University of Cambridge

Measurements of $B \to \mu^+\mu^-$ decays using the LHCb Experiment

Hannah Mary Evans
Selwyn College

Supervisor
Prof. Val Gibson

Abstract

This dissertation documents studies of very rare B meson decays at the LHCb experiment on data taken during the first two experiment runs of the Large Hadron Collider (LHC).

The LHCb experiment was designed to test the Standard Model of particle physics and search for new physics effects that go beyond the scope of the Standard Model, through the decay of b-hadrons produced in high energy proton-proton collisions at the LHC. The measurements described in this dissertation were made using data samples of proton-proton collisions with integrated luminosities of 1.0, 2.0 and 1.4 fb^{-1}, collected at centre-of-mass energies of 7, 8 and 13 TeV respectively.

The branching fractions of the very rare $B^0_{(s)} \to \mu^+\mu^-$ decays and the effective lifetime of $B^0_s \to \mu^+\mu^-$ decays are sensitive to effects from new physics theories. New physics processes could influence the $B^0_s \to \mu^+\mu^-$ effective lifetime and branching fraction independently, and therefore the two observables are complementary in the search for new physics.

The $B^0_s \to \mu^+\mu^-$ decay is observed with a statistical significance of 7.8 σ and the branching fraction is measured as $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9}$. The $B^0_s \to \mu^+\mu^-$ effective lifetime is measured for the first time as 2.04 $\pm$ 0.44 (stat) $\pm$ 0.05 (syst) ps. An upper limit is placed on the branching fraction $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.4 \times 10^{-10}$ at the 95 % confidence level. All results are consistent with the predictions of the Standard Model.

Declaration

Hannah Evans
March 2017

Acknowledgements

Preface

Table of contents

Chapter 1

Introduction

Chapter 2

Theory of $B \to \mu\mu$ decays; the Standard Model and beyond

Chapter 3

The LHC and the LHCb experiment

The European Organisation for Nuclear Research (CERN) was founded in 1954, it began with 12 member states as an organisation to encourage European collaboration and to study nuclear physics. The collaborative nature of CERN has enabled large-scale expensive experiments to be built that individual member states would not have be able to afford. The Proton Synchrotron (PS) was CERN's flagship accelerator which was operational in 1959. It had a circumference of 628 m and accelerated protons up to a center-of-mass energy of 25 GeV makig the PS highest energy particle accelerator at that time. Now 62 years since its foundation, CERN has grown to include 22 member states[1] and is still at the forefront of high energy physics research. CERN's latest accelerator, the Large Hadron Collider (LHC), is most energetic particle accelerator ever built, with a 27 km circumference the LHC was designed to collide protons at a centre-of-mass energy of 14 TeV. This chapter introduces the LHC and the LHC beauty experiment, one of the experiments that studies the produces of particle collisions produced at the LHC.

3.1 The LHC

The LHC is a proton synchrotron designed to accelerate and collide two beams of protons with a centre-of-mass energy of 14 TeV. Although operation of the LHC began in 2010 it is yet to reach the design energy. The purpose of the LHC is to provide high energy proton-proton pp collisions, the products of which are used for precision

[1]Countries and organisations that are unable to become member states can still participate in scientific research as observer states [1].

tests of the Standard Model (SM) and to search for new physics effects that cannot be explained within the context of the SM. There are four interaction points on the LHC ring where the beams are brought to collide, at these points various experiments detect and study the products of particle collisions. In addition to protons, the LHC can also accelerate lead-nuclei up to 2.76 TeV per nucleon, however it is only the products from proton collisions that are studied in this dissertation.

The protons accelerated by the LHC originate from hydrogen gas, the hydrogen atoms are ionised to strip away the electrons and then the protons are accelerated through a chain of particle accelerators of increasing energy before being injected into the LHC. The chain of accelerators, shown in Fig. 3.1, consists of machines that were used in experiments throughout the second half of the last century and have been modified to meet the requirements needed to provide protons to the LHC. The protons leave the chain of accelerators with of energy of 450 GeV per proton and in bunches of 0^{11} protons, as the bunches are injected into the LHC they are split into two oppositely circulating beams. The LHC accelerates the protons to the desired centre-of-mass energy using supercooled radio frequency cavities and guides them around the ring with superconducting dipole magnets. Once the required energy has been reached, the bunches are focused using quadrupole magnets before being brought to collided at 4 interaction points around the LHC ring.

The centre-of-mass energy of a collider is an important measure of its performance as it describes the energy avaliable to create new particules during *pp* collisions, another important measure of collider performance is the instantaneous luminosity a collider can provide. The instantaneous luminosity, $\mathcal{L}$, is a measure of how many collisions occur per second, it is given by

$$\mathcal{L} = \frac{N^2 f n_b}{\mathcal{F}}. \tag{3.1}$$

where N is the number of protons per bunch, n_b the number of bunches per beam, f the bunch revolution frequency and $\mathcal{F}$ contains information about the beam geometry. The LHC is designed to operate at a maximum instantaneous luminosity of 10^{34} cm^{-2}s^{-1}. To reach this luminosity the LHC can have up to 2808 proton bunches per beam with a revolution frequency of 11.245 kHz and a speration of 25 ns between each proton bunch. The higher the luminosity, the more collisions happen in a second and the more particles will be produced, this can either be advantageous or disadvantageous depending on the physics process that is being studied. Therefore luminosity delivered at each interaction point can be tuned by the quadrupole magnets by altering the shape of each bunch to suit the experiments at each point.

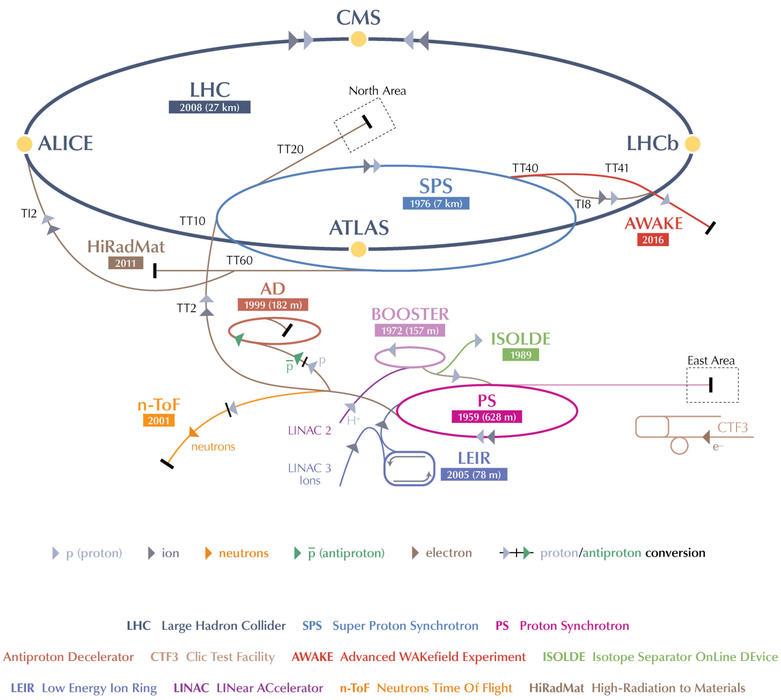


Fig. 3.1 The accelerator complex at CERN. The chain of accelerators used to inject protons into the LHC begins with the Linac 2 which accelerates protons to 50 MeV, the protons are passed to the Proton Synchrotron Booster that accelerates them to 1.4 GeV. The Proton Synchrotron is next in the chain, accelerating protons to 25 GeV and creating the desired spacing between proton bunches. Then finally the Super Proton Synchrotron accelerating protons to 450 GeV ready for injection into the LHC. Source: CERN.

Proton beams first circulated the LHC in 2008 and since then there have been two physics runs separated by a long shutdown period. Run 1 began in 2010 and continued until 2013, during this time protons were collided with a centre-of-mass energy of $\sqrt{s}$ = 7 TeV during 2010 and 2011, the energy was increased to $\sqrt{s}$ =8 TeV for operation during 2012. After Run 1 there was a period of long shut doen (LS1) during during work was done to prepare the LHC to operate at higher energies and renovation work was preformed on accelerators that provide the LHC with protons. Run 2 began in 2015 with proton collisions at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV, this Run will continue until 2018 when a second period of upgrades and maintenance, the Long Shutdown 2, will begin.

There are 7 experiments on the LHC that detect particles produced in proton and heavy ion collisions. There are two general purpose detectors, ATLAS and CMS, that were designed to search for the Higgs boson and new effects that are beyond the scope of the SM, these two experiments operate at the full instantaneous luminosity of the LHC. ALICE studies quark-gluon plasma produced in heavy ion collisions to understand conditions similar to those present in the early universe. The TOTEM experiment studies properties of protons as they collide head on at the LHC and the MOEDAL experiment is aims to detect magnetic monopoles. The LHCf experiment studies particles that are thrown forward in LHC collisions to understand similar processes that occur in cosmic rays.

The final experiment if the Large Hadron Collider Beauty experiment (LHCb) that will be described in the next section

3.2 The LHCb experiment

The LHCb experiment was build to study the SM and search for new physics phenonoma through the study of $\mathcal{CP}$-violating decays and rare b-hadron decays. At the LHC the dominant production mechanisms of $b\bar{b}$ pairs are gluon-gluon fusion, quark anti-quark annihilation and gluon-gluon splitting. The $b\bar{b}$ pairs produced travel at small angles relative to the beam pipe as shown in Figure 3.2 and hadronize to form a range of b-hadrons, including B^+, B_s^0 and Λ_b^0, that are studied by LHCb.

The LHCb experiment was built as a single arm forward spectrometer, with an angular coverage of 10 to 300 mrad in the vertical direction and 10 to 250 mrad in the horizontal direction relative the the beam pipe. The angular coverage was chosen to exploit the small angles at which $b\bar{b}$ pairs are produced. A cross-section of the LHCb detector is shown in Figure 3.3, where a right handed coordinate system is

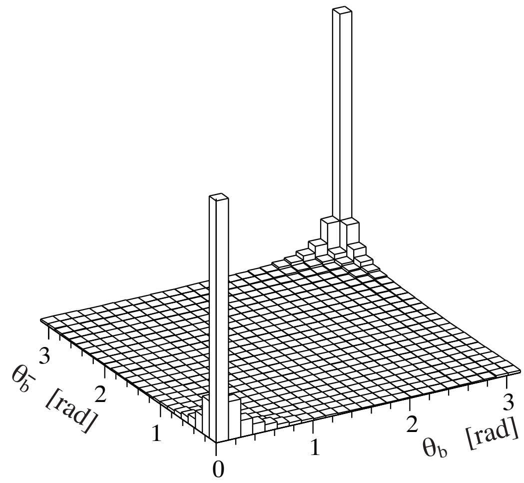


Fig. 3.2 Simulated angular distribution for $b\bar{b}$ production at the LHC, angles are relative the the beam pipe with $\theta = 0$ in the forward direction and $\theta = \pi$ in the backward direction [2].

used. Protons collide at the interaction point on the left hand side of the diagram, the products of the collisions travel through the detector leaving information in the sub-detectors along the length of the detector. The information deposited in the sub-detectors is reconstructed to determine what happened during the *pp* collisions.

The different sub-detectors have been chosen to exploit the characteristics of *b*-hadron decays and fall into 2 distinct categories; tracking detectors and particle identification detectors. Each sub-detector and its performance are described in the following sections along with the trigger system and software needed to analyse the data collected by the experiment. For a more detailed description of the detector and its performance during Run 1 see [4, 5].

3.2.1 Tracking

The tracking system within the LHCb experiment consists of the vertex locator (VELO), a dipole magnet and the tracking stations. Together the sub-detectors provide precise information on the passage of charged particles through the detector and the particle momentum. The tracking detectors work on the principle that the passage of high energy charged particles through silicon or ionised gas causes the excitation or ionisation atoms in the material. The release of this energy is recorded and translated into an electrical signal that reveals the path of a particle.

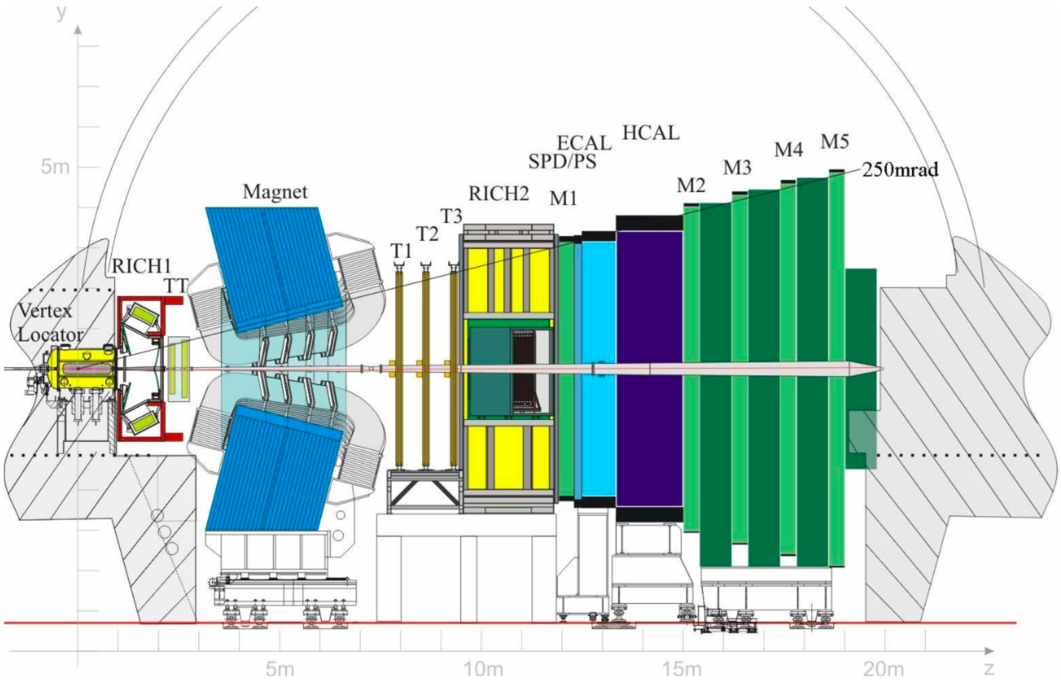


Fig. 3.3 Cross section of the LHCb detector [3].

3.2.1.1 The Vertex Locator

The vertex locator (VELO) is a silicon detector surrounding the interaction point. Its main goal is to provide precise information *pp* interaction vertices and secondary decay vertices of particles produced. Information the VELO provides enables precise measurements of particle lifetimes and impact parameters of particles tracks necessary for physics analyses.

The VELO is made of two identical halves, each half consists of 21 stations containing two silicon sensors arranged along the beam pipe. The two halves of the VELO slot together and there is a small gap in the centre for the beams to pass through. The arrangement of sensors along the z axis, shown in Figure 3.4, it designed so that the sensors cover the full LHCb acceptance and a charged particle within the detector acceptance will pass through at least three stations. In each station the two sensors measure different coordinates, one measures the r coordinates of charged particles and the other measures the ϕ coordinates as shown in Figure 3.5. The r, ϕ coordinates and the z placement of the sensors are used to reconstruct charged particle trajectories. Cylindrical coordinates were chosen to allow fast reconstruction of particle trajectories in the VELO.

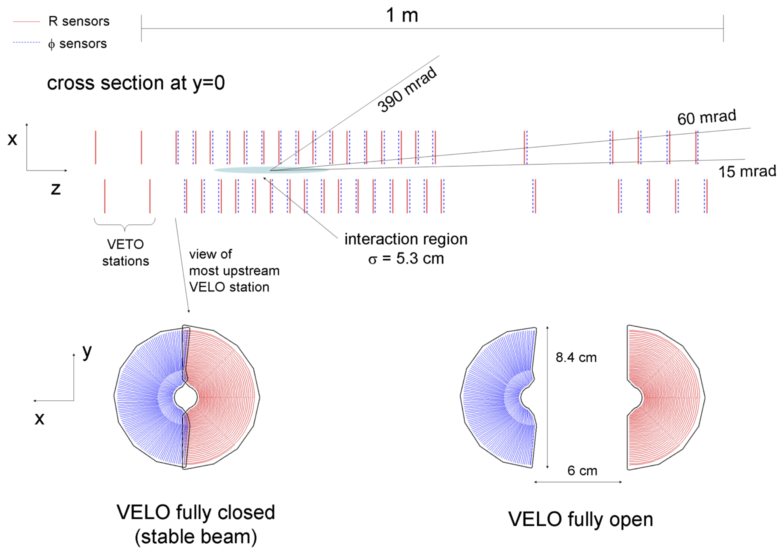


Fig. 3.4 The VELO layout and position of sensors along the beam axis [3].

The momentum resolution achievable for charged tracks by the LHCb experiment is limited by multiple scattering of particles as they travel through material in the detector. Therefore, to ensure good momentum resolution throughout the detector, the VELO is kept in a vacuum to reduce its material budget. Each half of the VELO is enclosed inside an aluminium box, which keeps it in a vacuum and shields the electronic readouts of the from radio frequencies generated by the beam. The overall material budget of the VELO comes to 17.5 % of a radiation length.

Excellent vertex resolution is required in the VELO, to acheive this the sensor need to be as close as possible to the interaction point. This is acheived by making the VELO out of two retractable halves and including the *pp* interaction point within the coverage of the VELO. During data taking, when the VELO is recording particle tracks the sensors are 8mm from the beam axis. However during the injection phase of the beam the width of the beam is much larger, therefore the halves of the VELO can retract to be 3 cm from the nominal beam axis. This keeps the VELO safe from unnecessary radiation damage. The two halves of the VELO are displaced by 150 mm in the z direction, as shown in Figure 3.4 so that when the VELO is closed, the sensors in each half overlap to help with detector alignment and reduced edge effects.

An additional purpose of the VELO is to identify high pile up events. There are 2 VELO sensors upstream of the interaction point that provide information to the trigger about how many *pp* interactions there were with each bunch crossing and this information can be used to identify events with high numbers of primary vertices.

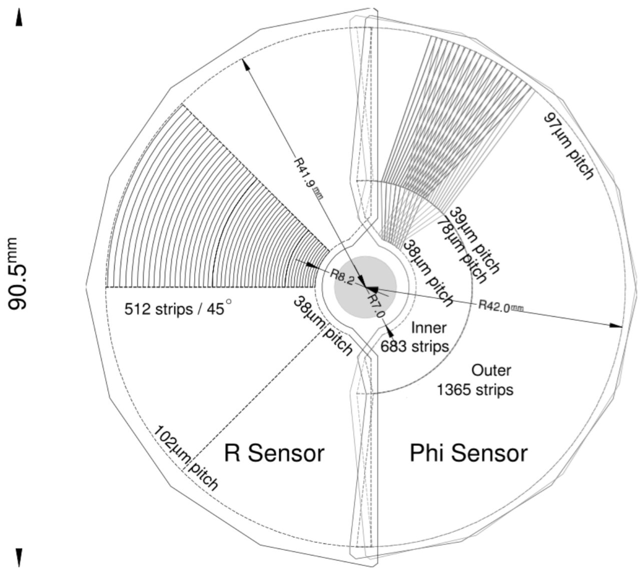


Fig. 3.5 Diagram of r and ϕ sensor layouts [3].

The VELO achieves a vertex resolution of 10 - 20 μm transverse to the z direction and 50 - 100 μm along the z direction, the resolution of each track depends on the number of tracks in each event as shown in Figure 3.6. The VELO also gives measurements on the impact parameters of particles tracks, which is the distance of closest approach between a particle track and the primary vertex. Figure 3.7 shows the IP resolution for 2012 data, for a track with transverse momentum of 1 GeV/c is has an impact parameter resolution of 35 μm.

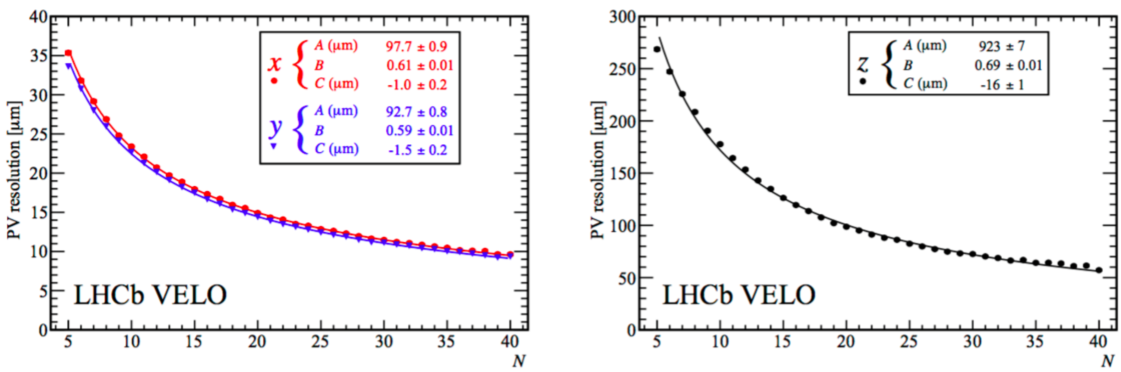


Fig. 3.6 Velo performance for primary vertex resolution perpendicular (left) and parallel (right) to the beam axis as a function of the number of tracks in an event for 2012 data [6].

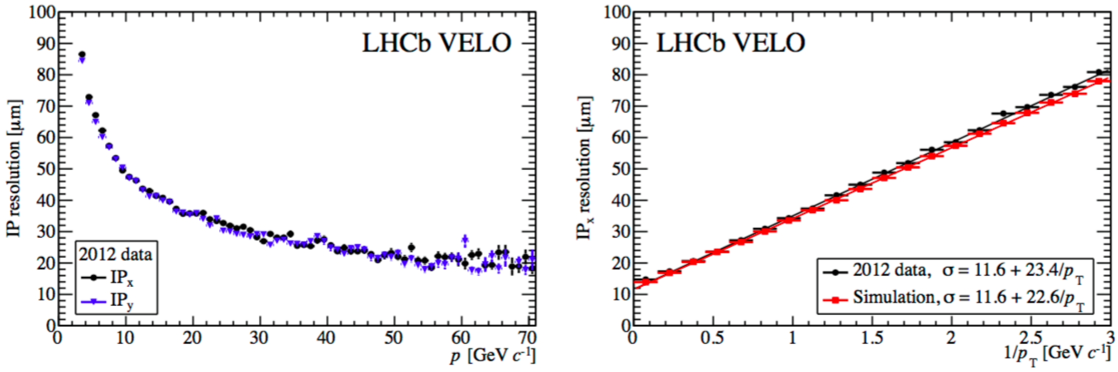


Fig. 3.7 Velo performance for impact parameter resolution as a function of momentum (left) and inverse transverse momentum (right) for 2012 data [6].

3.2.1.2 Tracking Stations

The LHCb experiment has 4 tracking stations in addition to the VELO, the Tracker Turicensis (TT) which is located upstream of the magnet and the T stations, T1-T3, located down stream of the magnet. These tracking stations provide complementary information to the VELO and the presence of the magnetic field allows the momentum of charged particles to be determined.

The TT is made up of 4 layers of silicon trackers spaced 27 cm apart that cover the full LHCb angular acceptance. The TT is located just within the influence of the magnetic field of the dipole magnet, which provides the detector with 2 main purposes. Firstly, the TT tracks the passage of charged particles with high momentum to enable good momentum resolution of tracks when the information is combined with that from other tracking stations. The TT has a resolution of 50 μm for a single hit, this resolution was chosen so that multiple scattering in the detector material rather than detector resolution is the limiting factor for the momentum resolution. The second purpose of the TT is to record tracks of low momentum particles that are then swept out of the detector acceptance as they continue through the magnetic field. These tracks will have a lower momentum resolution but help with pattern recognition within the RICH detectors.

The T stations, T1-3, are split into two sections, each composed of an Inner Tracker (IT) made of silicon and an Outer Tracker (OT) composed of straw drift tubes. There is a large increase in size of the tracking stations between the TT and the T3 so that all the detectors cover the full angular acceptance of the detector. The TT is 150 cm by 130 cm where as the T3 station is 600 cm by 490 cm, this is illustrated in Figure

3.8. The large size of the T stations meant that the high cost of silicon prevented it being used for the full coverage of each station.

