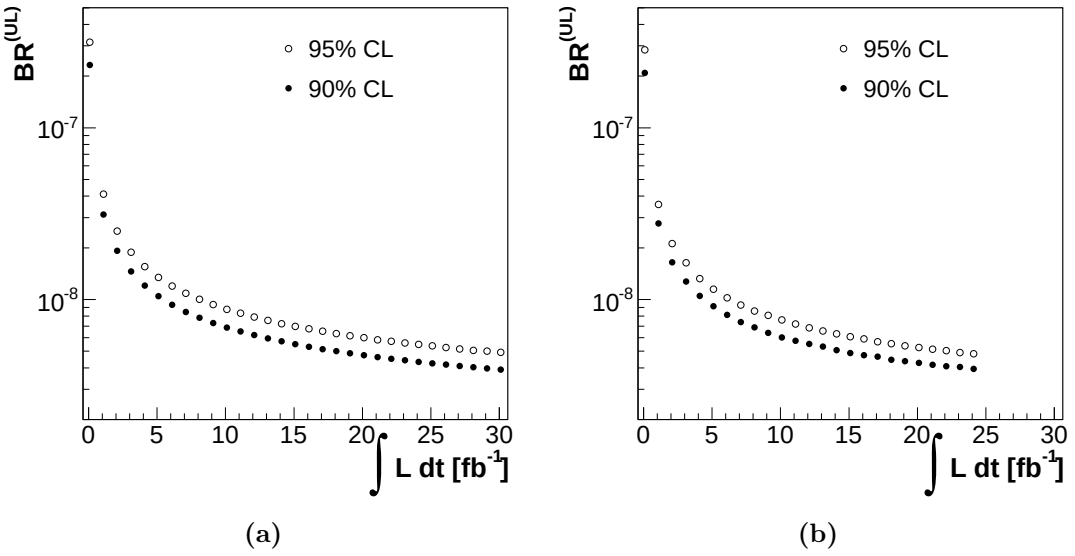


Figure 9.6: Expected upper limit on the branching ratio as a function of integrated luminosity. Upper limits are computed using a 95% (empty circles) or a 90% (filled circles) confidence level. Calculations are performed with the Bayesian (a) or the Feldman Cousins (b) methods.

The input quantities (b and ses) used in the computation of each upper limit shown in Fig. 9.6 are then varied by their errors. A best (worst) case upper limit on the branching ratio is computed for $b-\sigma_b$ and $ses-\sigma_{ses}$ ($b+\sigma_b$ and $ses+\sigma_{ses}$). The results of these calculations represent the edges of the bands displayed in Fig. 9.7. Calculations are performed using a 90% or 95% confidence level and using the Bayesian and the Feldman Cousins methods.

The expected upper limit on the branching ratio is compared with the current measurements. CDF and DØ, the two experiments located at the Tevatron accelerator, report an upper limit of 4×10^{-8} with 7 fb^{-1} recorded data [Aal11] and 5.1×10^{-8} with 6.1 fb^{-1} recorded data [Aba10]. Both computations are performed at a 95% confidence level. The sensitivity curve in Fig. 9.6 (with the Feldman Cousins method) shows that ATLAS will reach the same sensitivity with an integrated luminosity about 1 fb^{-1}. This represents roughly the total recorded data from the first collisions in 2010 up to the end of June 2011. This would be a remarkable achievement, taking into account that the same sensitivity will be

reached in a much shorter time scale. At the current rate of data taking, ATLAS will likely outperform CDF and DØ with the data recorded by the end of 2011.

LHCb and CMS, two of the other major experiments located at the LHC, are also measuring the $B_s^0 \to \mu^+\mu^-$ branching ratio. The best result reported by LHCb is an upper limit of 1.3×10^{-8} obtained with 300 pb^{-1} of recorded data [Ser11] at a 95% confidence level. The previous published result is an upper limit of $5.6{\times}10^{-8}$ obtained with 37 pb^{-1} [Aai11b]. In comparison, ATLAS will need a larger amount of data to reach the same sensitivity level of the best LHCb result (about $3-5$ fb^{-1}). The reason is that LHCb is designed for the study of B-hadrons, which are predominantly produced in the forward region. Therefore, LHCb benefits from increased statistics in the first LHC running period. However, with a planned increase of the LHC instantaneous luminosity by a factor 10, ATLAS might become competitive in the long turn (several years), due to the limitations of LHCb in dealing with a very high instantaneous luminosity (LHCb has already reached its peak instantaneous luminosity). CMS also reports an upper limit, which is computed to 1.9×10^{-8} with 1.14 fb^{-1} recorded data [Cha11] at a 95% confidence level. The upper limit estimates given in Fig. 9.6 show that ATLAS will reach the same sensitivity with about $2-3$ fb^{-1}. This is much closer than compared to LHCb. The slight advantage of CMS may be partly explained by the better mass resolution of reconstructed B_s^0 candidates.

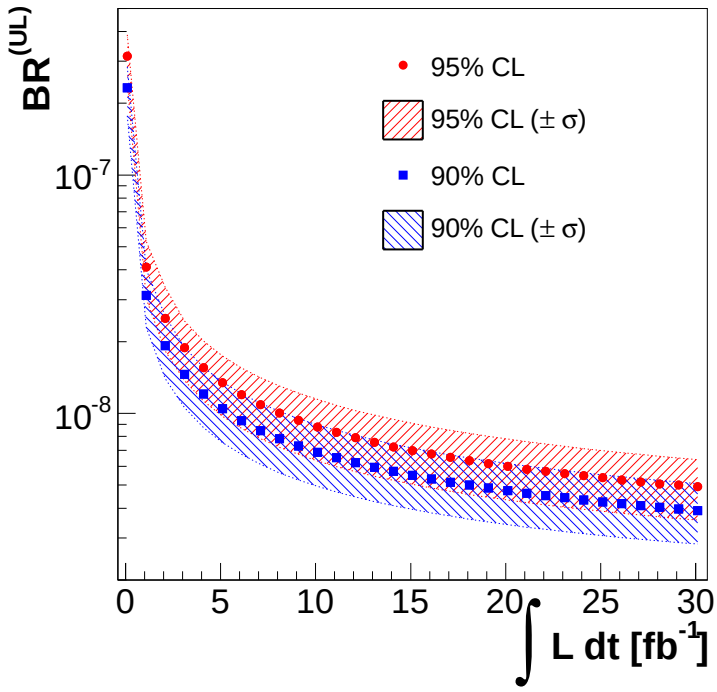
A possible improvement of the measurement based on recorded data might arise from lower uncertainties. The effect of the total computed errors of the expected number of background events and of the single event sensitivity is an increase of the upper limit by about a factor $1.5-2$. This effect has already been included in the calculations. However, several sources of systematic uncertainties are specific to the MC samples used in this analysis and are likely to be diminished in a future analysis. For example, the systematic uncertainty of the ratio of acceptances (about 12%) is an artifact of the selection cuts at generator level and will be minimised with future simulations. Another such systematic uncertainty is the error assumed to the cut factorisation due to the correlations between variables (about 13%). The analysis on recorded data will benefit from from higher statistics and the expected number of background events will be determined from extrapolating the sidebands of the invariant mass spectrum to the B_s^0 signal region. This eliminates the need to apply the cut factorisation. Another systematic uncertainty, which might be improved, is that of the f_u/f_s ratio (about 15%). For example, the LHCb measurement of the f_s/f_d ratio [Aai11a] has currently a comparable uncertainty, which will be reduced to a few percent with more statistics. Using a multi-variate analysis for separating signal from background events might also be more efficient than a cut chain analysis, which might result in a better sensitivity.

9.5 Summary

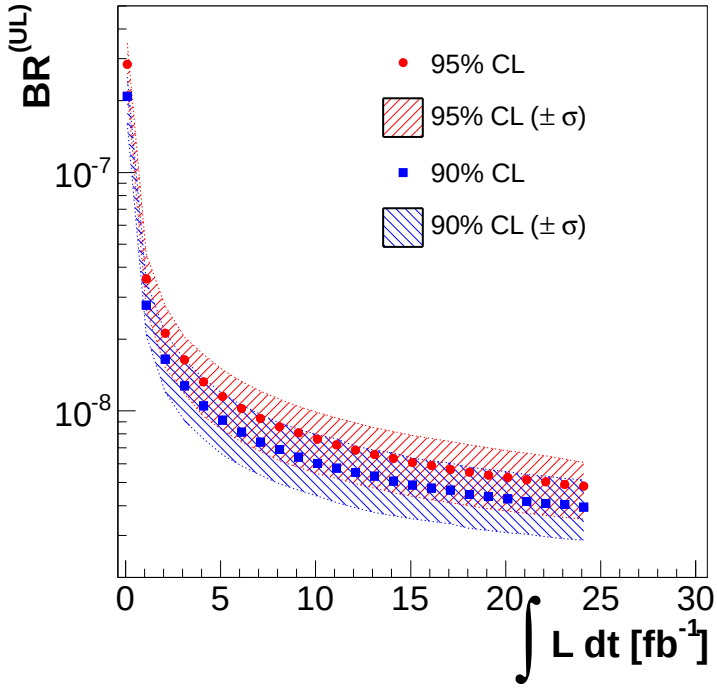
This chapter presents estimates of the upper limit on the $B_s^0 \to \mu^+\mu^-$ branching ratio. An optimisation of the cuts applied on the selection variables is performed with the aim of finding the smallest expected upper limit. The optimised cuts are

then used with independent data sets, with an improvement of the upper limit by about a factor 0.6.

In a next step, the expected upper limit is computed as a function of integrated luminosity and the expected ATLAS sensitivity to the $B_s^0 \to \mu^+\mu^-$ branching ratio is compared with the current experimental measurements. ATLAS needs about 1 fb^{-1} recorded data (already collected) to reach the sensitivity of the CDF and DØ experiments and will likely outperform them with the data collected in the near future. The LHCb experiment currently provides the best measurement. ATLAS will reach the same sensitivity with about $3 - 5$ fb^{-1}. It will become competitive in the long run, once the instantaneous luminosity reaches the design value. The CMS experiment is slightly more sensitive than the ATLAS predictions, but the future analysis on recorded data should add some improvements (e.g., lower systematic uncertainties or a multi-variate analysis), which will improve the ATLAS sensitivity.



(a)



(b)

Figure 9.7: Expected upper limit on the branching ratio as a function of integrated luminosity. Upper limits are computed using a 95% (red circles) or a 90% (blue squares) confidence level. The bands show the interval in which the upper limit can vary, due to the uncertainties of the input parameters. Calculations are performed with the Bayesian (a) or the Feldman Cousins (b) methods.

Chapter 10

Summary

In this thesis, prospects for measuring the branching ratio of the $B_s^0 \to \mu^+\mu^-$ decay with the ATLAS experiment are investigated. The study is performed with fully reconstructed Monte Carlo (MC) data. The centre-of-mass energy of simulated collisions is $\sqrt{s} = 10$ TeV, higher than that of current collisions at the LHC. At the time, this analysis was started, LHC operation at $\sqrt{s} = 10$ TeV was planned.

The $B_s^0 \to \mu^+\mu^-$ decay is highly suppressed in the Standard Model (SM) of particle physics, with an estimated branching ratio of $(3.2 \pm 0.2) \times 10^{-9}$. This decay is forbidden at the tree level by the GIM mechanism. In the lowest approximation, it is only possible through one-loop penguin or box diagrams. The branching ratio is sensitive to New Physics effects. Therefore, $B_s^0 \to \mu^+\mu^-$ is an excellent candidate to probe phenomena beyond the Standard Model. For example, in the Minimal Supersymmetric Standard Model, $b \to s$ transitions are enhanced by the coupling of quarks to the scalar and pseudoscalar Higgs field components. The $B_s^0 \to \mu^+\mu^-$ branching ratio has, therefore, a strong dependence on $\tan\beta$ and m_{A^0}, which may produce a significant deviation (either an increase or decrease) of the branching ratio value from the value expected in the SM.