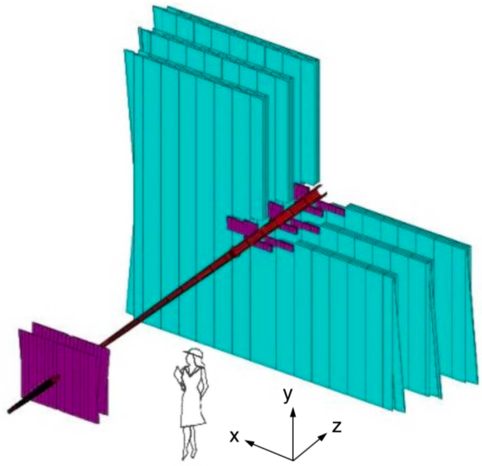


Fig. 3.8 Sizes of the TT and T stations [3].

The IT has very similar in design to the TT, each station is made of 4 layers of silicon trackers with an overall track resolution of 50 μm. The silicon trackers are arranged in a cross shape around the beam pipe, as shown in Figure 3.8, although the IT covers less than 2% of the T stations, 20% of tracks pass through it. This allows the occupancy of the OT to be less than 10% enabling a good overall track resolution from the OT despite it not being made of silicon. The OT of each tracking station is made of 2 staggered layers of straw tubes, they cover the remaining area required for cover the full LHCb angular acceptance which includes tracks bent by the magnetic field. The straw tubes have a fast drift time of 50 ns giving a better than 200 μm track resolution.

3.2.1.3 Dipole magnet

A warm dipole magnet is used to measure the momentum of charged particles travelling through the LHCb detector. In a magnetic field the trajectories of charged particles are bent and the particle momentum can be measured from the curve of the track.

The magnet is located between the TT and the T stations and its field covers the full LHCb acceptance. The field is in the vertical direction therefore bending tracks in the horizontal direction. The magnet was designed so that the field strength in the RICH detectors is negligible (less than 2 mT) and to have the largest strength possible between the TT and T stations. Figure 3.9 shows a plot of the magnet strength alongside the detector layout. A small magnetic field is achieved in the RICH detectors by iron shielding. The magnet was designed to have an integrated field strength is 4 Tm for track that travels 10 m through the detector.

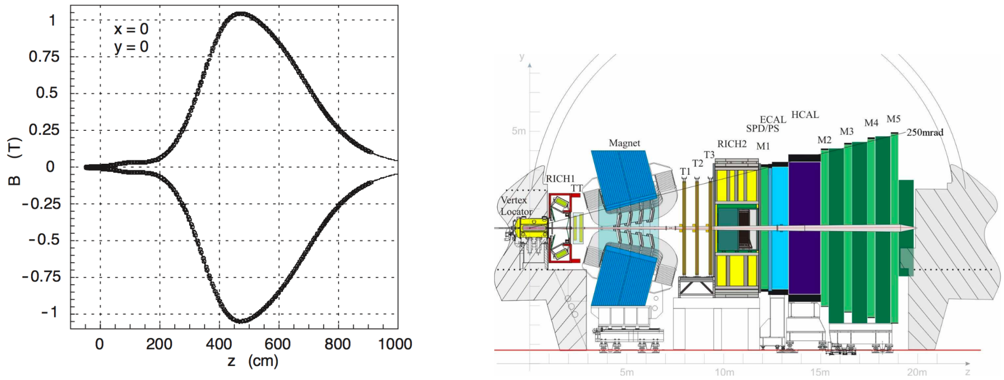


Fig. 3.9 Magnet field of the dipole magnet along the length of the LHCb detector (left) and the layout to the LHCb detector [3]. The peak strength of the field occurs between the TT and T1-3 station.

The polarity of the magnetic field is periodically switched so that is bends charged tracks in opposite directions. This is done so measure left-right detection asymmetries and to help understand systematic uncertainties of CP violation measurements.

3.2.1.4 Track reconstruction and performance

The information left by the passage of charged particles through the VELO, TT and T stations is combined using track reconstruction algorithms to find trajectories of charged particles though the length of the LHCb detector and the particle momentum. The algorithms start with either segments of tracks in the VELO or the T stations and extrapolate from these segments into the other tracking detectors in specific search windows. Once the segments of the track have been found the trajectory is fitted with a Kalman Filter with takes into account multiple scattering and energy loss within the detector. For each track the filter returns the χ^2 per degree of freedom, this is a measure of quality for the track. In LHCb this parameter is used to ensure that only good quality tracks are used in physics analyses. The reconstructed tracks are classified into five types depending on which detectors they travelled through, as shown in Figure 3.11.

The different track classifications are:

- **VELO tracks** are formed by particles produced at large angles to the beam axis or travelling in the negative z direction from the interaction point, these particles

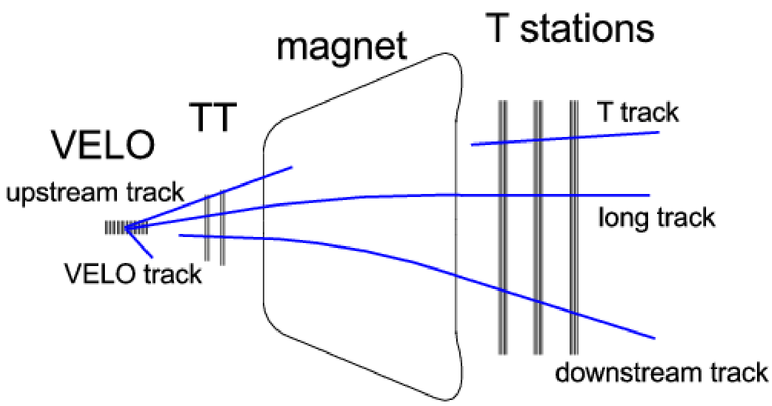


Fig. 3.10 Different types of tracks that are reconstructed at LHCb [7].

only leave tracks in the VELO. VELO tracks are useful for reconstructing primary vertices.

- **Upstream tracks** are made by low momentum particles that only leave hits in VELO and TT which are upstream of the magnet. The absence of tracks further down the detector is because the magnetic field sweeps the particles out of the detector acceptance. Upstream tracks have poor momentum resolution but are useful for understanding backgrounds and pattern recognition in the RICH 1 located between the VELO and the TT.

- **Downstream tracks** are produced by the decays of long lived neutral particles, that travel out of the VELO before decaying. These particles only leave tracks the TT and T stations.

- **T tracks** are tracks that only cross the T1-3 stations and are formed from particles created in interactions with the detector material. Similarly to upstream tracks, T tracks can help to understand backgrounds and pattern recognition in the RICH 2 located just before the T stations.

- **Long tracks** are the most useful for physics analyses because they are formed by particles that travel through the VELO, TT and T1-3 stations. Information from all the tracking stations is combined so these tracks have the best momentum resolution.

The efficiency to correctly reconstruct tracks varies with the particle momentum and the number of tracks present in an event, as shown in Figure 3.11 for 2012 data. In Run 1 long tracks were correctly reconstructed on average of 96 % of the time.

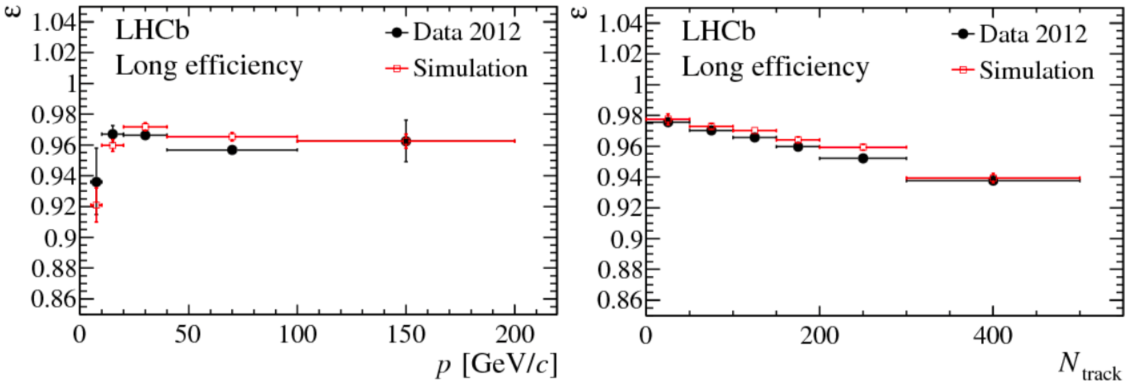


Fig. 3.11 Long track reconstruction efficiency as a function of momentum (left) and number of track in the event (right) for 2012 data [7].

Inevitably not all tracks that are reconstructed are correct, there are two main types of incorrectly reconstructed tracks. The first are clone tracks that occur when two tracks have many hits in common, when this happens the track with the highest number of total hits it used and the other is discarded. The second type of incorrect tracks are ghost tracks that are formed when track segments in different detectors are incorrectly joined together. This most offen occurs with segments in the VELO and T1-3 stations, the number of ghost tracks in an event depends on the event multiplicity. These tracks are removed by cutting on the output of a neural network that returns a probability of how likely a track is to be fake.

Once the tracks have been reconstructed, parameters that are necessary for the identifying and measuring different particles decays in an event can be computed from the tracks. The combined tracking systems achieve a momentum resolution of $\delta p/p = 0.5\%$ for particles with $p = 20$ GeV/c and a resolution of $\delta p/p = 0.8\%$ for particles with $p = 100$ GeV/c. This momentum resolution, when combined with vertex information from the VELO, gives a decay time resolution of around 50 ns.

3.2.2 Particle identification

In LHCb the particle identification (PID) detectors consist of two ring imaging Cherenkov (RICH) detectors, electromagnetic and hadronic calorimeters and muons stations. Together these detectors distinguish between different charged leptons and hadrons and between neutral particles such as photons and neutral pions. Good particle identification is necessary to determine which b-hadron decayed and to distinguish between topologically similar decays, such as $B^0 \to K^+\pi^-$, $B_s^0 \to K^+K^-$ and $B_{(s)}^0 \to \mu^+\mu^-$.

3.2.2.1 Ring Imaging Cherenkov detectors

RICH detectors are used at LHCb to distinguish charged hadrons and leptons that have a momentum between 2 and 100 GeV/c. The RICH detectors are vital to distinguish between pions, kaons and protons frequently produced in b-hadron decays. The energy range of the RICH detectors was chosen because the typical decay products of 2-body b-hadron decays is around 50 GeV.

The RICH detectors are based on the following principle; when a charged particle travels with velocity v through a dielectric medium with a refractive index n, the atoms excited by its passage are polarised, if the particle is travelling faster than the speed of light in the medium the excitation energy is released as a coherent wavefront. The angle, θ_c, the wavefront travels at relative to the particle trajectory depends on the speed at which the particle was travelling as $cos(\theta_c) = c/nv$. The light is produced in a ring and is called Cherenkov radiation. The angle at which Cherenkov radiation is produced gives a measurement of a particle's speed which when combined with the particle's momentum, the particle mass and consequently its identity can be determined. However many particles travel through the RICH detectors and create overlapping rings of light making particle identification complex. Particle trajectories through the RICH detectors are inferred from information in the tracking stations and the expected pattern of Cherenkov radiation is calculated for each possible particle type. The expected patterns of light are compared to the observed pattern to find the likelihood for each particle type, all possible particle types are compared to maximise the likelihood. An in depth description of the reconstruction algorithm used in the RICH detectors can be found in [8].

The two RICH detectors cover complimentary momentum regions. The RICH 1 detector is located between the VELO and the TT station, it covers the full LHCb angular acceptance and provides PID information on particles in the momentum range 1 to 40 GeV/c. The RICH 1, is illustrated in Figure 3.12, it contains two different radiator materials; at the front of the detector is a aerogel sensitive to particles with a momentum between 2 and 10 GeV/c, behind the aerogel is a gas radiator sensitive to particles in the momentum range 10 to 40 GeV/c. The aerogel radiation was removed after Run 1, therefore the RICH 1 is only sensitive to particles in the momentum range 10 to 40 GeV/c in Run 2. As charged particles travel through the RICH 1, the rings of light produced are focused by spherical mirrors onto Hybrid Photon Detectors (HPDs),

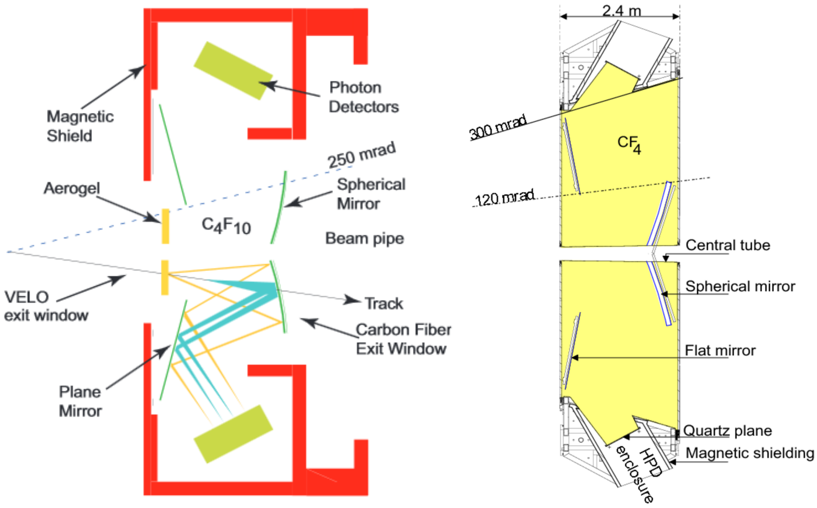


Fig. 3.12 Diagram of the RICH1 detector (left) and the RICH 2 detector (right) [3]. For Run 2 the aerogel radiator in the RICH 1 detector was removed.

the radii of the detected rings provides information about how fast the particle was travelling.

The RICH 2 detector is located upstream of the RICH 1, between the last tracking station and the first muon station. The RICH 2 consists of a gas radiator sensitive to particles with a momentum range 15 - 100 GeV/c and the detection of the light produced is the similar to the RICH 1 as illustrated in Figure 3.12. Unlike the RICH 1, the RICH 2 detector does not cover the full LHCb angular acceptance but only $\pm$ 120 mrad in the horizontal and $\pm$ 100 mrad in the vertical direction. This area contains the higher momentum particles the RICH 2 is sensitive to, the low momentum particles have been bent out of the acceptance by the magnetic field.

Both RICH detectors use HPDs that are sensitive to magnetic fields, the HPDs are shielded from the magnet field using iron sheets ensuring the field is less than 2mT across them. This allows accurate detection of light created within the RICH detectors.

The rings of light collected by the RICH detectors when combined with information about particle momentum and tracks from the tracking stations enables the particle type to be identified. Figure 3.14 shows how the Cherenkov angle and momentum can be combined to identify different types of particles in the RICH 1 detector, there are distinct bands for each particle mass. Figure 3.13 shows what is expected for the different radiators.

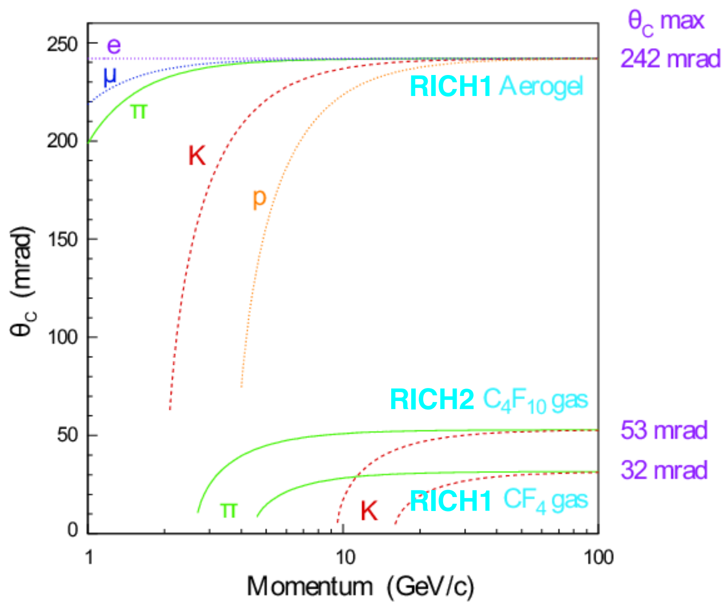


Fig. 3.13 Expected Cherenkov angles produced by different particles travelling through the radiators in the RICH detectors [3].

3.2.2.2 Calorimeters

The calorimeter system consists of the Scintillating Pad Detector (SDP), Pre-Shower (PS), electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). Information from the calorimeters is used to identify electrons, photons and hadrons with high transverse momentum to be used in the first level of the trigger and to help with the reconstruction and identification of these particles. The ECAL is the only part of the LHCb detector that measures the position and energy of photons and neutral pions.

The calorimeters in LHCb are sampling calorimeters that consist of layers of lead absorbers and scintillating material. In lead, incident particles create showers of secondary particles, the charged particles produced in the absorbers create light as they pass through the scintillators. The light travels through wavelength shifters where it is collected by photon multiplier tubes and turned into an electrical signal. In the ECAL showers are started by ionisation, bremsstrahlung radiation or pair production depending on the energy of the incident particle and whether is it a $e^{\pm}$ or a photon. In the HCAL it is interaction via the strong force that leads to showers of secondary particles. The showers produced in the calorimeters are along the direction of flight of the incident particle. Unlike other sub-detectors in LHCb, the calorimeters change the particle as it moves through the detector in order to measure the energy.

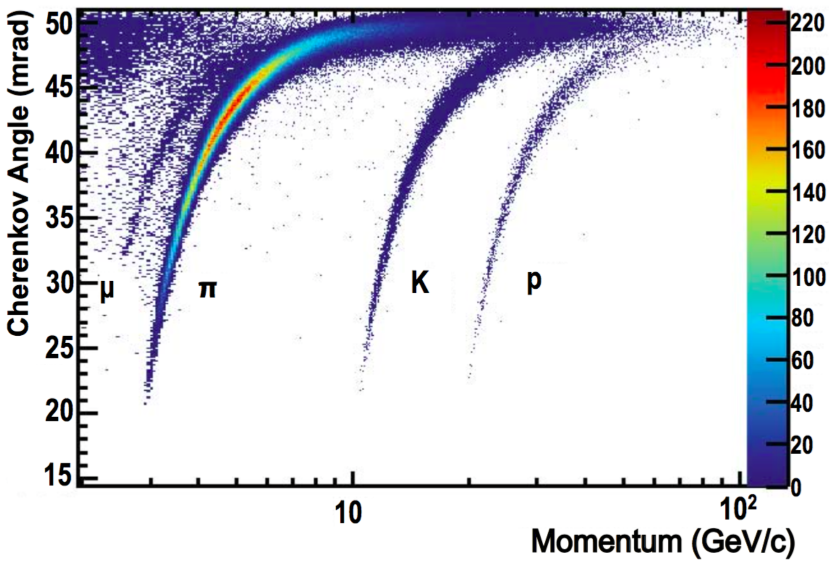


Fig. 3.14 Cherenkov angles for isolated tracks as a function of momentum in the RICH 1 detector for 2011 data [9].

The SPD, PS and ECAL identify electrons, positrons and photons. The SPD is a layer of scintillating material at the start of the calorimeter system, it separates electron and photon showers created later in the calorimeter because only charged particles will produce light in the SPD. Next in the calorimeter system is the PS, it consists of a lead absorber followed by another scintillator similar to the SPD, the length of the lead absorber is chosen so that electrons will start showers in the absorber but charged pions will not. There is only a 1% chance of a pion creating shower in the PS. Information collected by the PS enables showers created by pions in the ECAL to be separated from those created by electrons and positrons. The ECAL is designed to contain the entire shower of high energy photons so that it can provide good energy resolutions of photons passing through the detector. The ECAL has an energy resolution of $\delta E/E = 9\%/\sqrt{(E)} \oplus 0.8\%$ provided information from the PS and SPD are used.

The HCAL is predominately designed for use in the trigger and there is no requirement that the HCAL contains the full hadronic showers, therefore it was designed with a lower energy resolution of $\delta E/E = 69\%/\sqrt{(E)} \oplus 9\%$.

3.2.2.3 Muon stations

The muons stations are designed to identify highly penetrating muons, for use in the trigger and offline analyses. Muons are produced in many b-hadron decays, good muon identification is necessary trigger events containing muons and to distinguish topologically similar decays such as $B^0_{(s)} \to \mu^+\mu^-$ and $B^0 \to K^+\pi^-$ in physics analyses. Compared to other particles muons have a high penetrating power due to their relatively large mass and because muons do not interact via the strong force, these properties are exploited in the muon detectors.

There are 5 muon stations, M1-5, shown in Figure 3.15 that track and identify muons. The first muon station is located before the calorimeters, the inner section where the fluence is greatest, is made of gas electron multiplier foils and the outer section is made from multiwire proportional chambers (MWPCs). Stations M2-5 are located after the HCAL, by which point most other particles have been absorbed by the calorimeters. These stations are made from MWPCs and between each station is 80cm of lead absorber ensuring only high energy muons pass through the muon detector. A muon must have a momentum of at least 3 GeV/c to pass through the calorimeters and the M2 and M3 stations, to travel through all the muons stations a muon must have a momentum of 6 GeV/c. The first 3 stations have a high spatial resolution and provide track and transverse momentum information to be used the the trigger. M1 is located before the calorimeters to improve the transverse momentum measurement of the muons. The last two stations have a lower spatial resolution and are designed to identify muons with the greatest transverse momentum. After the muon stations there is an iron wall to stop any particles from travelling downstream of the detector. The size of the muon stations increases with distance from the interaction point to ensure the full angular acceptance of the detector is covered. Tracking information collected in the muon stations can be used in the trigger because the stations lie outside the magnetic field which allows for fast reconstruction of the tracks and a muons.

3.2.2.4 Particle identification and performance

The information collected in the PID detectors is combined to provide several discriminating variables that can be used to identify muons, protons, kaons, pions and electrons.

The muon stations are used, along with information from the tracking system, to produce a binary selection (isMuon) to identify muons. The tracking system is used to extrapolate a field of interest within the muon stations, a muon is identified if hits in the muon stations can be combined with those from the tracking system within the field of interest. The number of the hits required in the muon stations depends on the momentum of the muon. Muons with momentum in the range $3<p<6$ GeV must leave hits in M2-3, those in the momentum range $6<p<10$ muon leave

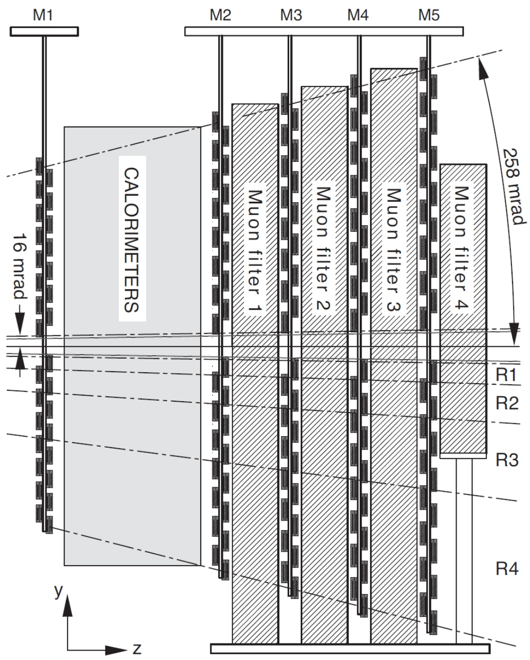


Fig. 3.15 Layout of the muon stations [3].

hits in M2-3 and either M4 or M5 and finally muons with momentum above 10 GeV must be observed in all the muon stations. Figure 3.16 shows the efficiency for the isMuon selection at selecting muons and probabilities of mis-identifying hadron are muons. The efficiencies and mis-identification probabilities are computing using the *tag and probe technique*, this technique uses two tracks from a decay and particle identification requirements are applied to one track, the tag track, and the other track, the probe track, is used to evaluate the efficiency or mis-identification probability. The muon efficiency uses $J/\psi \to \mu^+\mu^-$ decays, proton mis-identification probabilities are computed using $\Lambda^0 \to p\pi^-$ and pion and kaon mis-identification probability are computed from $D^{*+} \to \pi^+ D^0(\to K^-\pi^+)$ decays. The mis-identification rate is higher for lower momentum particles, which is expected given there are less hits in the muons detectors. The main contribution to misidentifying hadrons as muons comes from the kaons and pions that decay in flight, the muons from these decays are then detected in the muon stations.

The information from all the PID detectors is combined using two different methods to provide global particle identification variables. One method is based on likelihood fits and the other is based on Neural Networks. In the first method, likelihood fits are performed in each sub-detector comparing each charged particle track to different particle hypotheses. The information from the likelihood fits in each sub detector are

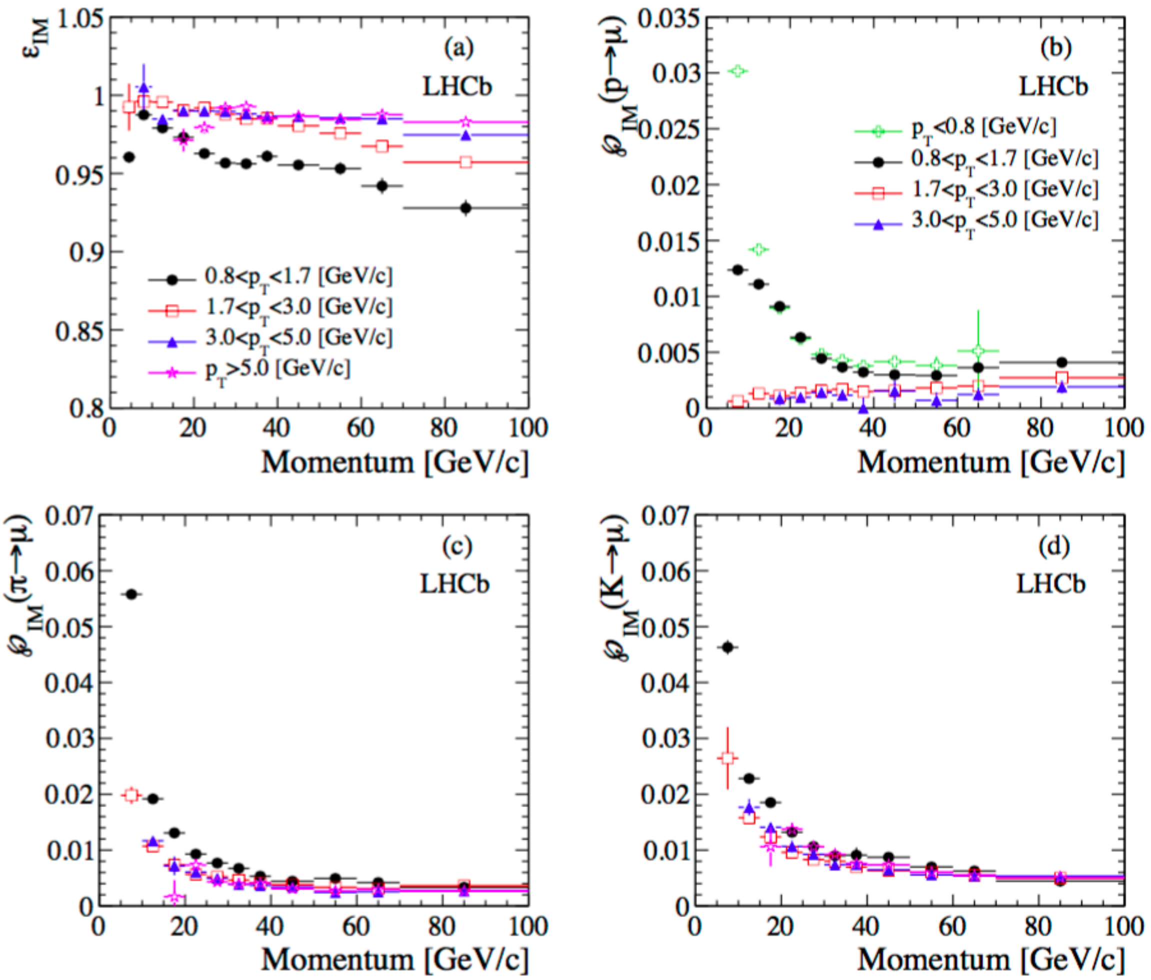


Fig. 3.16 Muon efficiency (top left) and misidentification probabilities for protons (top right), pions (bottom left) and kaons (bottom right) for isMuon criteria [10].

combined into a global variable. The final variable is the difference in the log-likelihoods between the track corresponding to a pion and a different particles hypothesis (kaon, proton, muon, electron), giving a measure of how likely each particle hypothesis is compared to that of a pion. These variables are known as DLL variables where the difference in log-likelihoods between the track corresponding to a pion and a kaon would be given by $\mathrm{DLL}_{K\pi}$.

The second method uses information from the PID detectors and the tracking system in Neural Networks to provide a global probability of a track having a particular particle hypothesis. This method takes into account correlations between detector systems and extra detector information that are not considered in the likelihood method. The Neural Networks are trained on simulated inclusive b decays and can be tuned to suit different situations, such as the data taking year. The variables produced by the Neural Networks are known as ProbNN variables where the probabilty of a particle being a muon is given by ProbNNμ and the probabilty of a particle being a pion is given by ProbNNπ.

Figure 3.17 shows a comparison of the performance of the DLL and ProbNN variables in selecting protons and muons. Although the performance to the two types of variables are quite different, the efficiencies of each variable varies with different kinematic properties of the decay. The most appropriate PID variable type to use depends on the physics analysis it is being used in.

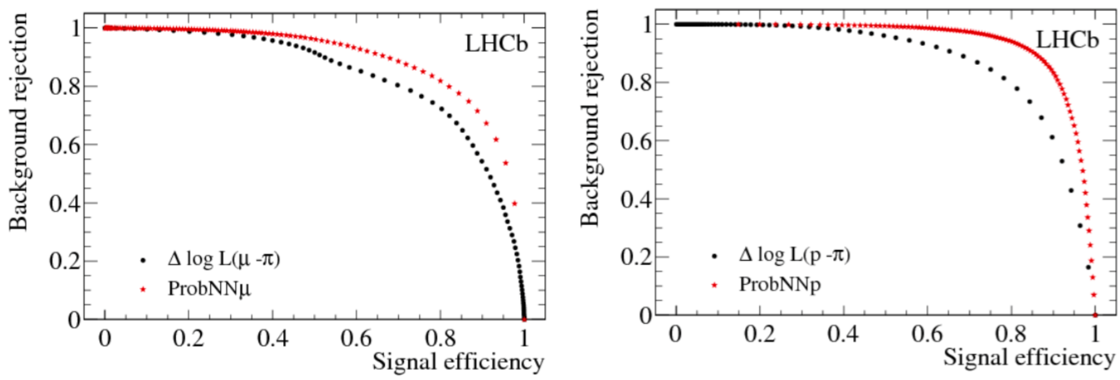


Fig. 3.17 Muon (left) and proton (right) signal efficiency vs background rejection for DLL and ProbNN PID variables [3].

3.2.3 Trigger

The LHC was designed to collide protons at a rate of 40 MHz, this rate is too high for information to be read out of the LHCb detector. However most *pp* collisions do not produce particles within the detector acceptance that are interesting for physics analyses at LHCb. A trigger system is used to identify *pp* collisions that contain potentially interesting physics processes, the information from these events are saved for later use in physics analyses. The trigger has been designed to select interesting physics events with a high efficiency whilst reducing the event rate to one where information from the full detector can be read out. There are two levels to the LHCb trigger; the hardware trigger and the software trigger. The hardware trigger is know as the level-zero (L0) trigger and reduces the 40 MHz collision rate to 1 MHz at which the full detector can be read out. The software trigger is know as the High-Level-Trigger (HLT), it has two stages and runs on the output of the L0 further reducing the event rate by utilising information for all the detector sub-systems. Each level of the trigger is composed of trigger ‘lines’; these lines are made up of reconstruction and selection algorithms and either accept or reject each event. Only events that are acceptance by a trigger line at both the L0 and HLT are available for use in physics analyses.

3.2.3.1 L0 trigger

The L0 trigger runs synchronously to the LHC bunch crossing. Its purpose is to reduce the events rate to 1 MHz, where information from the full detector can be read out. Therefore the L0 is limited to use information from the detector that can be read at the same rate as the LHC collision rate. The L0 uses information from 3 parts of the detector, the VELO, calorimeters and the muon stations, to make decisions about the relevance of each event.

The pileup veto stations in the VELO are used in L0 pileup trigger lines, these lines identity the number of collisions in an event and are predominatley used for luminosity measurements [11].

The other L0 trigger lines are based on the kinematic properties of b-hadron decays. The heavy masses of b-hadrons means that their decays are characterised by the production of daughter particles with large transverse momentum (p_T) and transverse energy (E_T). The calorimeters are used in trigger lines that select events containing high E_T electrons, photons or hadrons. Information from the PS, SPD, ECAL and HCAL is used to identify electrons, photons and hadrons in each event. Events are then accepted by the trigger lines if there is an electron, photon or hadron with E_T above a threshold value provided the event multiplicity is not too high. The E_T thresholds are different for each particle type. Events with high multiplicity take a long time to reconstruct and process in the HLT, therefore it is not efficiency to keep these events. The multiplicity is measured by the number of hits in the SPD detector (nSPD), only events with nSPD lower than a specified value can pass an L0 trigger line.

In a similar way to the calorimeters, the muon stations are used to identify muons with high p_T for trigger lines. There are two L0 trigger lines for muons that accept events based on muon p_T if either a single muon has a p_T above a threshold value or if the two muons with this highest p_T have $\sqrt{p_{T1} \times p_{T2}}$ above a threshold value, provided the event multiplicity is not too high.

The E_T and p_T thresholds and the multiplicity limit for the L0 trigger lines vary for each year of data taking depend on the bandwidth available for the trigger.

3.2.3.2 HLT trigger

Events that are accepted by trigger lines in the L0 are moved to the Event Filter Farm where the HLT is run. The HLT is a software trigger that is split into two levels that are run successively. During the long shut down between Run 1 and Run 2 of the LHC significant changes were made to the reconstruction of particle decays used to make decisions within the HLT.

The HLT1 is the first level of the HLT. It runs on the output of the L0 checking the decisions make by the L0 trigger lines and reducing the event rate. The HLT1 trigger lines are composed of generic selection criteria, making decisions that confirm those made in the L0 about particular particle types and also identify generic types of particle decays such as inclusive b-hadron decays. The second level of the HLT, the HLT2, runs on the output of the HLT1 trigger and consists of trigger lines designed to select decays relevant to specific physics analyses or particle decay topologies.

During Run 1 time constraints in the HLT1 trigger to process the output of the L0 did not allow for full event reconstruction using all LHCb sub-detectors, instead the HLT1 ran reconstruction and selection algorithms on event information only from the VELO and tracking stations. The reduced output of the HLT1 then provided an event rate that was low enough to allow event reconstruction that includes all detector subsystems to be used in the HLT2. However the reconstruction used in the HLT2 was different to the offline reconstruction that is used in physics analyses. Significant changes were made in the reconstruction used in the HLT between Run 1 and Run 2, the details of the changes made can be found in [12]. The majority of the changes to the HLT for Run 2 are not relevant for the analysis discussed in this dissertation, but the overall change is that the same reconstruction is used in the HLT and the offline reconstruction.

Just like the L0 trigger, trigger lines in the HLT vary for each year of data taking both the selection criteria used in the lines and also new trigger lines are introduced. The number of HLT2 lines increases with each year of data taking as understanding of the capabilities of the experiment increases; there were about 100 HTL2 lines in 2011, 200 in 2012 and 450 in 2015.

3.2.4 Software and simulation

The data that is read out of the LHCb experiment needs further processing before it can be used in physics analyses. The GAUDI framework [13] is a C++ framework that is the basis for the software applications needed to process the data at LHCb [14]. This framework ensures that the necessary software is available to all users and changes to the software are implemented across all applications, it is suited to the distributed computing system used in LHCb [15].

Once events have been accepted by the trigger, the first step in processing the output of the detector is reconstructing events, this is done by the Brunel application. It takes the digitised detector read out and reconstructs hits in the tracking stations to find particle trajectories and momenta and combines information from the RICH detectors, calorimeters and muon stations to compute PID variables. The output of processing by the Brunel application are stored in 'Data Summary Type' (DST) files.

Next the DaVinci application is used to fit the tracks reconstructed in Brunel with primary and secondary vertices. This application assigns particle hypotheses to each track and reconstructs the decay trees of particles in the detector, computing the kinematic properties that are needed for physics analyses. The the reconstructed output of the trigger is too large to be stored in one place and to be used by all analysts, therefore a 'stripping' procedure is used to break up the data into a manageable size for each physics analysis. Each physics analysis designs a set of loose selection requirements, called stripping lines, specific to their decays of interest, the selections are applied centrally to the reconstructed events and are designed to keep as much of the signal relevant to the analysis as possible but reduce the number background events. Only events that pass a stripping line selection are avaliable to be used in physics analyses. The output of this process are smaller DST files, events passing the stripping selections can either be saved with the full event information or with just the tracks related to the signal candidate. The choice depends on the physics process the stripping line is relevant for. The stripping selection is run a limited number of times and is applied seperately to data collected in different years. Requirements are imposed on the amount of data each stripping line can retain, typically the output of a line must be less than 0.05 % of the original data set size if the full event information is saved. Each analyst then uses the DaVinci application one last time to produce Root [16] files from the output of their stripping lines, these files display the data in histograms and are used for physics analyses.

As well as data collected by the experiment, simulated data that mirrors what is expected in the experiment is needed to understand the detector performance and for physics analyses. There is a set of software applications that are dedicated to the production of Monte Carlo simulated events within the Gaudi framework. Events are generated using the Gauss application [17, 18], this package uses Pythia [19, 20] to model *pp* collisions and the production of particles, then the EvtGen [21] application to calculate the decays of these particles. Final state radiation is modelled using PHOTOS [22]. Both PYTHIA and EVTGEN have been tuned for the production and decay of particles within the LHCb detector. The GEANT4 [23, 24] toolkit is used to model the interaction of particles as the travel through the LHCb sub-detectors and the hits made by particles in the detector. In the simulation the type of particles generated and how they decay can be specified so that the simulated events are relevant to particular physics decays. The BOOLE application then produces the digitised detector read out based on the information from GEANT4 that mimics the detector read out when data is recored. The output of BOOLE encompasses the detector response to the different hits, the electronic read out and the L0 hardware trigger as well as including additional hits from event spillover and LHC backgrounds. The digitised response of the detector is then processed by BRUNEL and DAVINCI in the same way as the real data to produce ROOT files that are used in physics analyses.

The LHCb software framework is set up so that it can be used on the Worldwide LHC Computing Grid [25, 26], the Grid is made up of computers across the world that each store part for the LHCb data set and simulation data. Despite the stripping process the data produced at LHCb is too large to be stored in one place. The DIRAC [27] system manages grid sites and the GANGA project allows the submission analysis code to different grid sites. The grid enables analysts to process and study the large amounts of data produced by LHCb without having to store the data where the analyst is.

3.3 Summary

The data taking periods of the LHC can be split up into different ‘Runs’ which are separated by Long Shut Down periods when maintenance and upgrades are performed on the LHC, the detectors and the accelerator chain that delivers protons to the LHC. Run 1 began in 2010 and ended in 2013, during this Run the LHC operated at two different centre-of-mass energies. In 2010 and 2011 the LHC delivered proton collisions a a centre-of-mass energy of 7 TeV, this was increased to 8 TeV in 2012. The luminosity recored by LHCb in each was; 0.04 fb^{-1} in 2010, 1.10 fb^{-1} in 2011 and 2.08 fb^{-1} in 2012. After Run 1 the LHC entered the Long Shutdown 1 (LS1) when the machine and experiments were prepared to deliver and detect proton collisions at $\sqrt{s} = 13$. Run 2 began in early 2015 and is still on going, so far LHCb has recorded 0.32 fb^{-1} in 2015 and 1.67 fb^{-1} in 2016 both at a centre-of-mass energy of 13 TeV. Figure 3.18 shows the integrated luminosity collected by LHCb in each year of data taking. The recorded luminosity of Run 2 is currently less that what was recorded in Run 1, however the production cross section for b-hadrons approximately doubled with the increase in centre-of-mass energy between Run 1 and Run 2 therefore the Run 2 data set will already contain more b-hadrons useful for physics analyses than the Run 1 data set.

The expected end of Run 2 is 2018 by which time LHCb is expected to have recorded 5 fb^{-1} luminosity during the Run. Run 2 will be followed by a second long shut down period (LS2) in which LHCb shall be upgraded ready to record proton collisions at 14 TeV during Run 3. This run of data taking is expected to be from 2021 - 2024 and by the end of Run 3 LHCb is expected to have collected an integrated luminosity of 23 fb^{-1} over all the runs.

The physics analysis described in this thesis uses the full data sets from Run 1 and 2015 and data taken up to September during 2016. The 2016 data set is therefore reduced to 1.1 fb^{-1}.

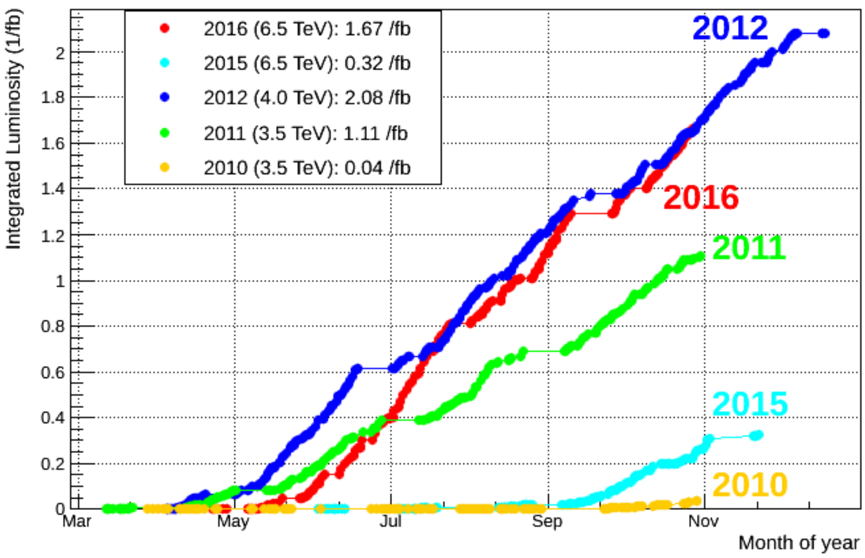


Fig. 3.18 Integrated luminosity collected by the LHCb experiment in each year of data taking. Source: LHCb.

Chapter 4

Event selection

This chapter describes the selection criteria used to identify $B^0_{(s)} \to \mu^+\mu^-$ candidates in data to measure the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions and $B^0_s \to \mu^+\mu^-$ effective lifetime. As well as identifying $B^0_{(s)} \to \mu^+\mu^-$, several other decays are needed, the branching fraction and effective lifetime measurements both require $B^0_s \to J/\psi\phi$ and $B \to h^+h^-$ decays, where $h = K, \pi$ and the branching fraction analysis also uses $B^+ \to J/\psi K^+$ decays.

Although $B^0_s \to \mu^+\mu^-$ decays leave a clear two muon signature in the detector, the identification of these decays is challenging because $B^0_{(s)} \to \mu^+\mu^-$ decays occur very rarely and there are many other processes that can mimic a $B^0_{(s)} \to \mu^+\mu^-$ decay in the detector, creating backgrounds in the data. The different background sources present in the data set are discussed in Section 4.1. The development of the selection of $B^0_{(s)} \to \mu^+\mu^-$ decays and analysis strategies to study these decays uses information from simulated particles decays, the details of the simulated decays are given in Section 4.2. The selection criteria are different for the two analyses. The selection for the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction analysis has been continuously developed over many years and the latest version is detailed in Section 4.3. This selection has been adapted for the measurement of the $B^0_s \to \mu^+\mu^-$ effective lifetime, the changes in the selection are given in Section 4.4.

The LHCb collaboration has published a number of papers studying the $B^0_s \to \mu^+\mu^-$ decay, the selection described in this Chapter has been built up over a number of years by a range of different collaboration members. The selection detailed in Sections 4.3.2.2, 4.4 were completed for this thesis as well as all Figures and quoted efficiencies.

4.1 Backgrounds

The reconstruction process (Sect. 3.2.4) produces numerous $B^0_{(s)} \to \mu^+\mu^-$ candidates from pairs of muons created during pp collisions. Some candidates will have come from real $B^0_{(s)} \to \mu^+\mu^-$ decays but there are other background processes that occur during pp collisions which leave a signature in the detector that can be reconstructed incorrectly as a $B^0_{(s)} \to \mu^+\mu^-$ decay. The selection aims to separate real $B^0_{(s)} \to \mu^+\mu^-$ decays from the backgrounds to produce a set of $B^0_{(s)} \to \mu^+\mu^-$ candidates with a high signal purity from which the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction and $B^0_s \to \mu^+\mu^-$ effective lifetime can be measured. The main background sources that mimic $B^0_{(s)} \to \mu^+\mu^-$ decays are:

- Elastic collisions of protons that produce a pair of muons via the exchange of a photon, $pp \to p\mu^+\mu^-p$. The protons travel down the beam pipe and are undetected leaving the muons to be reconstructed as $B^0_{(s)} \to \mu^+\mu^-$. Typically the muons produced in this way have low transverse momentum.

- Inelastic proton collisions that create two muons at the primary vertex. The muons form a good vertex and can be combined to form a $B^0_{(s)}$ that decays instantaneously. This type of background is prompt combinatorial background.

- $B^0_s \to \mu^+\mu^-\gamma$ decays where the photon is not reconstructed. The presence of the photon in the decay means that $B^0_s \to \mu^+\mu^-\gamma$ decays are not helicity suppressed and could therefore be a sizable background, however the photon gains a large transverse momentum resulting in the reconstructed $B^0_{(s)}$ mass being much lower than expected.

- Random combinations of muons produced by separate semi-leptonic decays. The $B^0_{(s)} \to \mu^+\mu^-$ candidates formed in this way are long lived combinatorial background because the reconstructed $B^0_{(s)}$ will not decay instantaneously.

- Semi-leptonic decays where one of the decay products is mis-identified as a muon and/or is not detected. The resulting mass of the $B^0_{(s)}$ candidate is lower than expected due to the missing particle information. The semi-leptonic decays that contribute to $B^0_{(s)} \to \mu^+\mu^-$ backgrounds in this way are $B^0 \to \pi^-\mu{+}\nu_\mu$, $B^0_s \to K^-\mu{+}\nu_\mu$, $\Lambda^0_b \to p\mu^-\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$ and $B^+_c \to J/\psi\mu^+\nu_\mu$ where $J/\psi \to \mu^+\mu^-$.

- $B \to h^+h^-$ decays, where $h = K, \pi$, when both hadrons are mis-identified as muons. This usually occurs when the hadrons decay whilst travelling through the detector. Similarly to mis-identified semi-leptonic decays the reconstructed $B^0_{(s)}$ candidate mass is lower than expected.

The separation of $B^0_{(s)} \to \mu^+\mu^-$ decays from the backgrounds is challenging because $B^0_{(s)} \to \mu^+\mu^-$ decays are highly suppressed decays therefore reconstructed candidates are predominately made from background decays. The removal of some background decays is straight forward by taking advantage of obvious differences between the $B^0_{(s)} \to \mu^+\mu^-$ and the backgrounds, however background from mis-identified semi-leptonic and $B \to h^+h^-$ decays and long lived combinatorial background are more challenging to remove.

4.2 Simulated particle decays

Simulated particle decays, as described in Section 3.2.4, are used to develop the selection and analysis of $B^0_{(s)} \to \mu^+\mu^-$ decays. Large clean samples of simulated decays are needed to separate signal decays from background decays and to understand the impact of selection criteria on decays present in data. Many different simulated decay types have been used for the development of the selection and analysis of $B^0_{(s)} \to \mu^+\mu^-$ decays, however only the simulated decays used for studies performed for this thesis are listed in Table 4.1 along with the data taking conditions and simulation versions used to generated the decays.

There exist multiple versions of the simulation because it is updated as understanding of the detector increases and to incorporate differences in data taking conditions, such as the trigger lines or center-of-mass energy, used each year data is collected. Similar simulation versions must be used to compare different types of simulated decays or data taking conditions so that differences are not masked by variations in the simulation of the decays.

Simulated $b\bar{b} \to \mu^+\mu^- X$ decays are used to understand the combinatorial background of $B^0_{(s)} \to \mu^+\mu^-$ decays, however producing a large enough sample of these decays to be useful is computational expensive and produces large output files to save generated decays. Therefore cuts are applied at the generation level for $b\bar{b} \to \mu^+\mu^- X$ decays to reduce the size of the samples that are saved and to speed production. The cuts, listed in Table 4.1, are applied on the muon momenta, the reconstructed mass of the muon pair, the product of the momenta of the muons and the distance of closest approach of the two muon.