The strategy employed in this thesis is to compute the branching ratio of the $B_s^0 \to \mu^+\mu^-$ decay (the signal channel) normalised to the branching ratio of the $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decay (the reference channel). The latter decay has a similar event topology (two muons among the final state particles). It is, therefore, expected that its reconstruction chain is similar to that of the $B_s^0 \to \mu^+\mu^-$ channel. The major differences are given by an additional particle among the decay products (the kaon) and an additional decay vertex. The branching ratio of the reference channel is expected to be measured with high precision, due to the large amount of decay candidates. The formula for computing the $B_s^0 \to \mu^+\mu^-$ branching ratio (4.1) involves estimating the number of observed candidates corresponding to the $B_s^0 \to \mu^+\mu^-$and $B^+ \to J/\psi K^+$ decays and applying corrections to take the different values of the acceptances and total reconstruction efficiencies into account. The major background channel considered in this thesis is the $b\bar{b} \to \mu^+\mu^- X$ combinatorial background. Several exclusive decay channels have been previously analysed and their contributions in the B_s^0 mass region are negligible.

Several Monte Carlo (MC) data sets are used for the analysis presented in this thesis. These correspond to the $B_s^0 \to \mu^+\mu^-$ decay (50,000 events), to the

$B^+ \to J/\psi K^+$ decay (150,000 events) and to the $b\bar{b} \to \mu^+\mu^- X$ background channel (1 million events). All data samples are generated, simulated and reconstructed centrally and are part of the official ATLAS MC production. All data sets have generator-level cuts on the muon p_T of 6 (4) GeV for the hardest (second hardest) muon. In addition, two more samples corresponding to the signal and the combinatorial background channels are used. These have lower threshold cuts on the muon p_T at generator level (i.e., 2.5 GeV for both muons) and are employed to study trigger effects around the threshold.

The kinematic and geometric acceptances of the signal and reference channels are determined using several generated samples, such that all generated particles are roughly contained within the geometrical acceptance of the ATLAS detector (i.e., a pseudorapidity range of $|\eta| < 2.5$). The acceptance is then determined from the values of the cross-sections (computed before and after applying the kinematic and geometric cuts) returned by the event generator program. The ratio of the acceptances is computed to $\alpha_{B^+}/\alpha_{B_s^0} = 0.349 \pm 0.002\,(stat) \pm 0.042\,(syst)$. The considerable systematic uncertainty accounts for slightly different cuts applied at generator level.

The trigger used in this analysis is based on a di-muon trigger with threshold cuts of 6 and 4 GeV on the muon p_T. The performance of the trigger system is determined using a software tool, which simulates the trigger operation. The ratio of trigger efficiencies observed for the $B^+ \to J/\psi K^+$and the $B_s^0 \to \mu^+\mu^-$ channels is determined to $\varepsilon_{B^+}^{trigger}/\varepsilon_{B_s^0}^{trigger} = 1.042 \pm 0.007\,(stat) \pm 0.156\,(syst)$.

In the offline analysis, the number of reconstructed candidates for the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ is extracted from the same data sample (i.e., from the combinatorial background sample). Since true $B_s^0 \to \mu^+\mu^-$ decays cannot be distinguished from the larger amount of background events, several criteria are used to suppress this background contribution. The total reconstruction and selection efficiencies are determined from dedicated samples corresponding to the signal and reference channels.

Two offline software algorithms are developed for the reconstruction of candidates for the $B_s^0 \to \mu^+\mu^-$ and the $B^+ \to J/\psi K^+$ decays. In each case, a vertex finder algorithm is employed to find pairs of muon particle tracks originating in the same decay vertex. An additional particle track (the kaon) is used to reconstruct the 3-prong vertex corresponding to the B^+ decay. All candidates are selected using a set of baseline quality cuts, including a $\chi^2/ndf < 10$ cut for reconstructed vertices. The observed purity of reconstructed $B_s^0 \to \mu^+\mu^-$ ($B^+ \to J/\psi K^+$) candidates passing the baseline selection is about 98% (90%). The invariant mass resolution of reconstructed $\mu\mu$ ($\mu\mu K$) candidates is estimated to 108 ± 3 MeV (43.3 ± 0.1 MeV).

The number of reconstructed background events misidentified as $B_s^0 \to \mu^+\mu^-$ decays is about three orders of magnitude larger than the number of true signal events. In order to reduce this contribution, several discriminating variables are used: the transverse decay length, the pointing angle, the muon isolation and the $\mu\mu$ invariant mass. The distributions of these variables differ significantly in the $B_s^0 \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ samples. Therefore, cuts applied

on these quantities can be tuned such that a high background rejection is obtained with a relatively high signal retention rate. Since the number of events in the $b\bar{b} \to \mu^+\mu^- X$ data sample is too low to apply all selection cuts simultaneously, a cut factorisation approach is employed to estimate the number of background events remaining after the selection. The overall efficiency is computed to $\varepsilon_{b\bar{b}} = (9.3 \pm 4\,(stat) \pm 1.3\,(syst)) \times 10^{-9}$. The systematic uncertainty of about 13% is an effect of the correlations between the selection variables. In comparison, the signal selection efficiency for the $B_s^0 \to \mu^+\mu^-$ sample is about 2.6%.

The number of $B^+ \to J/\psi K^+$ candidates is extracted from the $b\bar{b} \to \mu^+\mu^- X$ data sample. All events are selected with criteria similar to those applied for $B_s^0 \to \mu^+\mu^-$ candidates. The number of candidates is estimated from the $\mu\mu K$ invariant mass distribution by statistically subtracting sideband contributions in the signal region.

The final step in the analysis is to estimate the ATLAS sensitivity to the $B_s^0 \to \mu^+\mu^-$ branching ratio. Upper limits on the branching ratio are computed using two methods, a Bayesian method and a frequentist method based on the Feldman Cousins approach. Both methods take the uncertainties of the input parameters into account (i.e., the uncertainties of the single event sensitivity and of the expected number of background events). The Feldman Cousins method always gives less conservative results. The selection cuts are tuned by finding the optimum (i.e., smallest) value of the branching ratio. The optimisation is performed with independent data sets. All data samples are split into two subsets of roughly equal size. The first is used for the optimisation and the second is used for computing the upper limit on the branching ratio. The input parameters needed in the calculation of upper limits are summarised in Table 9.1 and Table 9.3. Finally, upper limits on the branching ratio are calculated using the optimised selection cuts. The expected number of events and the single event sensitivity are extrapolated to different values of integrated luminosity. The ATLAS sensitivity to the $B_s^0 \to \mu^+\mu^-$ branching ratio is evaluated for different integrated luminosities. As an example, the upper limit on the branching ratio is estimated to $BR(B_s^0 \to \mu^+\mu^-) < 3.8 \times 10^{-8}$ at a 95% confidence level, for an amount of data equivalent to 1 fb^{-1} (with the Feldman Cousins method). The sensitivity as a function of integrated luminosity is given in Fig. 9.7. It is expected that the sensitivity will reach a level comparable to the best current experimental measurements with about $2 - 5$ fb^{-1} of recorded data.

Appendix A

Kinematic Truth Variables of Generated Data Samples

This chapter presents in detail some of the kinematic truth variables of the generated data samples. These samples have been introduced in Chapter 5.

A.1 Correlations between $p_T(\bar{b})$ and $\hat{p}_\perp$

This section presents the correlations between the transverse momentum of the products of the hard scattering process, $\hat{p}_\perp$, and the transverse momentum of the $\bar{b}$-quark, $p_T(\bar{b})$, as described in Chapter 5. The datasets analysed are `Bs_mu6mu4` and `B+_mu6mu4K+`, presented in Table 5.2, each consisting of 200 000 events. Each event contains at least one decay of the corresponding channel.

In order to study these correlations, events are split into three categories:

- events where the $\bar{b}$-quark belongs to the decay products of the hard scattering process,
- events where the $\bar{b}$-quark does not appear among the decay products of the hard scattering process, but is produced in the decay tree of particles produced in the hard scattering process,
- events which do not fall into either of the two previous categories.

The correlations between $p_T(\bar{b})$ and $\hat{p}_\perp$are shown in Fig. A.1 for the `Bs_mu6mu4` sample and in Fig. A.2 for the `B+_mu6mu4K+` sample. The plots labeled (a)-(c) in each figure show the correlation for each category as explained above. The plot labeled (d) in each figure shows the correlation between $\hat{p}_\perp$ and the maximum of $p_T(\bar{b})$, for all events. In all plots the threshold imposed on the $\hat{p}_\perp$ variable is clearly visible at 10 GeV for `Bs_mu6mu4` and at 15 GeV for `B+_mu6mu4K+`. The threshold on the minimum value of $p_T(\bar{b})$, at 10 GeV for both datasets, is also seen in the bottom plots in each figure.

In the first category of events, where the $\bar{b}$-quark originated in the hard scattering process, the correlation with $\hat{p}_\perp$is the highest, as expected. This correlation

is lowest in the third category, when the $\bar{b}$-quark has no direct or indirect connection with the hard scattering process. In the first category, several events exist below the 10 GeV threshold for $p_T(\bar{b})$. This is possible in events where more than one $b\bar{b}$ pairs exist. In the second category, the number of events below the 10 GeV threshold is larger, due to energy loss inherent in the decay tree where the $\bar{b}$-quark is produced. Finally, this number is the largest in the third case, due to the low correlation between $p_T(\bar{b})$ and $\hat{p}_\perp$.