On the whole simulated decays accurately model what occurs in data, however there are a couple of areas where the simulation falls short of reality. The distributions of

Decay	Data taking conditions	Simulation version	Generated events
	Stripping selection studies selection		
$B_s^0 \to \mu^+\mu^-$	2012	sim06b	2 M
$B^0 \to \mu^+\mu^-$	2012	sim06b	2 M
$B^0 \to K^+\pi^-$	2012	sim06b	1 M
$B^+ \to J/\psi K^+$	2012	sim06b	1 M
	Multivariate classifier training		
$b\bar{b} \to \mu^+\mu^- X$, p > 3 GeV/c, $4.7 < M_{\mu^+\mu^-} < 6.0$ GeV/c^2, DOCA < 0.4mm, $1 <$ PtProd < 16 GeV/c	2012	sim06b	8.0 M
$b\bar{b} \to \mu^+\mu^- X$, p > 3 GeV/c, $4.7 < M_{\mu^+\mu^-} < 6.0$ GeV/c^2, DOCA < 0.4mm, PtProd > 16 GeV/c	2012	sim06b	6.6 M
$B_s^0 \to \mu^+\mu^-$	2012	sim06b	2 M
	Analysis method development		
$B_s^0 \to \mu^+\mu^-$	2011	sim08a	0.6 M
	2012	sim08i	2 M
	2015	sim09a	2 M
	2016	sim09a	2 M ?
$B^0 \to K^+\pi^-$	2011	sim08b	0.8 M
	2012	sim08g	8.6 M
	2015	sim09a	4 M
	2016	sim09a	8.2 M
$B_s^0 \to K^+K^-$	2012	sim08g	7.2 M
	2015	sim09a	4 M

Table 4.1 Simulated decays used for developing the selection and the analysis of $B^0_{(s)} \to \mu^+\mu^-$ listed according to the studies the decays are used in. Cuts are applied to $b\bar{b} \to \mu^+\mu^- X$ decays as they decays are generated, these cuts are included alongside the decay type and are applied to the muon momenta, invariant mass of the muons, the distance of closest approach of the muons and the product of the transverse momenta of the muons.

particle identification variables and properties of the underlying proton-proton collision, such as the number of tracks in an event, are not well modelled in the simulation. The mis-modelling of particle identification variables can be corrected for using the PIDCalib package [28] and simulated decays can be re-weighted using information from data to accurately model the under lying event, this re-weighting is described in Section 6.3.1.

4.3 Selection for branching fraction measurements

The selection of $B^0_{(s)} \to \mu^+\mu^-$ decays occurs in several steps, the first step is choosing what requirements to place on the trigger (Sect. 4.3.1) which is followed by a cut based selection to remove obvious background events (Sect. 4.3.2. Then particle identification variables (Sect. 4.3.3) are used to reduced backgrounds from mis-identified semi-leptonic and $B \to h^+h^-$ decays, finally multivariate classifiers (Sect. 4.3.4) are used as the last step in the selection to reduced the backgrounds to a low enough level so that the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions can be measured.

The measurement of the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ branching fractions are described in Chapter 5 and requires $B^+ \to J/\psi K^+$ and $B \to h^+h^-$ decays to determine the branching fractions from the observed number of $B^0_{(s)} \to \mu^+\mu^-$ decays in data. The selection criteria for these decays are kept as similar as possible to the selection of $B^0_{(s)} \to \mu^+\mu^-$ decays and will be described alongside the signal selection. Furthermore $B^0_s \to J/\psi\phi$ decays used to verify steps of measurement process.

4.3.1 Trigger requirements

The trigger, described in Sect. 3.2.3, is the first step in the selection, it selects events that could contain an interesting particle decays and these events are saved to be used in physics analyses. Candidates from different particle decays are reconstructed from events that have passed the trigger. For each candidate it is useful to know whether is was a component in that candidate that caused the event to be selected by a trigger line or if it was another part of the event. There are several different decisions that identify this;

- TOS, triggered on signal - a candidate is identified as TOS if only information from the candidate was enough to cause a trigger line to select the event

- TIS, triggered independent of signal - a candidate is identified as TIS if part of the event independent of the candidate was enough to cause a trigger line to select the event

- DEC - a candidate is identified as DEC if anything in the event caused a trigger line to select an event. This includes TIS and TOS decisions and also when a combination of information from the candidate and something else in the event caused a trigger line to select the event

$B_s^0 \to \mu^+\mu^-$ decays are very rare decays and therefore trigger requirements used are chosen to keep a high efficiency at this step of the selection. The trigger lines L0Global, Hlt1Phys and Hlt2Phys are used and candidates are required to be identified as DEC at each level of the trigger. These trigger lines combine the decisions of many individual lines which allows a high efficiency to be achieved for selecting $B_s^0 \to \mu^+\mu^-$ decays. The L0Global trigger combines all trigger lines present in the L0 trigger, it selects an event provided at least one L0 trigger line selects it and rejects an event if no L0 trigger selects it. The Hlt1Phys and Hlt2Phys triggers are very similar to the L0Global trigger except that decisions are based only trigger lines related to physics processes and HLT trigger lines used for calibration are excluded.

The trigger requirements to identify $B_{(s)}^0 \to \mu^+\mu^-$ decays are also used to select $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ decays and slightly different trigger requirements are used for $B \to h^+h^-$ decays. $B \to h^+h^-$ decays are required to be triggered independent of signal by the L0Global and Hlt1Phys trigger lines and triggered on signal by at the HLT2 level by specific trigger lines designed to select $B \to h^+h^-$ decays. The TIS decision is used for $B \to h^+h^-$ decays to reduce the difference between the dominant lines that trigger $B \to h^+h^-$ and $B_{(s)}^0 \to \mu^+\mu^-$ decays, however the efficiency of TIS decisions is quite low at the Hlt2 level therefore TOS decisions are used there to high a high enough number of decays.

The requirements imposed on the trigger to select $B_s^0 \to \mu^+\mu^-$, $B \to h^+h^-$, $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ decays are shown in Table 4.2.

4.3.2 Cut based selection

The $B_{(s)}^0 \to \mu^+\mu^-$ candidates that passed the required trigger decisions are refined by a cut based selection. These selection cuts are aimed at removing obvious backgrounds by exploiting the differences between real $B_{(s)}^0 \to \mu^+\mu^-$ decays and the backgrounds that mimic them. The selection of $B \to h^+h^-$, $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ decays

Trigger Line	Trigger decision
Select $B^0_s \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$, $B^0_s \to J/\psi\phi$ *decays*	
L0Global	DEC
Hlt1Phys	DEC
Hlt2Phys	DEC
Select $B \to h^+h^-$ *decays*	
L0Global	TIS
Hlt1Phys	TIS
Hlt2B2HHDecision	TOS

Table 4.2 Trigger decisions used to select $B^0_s \to \mu^+\mu^-$, $B \to h^+h^-$, $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ decays.

is kept as close as possible to that of $B^0_{(s)} \to \mu^+\mu^-$ decays. The cut based selection is composed of two parts; the stripping selection and the offline selection.

The stripping selection, as described in Section **??**, is applied to all events that pass the trigger. It consists of individual stripping lines that select reconstructed candidates for specific decays, the development of the stripping selection is described in Sections 4.3.2.1 and 4.3.2.2.The primary purpose of the stripping selection is to reduce to size of the data set produced from *pp* collisions to a manageable size from which properties of particle decays can be measured.

The offline selection cuts are applied to the output of the stripping selection. Overall the stripping selection imposes loose selection requirements onto $B^0_{(s)} \to \mu^+\mu^-$ candidates so that as much information as possible is still available to develop the analysis and understand background events after the stripping selection. Therefore the offline selection further refines the data, removing background candidates. The offline selection cuts are presented in Section 4.3.2.3.

4.3.2.1 Development of the stripping selection

The stripping selections used to selection all decays necessary for the measurement of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions were designed at the start of Run 1 by studying the efficiencies of different selection cuts from simulated events [29]. However since then improvements have been made to the simulation of particle decays at LHCb, therefore it is prudent to check the accuracy of the selection efficiencies with updated simulated events and also to investigate where improvements can be made to the efficiency of the stripping selection used to select $B^0_{(s)} \to \mu^+\mu^-$ events.

There are four separate stripping lines that select $B^0_{(s)} \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$, $B^0_s \to J/\psi\phi$ and $B \to h^+h^-$ candidates, the selection of the all decays is kept as similar as possible to the signal selection to avoid introducing systematic uncertainties in the normalisation procedure of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions described in Chapter 5. However, the selection of $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ decays must diverge from the $B^0_{(s)} \to \mu^+\mu^-$ selection due to the additional particles in the final state of the decay. Any changes made to the $B^0_{(s)} \to \mu^+\mu^-$ stripping selection to improve the selection efficiency must be included in the selection of the other decays to keep the systematic uncertainties under control, this is particularly important for $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ decays. The stripping selection cuts applied for the Run 1 Branching Fraction analysis [30, 31] to select $B^0_{(s)} \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$, $B^0_s \to J/\psi\phi$ and $B \to h^+h^-$ candidates are listed in Table 4.3 and 4.4 .

Particle	$B^0_s \to \mu^+\mu^-$	$B \to h^+h^-$
$B^0_{(s)}$	|M - M$_{PDG}$| < 1200 MeV/c^2	|M - M$_{PDG}$| < 500 MeV/c^2
	DIRA > 0	DIRA > 0
	FD χ^2 >225	FD χ^2 >225
	IP χ^2 < 25	IP χ^2 < 25
	Vertex χ^2/ndof < 9	Vertex χ^2/ndof < 9
	DOCA < 0.3 mm	DOCA < 0.3 mm
		τ < 13.248 ps
		p_T > 500 MeV/c
Daughter μ or h	Track χ^2/ndof < 3	Track χ^2/ndof < 3
	isMuon = True	
	Minimum IP χ^2 > 25	Minimum IP χ^2 > 25
	p_T > 0.25 GeV/c	0.25 GeV/c < p_T < 40 GeV/c
		ghost probability < 0.3

Table 4.3 Selection requirements applied during the stripping selection for Run 1 data used in the $B^0_{(s)} \to \mu^+\mu^-$ Branching Fraction analysis [30, 31] to select $B^0_{(s)} \to \mu^+\mu^-$ and $B \to h^+h^-$ decays. M_{PDG} corresponds to the Particle Data Group [32] mass of each particle.

The variables used in the stripping selection are:

- the reconstructed mass, M - the mass and momenta of the decay products of the B meson (or J/ψ) are combined to provide its reconstructed mass. Cuts on

Particle	$B^+ \to J/\psi(\mu^+\mu^-)K^+$	Particle	$B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$
B^+	$\lvert$M - M$_{PDG}\rvert$ < 500 MeV/c^2	B_s^0	$\lvert$M - M$_{PDG}\rvert$ < 500 MeV/c^2
	Vertex χ^2/ndof < 45		Vertex χ^2/ndof < 75
	IP χ^2 < 25		IP χ^2 < 25
J/ψ	$\lvert$M - M$_{PDG}\rvert$ < 100 MeV/c^2	J/ψ	$\lvert$M - M$_{PDG}\rvert$ < 100 MeV/c^2
	DIRA > 0		DIRA > 0
	FD χ^2 > 225		FD χ^2 > 225
	Vertex χ^2/ndof < 9		Vertex χ^2/ndof < 9
	DOCA < 0.3 mm		DOCA < 0.3 mm
$\mu^\pm$	Track χ^2/ndof < 3	μ	Track χ^2/ndof < 3
	isMuon = True		isMuon = True
	Minimum IP χ^2 > 25		Minimum IP χ^2 > 25
	0.25 GeV/c < p_T		0.25 GeV/c < p_T
K^+	Track χ^2/ndof < 3	ϕ	$\lvert$M - M$_{PDG}\rvert$ < 20 MeV/c^2
	p_T > 0.25 GeV/c		Minimum IP χ^2 > 4
	Minimum IP χ^2 > 25	K$^\pm$	Track χ^2/ndof < 3
			p_T > 0.25 GeV/c

Table 4.4 Selection requirements applied during the stripping selection for Run 1 data used in the $B_{(s)}^0 \to \mu^+\mu^-$ Branching Fraction analysis [30, 31] to select $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ decays. M_{PDG} corresponds to the Particle Data Group [32] mass of each particle.

the mass remove candidates with a reconstructed mass far from the expected mass that are clearly backgrounds. Loose mass requirements are made on for the $B^0_{(s)} \to \mu^+\mu^-$ selection to allow for the study of semi-leptonic backgrounds that have a mass less than the $B^0_{(s)}$ mass when mis-identified as $B^0_{(s)} \to \mu^+\mu^-$ decays;

- the "direction cosine", DIRA - this is the cosine of the angle between the momentum vector of the particle and the vector connecting the production and decay vertices[1] of the particle. For correctly reconstructed candidates the direction cosine should be very close to one, requiring candidates to have positive value ensuring events are travelling in the incorrect direction are removed;

- the flight distance (FD) χ^2 - this is computed by performing the fit for the production vertex of a particle but including the tracks from its decay products that originate from the decay vertex in the fit as well. For a B meson the FD χ^2 is likely to be large because B mesons have long lifetimes therefore the tracks of its decays products will not point towards the production vertex;

- track fit $\chi^2/ndof$ - provides a measure of the quality of a fitted track, placing an upper limit removes poor quality tracks and backgrounds composed of poorly reconstructed decays;

- vertex fit $\chi^2/ndof$ - provides a measure of how well tracks can be combined to form a vertex, placing an upper limit removes poorly constrained vertices and backgrounds composed of poorly reconstructed decays;

- distance of closest approach (DOCA) - this is the distance of closest approach of two particles computed from the straight tracks in the VELO. For the decay products of a particle, for example the muons from $B^0_{(s)} \to \mu^+\mu^-$, this distance would ideally be zero because the muons originate from the same vertex;

- decay time, τ - is the length of time a particle lives as it travels from its production vertex to its decay vertex. Applying an upper decay time cut removes unphysical background decays;

- isMuon - particle identification variable defined in Section 3.2.2 that returns True for muons and False for other particles;

[1] The production vertex of the B or the primary vertex is identified by extrapolating the B meson momentum vector towards the beam axis. The closest vertex to the intersection of the B momentum and the beam axis is assigned as the primary vertex.

- transverse momentum, p_T - the component of a particle's momentum perpendicular to the beam axis. Decay products of B mesons are expected to have relatively high p_T due to the heavy B meson masses however an upper limit removes unphysical backgrounds;
- momentum, p - an upper limit on the momentum of a particle removes unphysical backgrounds;
- ghost probability - defined in Section **??** provides the probability of a tracking being composed on random hits in the detector, tracks from the passage of real particles will have a low ghost probability;
- impact parameter (IP) χ^2 - this is the change in the fit for a primary vertex (PV) caused by removing one track in the fit. In a $B^0_s \to \mu^+\mu^-$ decay, the $B^0_{(s)}$ is produced at the PV therefore it should have a small IP χ^2 value whereas the muons will be displaced from the PV because of the relatively long lifetime of the $B^0_{(s)}$ and therefore will have a large IP χ^2;
- minimum impact parameter (IP) χ^2 - this is the IP χ^2 of the muons with respect to all PVs in the event, this parameter is used to remove prompt muons created at any PV in the event and therefore reduce the prompt combinatorial background.

The stripping selection imposes a greater number cuts to select $B \to h^+h^-$ decays compared to $B^0_s \to \mu^+\mu^-$ because $B \to h^+h^-$ decays are much more abundant therefore extra cuts are needed to reduce the number of events passing the stripping to an acceptable level. The cuts applied to only $B \to h^+h^-$ decays in the stripping are the later applied to $B^0_s \to \mu^+\mu^-$ candidates in the offline selection.

4.3.2.2 Optimisation of $B^0_s \to \mu^+\mu^-$ stripping selection

The efficiency of the cuts used in the stripping lines to selecting $B^0_{(s)} \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ decays are shown in Table 4.5, only cuts that are in common with the $B^0_{(s)} \to \mu^+\mu^-$ stripping lines are listed. The efficiencies of $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ are studied as well as $B^0_{(s)} \to \mu^+\mu^-$ because the selection of these channels must be kept similar to reduce systematic uncertainties in the normalisation procedure described in Chapter 5. The efficiencies are evaluated using 2012 sim06 simulated events that have the minimum track p_T, track χ^2 and isMuon requirements imposed. These cuts are applied during the reconstruction and particles that do not pass these requirements are not included in the samples of simulated decays. No trigger

	Efficiency			
Requirement	$B_s^0 \to \mu^+\mu^-$	$B^0 \to \mu^+\mu^-$	$B \to h^+h^-$	$B^+ \to J/\psi K^+$
B |M - M$_{PDG}$|	$(100.00 \pm 0.00)\%$	$(100.00 \pm 0.00\ )\%$	$(98.25 \pm 0.02)\%$	$(99.73 \pm 0.02)\%$
$B^0_{(s)}$ or J/ψ DIRA	$(99.41 \pm 0.01)\ \%$	$(99.47 \pm 0.01)\%$	$(99.47 \pm 0.01)\%$	$(95.83 \pm 0.08)\%$
$B^0_{(s)}$ or J/ψ FD χ^2	$(83.74 \pm 0.06)\ \%$	$(83.96 \pm 0.06)\%$	$(83.83 \pm 0.06)\%$	$(82.90 \pm 0.15)\%$
$B^0_{(s)}$ or J/ψ IP χ^2	$(96.78 \pm 0.03)\ \%$	$(96.93 \pm 0.03)\%$	$(97.44 \pm 0.03)\%$	$(97.52 \pm 0.06)\%$
$B^0_{(s)}$ or J/ψ vertex χ^2/ndof	$(97.21 \pm 0.03)\ \%$	$(97.18 \pm 0.03)\%$	$(97.68 \pm 0.02)\%$	$(96.78 \pm 0.07)\%$
$B^0_{(s)}$ or J/ψ DOCA	$(99.82 \pm 0.01)\ \%$	$(99.80 \pm 0.01)\%$	$(99.83 \pm 0.01)\%$	$(99.58 \pm 0.03)\%$
μ, h, K^+ minimum IP χ^2	$(80.16 \pm 0.06)\ \%$	$(80.62 \pm 0.06\ )\%$	$(79.66 \pm 0.07)\%$	$(86.98 \pm 0.14)\%$
Total after above cuts	$(71.29 \pm 0.07)\ \%$	$(71.82 \pm 0.07)\%$	$(70.97 \pm 0.07)\%$	$(71.30 \pm 0.18)\%$
Total after all cuts	-	-	$(70.70 \pm 0.07)\%$	$(62.25 \pm 0.20)\%$

Table 4.5 Stripping line cut efficiencies for $B_s^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ 2012 simulated decays. Selection cuts applied are listed in Table 4.3 and 4.4. Efficiencies have been calculated only for cuts that are present in the $B_s^0 \to \mu^+\mu^-$ stripping, each cut separately and the total efficiencies are given for the listed cuts and the complete set of cuts present in each stripping line.

requirements have been applied so that only the effect of the stripping selection on the efficiencies can be assessed. During the simulation of particle decays the trigger is run in *pass through* mode so that all reconstructed are saved not just those that have passed a trigger line.

The selection efficiencies are very similar for each stripping cut across the different decays which fits the requirement that the selection of signal and normalisation decays used in the branching fraction measurement are as similar as possible. The similarity of the selection efficiencies for the signal and normalisation decays is further illustrated in Figures 4.2 and 4.1 which show the ratio of selection efficiencies for $B_s^0 \to \mu^+\mu^-$ decays to $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ decays for a range of selection cuts. With the exception of the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ IP χ^2 cuts on the daughter particles, the ratio of efficiencies is well within 3% of 1 for the range of cuts values shown. The ratio of the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ efficiencies for the daughter particle IP χ^2 markedly deviates from unity, showing that the IP χ^2 distribution of the muons and kaon are very different as seen previous in [29]. If the FD χ^2, B_s^0 or J/ψ IP χ^2 and vertex χ^2 selection cuts are applied to the simulated events before the daughter IP χ^2 requirement the ratio of $B_{(s)}^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ efficiencies is much closer to 1. The stability of the ratios of selection efficiencies across a large range of cuts values shows that changing a cut value in the $B_{(s)}^0 \to \mu^+\mu^-$ selection will have a similar impact on the efficiencies of the normalisation decays.

The efficiencies for most of the stripping cuts are $\sim 97\%$ or higher, however, the efficiencies of the cuts on the FD χ^2 of the $B_{(s)}^0$ or J/ψ and the daughter IP χ^2 of the muon or hadron pair are lower at 83% and 80%, respectively. Therefore improvements to the stripping selection efficiencies could be achieved by altering these two selection requirements.

The set of events removed by each cut in the stripping selection is not independent. Therefore the effect of changing one cut on the total efficiency of a stripping selection must be considered. Figure 4.3 shows the total efficiency of the $B_s^0 \to \mu^+\mu^-$ stripping line on simulated $B_s^0 \to \mu^+\mu^-$ decays for a range of FD χ^2 and daughter IP χ^2 cut values. As expected the lower the cut values the more efficient the stripping line becomes. It is important that any increase in $B_s^0 \to \mu^+\mu^-$ selection efficiency from the stripping is not removed when the trigger requirements are applied, Figure 4.4 shows that the trigger efficiencies are relatively flat across a large range of FD χ^2 and daughter IP χ^2 cut values therefore the efficiency gained by a change in the stripping selection is not lost when trigger requirements are imposed. The selection efficiency

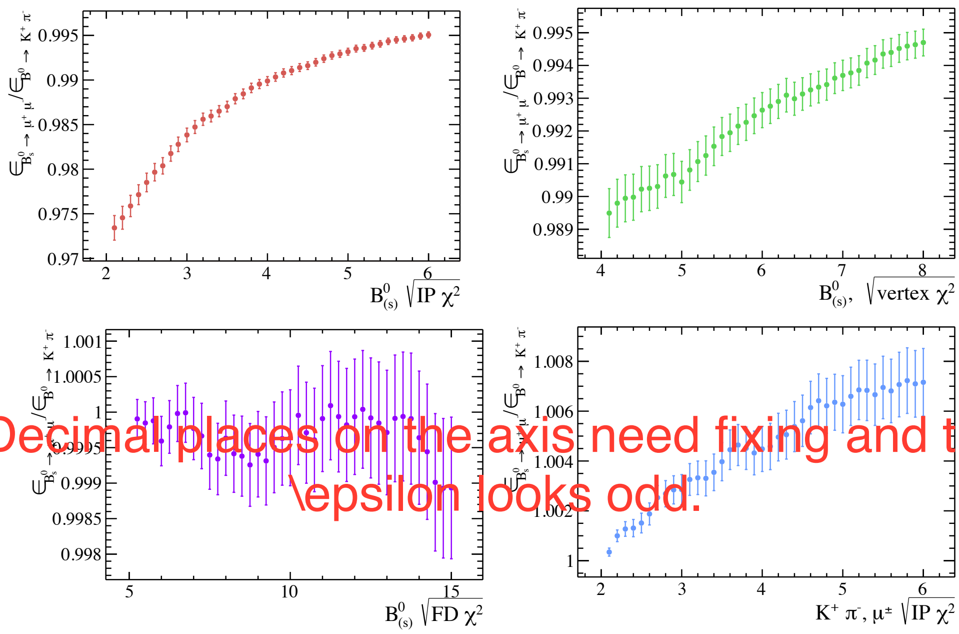


Fig. 4.1 The ratio of $B_s^0 \to \mu^+\mu^-$ to $B^0 \to K^+\pi^-$ stripping efficiencies when each cut has been applied independently of all other cuts. The current cut values are marked by the blue lines.