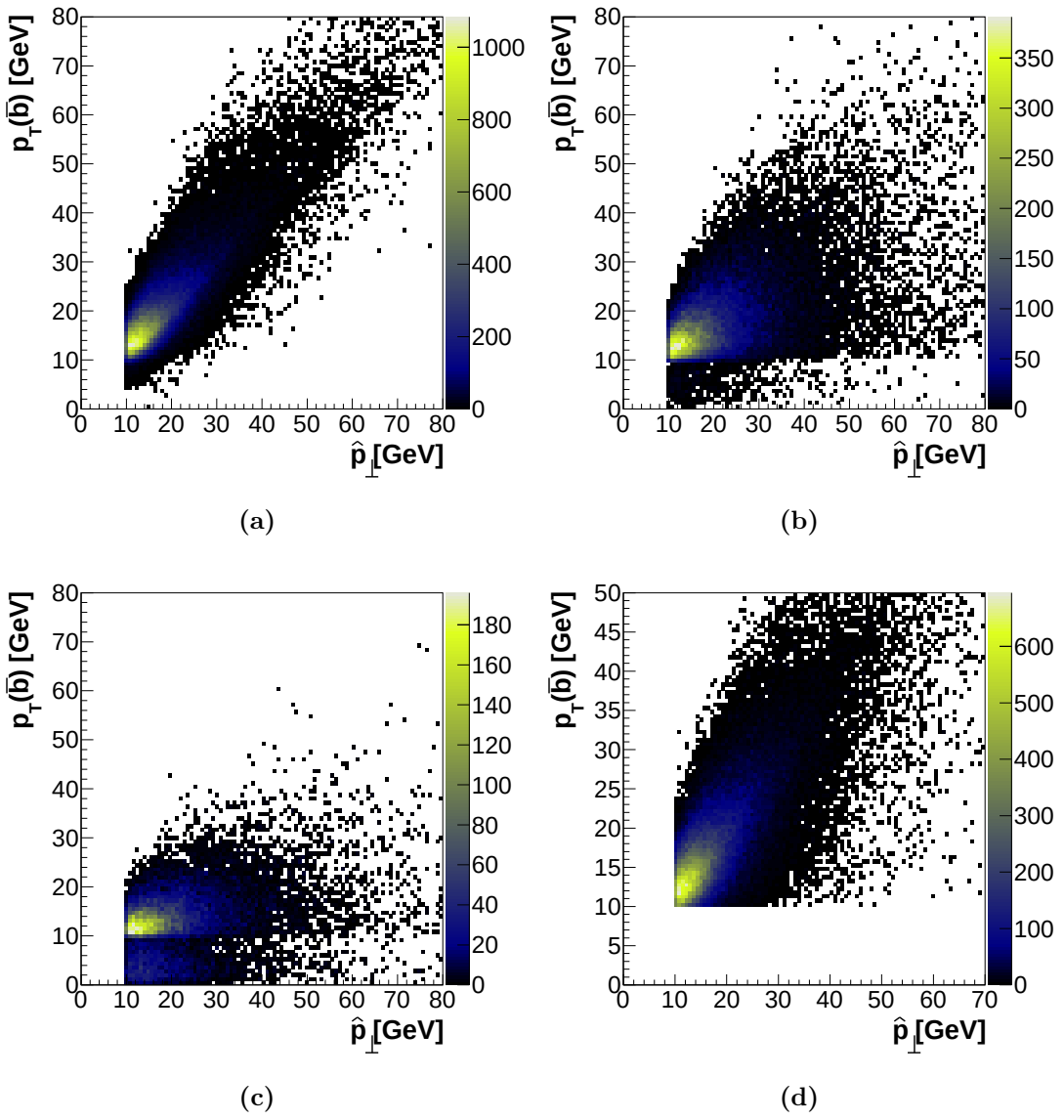


Figure A.1: Correlations between $p_T(\bar{b})$ and $\hat{p}_\perp$ for the `Bs_mu6mu4` data set. The $\bar{b}$-quark is among the products of the hard scattering process (a), in the decay tree of the hard scattering process (b) or unrelated to the hard scattering process (c). The hardest p_T $\bar{b}$-quarks are shown in (d).

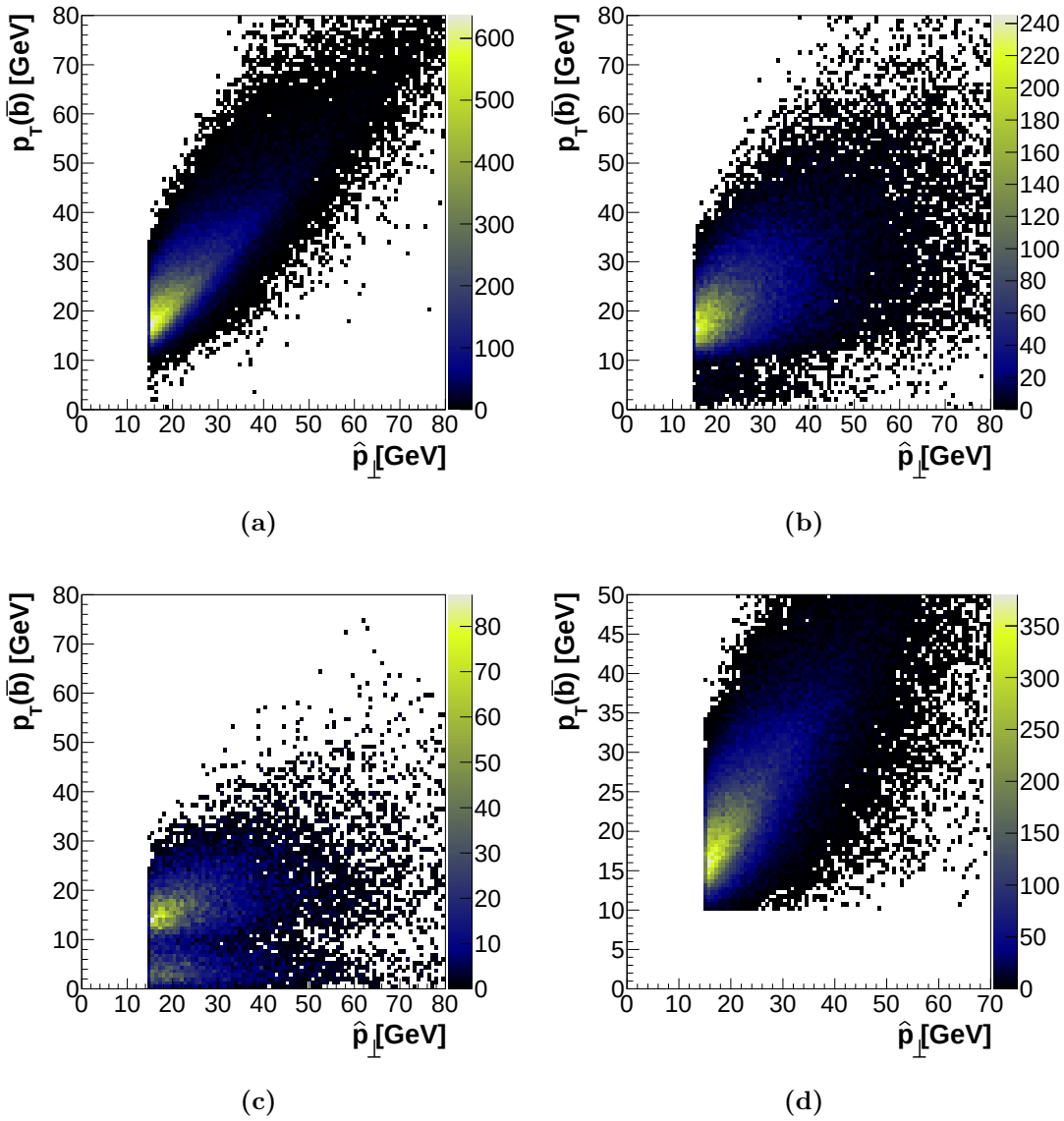


Figure A.2: Correlations between $p_T(\bar{b})$ and $\hat{p}_\perp$ for the `B+_mu6mu4K+` data set. The $\bar{b}$-quark is among the products of the hard scattering process (a), in the decay tree of the hard scattering process (b) or unrelated to the hard scattering process (c). The hardest p_T $\bar{b}$-quarks are shown in (d).

A.2 Kinematic Variables of Generated Particles

Several other kinematic variables are presented, in addition to those shown in Chapter 5. Fig. A.3 shows the η and ϕ distributions for the decay products of the B_s^0 (top plots) and B^+ (bottom plots) mesons. The p_T cuts at 4 and 6 GeV on the two muons are shown. A slight difference around $\eta = 0$ is seen in the η distributions of muons and kaons produced in the B^+ decay. This is due to the harder p_T spectrum of the muons. No differences are seen in the ϕ distributions.

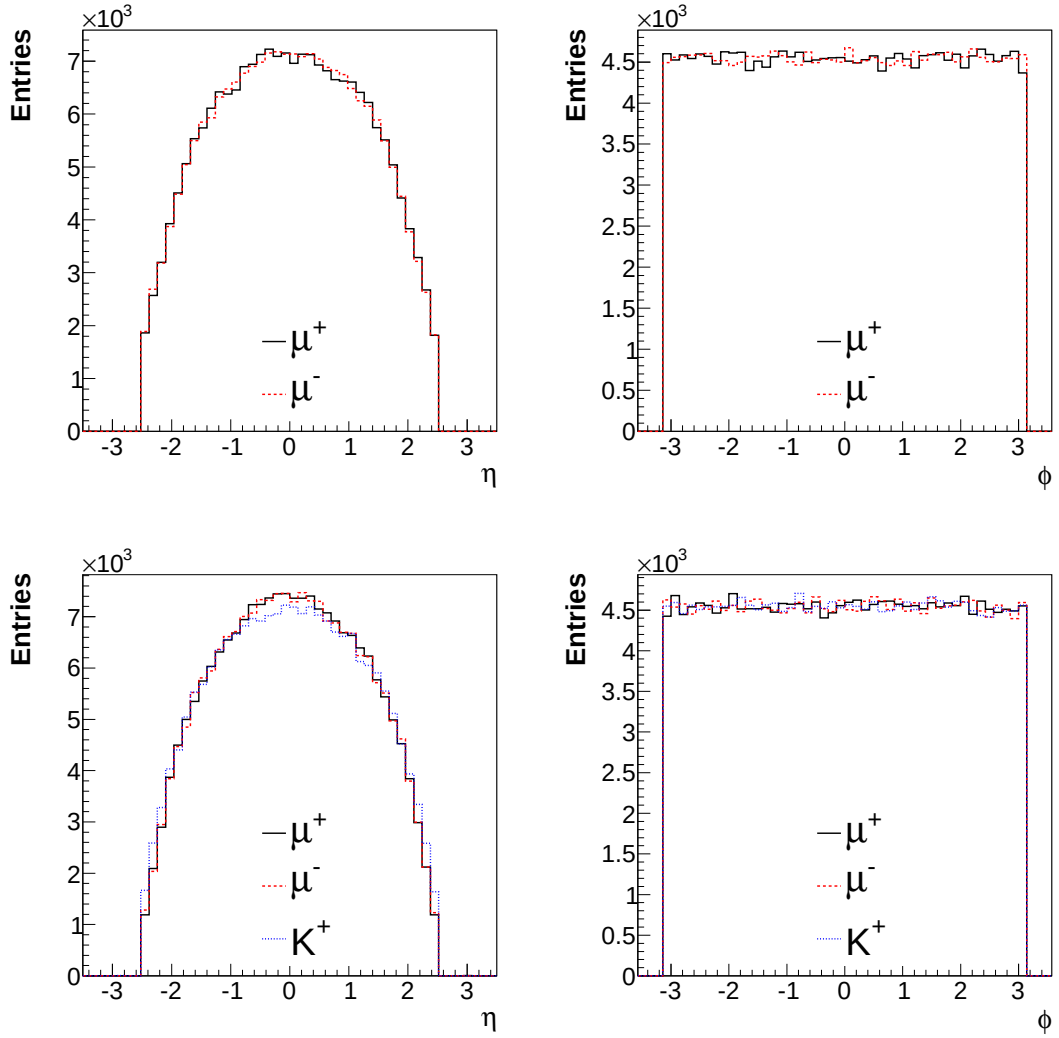
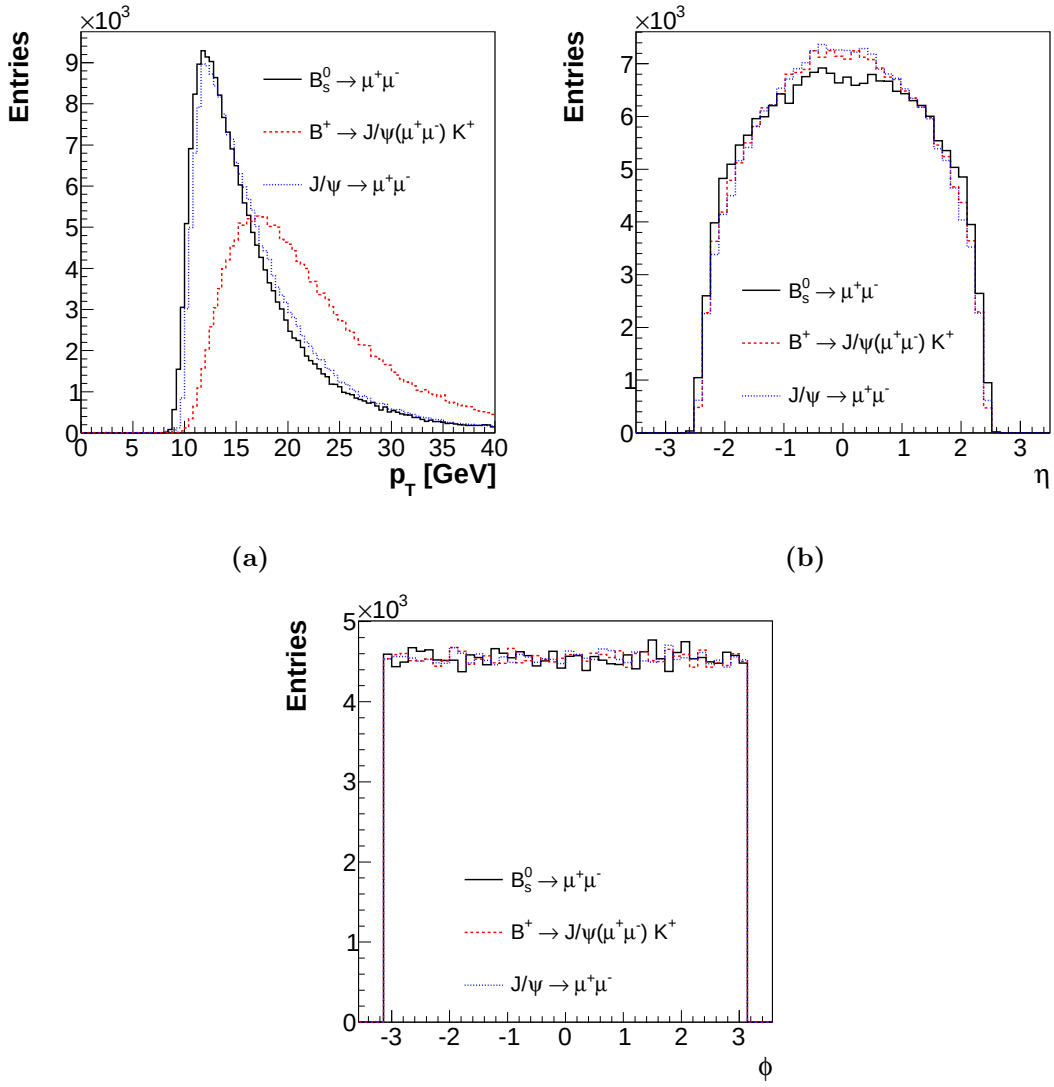


Figure A.3: Kinematic variables of generated B_s^0(top) and B^+(bottom) decay particles. Distributions of the p_T(left), η (centre) and ϕ (right) variables are shown.

Fig. A.4 shows the p_T, η and ϕ distributions of the generated B_s^0, B^+ and J/ψ particles. The p_T thresholds imposed on the muons are reflected in the steep rise of the B_s^0 and J/ψ p_T spectra around 10 GeV. The p_T distribution is shifted to higher valued due to the additional particle present in the decay.



(a) **(b)**

(c)

Figure A.4: p_T (a), η (b) and ϕ (c) distributions of generated B_s^0 (full line), B^+ (red dashed line) and J/ψ (dotted blue line) particles.

Appendix B

Kinematic Variables Used in the Acceptance Studies

This chapter presents additional details concerning the study of the acceptance ratio described in Chapter 6.

B.1 b-quark Fragmentation Ratio

The method to compute the ratio of probabilities that a b-quark hadronises into an B_s^0 or B^+ meson and the data sets used for this calculation are presented in detail in Section 6.2. The dependence of this ratio on additional kinematic variables of the B meson is presented in this section. The dependence on the B ($\bar{B}$) meson pseudorapidity is shown in Fig. B.1 (Fig. B.2). Plots on the left (right) are obtained from data samples with a $p_T(\bar{b})$ threshold of 10 (2.5) GeV. Similarly, the dependence on the B ($\bar{B}$) azimuthal angle is shown in Fig. B.3 (Fig. B.4). No systematic variations across the ranges of these quantities are observed.

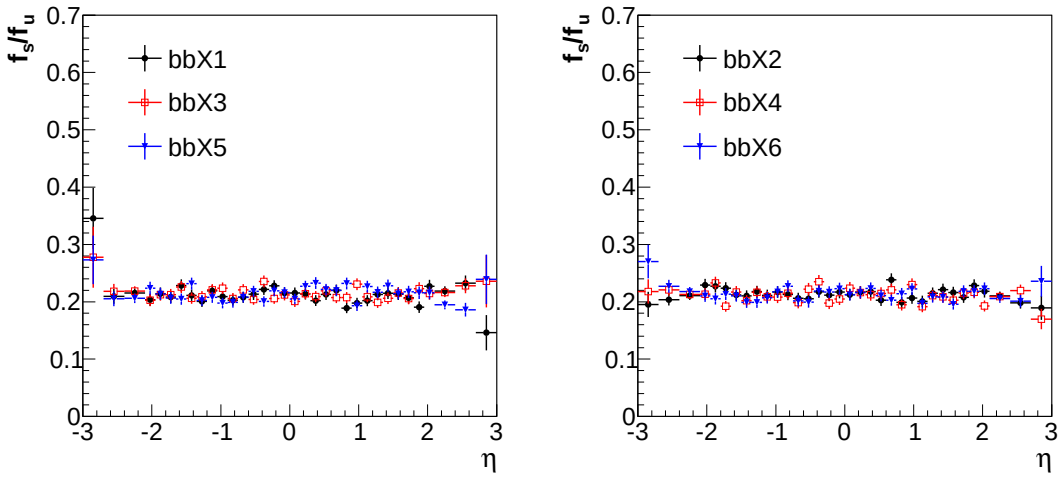


Figure B.1: Dependence of f_s^{Py}/f_u^{Py} on the B meson pseudorapidity determined on the data samples with a threshold on $p_T(\bar{b})$ of 10 (left) or 2.5 GeV (right).

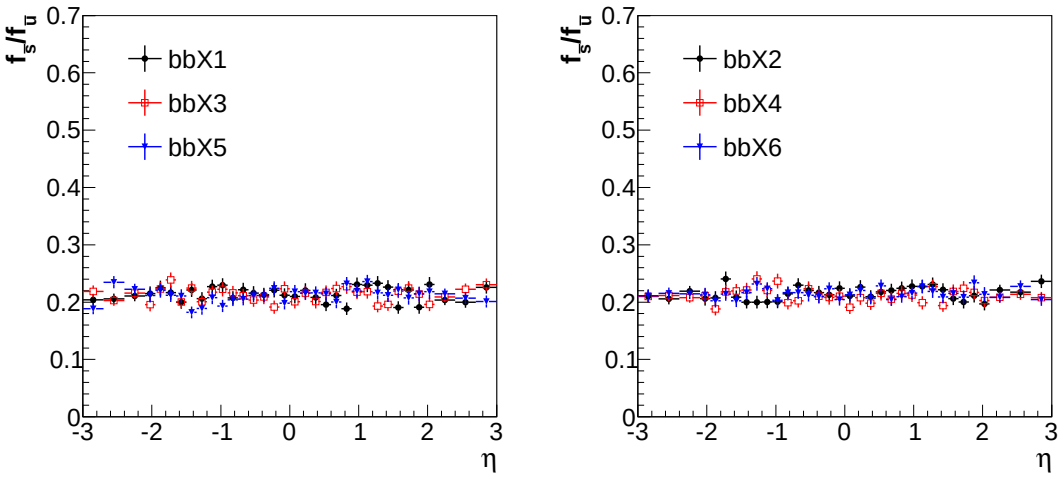


Figure B.2: Dependence of $f_{\bar{s}}^{Py}/f_{\bar{u}}^{Py}$ on the B meson pseudorapidity determined on the data samples with a threshold on $p_T(\bar{b})$ of 10 (left) or 2.5 GeV (right).

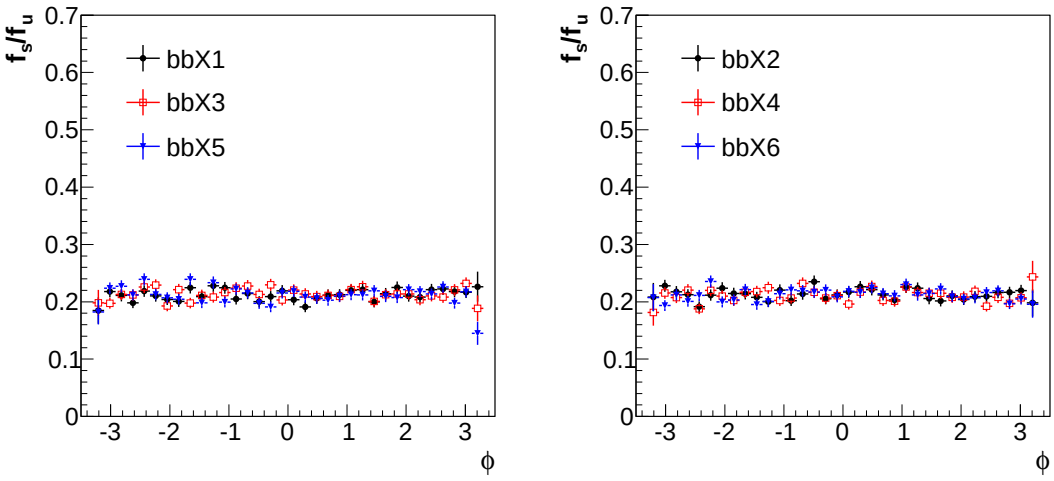


Figure B.3: Dependence of f_s^{Py}/f_u^{Py} on the B meson azimuthal angle determined on the data samples with a threshold on $p_T(\bar{b})$ of 10 (left) or 2.5 GeV (right).

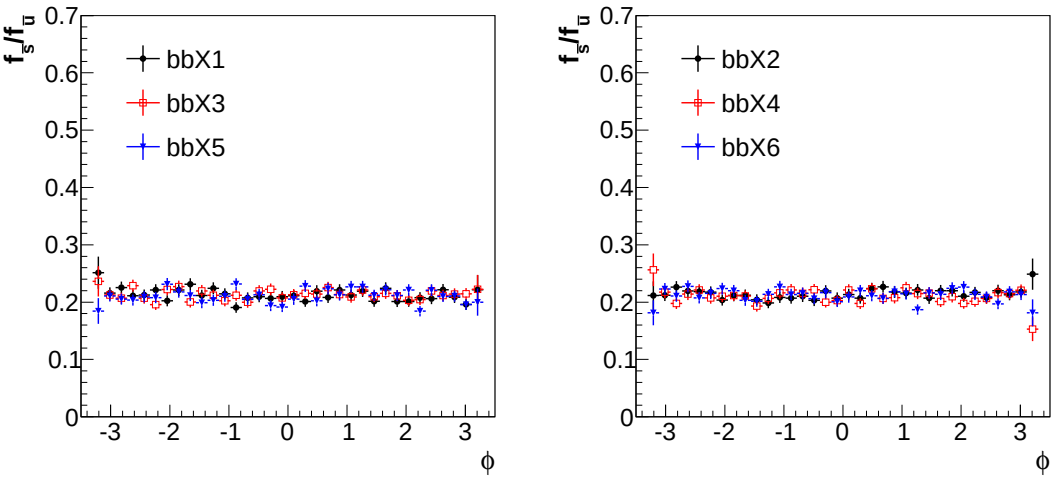


Figure B.4: Dependence of $f_{\bar{s}}^{Py}/f_{\bar{u}}^{Py}$ on the B meson azimuthal angle determined on the data samples with a threshold on $p_T(\bar{b})$ of 10 (left) or 2.5 GeV (right).

Appendix C

Estimating Systematic Uncertainties due to Correlations Between Selection Variables

In Chapter 8 the concept of cut factorisation was introduced to estimate the number of $B_s^0 \to \mu^+\mu^-$ candidates in the $b\bar{b}$ combinatorial background data sample, `bb_mu6mu4X`. Each cut is applied independently on those events passing the baseline selection and the corresponding efficiency is calculated. The total efficiency is then determined as the product of individual efficiencies and can be later used to estimate the number of decays for a given integrated luminosity. The cut factorisation is correct only if there are no correlations between the selection variables. It was shown that some correlations are not negligible. A systematic uncertainty can be assigned to describe the effects of correlations. The procedure to determine this uncertainty is presented in detail in this appendix.