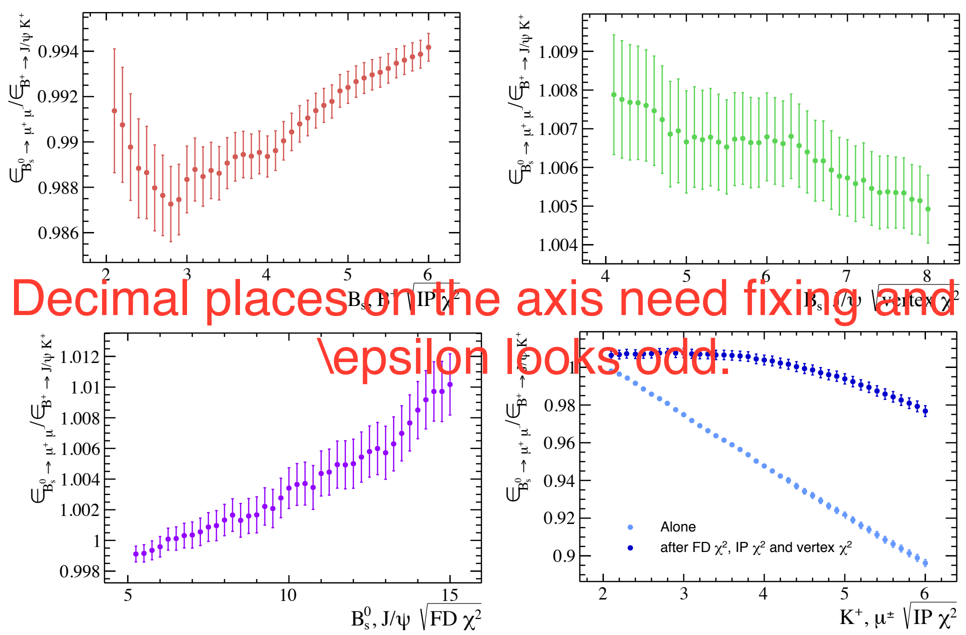


Fig. 4.2 The ratio of $B_{(s)}^0 \to \mu^+\mu^-$ to $B^+ \to J/\psi K^+$ stripping efficiencies when each cut has been applied independently of all other cuts. The current cut values are marked by the blue lines.

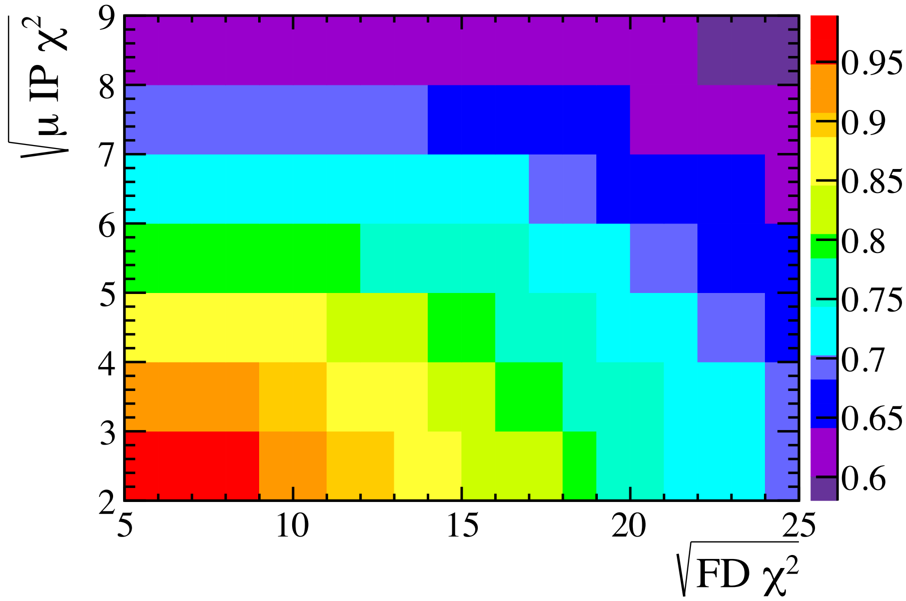


Fig. 4.3 Efficiency of the $B^0_{(s)} \to \mu^+\mu^-$ stripping selection for $B^0_s \to \mu^+\mu^-$ simulated decays for a range of cuts on the B^0_s FD χ^2 and the minimum muon IP χ^2.

for $B^0 \to \mu^+\mu^-$ is very similar to $B^0_s \to \mu^+\mu^-$ as seen in Table 4.5, therefore only $B^0_s \to \mu^+\mu^-$ have been studied for different stripping selection cut values.

One of the main purposes of the stripping selection, as described in Section 3.2.4, is to reduce the size of the data set, therefore the cuts cannot be set as loose as possible and the amount of data passing the selection must be considered. Also, any change applied to the $B^0_{(s)} \to \mu^+\mu^-$ stripping line must be propagated through into the stripping lines for $B \to h^+h^-$, $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ decays therefore the retention of all stripping lines must be evaluated.

Table 4.6 shows the total efficiency of the $B^0_s \to \mu^+\mu^-$ stripping line along side the amount of data retained for the set of cuts on the FD χ^2 and daughter IP χ^2 for the $B^0_{(s)} \to \mu^+\mu^-$, $B \to h^+h^-$, $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ stripping lines for a set of IP χ^2 and FD χ^2 cuts. The set of chosen cuts used in table 4.6 aims keep both cuts as high as possible for a certain $B^0_s \to \mu^+\mu^-$ efficiency.

The data retention is computed by applying the stripping selection to a sub-set of 2012 data to find the number of events that pass the stripping lines for each pair of FD χ^2 and daughter IP χ^2 cuts. No trigger requirements are imposed on trigger lines because the stripping selection run on the full output of the trigger. The number of events for each set of cuts is normalised to the number of events passing the original

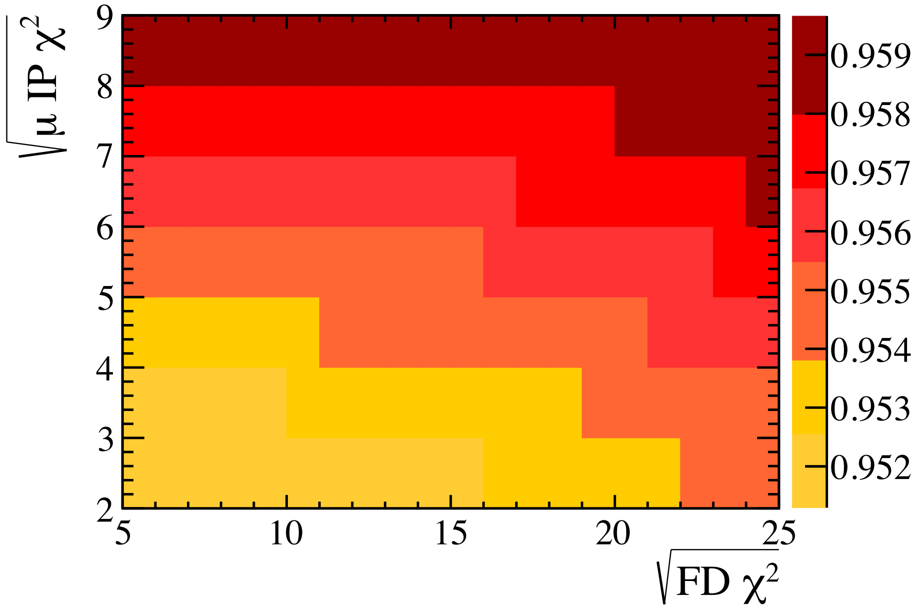


Fig. 4.4 The trigger efficiencies of $B_s^0 \to \mu^+\mu^-$ simulated decays across a range of B_s^0 FD χ^2 and the minimum muon IP χ^2 cut values for the trigger requirements used to select $B_{(s)}^0 \to \mu^+\mu^-$ decays for the branching fraction measurement.

Stripping cut		Stripping line efficiency	Stripping line retention			
$\sqrt{\mathrm{FD}\chi^2}$	Daughter $\sqrt{\mathrm{IP}\chi^2}$	$B_s^0 \to \mu^+\mu^-$	$B_{(s)}^0 \to \mu^+\mu^-$	$B^0 \to K^+\pi^-$	$B^+ \to J/\psi K^+$	$B_s^0 \to J/\psi\phi$
15	5.00	(71.29 ± 0.07) %	1.0	1.0	1.0	1.0
14	4.25	(74.91 ± 0.07) %	1.5	1.3	1.1	1.3
13	4.00	(76.84 ± 0.07) %	1.8	1.5	1.2	1.4
12	3.50	(79.76 ± 0.07) %	2.6	1.8	1.3	1.7
11	3.00	(82.72 ± 0.06) %	3.7	2.4	1.6	1.9
10	2.75	(84.86 ± 0.06) %	4.7	3.0	1.7	2.1
9	2.50	(86.96 ± 0.06) %	6.8	3.9	2.0	2.2

Table 4.6 The efficiency of the $B_{(s)}^0 \to \mu^+\mu^-$ stripping line to select $B_s^0 \to \mu^+\mu^-$ decays and the change in the data retention for $B_{(s)}^0 \to \mu^+\mu^-$, $B \to h^+h^-$, $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ stripping lines for a range of FD χ^2 and daughter IP χ^2 cut values. The amount of data passing each selection has been normalised to the original set of stripping select cuts. The fractional uncertainty on the retention is less than 1 %.

Run 1 stripping line requirements to show the fractional increase caused by loosening the cut values.

An increase of 15 % can be gained in the stripping selection efficiencies by using the loosest cuts in Table 4.6 however the loosest cuts increases the amount of data passing the $B^0_{(s)} \to \mu^+\mu^-$ stripping selection by a factor of 7 and the $B \to h^+h^-$ stripping selection by a factor of 4. Table 4.7 shows the number of Run 1 candidates passing the original stripping selection listed in Tables 4.3 and 4.4 for the last published analysis. The $B \to h^+h^-$ stripping line lets through the most candidates where as the $B^0_{(s)} \to \mu^+\mu^-$ stripping line saves far fewer candidates, therefore a chance in the retention of the $B \to h^+h^-$ line is more significant than the $B^0_{(s)} \to \mu^+\mu^-$ line.

The final set of cuts used in the stripping selection must be a compromise between the selection efficiency and the amount of data that passes the selection. The selection cuts of B^0_s FD $\chi^2 > 121$ and minimum muon IP $\chi^2 > 9$ would increase the $B^0_{(s)} \to \mu^+\mu^-$ selection efficiency by from 71 % to 82 % and the amount of data retained would be doubled. The increase of the data retained by the $B \to h^+h^-$, $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ lines is smaller and the efficiencies are similar to the $B^0_{(s)} \to \mu^+\mu^-$ selection efficiencies. Therefore these cuts are applied in the stripping selection for this analysis.

Stripping Lines	Events	Retention / %
$B^0_{(s)} \to \mu^+\mu^-$	898880	0.0022
$B \to h^+h^-$	14502295	0.0831
$B^+ \to J/\psi K^+$	3344568	0.0087
$B^0_s \to J/\psi\phi$	456787	0.0011
Total	18745743	-

Table 4.7 The number of events passing stripping lines used for the $B^0_s \to \mu^+\mu^-$ analysis from the selection listed in Table **??** and the percentage of the total LHCb data set that they correspond to. The total does not include correlation between lines.

4.3.2.3 Additional offline cuts

Additional selection requirements are applied after the stripping to remove specific backgrounds. A lower bound is placed on the B meson transverse momentum to remove pairs of muons originating from $pp \to p\mu^+\mu^-p$ decays and a J/ψ veto is used to remove backgrounds from $B^+_c \to J/\psi\mu^+\nu_\mu$ decays. Semi-leptonic $B^+_c \to J/\psi\mu^+\nu_\mu$ decays, where $J/\psi \to \mu^+\mu^-$, contribute to the background of $B^0_{(s)} \to \mu^+\mu^-$ decays when a muon from the J/ψ forms a good vertex with the muon from the B^+_c decay.

Due to the high mass of the B_c^+ this could place mis-reconstructed candidates within the B_s^0 mass window. A 'J/ψ veto' can be used to remove background events from $B_c^+ \to J/\psi\mu^+\nu_\mu$ decays. The veto works by removing events where one muon from the $B^0_{(s)} \to \mu^+\mu^-$ candidate combined with any other oppositely charged muon in the event has $|m_{\mu^+\mu^-} - m_{J/\psi}| < 30$ MeV/c^2.

The offline selection of $B^0_{(s)} \to \mu^+\mu^-$ decays includes the momentum, ghost track probability and decay time cuts made in the $B \to h^+h^-$ stripping line, but were absent in the $B^0_{(s)} \to \mu^+\mu^-$ stripping line. Also a narrower mass range of 4900 - 6000 MeV/c^2 us imposed to $B_s^0 \to \mu^+\mu^-\gamma$ backgrounds. The stripping selection for $B^0_{(s)} \to \mu^+\mu^-$ decays is kept loose to allow for the study of background decays in data.

The selection applied to Run 1 and Run 2 data is the same for all variables expect the track ghost probability and track χ^2/ndof. Slightly looser cuts of track ghost probability < 0.4 and track $\chi^2/\text{ndof} < 4$, are used for Run 2 to take advantage to changes in the reconstruction that were introduced for Run 2.

Table 4.12 summaries all selection cuts used to identify $B^0_{(s)} \to \mu^+\mu^-$ decays at the end of this section.

4.3.3 Particle identification

In the selection of $B^0_{(s)} \to \mu^+\mu^-$ decays particle identification variables are particularly useful to reduce the backgrounds coming from mis-identified semi-leptonic decays and $B \to h^+h^-$ decays and also help to reduce the number of combinatorial background. On top of the isMuon requirement used in the stripping selection, a linear combination of ProbNN variables

$$\text{PID}_\mu = \text{ProbNN}\mu \times (1 - \text{ProbNN}K) \times (1 - \text{ProbNN}p) \tag{4.1}$$

is used to refine the selection of $B^0_{(s)} \to \mu^+\mu^-$ candidates. The ProbNNK variable is effective at removing mis-identified $B \to h^+h^-$ backgrounds and the ProbNNp variable is effective at removing $\Lambda_b^0 \to p\mu^-\nu_\mu$ backgrounds.

Different tunings of the algorithms used in the ProbNN variables are used to select candidates in Run 1 and 2015 data compared to 2016 data. The tunings perform differently therefore the cut values placed on PID$_\mu$ are different for each tuning. The cuts applied to data are

$$\begin{aligned} \text{Run 1 and 2015: PID}_\mu &> 0.4 \\ \text{2016: PID}_\mu &> 0.8. \end{aligned} \tag{4.2}$$

The cut value on PID_μ for the Run 1 and 2015 tuning was optimised to sufficiently reduce the background decays and give the highest sensitivity to the $B^0 \to \mu^+\mu^-$ decays. Accurate particle identification is most important for $B^0 \to \mu^+\mu^-$ because the background decays from $B \to h^+h^-$ and $\Lambda_b^0 \to p\mu^-\nu_\mu$ pollute the B^0 mass window. The cut value for 2016 was chosen to have the same or lower background rejection as the Run 1 and 2015 cut, however the 2016 tuning has a better performance therefore the final cut choice has a high efficiency for selecting $B^0_{(s)} \to \mu^+\mu^-$ decays.

4.3.4 Multivariate Classifiers

The selection described so far removes a large number of background candidates however because $B^0_{(s)} \to \mu^+\mu^-$ decays occur very rarely the data is still dominated by long lived combinatorial backgrounds. To improve the separation of signal and background decays two multivariate classifiers are used.

A multivariate classifier is an algorithm that learns differences between signal and background decays. The classifier is given two input samples, one contain only signal decays and the other containing just background decays and a set of input variables. These input variables have different distributions for signal and background decays. The classifier uses the distributions of the input variables along with its knowledge of which decays are signal and background to learn the difference between the two types. The algorithm can then be applied to a data set containing an unknown mixture of signal and background decays to separate them. For each decay the algorithm produces a number, typically between -1 and +1, where high numbers indicate signal-like decays and low numbers indicating background-like decays.

Two multivariate classifiers are used to identify $B^0_s \to \mu^+\mu^-$ decays. Both classifiers are a type called a Boosted Decision Tree (BDT) that are described in Section 4.3.4.1.

The first classifier (Sect. 4.3.4.2), called the BDTS, is used to remove candidates that are very unlikely to be signal by placing a cut on the BDTS output. The second classifier (Sect. 4.3.4.3), called the global BDT, is used to classify candidates into bins containing increasing proportions of signal candidates, no candidates are removed based on the output of the second BDT. The the BDTS is necessary to reduce the number of background events to a more manageable level for the global BDT.

4.3.4.1 Boosted Decision Trees

A BDT is made up of the combined outputs of separate decision trees. A decision tree begins with a data sample, where each decay is know to be signal or background and a set of variables describing them. The decision tree applies a cut on a variable that will be the most effective at separating the signal and background in the sample and creates two sub-samples. Another cut is then applied to each of the sub-samples to further separate signal from background. This process is repeated until either a certain number of cuts, defined as the depth of the tree, or the number of candidates in each sub-sample has reached a minimum number. Each sub-sample produced at the end of the tree is called a leaf. The tree uses the knowledge of whether decays are signal or background to assign a value of +1 or -1 to every decay. A decay is given a value +1 if it is in a leaf where the majority is signal and the value -1 if it is in a leaf that has a majority of background decays. The final decisions made by the tree are not prefect, some signal (background) decays will be mis-classified as background and given the value of -1 (+1).

One decision tree on its own is often not particularly good at classifying events, there is no way to correct mis-classified events in the leaves, and it is particularly sensitive to statistical fluctuations in the training samples. A BDT combines the output of numerous decision trees to improve the classification of events and reduce the dependence of the final decisions on statistical fluctuations. A BDT starts with one decision tree and assigns weights to decays in the signal and background samples depending on whether the output of the decision tree classified the events correctly or incorrectly. The weighted sample is then used as the input for the training of the next decision tree. The weights are designed so that the next tree is more likely to correctly classify previously mis-classified events. This process is repeated until a certain number of trees have been trained. The re-weighting process is known as boosting and the weights applied to the samples are taken into account when combining the output of each decision tree into the overall output of the BDT. The output of a BDT will be a number between -1 and +1 where high numbers indicate signal and low numbers indicating background.

The TMVA package [33] is use to develop and train the BDTs, the package provides several different methods of boosting that can be used. The adaptive boosting method was found to produce the most effective BDT. This method of boosting assigns decays incorrectly classified by one tree the weight, w, before being used as the input to the next decision tree. The weights assigned are given by

$$w = \frac{1-f}{f}, \text{ where } f = \frac{\text{total misclassified events}}{\text{total events}}. \tag{4.3}$$

Therefore incorrectly assigned candidates are given a higher weight than correctly classified candidates. The 'speed' at which the boosting occurs is controlled by a the parameter β where $w \rightarrow w^{\beta}$, this can be specified in the training of the decision tree and a large number of boosting steps can improve the performance of the BDT.

The ability of a BDT to correctly identify signal and background candidates depends on three main factors;

- the size of the training samples - a large training sample is useful to prevent the BDT from being sensitive to statistical fluctuations and contains more information the classifier can use to learn the difference between signal and background
- the input variables - different distributions in the input variables for signal and background candidates enable the classifier to easily separate the types of candidates, the overall performance is insensitive to poorly discriminating variables that are included
- parameters that dictate the BDT training - the training of a BDT is specified by several parameters; the number of trees (NTrees), the tree depth (MaxDepth), the minimum number of events a leaf can contain (nEventsMin or MinNodeSize[2]);, the 'speed' at which the boosting occurs (β) and the number of cut values that a tree tries for a variable before making a decision (nCuts).

These three factors affect the performance of the BDT however the importance of each varies. Together they can be used to prevent the BDT being very sensitive to the statistical fluctuations in the training sample. This is called overtraining, an overtrained BDT is extremely accurate at classifying the candidates in the training sample but preforms poorly at classifying candidates in a statistically independent sample. Although this is less common in BDTs than single decision trees, it can be avoided by having a sufficiently large training sample or by limiting the depth of trees or the number of trees in the BDT.

4.3.4.2 The BDTS

The BDTS uses input variables similar to those in the stripping selection to classify events;

- impact parameter χ^2 of the B_s^0

[2]nEventsMin is the minimum number of decays in a lead where as MinNodeSize is the number of decays in a leaf given as a percentage of the training sample size. The parameter specified in the training depends on the version of the TMVA package used.

- vertex χ^2 of the B_s^0
- direction cosine of B_s^0
- distance of closest approach of the muons
- minimum impact parameter χ^2 of the muons with respect to all primary vertices in the event
- impact parameter of the B_s^0, this is the distance of closest approach of the B to the primary vertex

The signal and background samples used to train the BDTS are simulated $B_s^0 \to \mu^+\mu^-$ decays and $B_s^0 \to \mu^+\mu^-$ candidates in a sample of Run 1 data from the mass ranges 4800 - 5000 MeV/c^2 and 5500 - 6000 MeV/c^2. The selection cuts listed in Table 4.8 are applied to the training samples and the training parameters used are listed in Table 4.9. The output of the BDTS is flattened between 0 and 1 so that signal is uniformly distributed across the range and background is peaked at zero as illustrated in Figure 4.5. The BDTS is applied to all candidates passing the $B_{(s)}^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ stripping lines, and candidates are required to have a BDTS value above 0.05. When the BDTS is applied to $B^+ \to J/\psi K^+$ decays the distance of closest approach of the muons refers to the muons in the J/ψ and the vertex χ^2 is of the J/ψ. The performance of the BDTS at removing backgrounds is illustrated in Figure 4.6.

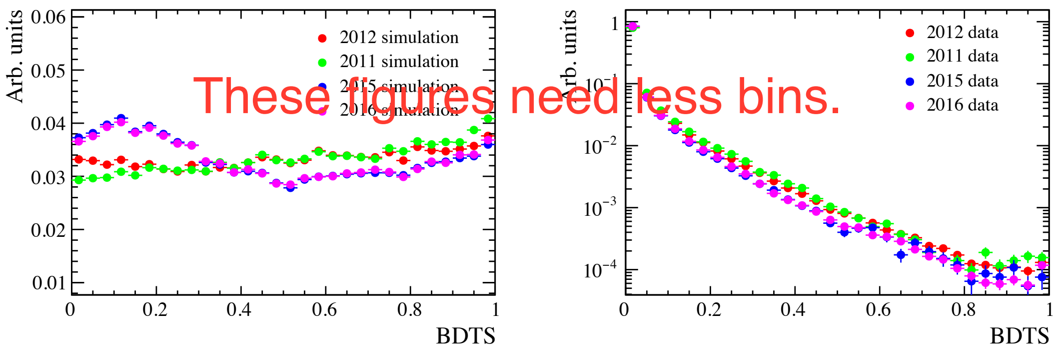
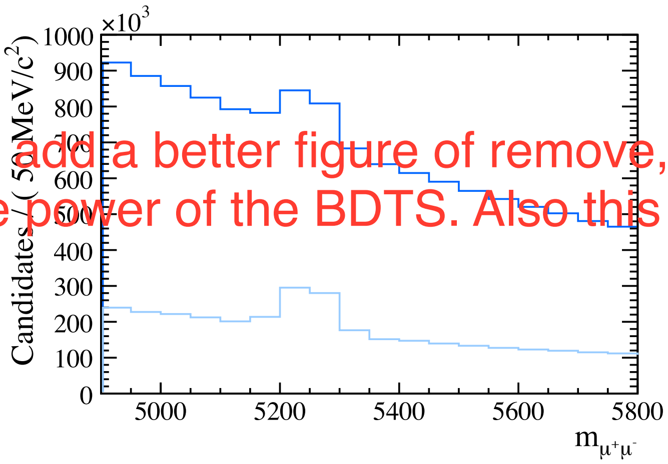


Fig. 4.5 BDTS response for simulated $B_s^0 \to \mu^+\mu^-$ decays (left) and data with a mass above 5447 MeV/c^2 consisting on $b\bar{b} \to \mu^+\mu^- X$ decays.

4.3.4.3 Global BDT

The global BDT is the final step in identifying $B_{(s)}^0 \to \mu^+\mu^-$ decays and it is very effective at separating them from long lived combinatorial background decays. The



This is poor, either add a better figure of remove, although it is nice to illustrate the power of the BDTS. Also this is 2016 data?

Fig. 4.6 Invariant mass spectrum for $B \to h^+h^-$ decays in 2016 data passing the selection requirements in Table 4.8 before and after the BDTS cut is applied.