C.1 Procedure for Determining Uncertainties

The procedure adopted for estimating uncertainties has been introduced in Chapter 8. The goal is to use the probability density functions (p.d.f.) and the correlations of the four discriminating variables, L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$, as input of a toy Monte Carlo (MC) simulation tool. This tool is used to generate events containing these four variables, such that they describe as accurately as possible the initial data. With increased statistics available, it is possible to apply the entire chain of selection cuts and the overall selection efficiency can be computed. This efficiency can then be compared to the efficiency calculated by cut factorisation using the toy MC sample. The difference represents a systematic shift from the true value and is assigned to the systematic uncertainty.

A method to take correlations into account is the "projection and correlation approximation" (PCA) [Kar98]. The distributions of each of the four input variables can be transformed to a Gaussian distribution, according to (8.41). This transformation should preserve the correlations between variables. In a next step, a 4-dimensional p.d.f. is build using (8.40), incorporating correlations via the co-

variance matrix of the transformed distributions. The PCA method assumes that the 4-dimensional Gaussian function correctly describes the data, given the four projections follow a normal distribution.

In order to obtain an accurate simulation of the efficiency cuts, two conditions must be fulfilled: the initial correlations must be preserved in the simulated data and the p.d.f.'s of the initial and simulated data must be identical. These two aspects are detailed in the next sections.

C.2 Preserving the Correlations

In a first attempt, all events containing a B_s^0 candidate (passing the trigger and baseline requirements) are studied. In total 2 049 events are available. Scatter plots showing the correlations between the four variables are shown in Fig. C.1. The toy MC tool is used to generate 1 million events, which are compared with the initial data. The mechanism of the toy MC tool is described in detail in the following sections. For now, only the correlation coefficients are investigated and their fluctuations under the simulation process are studied. The main concern is that the PCA method does not work properly in all situations, as pointed out in [Kar98]. The performance is poor in those cases when large variations of a correlation coefficient are present in different space regions. The linear correlation coefficients between the selection variables in the initial and simulated data samples are shown in Table C.1. A significant difference is observed for the linear correlation factor between L_{xy} and α ($\rho = -0.49$ from the initial value of -0.33).

	L_{xy}	α	$I_{\mu\mu}$	$M_{\mu\mu}$
L_{xy}	1	-0.33	-0.08	-0.15
α		1	-0.13	0.17
$I_{\mu\mu}$			1	-0.01
$M_{\mu\mu}$				1

	L_{xy}	α	$I_{\mu\mu}$	$M_{\mu\mu}$
L_{xy}	1	-0.49	-0.06	-0.16
α		1	-0.14	0.19
$I_{\mu\mu}$			1	0.01
$M_{\mu\mu}$				1

Table C.1: Linear correlation coefficients between the L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$ variables. The coefficients in the left table are computed from the 2 049 input events obtained after the trigger and baseline selections with a statistical uncertainty of 0.02. The coefficients in the right table are computed from a simulated sample of 1 million events. The statistical error in this case is 0.0009.

The correlations in different L_{xy}-α regions are investigated. A low correlation is found in the region at low values of α (e.g., $\rho \simeq -0.1$ in the region delimited by $0 < \alpha < 0.2$), while in the outer region the correlation becomes much higher (e.g., $\rho \simeq -0.35$ for $\alpha > 0.2$). Therefore, in order to avoid an artificial bias introduced by the PCA method, a preselection of events in the low α region is investigated. Preselecting events by applying a loose cut on α results in a higher statistical uncertainty, but offers the advantage of a better estimation of the systematic uncertainty. Several cuts are tested, with the aim of finding the loosest cut, which

still preserves the correlation coefficient after the simulation process. Cut values in the range $0.16 - 0.21$ are studied. Outside this range either the cut is too tight (i.e., the statistical error becomes too large), or the bias on the correlation coefficients becomes too large. The optimal situation was found for $\alpha < 0.18$, in which case ρ is kept within one standard deviation. 673 events are available in the input sample after the preselection is applied. These events are used as input for the toy MC tool and another sample of 1 million events is simulated. The correlations computed from the initial preselected events as well as from the simulated sample are shown in Table 8.9. The simulation preserved all correlations within the statistical uncertainties.

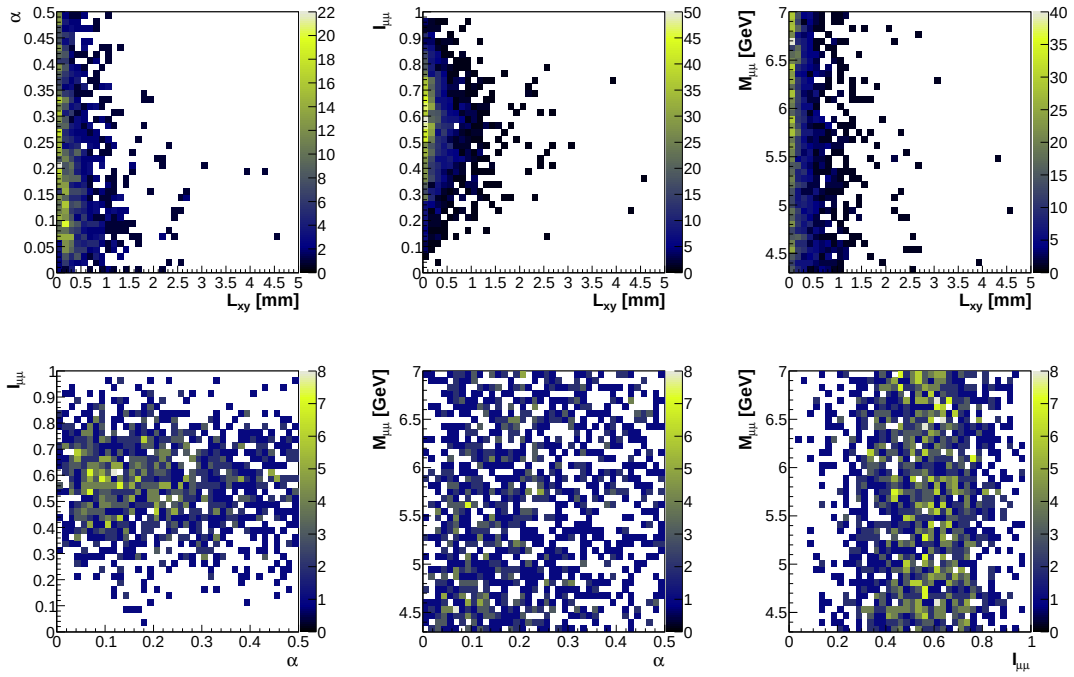


Figure C.1: Scatter plots showing the correlations between the L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$ variables for B_s^0 candidates observed in the `bb_mu6mu4X` data sample.

C.3 Transformation to a Gaussian Shape

In this section the transformation of the initial p.d.f's to Gaussian distributions is described. It has already been proven that, in order to preserve the initial correlations between the selection variables, a preselection of events is needed. Therefore, only the 673 events preselected by $\alpha < 0.18$ are used here.

The p.d.f.'s of the L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$ variables are shown in Fig. C.2. The distributions are fitted to functions which accurately describe their shapes. The statistical fluctuations in the distributions are thus smoothened by the parametrized functions. These functions are later used in the toy MC simulation. The L_{xy} distribution is fitted to the sum of two exponential functions, the α distribution is

fitted to a polynomial of third order, the $I_{\mu\mu}$ distribution is fitted to a normal distribution and the $M_{\mu\mu}$ is fitted to a polynomial of second order. The resulting functions are overlaid in Fig. C.2.

The next step is to construct the cumulative distribution functions (c.d.f.) for each variable. The c.d.f. is computed as:

$$F(x_i) = \frac{\int_{x_i^{min}}^{x_i} P(t)dt}{\int_{x_i^{min}}^{x_i^{max}} P(t)dt}, \tag{C.1}$$

where x_i stands for any of the L_{xy}, α, $I_{\mu\mu}$ or $M_{\mu\mu}$ variables, $F(x_i)$ is the corresponding c.d.f. and $P(x_i)$ is the p.d.f. computed from the parametrization of the distributions as described above. For convenience, this is computed numerically. The range of each variable x_i is split in 1000 equal intervals and $F(x_i)$ is then computed for each point. This is shown in Fig. C.3 for each selection variable. The error bands are obtained after propagating the covariance matrix of the parameters of the fit functions.

The transformation to a Gaussian distribution is done according to (8.41). This is also computed numerically for each variable x_i. The transformation functions $y_i(x_i)$ are displayed in Fig. C.4. Errors are computed by propagating the covariance matrix of the parameters of the fit functions. In order to verify that the transformation produces indeed a Gaussian distribution, the y_i variables are plotted as histograms and Gaussian functions are fitted to the distributions. The distributions as well as the fitted functions are shown in Fig. C.5. All distributions are indeed described by normal distributions. The resulting values for the mean and sigma are found in Fig. C.5.

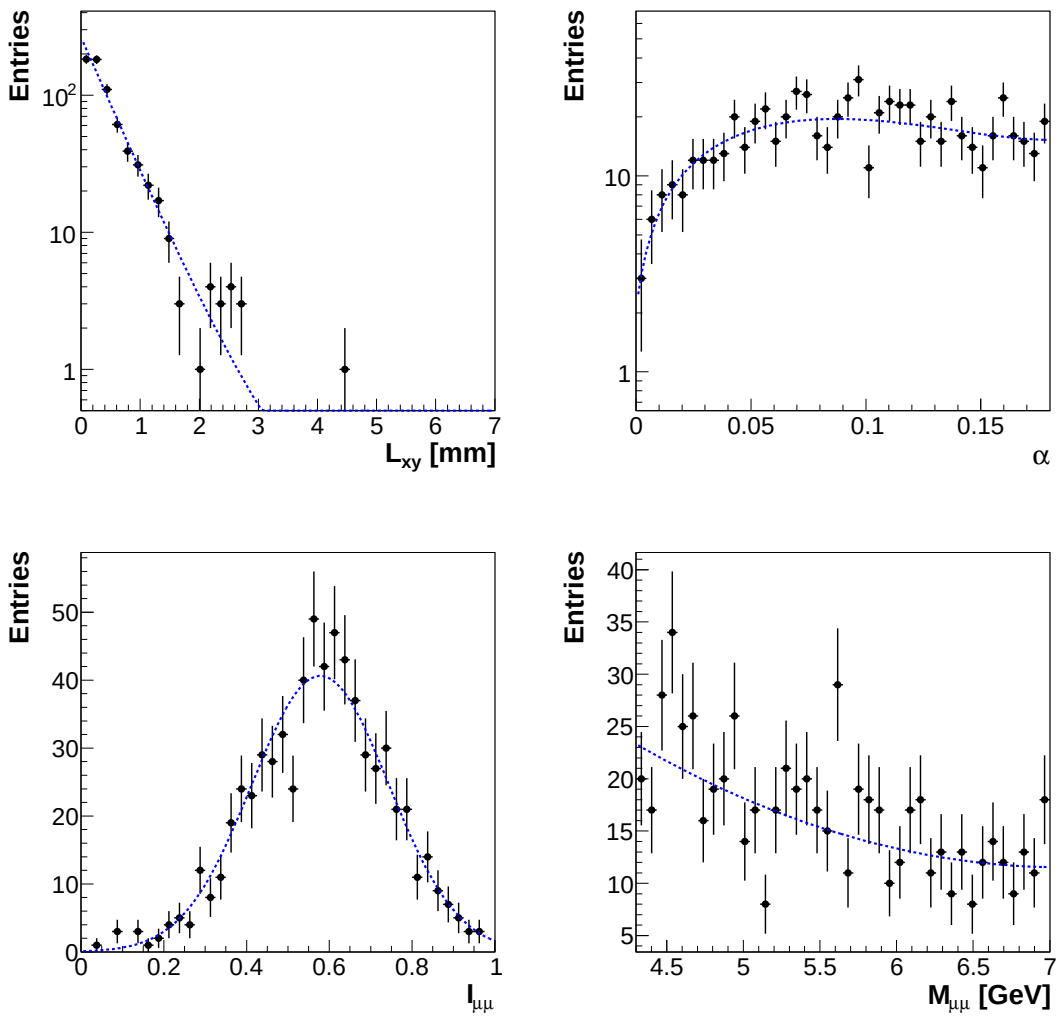


Figure C.2: Distributions of the L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$ variables for preselected events (dots). The dashed lines show the functions which are fitted to the distributions.

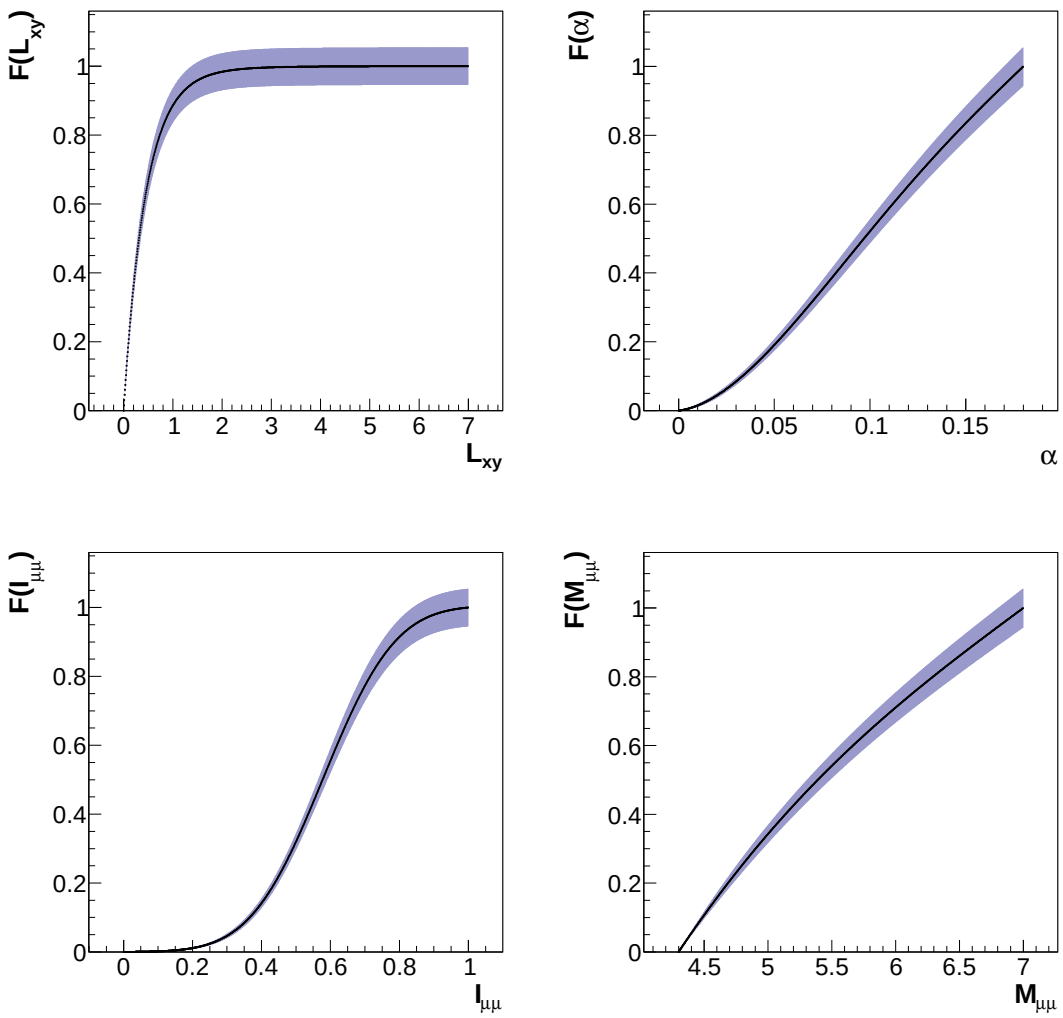


Figure C.3: Cumulative distribution functions of L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$. The error band is computed by propagating the covariance matrix of the fit parameters.

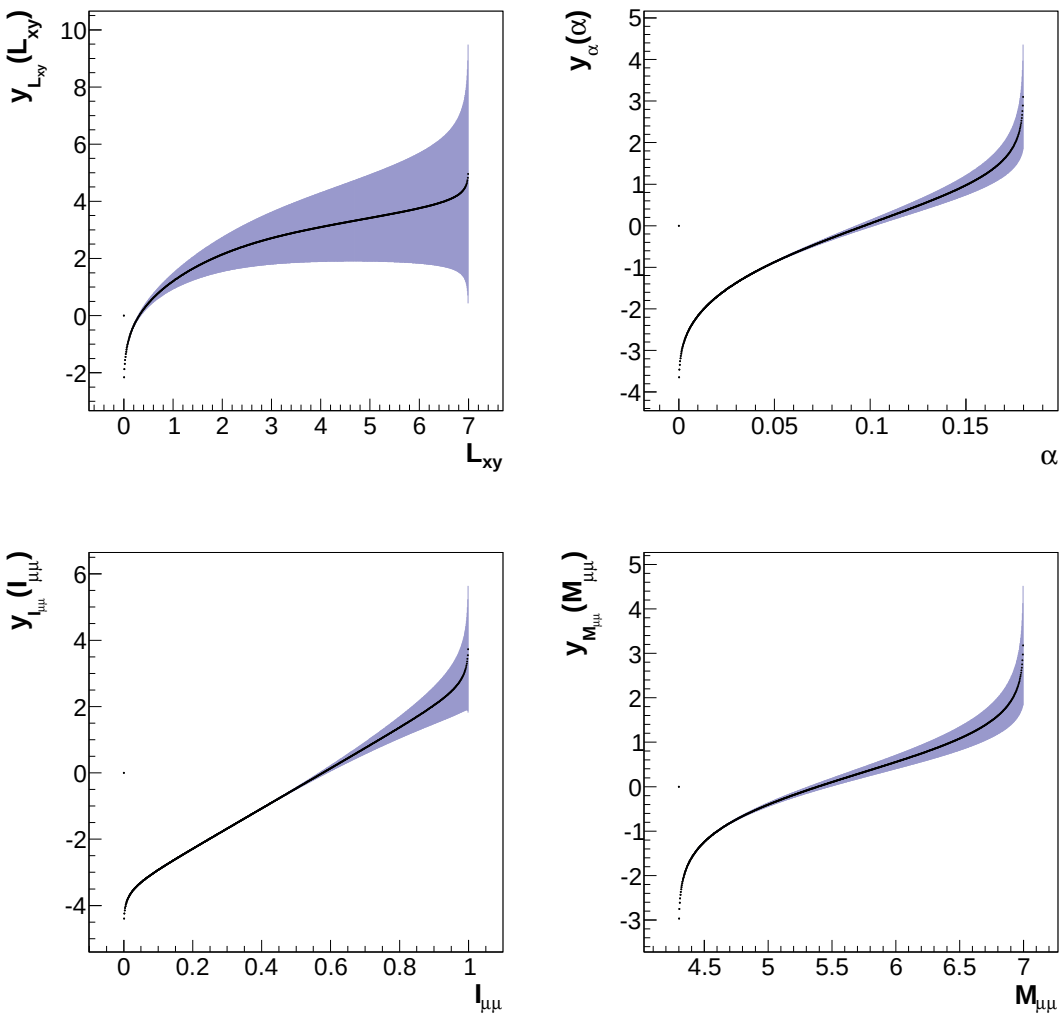


Figure C.4: Transformation functions of L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$ to Gaussian functions. The error band is computed by propagating the covariance matrix of the fit parameters.

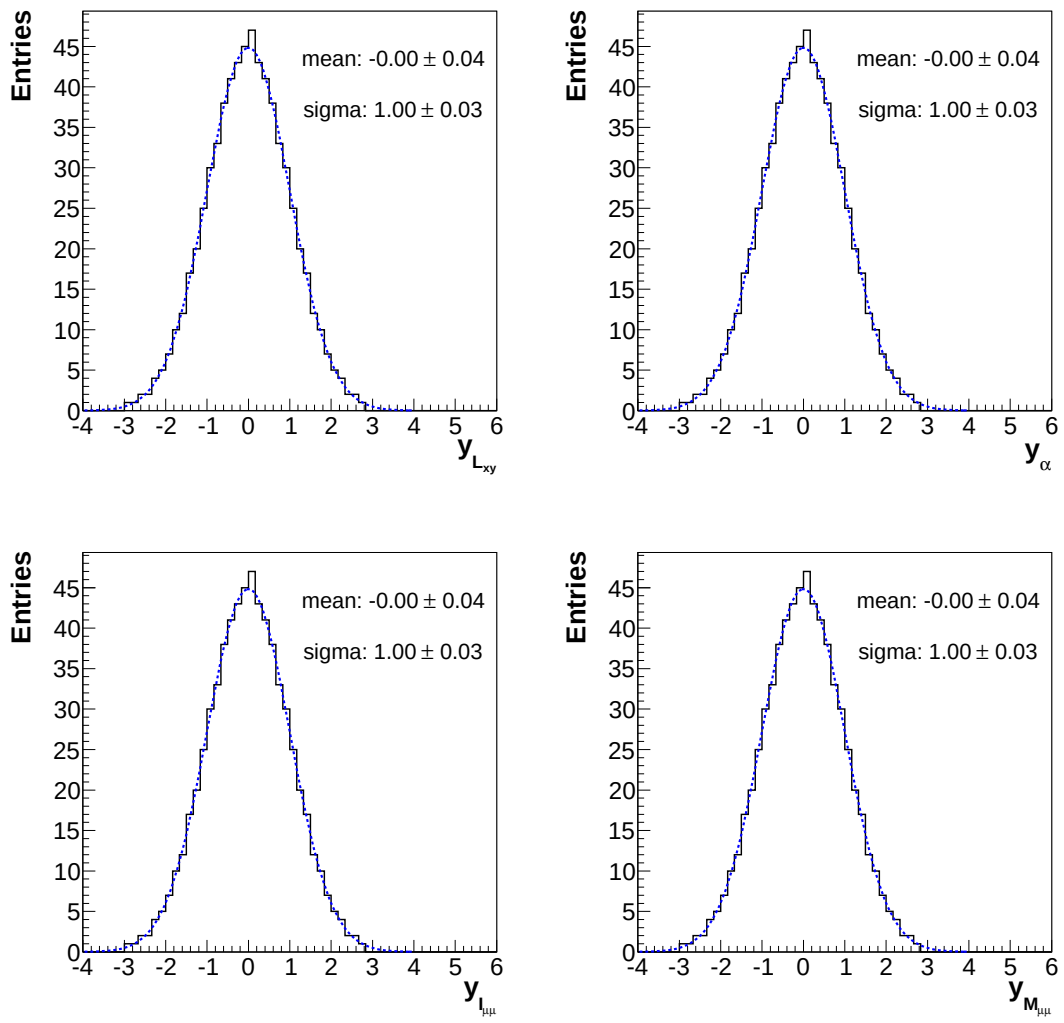


Figure C.5: Transformed (Gaussian) distributions corresponding to L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$. Gaussian functions fitted to the distributions are shown with a dashed line.