Selection applied to BDTS training samples.	
B_s^0	$\mu^\pm$
FD $\chi^2 > 225$	$p_T > 500$ MeV$/c$
IP $\chi^2 < 25$	track χ^2/ndof < 3
Vertex χ^2/ndof < 9	minimum IP $\chi^2 > 25$
DOCA < 0.3 mm	0.25 GeV$/c < p_T < 40$ GeV$/c$
$\tau < 13.248$ ps	$p < 500$ GeV$/c$
$p_T > 500$ MeV$/c$	
DIRA > 0	
Trigger requirements	
L0Global	DEC
Hlt1Phys	DEC
Hlt2Phys	DEC

Table 4.8 Selection cuts applied to select candidates for signal and background samples used to train the BDTS. The isMuon requirement is not applied to the muons so that the BDTS can be used on $B \to h^+h^-$ decays.

Parameter	Value
nTrees	250
nEventsMin	400
MaxDepth	3
β	1.0
nCuts	20

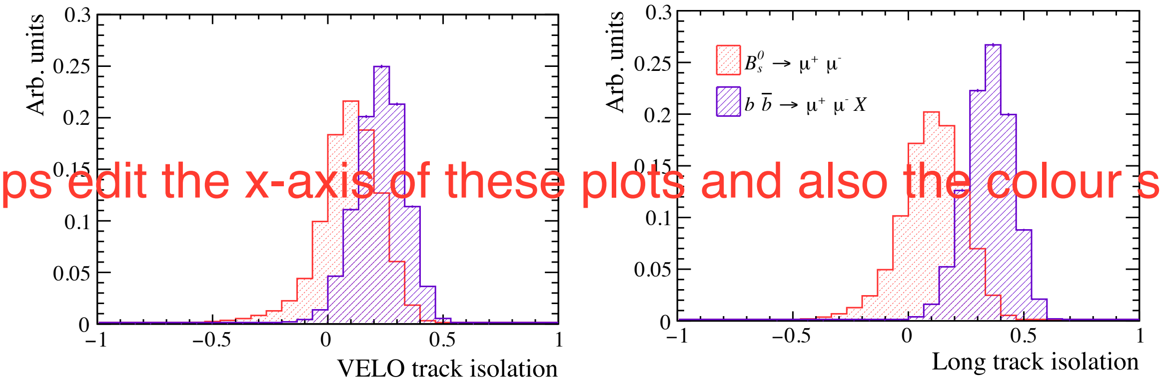
Table 4.9 Training parameters used to specify the training of the BDTS.

discriminating power achieved by the global BDT is mostly dependant on isolation variables. Isolation variables, or just isolations, provide a measure of how far away each muon from a $B^0_{(s)} \to \mu^+\mu^-$ candidate is from other tracks in the event. The tracks of the muons from a real $B^0_{(s)} \to \mu^+\mu^-$ decays will be, in general, far from other tracks in the event because the $B^0_{(s)} \to \mu^+\mu^-$ decays tree contains no other tracks apart from the muons. However long lived combinatorial background arises from semi-leptonic decays therefore muon tracks are likely to be close to other tracks that have originated from the same decay tree as the muon. Isolations are very useful in the selection of very rare decays like $B^0_s \to \mu^+\mu^-$ because they enable background to be removed whilst keeping a high efficiency for signal decays.

Two isolation variables are used in the global BDT, one compares long tracks in the event to the muons in $B^0_{(s)} \to \mu^+\mu^-$ candidates and the other compares VELO tracks in the event to the muons. The definition of the track types can be found in Section **??**. The isolation variables are built from the output of BDTs. For each type of track a BDT is trained on simulated $B^0_s \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ decays using a set of input variables that describe track and vertex properties. The BDT compares the μ^+ from a $B^0_s \to \mu^+\mu^-$ candidate with all other tracks in the event, excluding the track of the μ^-, and gives an output, iso_{μ^+} *(track)*, for each possible μ^+ and track pairing. The process is repeated for the μ^-. The BDT is designed to produce high output values for muons from $b\bar{b} \to \mu^+\mu^- X$ decays and a low value for muons from $B^0_s \to \mu^+\mu^-$ decays. The isolation variable of a $B^0_s \to \mu^+\mu^-$ candidate is then composed of the sum of the highest values of iso_{μ^+} *(track)* and iso_{μ^-} *(track)* for any tracks in the event. The separation power of these isolations are shown in Figure 4.7.

The isolations are used along with five other variables in the global BDT. The full list of input variables used are;

- Long track isolation
- VELO track isolation



Prehaps edit the x-axis of these plots and also the colour scheme.

Fig. 4.7 VELO track (left) and Long track (right) isolation distributions of simulated $B_s^0 \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ decays used to train the global BDT passing cuts in Table 4.11.

- $\sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ is the difference in azimuthal angles of the muons and $\Delta\eta$ the difference in the pseudo-rapidity of the muons
- the smallest IP χ^2 with respect to the primary vertex of the $B_s^0 \to \mu^+\mu^-$ of the muons
- vertex χ^2 of the B_s^0
- IP χ^2 of the B_s^0 with respect to the primary vertex
- angle between the momentum vector of the B_s^0 and the vector connecting the production and decay vertices of the B_s^0

A comparison of the signal and background distributions of the input variables are shown in Figures 4.7 and 4.8. These variables were chosen by training a BDT beginning with the most discriminating variable, the Long track isolation, and adding variables to determine which improved the performance to the classifier. Only variables that improved the performance were included in the global BDT. The training parameters used in the BDT are listed in Table 4.10. These parameters were chosen by scanning across a range of variables and choosing those that gave the best performance.

Simulated $B_s^0 \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ decays are used to provide large signal and background training samples for the global BDT. The simulated sample $b\bar{b} \to \mu^+\mu^- X$ decays corresponds to the background expected with 7 fb^{-1} of data from pp collisions at $\sqrt{s} = 8$ TeV. The production of such a large sample requires a lot of space to be saved, therefore several measures were taken to reduce the size needed to save the simulated $b\bar{b} \to \mu^+\mu^- X$ decays. The cuts, listed in Table 4.1, were applied to the

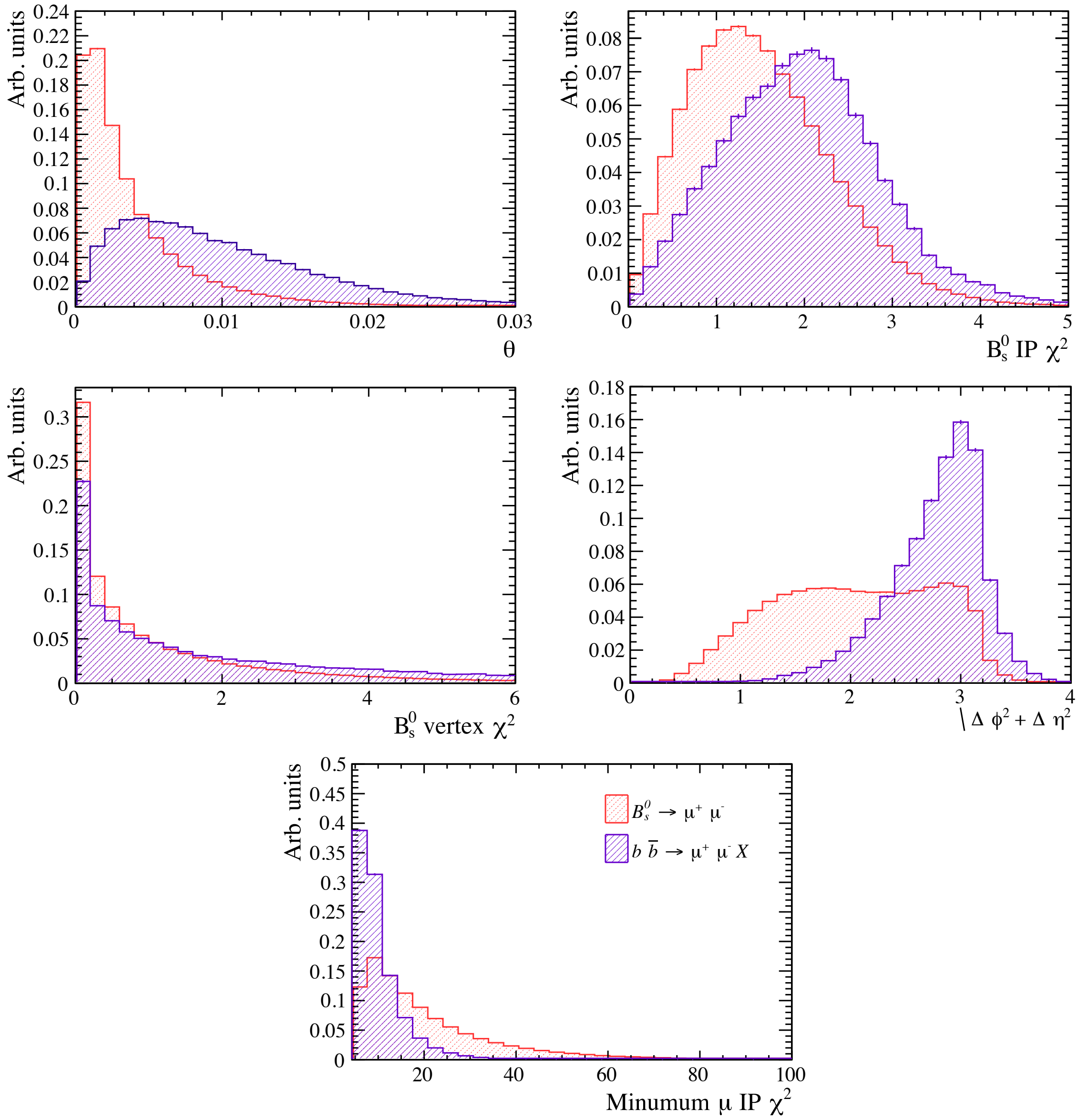


Fig. 4.8 Distributions of input variables of the global BDT from simulated $B_s^0 \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ decays used to train the global BDT passing cuts in Table 4.11.

Parameter	Value
nTrees	1000
MinNodeSize	1%
MaxDepth	3
β	0.75
nCuts	30

Table 4.10 Training parameters used to specify the training of the global BDT.

simulated decays as they were generated to reduce the number of events saved on disk. Also the stripping selection cuts in Table 4.3 were applied and candidates that did not pass the stripping selection were not saved. Unfortunately the $b\bar{b} \to \mu^+\mu^- X$ sample therefore does not include candidates that are selected by the looser stripping selection described in Section 4.3.2. In order to gain the best BDT performance on data the same cuts should be applied to data that are applied to the samples used to train the BDT. Therefore the original cuts on FD χ^2 and daughter IP χ^2 listed in Table **??** must be used to select $B^0_s \to \mu^+\mu^-$ candidates. The complete list of selection requirements applied to the training samples used to develop global BDT are listed in Table 4.11, the same selection is applied to $B^0_s \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ decays.

Selection applied to BDTS training samples.	
B^0_s	$\mu^\pm$
FD $\chi^2 > 225$	$p_T > 500$ MeV/c
IP $\chi^2 < 25$	track χ^2/ndof < 3
Vertex χ^2/ndof < 9	minimum IP $\chi^2 > 25$
DOCA < 0.3 mm	0.25 GeV/$c < p_T < 40$ GeV/c
$\tau < 13.248$ ps	$p < 500$ GeV/c
$p_T > 500$ MeV/c	isMuon = True
DIRA > 0	BDTS > 0.05
$4900 < \mathrm{M}_{\mu^+\mu^-} < 6000$ MeV/c^2	
Trigger requirements	
L0Global	DEC
Hlt1Phys	DEC
Hlt2Phys	DEC

Table 4.11 Selection cuts applied to select candidates for signal and background samples used to train the BDT.

The global BDT is applied to data taken in all years and in the same way as the BDTS the final output of the global BDT is flattened to have a response between 0 and 1 that is uniform for signal and the background peaks at zero, as shown in Figure 4.9 for each year of data taking. The flattening is important for the measurement of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions because a simultaneous fit is applied to the dimuon invariant mass in bins of BDT, flattening the BDT output enable bins containing equal proportions of signal decays to be easily created. The signal efficiency verses the background rejection of the global BDT is shown in Figure 4.10 for all years of data taking, the performance is similar across all the years but with Run 2 data having a slightly better background rejection for a given signal efficiency. A comparison of the input variables used in the global BDT for each year of data taking is given in Appendix A.

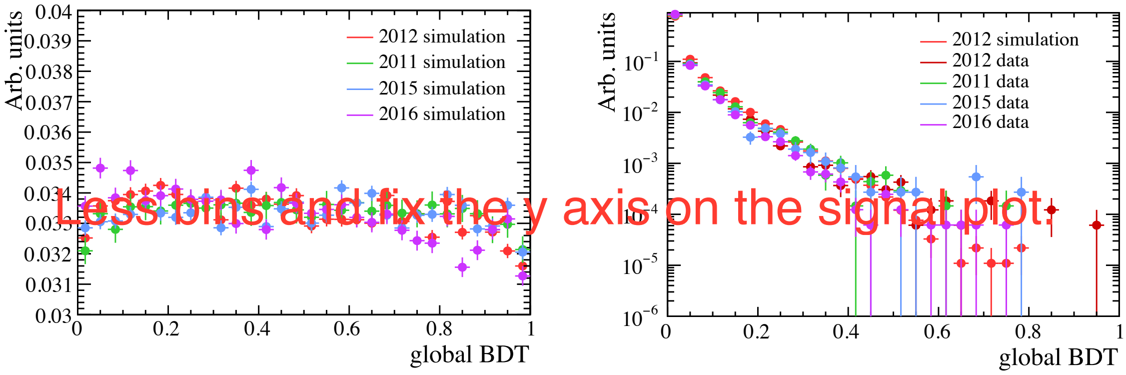


Fig. 4.9 Global BDT output distributions for $B_s^0 \to \mu^+\mu^-$ simulated decays (left) and $b\bar{b} \to \mu^+\mu^- X$ decays from simulation and data.

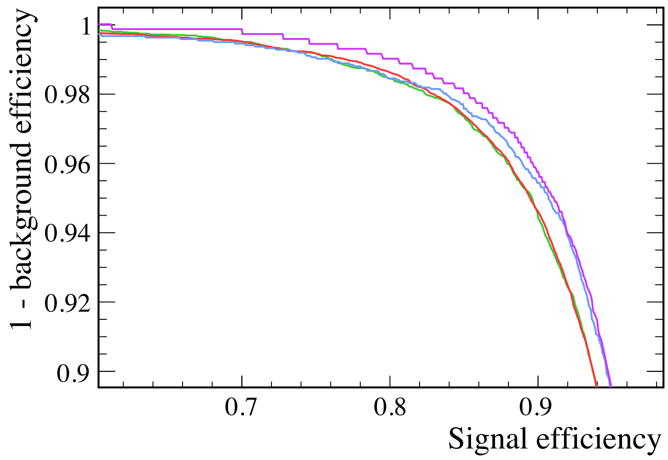


Fig. 4.10 Global BDT performance for 2011, 2012, 2015 and 2016 data taking conditions. Signal efficiency is calculated from $B_s^0 \to \mu^+\mu^-$ simulated decays and background rejection from data passing the $B_s^0 \to \mu^+\mu^-$ selection with $m_{\mu^+\mu^-} > 5447$ MeV$/c^2$. The performance is very similar for the different data taking years therefore only the most sensitive region is shown.

4.3.5 Summary

The complete set of selection criteria used for identify $B_{(s)}^0 \to \mu^+\mu^-$ decays in Run 1 and Run 2 data for the mesaurement of the $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions are listed in Tables **??**.The selection requirements do not remove all backgrounds decays from the data set but reduce them to a level at which the branching fractions can be measured. The selection criteria for $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ decays is kept as close as possible to that used to identify $B^0_{(s)} \to \mu^+\mu^-$ decays in order to reduce systematic uncertainties from selection efficiencies in the normalisation procedure described in Section 5.4.

Particle	$B^0_s \to \mu^+\mu^-$
B^0_s or B^+	$4900\ \mathrm{MeV}/c^2 < \mathrm{M} < 6000\ \mathrm{MeV}/c^2$
	DIRA > 0
	FD $\chi^2 > 225$
	IP $\chi^2 < 25$
	Vertex χ^2/ndof < 9
	DOCA < 0.3 mm
	$\tau < 13.248$ ps
	$p_T > 500\ \mathrm{MeV}/c$
Daughter μ or h	Track χ^2/ndof < 3 (4)
	Minimum IP $\chi^2 > 25$
	$0.25\ \mathrm{GeV}/c < p_T < 40\ \mathrm{GeV}/c$
	$p < 500\ \mathrm{GeV}/c$
	ghost probability < 0.3 (0.4)
	$\lvert m_{\mu\mu} - m_{J/\psi}\rvert < 30\ \mathrm{MeV}/c^2$
	isMuon = True
	$\mathrm{PID}_\mu > 0.4$ (0.8)
	BDTS > 0.05
Trigger requirements	L0Global = DEC
	Hlt1Phys = DEC
	Hlt2Phys = DEC

Table 4.12 Selection cuts applied to select $B^0_s \to \mu^+\mu^-$, where selection is different between Run 1 and Run 2 the Run 2 values are shown in parenthesis.

4.4 Selection for the effective lifetime measurement

The selection used to identify candidates for the measurement of the $B^0_s \to \mu^+\mu^-$ effective lifetime is based on the selection used to identify candidates for the measurement of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions. However several changes are made to account for the different measurement strategies that are descried in Chapters 5 and 6 and that only the B_s^0 decay mode is required for the effective lifetime mesaurement. $B \to h^+h^-$ and $B_s^0 \to J/\psi\phi$ decays are used to develop and validate the effective lifetime analysis strategy. The selection of $B_s^0 \to J/\psi\phi$ decays is kept the same as that used for the branching fraction measurement but there are a few differences in the selection of $B \to h^+h^-$ decays for the effective lifetime measurement.

The majority of the selection of $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays is kept the same as the selection for the branching fraction measurement, the same cut based selection in Section 4.3.2 is used and the BDTS requirement is applied. The differences in the selection are the trigger requirements, the mass ranges, the particle identification requirements and the use of multivariate classifiers. The differences are outlined in the following sections and the full selection criteria are summarised in Section 4.4.5.

4.4.1 Trigger

The trigger lines, L0Global, Hlt1Phys and Hlt2Phys, are used to selected $B_s^0 \to \mu^+\mu^-$ candidates however different decisions made by these lines are used compared to the selection of candidates for the measurement of the $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions. Candidates are required to be TOS or TIS at each level of the trigger. The change in trigger decisions used for the effective lifetime analysis is motivated by the dependence on simulated decays in the measurement. The selection used to identify candidates is not uniform across the decay time range and the selection efficiency as a function of decay time is needed to measure the lifetime as describe in Section 6.3. Simulated $B_s^0 \to \mu^+\mu^-$ decays are used in the determination of this efficiency and in simulated data candidates that are triggered by DEC decisions but are not included in TIS or TOS decisions are not well modelled. This arises due to the underlying event in simulated decays not accurately describing data. Therefore these candidates are not used in data to reduce the systematic uncertainties from the evaluation of the selection efficiency as a function of decay time. Candidates trigger by DEC decisions but not TIS or TOS do not pose the same problem for the branching fraction analysis because the selection and trigger efficiencies are evaluated using different methods as discussed in Section 5.4.

The analysis strategy used to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime is verified by measuring the lifetimes of the more abundant $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ decays. Slightly different trigger decisions are used to select $B \to h^+h^-$ decays but the same trigger lines are used. To be useful as a validation channel the efficiency of the trigger requirements as a function of the decay time will ideally be similar to the

$B_s^0 \to \mu^+\mu^-$ triggers, this is achieved by requiring $B \to h^+h^-$ decays to be TIS at each level of the trigger.

The requirements imposed on the trigger to select $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays is shown in Table 4.13.

Trigger Line	Trigger decision
Select $B_s^0 \to \mu^+\mu^-$ decays	
L0Global	TIS or TOS
Hlt1Phys	TIS or TOS
Hlt2Phys	TIS or TOS
Select $B \to h^+h^-$ decays	
L0Global	TIS
Hlt1Phys	TIS
Hlt2Phys	TIS

Table 4.13 Trigger lines used to select $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays for the $B_s^0 \to \mu^+\mu^-$ effective lifetime analysis.

4.4.2 Mass range

The mass of $B_s^0 \to \mu^+\mu^-$ candidates is restricted to the range 5320 - 6000 MeV/c^2, the motivation for the narrow mass window compared to the selection of candidates for the branching fraction analysis comes from the optimisation of the measurement strategy detailed in Section 6.4. The lower mass bound now lies on the low edge of the B_s^0 mass window, therefore $B^0 \to \mu^+\mu^-$ candidates and backgrounds from mis-identified $B \to h^+h^-$ and semi-leptonic decays are almost completely removed. The dominant background left in the data set is from combinatorial background.

Similarly $B \to h^+h^-$ decays used to verify the measurement strategy for the $B_s^0 \to \mu^+\mu^-$ effective lifetime have a reduced mass range compared to the selection of $B \to h^+h^-$ decays for the branching fraction measurements, $B \to h^+h^-$ decays must be in the mass range 5100 - 5500 MeV/c^2 in order to remove contributions from exclusive backgrounds.

4.4.3 Particle identification

The particle identification requirements used for selecting candidates for the measurement of the branching fraction analysis were optimised to give the greatest sensitivity to $B^0 \to \mu^+\mu^-$ decays. Backgrounds from $B \to h^+h^-$ and $\Lambda_b^0 \to p\mu^-\nu_\mu$ decays pollute the B^0 mass window and must be reduced as much as possible to enable good sensitivity of $B^0 \to \mu^+\mu^-$ decays. The requirement placed on the linear combination of ProbNN variables in Section 4.3.3 is a compromise between background rejection and signal efficiency. However for the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime, the B^0 mode is not relevant and the mass region of selected candidates removes the majority of $B^0 \to \mu^+\mu^-$ decays and backgrounds from $B \to h^+h^-$ and $\Lambda_b^0 \to p\mu^-\nu_\mu$ decays, therefore looser particle identification requirements can be used leading to a high efficiency to select signal decays.

The same linear combination of ProbNN variables, PID_μ, is used because there is still a small contribution from mis-identified $B \to h^+h^-$ and $\Lambda_b^0 \to p\mu^-\nu_\mu$ decays that can be reduced and in addition the particle identification requirements help to reduce the number of combinatorial background decays. Different ProbNN tunes and consequently cut values are used for Run 1 and 2015 data compared to 2016 data. The cuts are chosen to give similar efficiencies for each data set at selecting signal and removing background and are listed in Table 4.14. The cut values have not been optimised because there are too few candidates in data after the selection to enable this and simulated decays are not used because particle identification variables are not well modelled in simulation. However the particle identification requirement are still tight enough to make the expected number of mis-identified decays in the data set after the full selection negligible, as shown in Section 7.

The separation of $B \to h^+h^-$ decays into $B_s^0 \to K^+K^-$ and $B_s^0 \to K^+\pi^-$ decays is done using the $\mathrm{DLL}_{K\pi}$ variable. The DLL variables are useful to separate $B \to h^+h^-$ decays where h is either a pion or kaon because the variables compare different particle hypotheses with the pion hypotheses. The selection requirements used are given in Table 4.14 and are the same for each year of data taking.

Decay	Particle	PID requirements
$B_s^0 \to \mu^+\mu^-$ (Run 1 and 2015)	$\mu^\pm$	$\mathrm{PID}_\mu > 0.2$
$B_s^0 \to \mu^+\mu^-$ (2016)	$\mu^\pm$	$\mathrm{PID}_\mu > 0.4$
$B^0 \to K^+\pi^-$ and $B_s^0 \to K^+\pi^-$	K^+	$\mathrm{DLL}_{K\pi} > 10$
	$\pi-$	$\mathrm{DLL}_{K\pi} <$ -10
$B_s^0 \to K^+K^-$	K^+ and K^-	$\mathrm{DLL}_{K\pi} > 10$

Table 4.14 Particle identification requirements to select $B_s^0 \to \mu^+\mu^-$ decays and to separate the $B \to h^+h^-$ decays $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+\pi^-$ from $B_s^0 \to K^+K^-$.

4.4.4 Multivariate classifier

The selection of candidates for the measurements of the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction uses two multivariate classifiers to separate signal and combinatorial background decays. The BDTS is used first to remove candidates that are very unlikely to be signal and to reduce the size of the data set. The global BDT is then used to classify candidates in to bins of BDT and a simultaneous fit it then applied across the BDT bin to measure the branching fractions, this procedure is described in Chapter 5.