C.4 Toy Monte Carlo Simulation

Having checked that the distributions of the four variables are transformed to normal distributions, the PCA method is employed (i.e., it is assumed that a 4-dimensional normal distribution describes the data, given the 1-dimensional projections are distributed accordingly). A random generator is employed to simulate events consisting of a 4-dimensional vector $\mathbf{y}$ distributed according to (8.40). Here $G(\mathbf{y})$, the 4-dimensional Gaussian distribution, includes the covariance matrix V of the y_i variables. This preserves the correlations of the initial x_i variables, as already shown in Table 8.9. The toy MC tool is used to generate 500 million events.

Besides correlations, another requirement is that the simulated variables accurately describe the p.d.f.'s of the initial variables. Since, by construction, the $y_i(x_i)$ functions are bijective, each simulated y_i value corresponds to a unique x_i value. Therefore the inverse of transformation (8.41) is used to compute the corresponding 4-dimensional vector $\mathbf{x} = (x_i)$ for each event. The distributions of the simulated x_i variables are shown in Fig. C.6. The initial MC data and the data simulated with the toy MC tool are in good agreement.

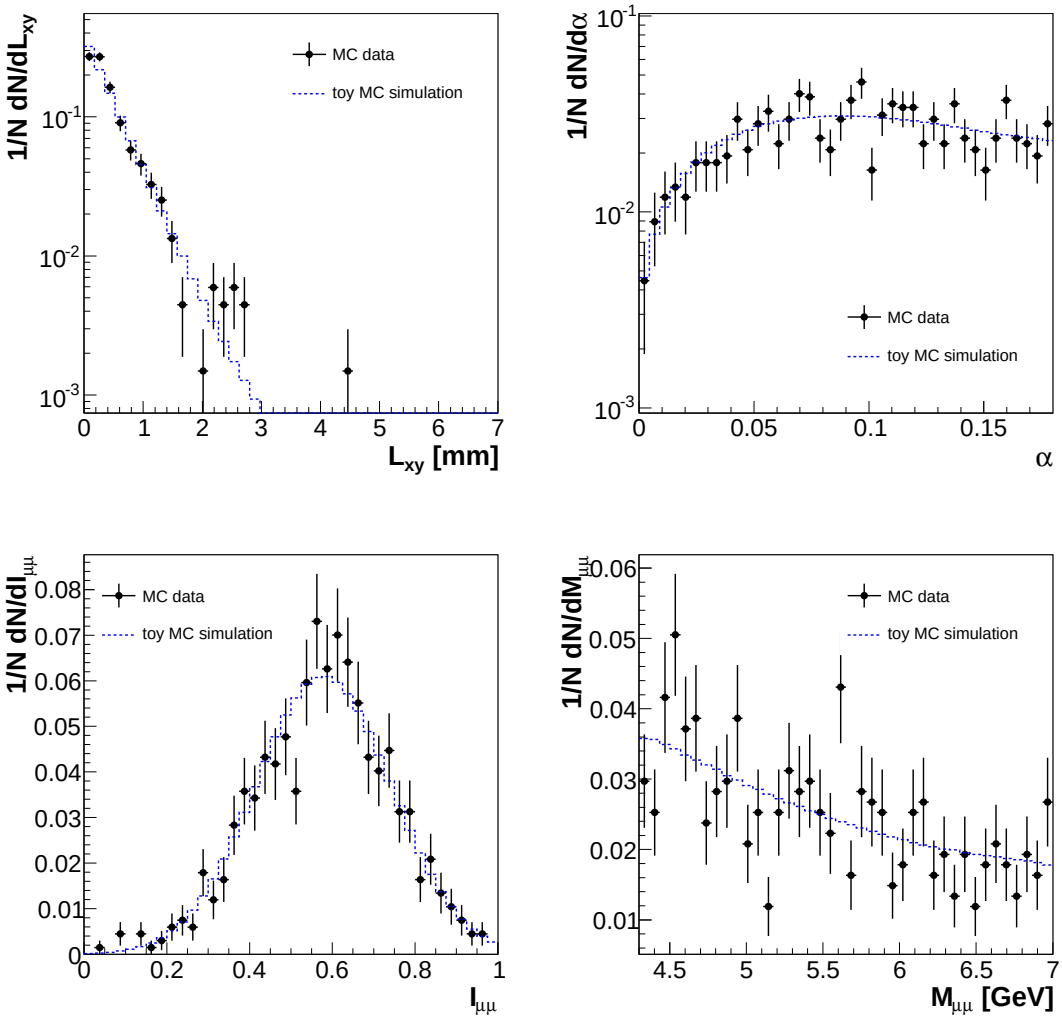


Figure C.6: Distributions of the L_{xy}, α, $I_{\mu\mu}$ and $M_{\mu\mu}$ variables for preselected events (dots). The dashed lines show the functions to which the distributions are fitted.

C.5 Efficiency Study

The last step is analysing the 500 million simulated events. The same selection criteria as those used in the analysis of the MC data are considered:

- $L_{xy} > 0.5\,\text{mm}$,
- $\alpha < 0.017$,
- $I_{\mu\mu} > 0.9$,
- $M_{\mu\mu} \in [5261\,\text{MeV}, 5585\,\text{MeV}]$.

The cut selection efficiencies are calculated and two cases are discussed. In the first case, each cut is applied independently and the corresponding efficiencies $\varepsilon^{L_{xy}}$, ε^{α}, $\varepsilon^{I_{\mu\mu}}$ and $\varepsilon^{M_{\mu\mu}}$ are determined. The total efficiency ε^{fact} is then computed as the product of each efficiency. This represents the situation, that cut factorisation is used in the analysis of the MC data, where the correlations are not taken into account. In the second case, all cuts are applied consecutively and the final efficiency ε^{total} is determined. This represents the efficiency for the case, that all correlations are taken into account. The values of these efficiencies are shown in Table C.2. The selection efficiencies computed for the MC data are also given in order to show that simulated data accurately describes the original MC data used in the analysis. All differences seen are within the statistical uncertainties.

	toy MC simulation (500 million events)	MC data (673 events)
$\varepsilon^{L_{xy}}$	0.33347 ± 0.00002	0.319 ± 0.019
ε^{α}	0.03275 ± 0.00001	$0.037\,^{+0.009}_{-0.007}$
$\varepsilon^{I_{\mu\mu}}$	0.01995 ± 0.00001	$0.019\,^{+0.007}_{-0.005}$
$\varepsilon^{M_{\mu\mu}}$	0.12218 ± 0.00001	$0.127\,^{+0.014}_{-0.013}$
ε^{fact}	$(2.661 \pm 0.001) \times 10^{-5}$	$(2.5 \pm 0.9) \times 10^{-5}$
ε^{tot}	$(3.02 \pm 0.02) \times 10^{-5}$	-

Table C.2: Efficiencies of selection variables for the simulated data and for the MC data available for analysis.

Finally, the relative difference $(\varepsilon^{fact} - \varepsilon^{tot})/\varepsilon^{fact}$ is computed to be -0.134. This represents the deviation of the efficiency computed from cut factorisation (which does not take correlations into account) from the ideal case when the efficiency is computed taking the correlations into account. The selection efficiency computed by cut factorisation is about 13% too small and a systematic uncertainty of 13% is assigned to the total efficiency.

Appendix D

Upper Limit on the $B_s^0 \to \mu^+\mu^-$ Branching Ratio as a Function of the Integrated Luminosity

In Chapter 9, the upper limit on the $B_s^0 \to \mu^+\mu^-$ branching ratio is represented as a function of integrated luminosity (Fig. 9.6). In this appendix, the individual values of the upper limits are detailed in Table D.1. The computations are performed with a 95% and a 90% confidence level. Both the Bayesian and the Feldman Cousins methods are considered. For the Bayesian approach the maximum integrated luminosity considered is about 30 fb^{-1}. For the Feldman Cousins method, the maximum integrated luminosity considered is about 24 fb^{-1} (discussed in Chapter 9).

$\int Ldt$	Bayesian approach		Feldman Cousins approach	
$[\,\text{fb}^{-1}]$	95%	90%	95%	90%
0.1	3.2×10^{-7}	2.3×10^{-7}	2.8×10^{-7}	2.1×10^{-7}
1.1	4.1×10^{-8}	3.1×10^{-8}	3.6×10^{-8}	2.8×10^{-8}
2.1	2.5×10^{-8}	1.9×10^{-8}	2.1×10^{-8}	1.6×10^{-8}
3.1	1.9×10^{-8}	1.5×10^{-8}	1.6×10^{-8}	1.3×10^{-8}
4.1	1.6×10^{-8}	1.2×10^{-8}	1.3×10^{-8}	1.0×10^{-8}
5.1	1.3×10^{-8}	1.0×10^{-8}	1.1×10^{-8}	9.1×10^{-9}
6.1	1.2×10^{-8}	9.3×10^{-9}	1.0×10^{-8}	8.1×10^{-9}
7.1	1.1×10^{-8}	8.5×10^{-9}	9.3×10^{-9}	7.4×10^{-9}
8.1	1.0×10^{-8}	7.8×10^{-9}	8.6×10^{-9}	6.9×10^{-9}
9.1	9.3×10^{-9}	7.3×10^{-9}	8.0×10^{-9}	6.4×10^{-9}
⋮	⋮	⋮	⋮	⋮

⋮	⋮	⋮	⋮	⋮
10.1	8.8×10^{-9}	6.9×10^{-9}	7.6×10^{-9}	6.0×10^{-9}
11.1	8.3×10^{-9}	6.5×10^{-9}	7.2×10^{-9}	5.7×10^{-9}
12.1	7.9×10^{-9}	6.2×10^{-9}	6.8×10^{-9}	5.5×10^{-9}
13.1	7.6×10^{-9}	5.9×10^{-9}	6.6×10^{-9}	5.3×10^{-9}
14.1	7.2×10^{-9}	5.7×10^{-9}	6.3×10^{-9}	5.1×10^{-9}
15.1	7.0×10^{-9}	5.5×10^{-9}	6.1×10^{-9}	4.9×10^{-9}
16.1	6.8×10^{-9}	5.3×10^{-9}	5.9×10^{-9}	4.7×10^{-9}
17.1	6.5×10^{-9}	5.1×10^{-9}	5.7×10^{-9}	4.6×10^{-9}
18.1	6.3×10^{-9}	5.0×10^{-9}	5.5×10^{-9}	4.5×10^{-9}
19.1	6.1×10^{-9}	4.9×10^{-9}	5.4×10^{-9}	4.4×10^{-9}
20.1	6.0×10^{-9}	4.7×10^{-9}	5.3×10^{-9}	4.3×10^{-9}
21.1	5.8×10^{-9}	4.6×10^{-9}	5.1×10^{-9}	4.2×10^{-9}
22.1	5.7×10^{-9}	4.5×10^{-9}	5.0×10^{-9}	4.1×10^{-9}
23.1	5.6×10^{-9}	4.4×10^{-9}	4.9×10^{-9}	4.0×10^{-9}
24.1	5.5×10^{-9}	4.3×10^{-9}	4.8×10^{-9}	3.9×10^{-9}
25.1	5.4×10^{-9}	4.3×10^{-9}	-	-
26.1	5.3×10^{-9}	4.2×10^{-9}	-	-
27.1	5.2×10^{-9}	4.1×10^{-9}	-	-
28.1	5.1×10^{-9}	4.0×10^{-9}	-	-
29.1	5.0×10^{-9}	4.0×10^{-9}	-	-
30.1	4.9×10^{-9}	3.9×10^{-9}	-	-

Table D.1: Expected upper limits on the branching ratio as a function of integrated luminosity. Computations are performed using the Bayesian and the Feldman Cousins approaches at a 95% or a 90% confidence level.

List of Figures

List of Tables

Bibliography

[Aad08] G. Aad et al. (ATLAS Collaboration), *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008), S08003.

[Aad09] G. Aad et al. (ATLAS Collaboration), *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, 2009 [arXiv:hep-ex/0901.0512] .

[Aad10] G. Aad et al. (ATLAS Collaboration), *The ATLAS Simulation Infrastructure*, submitted to Eur. Phys. J. C (2010) [arXiv:physics.ins-det/1005.4568v1] .

[Aad11] G. Aad et al. (ATLAS Collaboration), *Luminosity Determination in pp Collisions at sqrt(s)=7 TeV Using the ATLAS Detector at the LHC*, Eur. Phys. J. **C71** (2011), 1630 [arXiv:hep-ex/1101.2185] .

[Aai11a] R. Aaij et al. (LHCb Collaboration), *Determination of f_s/f_d for 7 TeV pp collisions and a measurement of the branching fraction of the decay $B_d \to D^- K^+$* [arXiv:hep-ex/1106.4435] .

[Aai11b] R. Aaij et al. (LHCb Collaboration), *Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$*, Phys.Lett. **B699** (2011), 330–340 [arXiv:hep-ex/1103.2465v2] .

[Aal11] T. Aaltonen et al. (CDF Collaboration), *Search for $B_s^0 \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ Decays with CDF II* [arXiv:hep-ex/1107.2304] .

[Aam08] K. Aamodt et al. (ALICE Collaboration), *The ALICE experiment at the CERN LHC*, JINST **3** (2008), S08002.

[Aba10] V. M. Abazov et al. (D0 Collaboration), *Search for the rare decay $B_s^0 \to \mu^+\mu^-$*, 2010 [arXiv:hep-ex/1006.3469] .

[Ado08] R. Adolphi et al. (CMS Collaboration), *The CMS experiment at the CERN LHC*, JINST **3** (2008), S08004.

[Ago03] S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003), 250–303.

[All06] J. Allison et al., *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006), 270.

[Alv08] A. A. Alves et al. (LHCb Collaboration), *The LHCb Detector at the LHC*, JINST **3** (2008), S08005.

[Ams08] C. Amsler et al. (Particle Data Group Collaboration), *Review of Particle Physics*, Physics Letters **B667, 1** (2008) [and 2009 partial update for the 2010 edition] .

[Ant09] I. Antcheva et al., *ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization*, Comput. Phys. Commun. **180** (2009), 2499–2512.

[Art08] M. Artuso, D. Asner, P. Ball, E. Baracchini, G. Bell, et al., *B, D and K decays*, Eur.Phys.J. **C57** (2008), 309–492 [arXiv:hep-ph/0801.1833] .

[Atl10a] ATLAS Collaboration, *ATLAS internal document*, ATL-COM-PHYS-2010-902, 2010.

[Atl10b] ATLAS Collaboration, *Estimating Track Momentum Resolution in Minimum Bias Events Using Simulation and K^0_s in $\sqrt{s} = 900$ GeV Collision Data*, ATLAS conference note (ATLAS-CONF-2010-009), 2010.

[Atl10c] ATLAS Collaboration, *Muon Reconstruction Performance*, ATLAS conference note (ATLAS-CONF-2010-065), 2010.

[Atl10d] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment*, ATLAS conference note (ATLAS-CONF-2010-069), 2010.

[Atl10e] ATLAS Collaboration, *Performance of the ATLAS Muon Trigger in p-p collisions at $\sqrt{s} = 7$ TeV*, ATLAS conference note (ATLAS-CONF-2010-095), 2010.

[Atl10f] ATLAS Collaboration, *Tracking Results and Comparison to Monte Carlo Simulation at $\sqrt{s} = 900$ GeV*, ATLAS conference note (ATLAS-CONF-2010-011), 2010.

[Atl10g] ATLAS Collaboration, *Tracking Validation with Heavy Ion Collisions*, ATLAS internal document (ATLAS-COM-PHYS-2010-1028), 2010.

[Atl11a] ATLAS Collaboration, *A measurement of the ATLAS muon reconstruction and trigger efficiency using J/ψ decays*, ATLAS conference note (ATLAS-CONF-2011-021), 2011.

[Atl11b] ATLAS Collaboration, *Updated Luminosity Determination in $p-p$ Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector*, ATLAS conference note (ATLAS-CONF-2011-011), 2011.

[Ber] R. P. Bernhard, *Search for rare decays of the Bs meson with the D0 experiment*, FERMILAB-THESIS-2005-56.

[Buc96] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, Rev. Mod. Phys. **68** (1996), 1125–1144 [arXiv:hep-ph/9512380] .

[Bur10] A. J. Buras, *Minimal flavour violation and beyond: Towards a flavour code for short distance dynamics*, Acta Phys. Polon. **B41** (2010), 2487–2561 [arXiv:hep-ph/1012.1447] .

[Cab63] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys.Rev.Lett. **10** (1963), 531–533.

[Cas09] D. Casadei, *How to measure efficiency* [arXiv:data-an/0908.0130v1] .

[Cha11] S. Chatrchyan et al. (CMS Collaboration), *Search for B(s) and B to dimuon decays in pp collisions at 7 TeV* [1107.5834] .

[Con98] CDF Collaboration, *Upper Limits on Poisson Processes Incorporating Uncertainties in Acceptance and Background*, CDF public note (4476), 1998.

[Con03] J. Conrad, O. Botner, A. Hallgren, and C. Perez de los Heros, *Including systematic uncertainties in confidence interval construction for Poisson statistics*, Phys. Rev. **D67** (2003), 012002 [arXiv:hep-ex/0202013] .

[Dob01] M. Dobbs and J. B. Hansen, *The HepMC C++ Monte Carlo event record for High Energy Physics*, Comput. Phys. Commun. **134** (2001), 41–46.

[Duc] G. Duckeck, (Ed.) et al. (ATLAS Collaboration), *ATLAS Computing: Technical Design Report*, CERN-LHCC-2005-022.

[Eng64] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys.Rev.Lett. **13** (1964), 321–322.

[Eva08] L. Evans, (ed.) and P. Bryant, (ed.), *LHC Machine*, JINST **3** (2008), S08001.

[Fel98] G. J. Feldman and R. D. Cousins, *A Unified Approach to the Classical Statistical Analysis of Small Signals*, Phys. Rev. **D57** (1998), 3873–3889 [arXiv:physics/9711021] .

[Gla70] S. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys.Rev. **D2** (1970), 1285–1292.

[Gre02] W. Greiner, S. Schramm, and E. Stein, *Quantum chromodynamics*.

[Gre07] W. Greiner, S. Schramm, and E. Stein, *Quantum chromodynamics*.

[Gru] C. Grupen and B. Schwartz, *Particle detectors*, Cambridge, UK: Cambridge Univ. Pr. (2008) 651 p.

[Gur64] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys.Rev.Lett. **13** (1964), 585–587.

[Hag07] K. Hagiwara, A. D. Martin, D. Nomura, and T. Teubner, *Improved predictions for g-2 of the muon and* $\alpha_{\rm QED}(M_Z^2)$, Phys. Lett. **B649** (2007), 173–179 [hep-ph/0611102] .

[Has10] S. Hassani, C. Petridou, D. Iliadis, and K. Bachas, *The Tag and Probe method with* J/ψ *and* Υ *candidates in ATLAS*, ATLAS internal document (ATL-PHYS-INT-2010-050), 2010.

[Hei04] J. Heinrich et al., *Interval estimation in the presence of nuisance parameters. 1. Bayesian approach* [arXiv:physics/0409129] .

[Hig64] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys.Rev.Lett. **13** (1964), 508–509.

[Kar98] D. Karlen, *Using projections and correlations to approximate probability distributions*, Comput. Phys. **12** (1998), 380–384 [arXiv:physics/9805018] .

[Kob73] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog.Theor.Phys. **49** (1973), 652–657.

[Kos03] V. Kostiyukhin, *VKalVrt - package for vertex reconstruction in ATLAS*, ATLAS internal document (ATL-PHYS-2003-031), 2003.

[LHC11] CMS and LHCb Collaboration, *Search for the rare decay* $B_s^0 \to \mu^+\mu^-$ *at the LHC with the CMS and the LHCb experiments*, LHCb conference note (LHCb-CONF-2011-047), 2011.

[Lim10] A. Limosani (ATLAS Collaboration), *Commissioning and performance of the ATLAS inner detector with proton-proton collisions at the LHC*, PoS **ICHEP2010** (2010), 015.

[Lun10] J. Lundberg, J. Conrad, W. Rolke, and A. Lopez, *Limits, discovery and cut optimization for a Poisson process with uncertainty in background and signal efficiency: TRolke 2.0*, Comput. Phys. Commun. **181** (2010), 683–686 [arXiv:physics.data-an/0907.3450] .

[Met85] M. Metcalf, *Effective Fortran 77*, Oxford University Press, 1985.

[Nak10] K. Nakamura et al. (Particle Data Group Collaboration), *Review of particle physics*, J. Phys. **G37** (2010), 075021.

[Ney37] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*, Phil. Trans. R. Soc. Lond. A **236** (1937), 333–380.

[Pun03] G. Punzi, *Sensitivity of searches for new signals and its optimization* [arXiv:physics/0308063] .

[Ser11] J. Serrano (LHCb Collaboration), *Search for the Very Rare Decays $B_{s/d} \to \mu^+\mu^-$ at LHCb*, Talk at EPS-HEP, 2011.

[Sjo06] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006), 026 [arXiv:hep-ph/0603175] .

[Smi05] M. Smizanska, *PythiaB interface to Pythia6 dedicated to simulation of beauty events*, ATLAS internal document, 2005.

[Str97] B. Stroustrup, *The C++ Programming Language: Language Library and Design Tutorial*, Addison-Wesley Longman, 1997.

[Ull08] T. Ullrich and Z. Xu, *Treatment of Errors in Efficiency Calculations* [arXiv:physics.data-an/0701199] .

[vR09] G. van Rossum and F. L. Drake, *Python 3 Reference Manual*, Createspace, 2009.

[Wal09] W. Walkowiak, private communication, 2009.

[Yao06] W. M. Yao et al. (Particle Data Group Collaboration), *Review of particle physics*, J. Phys. **G33** (2006), 1–1232.

Acknowledgment

This page is dedicated to the people who made my doctoral dissertation possible and who encouraged and helped me in achieving my goals.

I would like to express my gratitude to my doctoral advisor, Prof. Dr. Peter Buchholz, for giving me the the wonderful opportunity to be part of one of the most exciting research environments. I am thankful for his trust and his guiding during my work as a PhD student. Thanks to him I could gain a lot of experience in the field of experimental particle physics, which will certainly be useful in my future activity.

I would like to thank Prof. Dr. Ivor Fleck, who accepted to be my second advisor for my doctoral dissertation and offered his support for achieving this goal.

Furthermore, I want to thank Dr. Wolfgang Walkowiak for his active support during my work. He has always had the openness to discuss interesting topics related to my research (and not only) and to give useful answers to my questions. I appreciate him sharing his experience as an experimental physicist, from which I could benefit to extend my own knowledge. And, not the least, thanks for taking the time to proofread the largest part of this thesis.

I would also like to thank Dr. Kai Grybel, with whom I shared the office for the last years. Thanks for the pleasant work atmosphere, for sharing a laugh or two on some occasions, for useful conversations on other occasions and for taking the time to proofread a part of my thesis.

I would like to thank the members of the B Physics Siegen group for the fruitful collaboration. I would like to mention Dr. Iskander Ibragimov, who was kind to proofread a part of this thesis, and Bakul Gaur.

I would also like to mention the members of the ATLAS collaboration and of the rare B decays subgroup, and especially Dr. Sergey Sivoklokov for the useful discussions on the topic of this thesis.

Finally, I would like to thank my family, especially my mother and sisters, who have continually and tirelessly offered me encouragement during the time spend as a PhD student.

This work was supported by the German Federal Ministry for Education and Research (BMBF) as part of the BMBF FSP 101 - ATLAS.