A different, simpler strategy is used to identify candidates for the measurement of the $B^0_s \to \mu^+\mu^-$ effective lifetime. Combinatorial backgrounds are reduced by placing a cut on the output of a multivariate classifier, then all candidates passing the selection cut are used to measure the effective lifetime. The measurement strategy that motivates this selection procedure is detailed in Chapter 6.

As a consequence of the different selection method, two classifiers may not be necessary for the measurement of the effective lifetime. Therefore alternative classifiers were developed in parallel to the development of the global BDT and the performance was computed to determine the most effective way to select candidates for the $B^0_s \to \mu^+\mu^-$ effective lifetime measurement.

4.4.4.1 Development of effective lifetime multivariate classifiers

Several different types of multivariate classifiers were investigated for the effective lifetime selection and boosted decision trees gave the best performance at separating signal and from background. A range of boosting methods for the decision trees were studied and the adaptive boosting method yielded the best results. However a boosting method of particular interest of the effective lifetime measurement was the uBoost technique **??**. The uBoost method produces a classifier output that has a uniform efficiency for a specified variable. The selection used to identify $B^0_s \to \mu^+\mu^-$ candidates uses variables that are correlated with the decay time of the B^0_s in the stripping selection and the most effective input variables for achieving good signal and background separation in BDTs are also correlated with the decay time. Therefore the overall selection efficiency varies as a function of decay time. The final measurement of the $B^0_s \to \mu^+\mu^-$ effective lifetime relies of the efficiency being well understood. This is described in detail in Section 6.3. If the output of a BDT is correlated with the B^0_s decay time the efficiency as a function of decay time may not have a smooth or easily understandable distribution, the uBoost method could provide a way to make modelling the efficiency as a function of decay time easier by requiring the algorithm output has a uniform efficiency across the decay time distribution.

Simulated 2012 $B_s^0 \to \mu^+\mu^-$ decays were used as the signal training sample for the BDTs and two different samples were used for the background training sample. One sample consisted of simulated 2012 $b\bar{b} \to \mu^+\mu^- X$ decays and the other was combinatorial background decays in Run 1 data. At the time of BDT development only Run 1 data was available. The selection requirements listed in Table 4.11, except the BDTS requirement, were applied to training samples of simulated decays and combinatorial background decays in data were identified as candidates that pass the selection requirements listed in Table 4.11, except the BDTS requirement and with by having a dimuon invariant mass in the range 5447 - 6500 MeV/c^2, outside to the B_s^0 mass window. The number of events in each training sample is given in Table 4.15.

Sample	Number of decays
Simulated $B_s^0 \to \mu^+\mu^-$	668292
Simulated $b\bar{b} \to \mu^+\mu^- X$	586586
Data	189077

Table 4.15 Number of candidates present in each training sample after the selection cuts have been applied. Simulated decays and decays in data a were identified as candidates that pass the selection requirements listed in Table 4.11, expect the BDTS cut was not applied and the decays in data must be in the mass range 5447 - 6000 MeV/c^2.

The input variables used in the adaptive boosting and uBoost BDTs were chosen separately starting from a large set of variables including kinematic, geometric and isolation variables. Initially the BDTs were trained using all input variables within the set and variables that had no impact on the BDT performance were removed until removing any of the remaining variables had a negative impact on the BDT performance. The final variable sets were different for the two boosting methods; the adaptive boosting BDT uses 11 input variables and the uBoost BDT uses 21 input variables. The full list of input variables used and the definition of those variables are given in Appendix **??**.

The performance of the BDTs with different boosting methods and trained using simulated decays and data as the background decays was evaluated using $B \to h^+h^-$ decays in data. No particle identification variables were used in the input variables of the BDTs due to the incorrect modelling of particle identification variables in simulated decays, therefore the performance on the BDTs on $B \to h^+h^-$ decays should be very similar to $B_s^0 \to \mu^+\mu^-$ decays. $B \to h^+h^-$ decays in data were identified by the same selection requirements used applied to the BDT training samples of simulated decays except the isMuon requirement was not applied and no particle identification requirements were used to separate different $B \to h^+h^-$ decays. An unbinned maximum likelihood fit was applied to the $B \to h^+h^-$ mass distribution, where all $B \to h^+h^-$ decays are reconstructed as $B_s^0 \to \mu^+\mu^-$, for a range cuts on the BDT outputs and the signal significance $\mathcal{S}$ was evaluated for each cut on the BDT output. In the maximum likelihood fit the $B \to h^+h^-$ mass distribution was modelled with a Gaussian function and the combinatorial background decays with an exponential, an example of the mass fit is given in Figure 4.11. The signal significance is given by

$$\mathcal{S} = \frac{S}{\sqrt{S+B}} \tag{4.4}$$

where S (B) are the number of signal (background) decays within 3 σ of the center of the $B \to h^+h^-$ mass peak. The signal significance as a function of the cut value placed on the BDT output for the different BDTs trained on the different background samples are shown in Figure 4.12. The outputs of the BDTs are not flattened, adaptive boosting BDT gives output values between -1 and +1 and uBoost BDT gives output values between 0 and +1. It is clear from Figure 4.12 that the performance of BDTs trained on simulated decays is better that that of the BDTs trained on data, this is due to the higher statistics available for simulated decays, as shown in Table 4.15. Furthermore the performance of adaptive boosting BDT is better that uBoost BDT which is expected because the adaptive BDT is not constrained to have a uniform efficiency across the decay time range. From now on only BDTs trained using simulated decays as the background training sample will be considered.

Both the adaptive and uBoost BDTs shown in Figure 4.12 were trained without applying the BDTS cut to the training samples, however the signal significance on $B \to h^+h^-$ decays has been reevaluated with the BDTS cut applied after the BDT training and before the BDT training. The improvement in the overall performances of the BDTs is small but applying the BDTS cut to $B \to h^+h^-$ after the BDT training gives the highest signal significances.

The training parameters of the adaptive BDT have been optimised by iterating over a large range of different values, whereas the training parameters of the uBoost were not optimised because changing the parameters has a small impact of the overall BDT performance []. The values used are given in Appendix **??**.

The understanding of the selection efficiency as a function of decay time was the main motivation for investigating the uBoost boosting method. The performance of

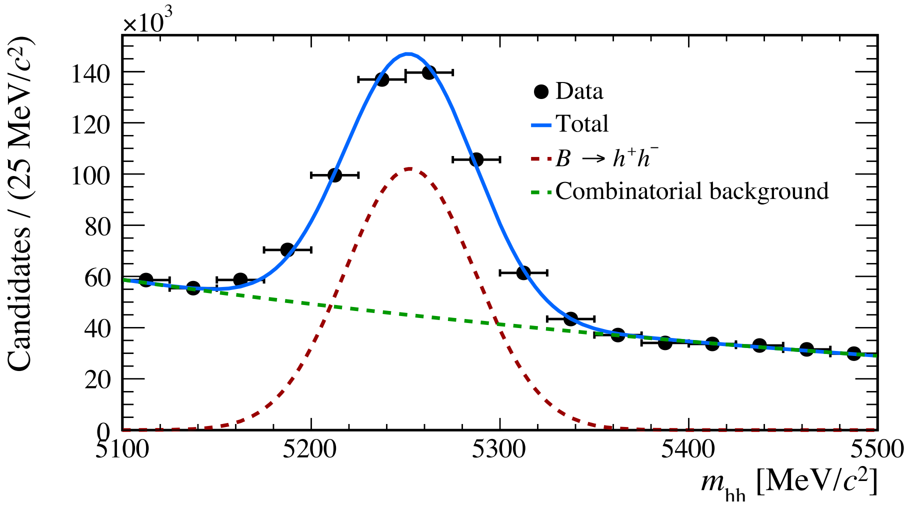


Fig. 4.11 Example of the mass fit to $B \to h^+h^-$ Run 1 data to find the signal significance for the Adaptive BDT with a cut value at 0.0 on the BDT output.

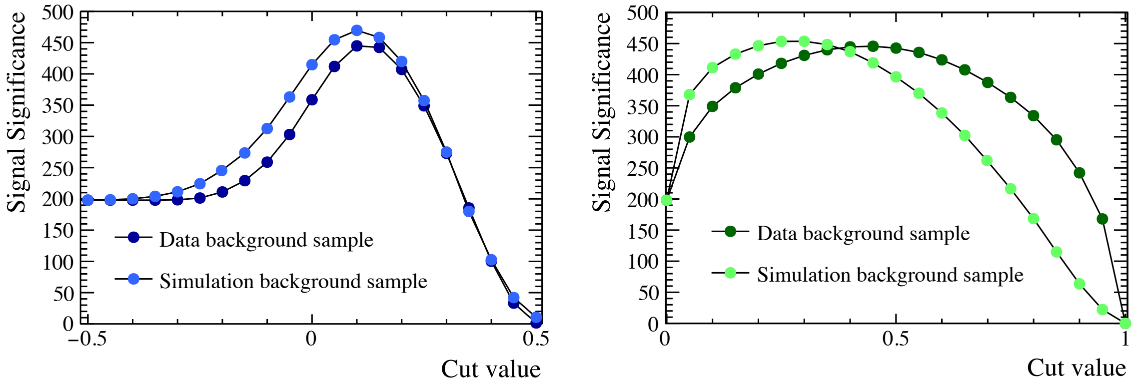


Fig. 4.12 Signal significance from $B \to h^+h^-$ decays in Run 1 data of the adaptive (left) and uBoost (right) BDTs trained using simulated decays and data as the background training samples.

the uBoost boosting method is worse than the adaptive BDT method. The selection efficiency as a function of decay time has been evaluated in simulated $B_s^0 \to \mu^+\mu^-$ decays after the all selection requirements and a range of different cut values on the outputs of the adaptive and uBoost BDTs trained on simulated decays. The cut values are chosen to have the same selection efficiencies for each algorithm. The efficiencies are shown in Figure 4.13, although the shapes are quite different the uniform efficiency of the uBoost BDT is evident. The adaptive BDT removes a greater proportion of decays at low decay time than the uBoost method. Ideally, to reduce systematic uncertainties on the measurement of the effective lifetime, the selection would not bias the decay time distribution, however with the expected statistics of the data set an unbiased selection would not be appropriate. However the efficiency of the adaptive boost BDT is smooth across the decay time range and therefore using that algorithm the efficiency as as function of decay time can be well modelled.

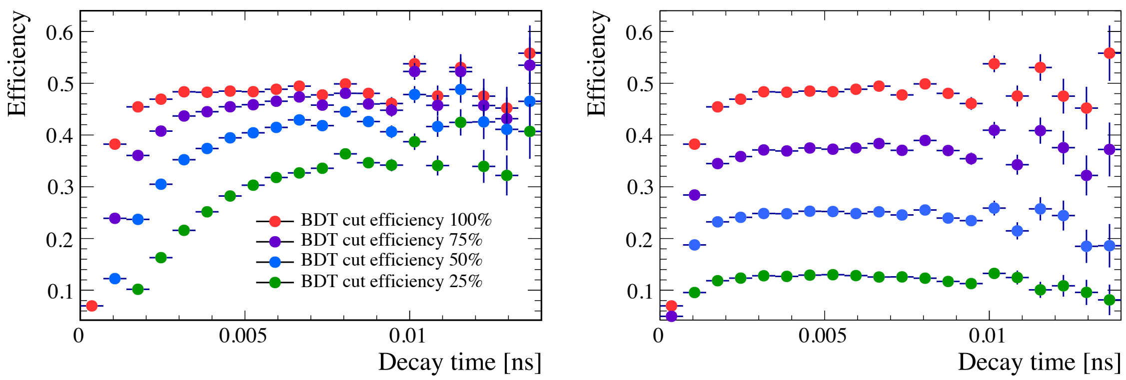


Fig. 4.13 Selection efficiency as a function of decay time of simulated 2012 $B_s^0 \to \mu^+\mu^-$ decays for the adaptive (left) and uBoost (right) BDTs. The selection requirements applied to the training sample are applied to the simulated decays and cuts are places on the BDT output so that the efficiency of the cut on decays passing the other selection requirements is 100 %, 75 %, 50 % and 25 %.

4.4.4.2 Classifier performance comparison

The final classifier used to select $B_s^0 \to \mu^+\mu^-$ candidates will be the BDT that has the greatest separation power between signal and combinatorial background decays consequently removing the most combinatorial background decays for a given signal efficiency. The performance of the two BDTs developed specifically for the effective lifetime measurement is compared to that of the global BDT used to classify candidates for the branching fraction measurements. Two different approaches are used to evaluate the performances. The output of the two BDTs developed for the effective lifetime selection have been flattened like the global BDT to enable easy comparison.

The signal significance for each BDT is evaluated on $B \to h^+h^-$ decays in Run 1 data and the maximum signal significance in found. The BDTS cut is applied in the selection process because the global BDT was designed to be used with the BDTS and the performance of the BDTs developed for the effective lifetime is best when the BDTS requirement is used. The results are shown in Figure 4.14, the global BDT produces the highest signal significance but is closely followed by the adaptive BDT developed for the effective lifetime measurement.

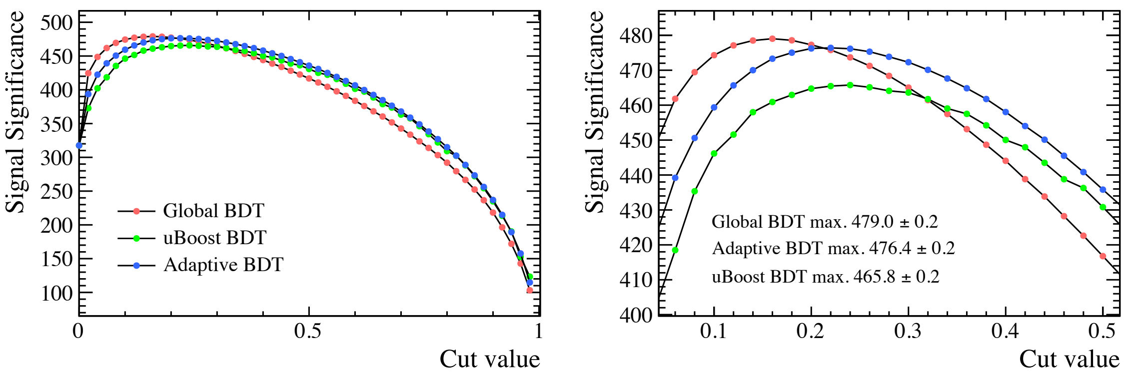


Fig. 4.14 Signal significance from $B \to h^+h^-$ decays in Run 1 data of the adaptive and uBoost BDTs trained using simulated decays as the background sample and the signal significance of the global BDT developed for the branching fraction measurement. The selection requirements listed in Table 4.11 are used, apart from the isMuon requirement and the mass range.

However since the purpose of the BDT is to remove combinatorial background decays passing the $B_s^0 \to \mu^+\mu^-$ selection an additional comparison of the different algorithms is made. The number of combinatorial background decays present in Run 1 data passing the $B_s^0 \to \mu^+\mu^-$ selection criteria for the effective lifetime measurement but in the mass range 5447 - 6550 MeV/c^2 are found for a range of cuts to the output of the three BDTs. The same cut values are applied to each BDT and since all the BDTs are flattened to have a uniform distribution of signal decays between 0 and 1 the cut values will have very similar efficiencies for each BDT. The results are given in Table 4.16 and the global BDT is most effective at removing background decays for a given signal efficiency.

Additionally all BDTs improve upon the performance of the classifier used in the last published analysis [30, 31].

BDT	Number events above BDT output value								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Global BDT	2261	597	229	89	34	13	4	1	0
Adaptive BDT	4623	1395	513	215	77	32	15	4	2
uBoost BDT	7904	3344	1535	630	268	92	27	7	0

Table 4.16 Number of candidates in Run 1 data passing the $B_s^0 \to \mu^+\mu^-$ effective lifetime selection and the BDTS cut in the mass range 5447 - 6000 MeV/c^2. The output of each BDT is flattened to have a uniform response between 0 and 1, therefore the cuts applied to each BDT will have approximately the same efficiency.

Although the global BDT combined with the BDTS performs best at separating signal from background decays, the efficiency as a function of decay time must also be evaluated for this algorithms to ensure it does not exhibit any strange behaviour which would make modelling the efficiency challenging. The decay time efficiency is shown in Figure 4.15 for several cut values on the BDT output and gives a smooth distribution as a function of decay time. For the data set used to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime the expected number of $B_s^0 \to \mu^+\mu^-$ decays is very low, therefore the benefits of using the uBoost method are outweighted by its poor performance. Therefore the global BDT developed for the the branching fraction is the best BDT to use for the selection of events for the $B_s^0 \to \mu^+\mu^-$ effective lifetime.

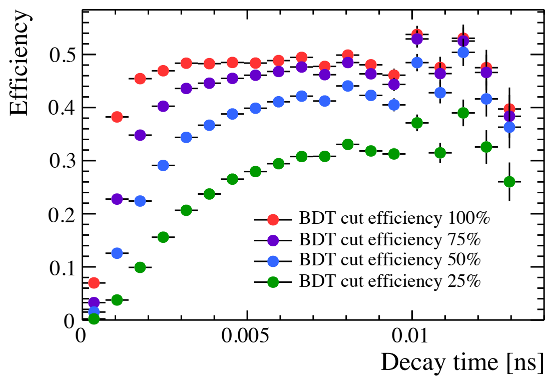


Fig. 4.15 Selection efficiency as a function of decay time of simulated 2012 $B_s^0 \to \mu^+\mu^-$ decays for the global BDT. The selection requirements applied to the training sample are applied to the simulated decays and cuts are places on the BDT output so that the efficiency of the cut on already selection event in 100 %, 75 %, 50 % and 25 %.

4.4.4.3 Optimisation of BDT cut choice

A cut is placed on the output of the global BDT to select $B_s^0 \to \mu^+\mu^-$ decays. The cut value has been optimised to give the smallest expected uncertainty on the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime, $\tau_{\mu\mu}$, and its inverse, $\tau_{\mu\mu}^{-1}$. This is done by using toy experiments for the expected number of $B_s^0 \to \mu^+\mu^-$ combinatorial background decays for different cuts on the global BDT output. The same cut value on the global BDT is used to select $B \to h^+h^-$ decays to verify the analysis strategy to measure $\tau_{\mu\mu}$.

The fit procedure to extract $\tau_{\mu\mu}$ from the data is described in depth in Chapter 6 along with a discussion of whether it is best to fit for $\tau_{\mu\mu}$ or $\tau_{\mu\mu}^{-1}$. The toy experiment used to optimise the global BDT cut value are preformed following the steps;

- the mass and decay time distribution for number of expected $B_s^0 \to \mu^+\mu^-$ and combinatorial background events are generated using the expected mass and decay time probability density functions

- an unbinned maximum likelihood fit is performed to the dimuon invariant mass spectrum, where the $B_s^0 \to \mu^+\mu^-$ and combinatorial background yields are free to float in the fit along with the slope of the combinatorial background mass distribution c

- the mass fit is used to compute sWeights using the sPlot method [34]

- a maximum likelihood fit is performed to the sWeighted decay time distribution to extract $\tau_{\mu\mu}$ and $\tau_{\mu\mu}^{-1}$.

Full details of the toy experiment set up and the probability density functions used are given in Appendix B.

The number of expected $B_s^0 \to \mu^+\mu^-$ and combinatorial background events for different BDT cut values is derived from the expected number of decays passing in the all the selection cuts and BDT > 0.55 for Run 1 and Run 2 data but in the mass range $4900 < m_{\mu^+\mu^-} < 6000$ MeV$/c^2$. These predictions assume the SM branching fraction for $B_s^0 \to \mu^+\mu^-$ and are given in Table 4.17. Since the output of the global BDT is flattened the number of $B_s^0 \to \mu^+\mu^-$ decays is evenly distributed across the BDT range, therefore the expected number of $B_s^0 \to \mu^+\mu^-$ decays is straight forward to calculated for each BDT cut value. The number of combinatorial background decays expected after each BDT is is computed from simulated $b\bar{b} \to \mu^+\mu^- X$ decays using the ratio

$$R = \frac{\epsilon(BDT > X)}{\epsilon(BDT > 0.55)} \tag{4.5}$$

where $\epsilon(BDT > X)$ is the efficiency of the cuts BDT > X. The $B_s^0 \to \mu^+\mu^-$ selection requirements are applied to the simulated decays before taking the efficiency. The ratios for the different cuts values are shown in Table 4.18. Simulated decays had to be used to compute the efficiencies rather than data because there were too few candidates left after the higher BDT cuts were applied to data to enable meaningful studies.

Decay	Expected number of candidates
$B_s^0 \to \mu^+\mu^-$	30.94
Combinatorial background	66.23
Total	97.17

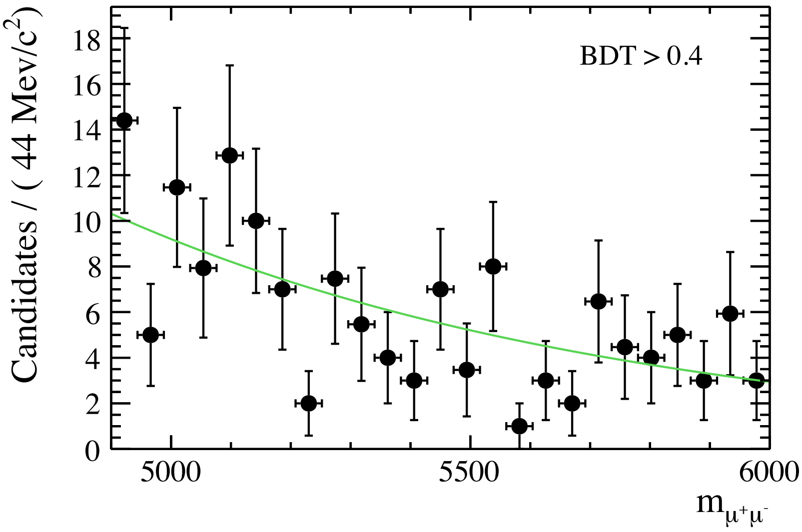
Table 4.17 Expected number of $B_s^0 \to \mu^+\mu^-$ and combinatorial background candidates after the $B_s^0 \to \mu^+\mu^-$ selection requirement and with a global BDT value greater than 0.55 in the mass range $4900 < m_{\mu^+\mu^-} < 6000$ MeV/c^2.

Global BDT cut	R_ϵ
0.40	8.69
0.45	3.91
0.50	1.91
0.55	1.00
0.60	0.55
0.65	0.32

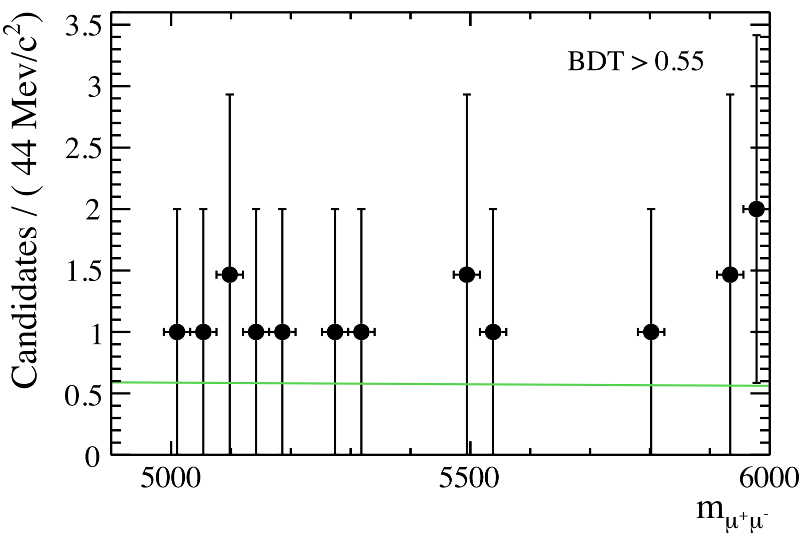
Table 4.18 The ratio of efficiencies of cuts on the global BDT to select $b\bar{b} \to \mu^+\mu^- X$ decays relative to a cut of 0.55 on the global BDT.

The mass distribution of the combinatorial background is a decaying exponential, it was observed from the simulated $b\bar{b} \to \mu^+\mu^- X$ decays that the slope of the mass distribution changed with the BDT cut value as illustrated in Figure 4.16. The change in slope is accounted for when generating events for the toy experiment by changing the slope parameter (λ) for each BDT cut. Table 4.19 shows the slope of the mass distribution for different BDT cuts values evaluated from $b\bar{b} \to \mu^+\mu^- X$ simulated decays.

The results from 10,000 toy experiments for BDT cut values every 0.05 in the range 0.4 - 0.65 are shown in Table 4.20 along with the expected number of $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays for each BDT cut value. The median uncertainty of the fit for $\tau_{\mu\mu}$ and $\tau_{\mu\mu}^{-1}$ are given along with the signal significance ($\mathcal{S} = S/\sqrt{S+B}$) for each BDT cut. The highest signal significance and lowest expected uncertainties occur



(a)



(b)

Fig. 4.16 Mass distribution of simulated decays after global BDT cuts of 0.4 and 0.55 and the $B_s^0 \to \mu^+\mu^-$ selection.

BDT cut	λ /c^{2}MeV^{-1}
0.40	-0.00114 $\pm$ 0.00028
0.45	-0.00129 $\pm$ 0.00041
0.50	-0.00132 $\pm$ 0.00060
0.55	-0.00004 $\pm$ 0.00089
0.60	-0.00000 $\pm$ 0.00114
0.65	-0.00024 $\pm$ 0.00122

Table 4.19 The slope of the combinatorial background mass distribution for different cut value on the global BDT evaluated from $b\bar{b} \to \mu^+\mu^- X$ simulated decays.

for a BDT cut of 0.55, therefore this cut value is used to select $B_s^0 \to \mu^+\mu^-$ decays. The same cut is applied to the global BDT to select $B \to h^+h^-$ decays.

Global BDT cut	$\frac{S}{\sqrt{S+B}}$	$\sigma\left(\tau_{\mu\mu}\right)$ / ps	$\sigma\left(\tau_{\mu\mu}^{-}\right)$ / ps $^{-1}$
0.40	3.87	0.345	0.128
0.45	4.51	0.309	0.114
0.50	4.85	0.291	0.108
0.55	4.94	0.285	0.106
0.60	4.86	0.297	0.109
0.65	4.65	0.309	0.115

Table 4.20 The signal significance for each cut value in the global BDT and the $\tau_{\mu\mu}$ and $\tau_{\mu\mu}^{-1}$ results from 10,000 toy experiment for the expected number of events.

4.4.5 Summary

The complete set of selection criteria used for identify $B_{(s)}^0 \to \mu^+\mu^-$ decays in Run 1 and Run 2 data for the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime are listed in Table 4.21.The selection requirements do not remove all backgrounds decays from the data set but reduce them to a level at which the effective lifetime can be measured. The selection criteria for $B \to h^+h^-$ to verify the measurement strategy is very similar to the selection used to identify $B_{(s)}^0 \to \mu^+\mu^-$ decays the differences are in the mass range used and the trigger and particle identification requirements.

Particle	$B_s^0 \to \mu^+\mu^-$
B_s^0 or B^+	5320 MeV/c^2 < M < 6000 MeV/c^2
	DIRA > 0
	FD χ^2 > 225
	IP χ^2 < 25
	Vertex χ^2/ndof < 9
	DOCA < 0.3 mm
	τ < 13.248 ps
	p_T > 500 MeV/c
Daughter μ or h	Track χ^2/ndof < 3 (4)
	Minimum IP χ^2 > 25
	0.25 GeV/c < p_T < 40 GeV/c
	p < 500 GeV/c
	ghost probability < 0.3 (0.4)
	$\lvert m_{\mu\mu} - m_{J/\psi}\rvert$ < 30 MeV/c^2
	isMuon = True
	PID$_\mu$ > 0.2 (0.4)
	BDTS > 0.05
	Global BDT > 0.55
Trigger requirements	L0Global = TIS or TOS
	Hlt1Phys = TIS or TOS
	Hlt2Phys = TIS or TOS

Table 4.21 Selection cuts applied to select $B_s^0 \to \mu^+\mu^-$, where selection is different between Run 1 and Run 2 the Run 2 values are shown in parenthesis.

Chapter 5

Measurement of $B^0_{(s)} \to \mu^+\mu^-$ branching fractions

This chapter presents the measurements of the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ branching fractions. Section 5.1 gives an overview of the analysis strategy and a description of how the $B^0_{(s)} \to \mu^+\mu^-$ yield is extracted from the data is given in Section 5.2. The estimation of the background decays present in that data set is detailed in Section 5.3 and the normalisation procedure to convert the number of observed $B^0_{(s)} \to \mu^+\mu^-$ decays in to the branching fractions for these decays is explained in Section 5.4. Finally the results are presented in Section 5.5.

The work presented in this Chapter was performed by the $B^0_{(s)} \to \mu^+\mu^-$ LHCb analysis group and is published here []. My contribution was maintaining the stripping selection used for this analysis and providing the ROOT files for contained the data and simulated events needed for the analysis development and measurements.

5.1 Analysis strategy

The $B^0_{(s)} \to \mu^+\mu^-$ branching fractions, $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$, are defined as the ratio of $B^0_{(s)} \to \mu^+\mu^-$ decays to the number of $B^0_{(s)}$ mesons created. However in reality not every $B^0_{(s)} \to \mu^+\mu^-$ decay produced in *pp* collisions will be within the LHCb detector acceptance or be reconstructed and pass the selection criteria of Chapter 4. Therefore the number of observed $B^0_{(s)} \to \mu^+\mu^-$ decays at LHCb is reduced by the efficiency, ϵ, of the detector, reconstruction and selection. The $B^0_{(s)} \to \mu^+\mu^-$ branching fractions can be given by

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = \frac{\mathcal{N}_{B^0_{(s)}\to\mu^+\mu^-}}{\mathcal{N}_{B^0_{(s)}}} = \frac{\mathcal{N}^{obs}_{B^0_{(s)}\to\mu^+\mu^-}}{\mathcal{N}_{B^0_{(s)}}} \tag{5.1}$$

where $\mathcal{N}_{B^0(s)\to\mu^+\mu^-(B^0_{(s)})}$ is the total number of $B^0_{(s)} \to \mu^+\mu^-$ decays ($B^0_{(s)}$ mesons) and $\mathcal{N}^{obs}_{B^0_{(s)}\to\mu^+\mu^-}$ the number of observed $B^0_{(s)} \to \mu^+\mu^-$ decays.

The number of $B^0_{(s)}$ created can be calculated from the integrated luminosity, $\mathcal{L}_{int}$, and the $b\bar{b}$ production cross-section, $\sigma_{b\bar{b}}$, via

$$\mathcal{N}_{B^0_{(s)}} = 2 \times \mathcal{L}_{int} \times \sigma_{b\bar{b}} \times f_{d(s)} \tag{5.2}$$

where $f_{d(s)}$ is the hadronisation factor, giving the probability for a b or $\bar{b}$ quark to form a B^0 (B^0_s) or a $\bar{B}^0$ ($\bar{B}^0_s$). The factor of 2 arises because no distinction is made between the $B^0_{(s)}$ and the $\bar{B}^0_s$. Although the number of $B^0_{(s)}$ can be computed this way the measured cross-section is not precisely known. Therefore to achieve a more precise branching fraction measurement an alternative approach is used. Another decay with a well know branching fraction is used to normalise the observed number of $B^0_{(s)} \to \mu^+\mu^-$ decays and obtain the branching fractions. The extraction of $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$ from the number of observed decays is therefore

$$\begin{aligned}\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) &= \frac{1}{\mathcal{B}_{norm}} \cdot \frac{f_{norm}}{f_{d(s)}} \cdot \frac{\epsilon_{norm}}{\epsilon_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{\mathcal{N}^{obs}_{B^0_{(s)}\to\mu^+\mu^-}}{\mathcal{N}^{obs}_{norm}} \\ &= \alpha_{d(s)} \cdot \mathcal{N}_{obs\,B^0_{(s)}\to\mu^+\mu^-}\end{aligned} \tag{5.3}$$

where *norm* indicates the normalisation channel. The normalisation factors can be combined into one normalisation parameter $\alpha_{d(s)}$ for each of the B^0_s and B^0 decays. The normalisation procedure removes the uncertainty from $\sigma_{b\bar{b}}$. Therefore the number of observed $B^0_{(s)} \to \mu^+\mu^-$ decays and the normalisation parameters, $\alpha_{d(s)}$, need to be evaluated to measure the branching fractions. The selection described in Chapter 4 allows $B^0_{(s)} \to \mu^+\mu^-$ candidates to be classified by their dimuon invariant mass and global BDT output. A simultaneous unbinned maximum likelihood fit is performed to the dimuon invariant mass distribution in 4 BDT bins to measure the observed number of $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ decays. The Run 1 and Run 2 data are kept separate and the fit is applied simultaneously to both data sets. To measure the number of $B^0_{(s)} \to \mu^+\mu^-$ decays knowledge is required of the mass shapes and the fraction of $B^0_{(s)} \to \mu^+\mu^-$ decays in each BDT bin and the number of background decays and their mass shapes in each bin. The mass shapes and fraction of $B^0_{(s)} \to \mu^+\mu^-$ decay in each BDT bin are described by probability density functions (*pdfs*). The evaluation of the $B^0_{(s)} \to \mu^+\mu^-$ mass *pdfs* and the fraction of decays in each BDT bin are described in Section 5.2. The expected number of background decays and their mass *pdfs* in each BDT bin are described in Section 5.3.

The binning choice used for the BDT is chosen to optimise both fit stability and sensitivity to the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions. The bin boundaries used are

$$[0.25, 0.4, 0.5, 0.6, 1.0]. \tag{5.4}$$

Candidates with BDT values between 0 and 0.25 are not included in the fit because this bin is dominated by backgrounds from random combinations of muons in the event. The inclusion of this bin does not improve the branching fraction sensitivity and reduces the stability of the fit.

The normalisation decay is chosen to be as similar as possible to $B^0_{(s)} \to \mu^+\mu^-$ decays to reduce systematic uncertainties introduced by different detection and selection efficiencies between signal and normalisation decays. Furthermore the chosen decay needs to be abundant so the the precision of the $B^0_{(s)} \to \mu^+\mu^-$ branching fraction measurements are not limited by the statistics available for the normalisation channel and it must have a precisely measured branching fraction, which is likely for abundant decays. Two decays are chose as normalisation channels; $B^+ \to J/\psi K^+$ where $J/\psi \to \mu^+\mu^-$ and $B^0 \to K^+\pi^-$. Both decays have large, precisely measured branching fractions and are similar to $B^0_{(s)} \to \mu^+\mu^-$ decays in complementary ways. The $B^+ \to J/\psi K^+$ decay has a very similar trigger efficiency to $B^0_{(s)} \to \mu^+\mu^-$ decays, due to the two muons from the J/ψ, but the extra particle in the final state leads to different selection and reconstruction efficiencies. The $B^0 \to K^+\pi^-$ decay has a very similar topology to $B^0_{(s)} \to \mu^+\mu^-$ therefore the selection and reconstruction efficiencies will be similar, but the trigger efficiencies for hadrons is quite different compared to muons.

The normalisation factors $\alpha_{d(s)}$ for $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ decays are evaluated independently for each normalisation channel and year of data taking, the factors are combined to produce an overall normalisation factor for Run 1 and Run 2. The evaluation of the normalisation factors is described in Section 5.4.

5.2 $B^0_s \to \mu^+\mu^-$ mass and BDT *pdfs*

5.2.1 Mass *pdfs*

The mass *pdfs* for $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ decays are modelled by a Crystal Ball function [35]. A Crystal Ball function is a Gaussian function that has an exponential tail on the low mass side to model radiative energy loss in the the final state. The parameters defining the function are the mean, μ, and resolution, σ of the Gaussian, the slope of the exponential, n, and a parameter α, defined in terms of σ, that determines the transition point between the Gaussian and the exponential function.

The parameters are evaluated using different methods:

- μ - the means of B^0 and B^0_s decays are evaluated separately from a fits to $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ decays in data
- σ - the resolution is extrapolated from the resolutions of quarkonia resonances. The resolutions for the J/ψ, $\Psi(2S)$ and $\Upsilon(1,2,3S)$ decaying into two muons are measured from a fits to data. The B^0 and B^0_s resolutions are then extrapolated from the observed relationship between quarkonia mass and resolution.
- n and α - these parameters are evaluated from the mass spectrum of $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ simulated decays where the mass distributions are smeared to have the same resolution as that measured from the quarkonia decays in data.

All parameters are evaluated separately for the B^0 and B^0_s for each year of data taking. The resulting parameter values are in good agreement across each year in the Run 1 and Run 2 data sets. The weighted average of the yearly parameters is used to produced the mass *pdfs* for Run 1 and Run 2 and are given in Tables 5.1 and 5.2 .

Parameter	$B^0 \to \mu^+\mu^-$	$B^0_s \to \mu^+\mu^-$
$\mu/\mathrm{MeV}/c^2$	$5284.73 \pm 0.15_{stat} \pm 0.27_{syst}$	$5372.05 \pm 0.16_{stat} \pm 0.36_{syst}$
$\sigma/\mathrm{MeV}/c^2$	$22.68 \pm 0.05_{stat} \pm 0.39_{syst}$	$23.07 \pm 0.05_{stat} \pm 0.39_{syst}$
n	1.141 ± 0.026	1.156 ± 0.013
α	2.054 ± 0.013	2.053 ± 0.007

Table 5.1 Parameter values for Crystal Ball functions used to describe the $B^0_{(s)} \to \mu^+\mu^-$ mass *pdf* for Run 1.

Parameter	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$
$\mu/\mathrm{MeV}/c^2$	$5279.95 \pm 0.13_{stat} \pm 0.08_{syst}$	$5367.34 \pm 0.14_{stat} \pm 0.35_{syst}$
$\sigma/\mathrm{MeV}/c^2$	$22.46 \pm 0.08_{stat} \pm 0.41_{syst}$	$22.85 \pm 0.08_{stat} \pm 0.42_{syst}$
n	1.118 ± 0.014	1.110 ± 0.017
α	2.063 ± 0.007	2.062 ± 0.008

Table 5.2 Parameter values for Crystal Ball functions used to describe the $B_{(s)}^0 \to \mu^+\mu^-$ mass *pdf* for Run 2.

5.2.2 BDT *pdfs*

The global BDT distribution for $B_{(s)}^0 \to \mu^+\mu^-$ decays is expected to be uniform between 0 and 1 as designed by the flattening procedure described in Section **??**. The fraction of $B_{(s)}^0 \to \mu^+\mu^-$ decays in a BDT bin should simply be proportional to the bin with. However the global BDT was trained and flattened using simulated decays, therefore to avoid differences between simulated decays and data affecting the expected fraction of $B_{(s)}^0 \to \mu^+\mu^-$ decays in each BDT bin, the BDT *pdf* is evaluated from data. This process is know as the BDT calibration. The global BDT is designed to use only kinematic and geometric information to classify candidates and includes no PID information. Therefore the BDT distributions of $B \to h^+h^-$ decays will be the same as $B_{(s)}^0 \to \mu^+\mu^-$ decays. $B^0 \to K^+\pi^-$ decays are used to calibrate the BDT response because it is the most abundant $B \to h^+h^-$ decay.

The number of $B^0 \to K^+\pi^-$ decays is extracted from data using maximum likelihood fits in each BDT bin for each year of data taking. The $B^0 \to K^+\pi^-$ candidates must pass the standard $B \to h^+h^-$ selection in Table **??** and are separated from other $B \to h^+h^-$ modes using $\mathrm{DLL}_{K\pi}$ variable. To reduce the difference in the trigger efficiency between and $B_{(s)}^0 \to \mu^+\mu^-$ decays, $B^0 \to K^+\pi^-$ candidates are required to be TIS at the L0 and Hlt1 but TOS at Hlt2 to ensure enough statistics.

The particle identification and trigger efficiencies are different for $B^0 \to K^+\pi^-$ and $B_{(s)}^0 \to \mu^+\mu^-$ decays. Therefore the $B^0 \to K^+\pi^-$ yields in each BDT bin are corrected for using by the different trigger and particle identification efficiencies. The same calibration is used for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays. The calibration is preformed for each year separately then combined to give the Run 1 and Run 2 fractions per BDT bin. Figure 5.1 shows the BDT distribution for $B_{(s)}^0 \to \mu^+\mu^-$ decays calibrated with $B^0 \to K^+\pi^-$ data for Run 1 and Run 2.

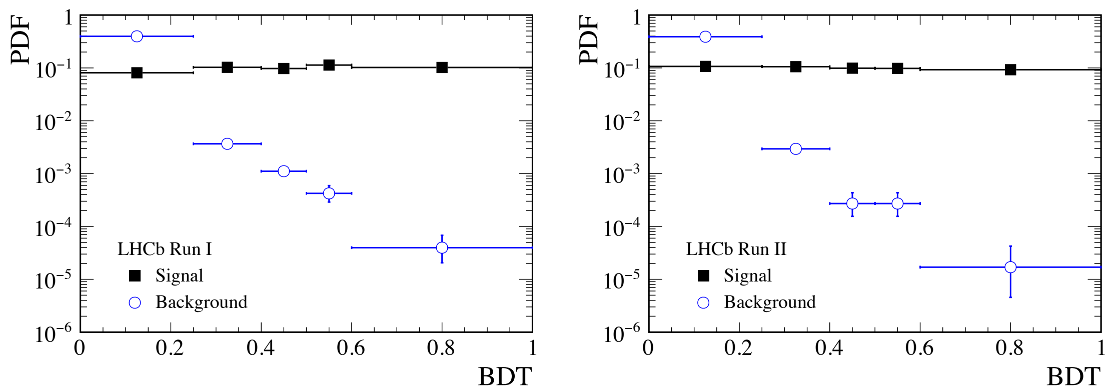


Fig. 5.1 $B^0_{(s)} \to \mu^+\mu^-$ BDT *pdfs* (black squares) for Run 1 and Run 2 data calibrated on $B^0 \to K^+\pi^-$ decay and the combinatorial background decays (blue circles) for $B^0_{(s)} \to \mu^+\mu^-$ candidates in data with a dimuon mass above 5477 MeV/c^2.

5.2.3 Decay time dependence

The output of the global BDT for $B^0_{(s)} \to \mu^+\mu^-$ is correlated with the $B^0_{(s)} \to \mu^+\mu^-$ decay time due the the choice of input variables used in the BDT as listed in Section **??**. This correlation will lead of slightly incorrect estimations of the $B^0_s \to \mu^+\mu^-$ BDT *pdf*. In the Standard Model the $B^0_s \to \mu^+\mu^-$ effective lifetime, $\tau_{\mu\mu}$, is equal to the lifetime of the heavy B^0_s mass eigenstate, τ_H, however in reality $\tau_{\mu\mu}$ could be somewhere in between the lifetimes of the heavy and light mass eigenstates. As described in Chapter (*the Theory Chapter*) the $B^0_s \to \mu^+\mu^-$ effective lifetime is related to the parameter $A_{\Delta\Gamma}$, where $A_{\Delta\Gamma}$ = +1 for $\tau_{\mu\mu} = \tau_H$ and $A_{\Delta\Gamma}$ = -1 for $\tau_{\mu\mu} = \tau_L$, where τ_L is the lifetime of the light $B^0_s \to \mu^+\mu^-$ mass eigenstate.

The simulated decays used to train and flatten the global BDT use as the $B^0_s \to \mu^+\mu^-$ lifetime the mean of the measured τ_H and τ_L values at the time of production. Therefore the lifetime used is different between simulation versions. Since the BDT output is correlated with the lifetime the fraction of $B^0_s \to \mu^+\mu^-$ decays in each BDT bin will depend on the lifetime used in the simulation. Numerical correction factors are computed for each year to scale the fraction of $B^0_s \to \mu^+\mu^-$ decays in each BDT bin for the situations where = -1, 0 or +1, so that the dependence on $A_{\Delta\Gamma}$ of the measured branching fractions can be evaluated.

No corrections are needed for $B^0 \to \mu^+\mu^-$ because the difference in lifetime of the heavy and light B^0 mass eigenstates is negligible and the need for correction cancels out with the BDT calibration that uses the B^0 decay $B^0 \to K^+\pi^-$.

5.3 Background mass *pdfs* and expected yields

The selection described in Chapter 4 is effective at reducing the backgrounds in the data set to a suitable level so that number of the $B^0_{(s)} \to \mu^+\mu^-$ decays can be measured. However background decays are present in the final data set, these cannot be completely removed without drastically reducing the efficiency to select signal decays. The backgrounds present in the data set must be included in the fit to the dimuon invariant mass in order to accurately measure the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions. The backgrounds present in the final data set come from;

- $B \to h^+h^-$ decays (where h = K, π) when both hadrons are mis-identified as muons because the hadrons decay during their flight through the detector after leaving the VELO. This background falls within the B^0 mass window but not the B^0_s mass window[1] due to the missing energy from the undetected neutrino.
- semi-leptonic decays where one hadron is mis-identified as a muon that include;
  - $B^0 \to \pi^-\mu+\nu_\mu$ and $B^0_s \to K^-\mu+\nu_\mu$ decays where the final state hadrons are mis-identified as muons. The mass of these backgrounds falls below the B^0 mass window in the left mass sideband
  - $\Lambda^0_b \to p\mu^-\nu_\mu$ decays when the proton is mis-identified as a muon. The large mass of the Λ_d means that this background pollutes the B^0_s and B^0 mass windows and below these windows
- semi-leptonic decays where muons in the decay form a good vertex that include;
  - $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ decays where the pion is not detected. The missing hadron means that these backgrounds fall well below the B^0 mass window.
  - $B^+_c \to J/\psi\mu^+\nu_\mu$ decays where $J/\psi \to \mu^+\mu^-$. The large mass of the B^+_c causes this background of cover the full mass range 4900 - 6000 MeV/c^2
- combinatorial background formed by the random combination of any two muons in the event, this background is distributed across the full mass range

In the fit to the dimuon invariant mass the combinatorial background is modelled by an exponential function. The combinatorial background yield is not constrained in the fit and the slope is constrained to have the same value across all BDT bins for each data set. These parameters are determined from a simultaneous fit to candidates

[1] B^0 and B^0_s mass windows are defined as $\pm$ 60 MeV/c^2 of the B^0 and B^0_s masses.

in data in BDT bins for the mass ranges [4900, $(m_{B^0} - 50)$] MeV/c^2 and [$(m_{B^0_0} + 60$, 6000] MeV/c^2, where the mass shapes and yields of the remaining backgrounds are constrained. The mass *pdfs* and yields of the background from $B \to h^+h^-$ and semi-leptonic decays are constrained in the fit around the expected values. The backgrounds that have lower masses than the B^0 and B^0_s must be accurately modelled in the fit to ensure the combinatorial background yield, that spans the full mass range, is accurate described within the signal mass windows. The approaches for finding the mass *pdfs* and expected yields differ for $B \to h^+h^-$ and semi-leptonic backgrounds, these procedures are described in the following sections.

5.3.1 $B \to h^+h^-$

The mass *pdf* describing mis-identified $B \to h^+h^-$ decays is formed of two Crystal Ball functions. The parameter values are evaluated from simulated decays for $B^0 \to K^+\pi^-$, $B^0_s \to K^+K^-$, $B^0 \to \pi^+\pi^-$ and $B^0_s \to K^+\pi^-$ the have the momenta of tracks smeared to model the hadrons decaying in flight. The parameters are evaluated separately for each decay and combined using the branching fractions and the particle identification efficiencies for each decay.

The number of mis-identified $B \to h^+h^-$ decays in each BDT bin, $\mathcal{N}_{B\to hh\to\mu\mu}$, is found using the relationship

$$\mathcal{N}_{B\to hh\to\mu\mu} = \epsilon^{TRIG}_{B^0_{(s)}\to\mu^+\mu^-} \cdot \frac{\mathcal{N}_{B\to hh}}{\epsilon^{TRIG}_{B\to hh}} \cdot \epsilon_{B\to hh\to\mu\mu} \tag{5.5}$$

where $\mathcal{N}_{B\to hh}$ is the number of TIS $B \to h^+h^-$ decays in data, $\epsilon^{TRIG}_{B^0_{(s)}\to\mu^+\mu^-, B\to hh}$ are the $B^0_{(s)} \to \mu^+\mu^-$ and $B \to h^+h^-$ trigger efficiencies and $\epsilon_{B\to hh\to\mu\mu}$ is the probability that a $B \to h^+h^-$ decays is mis-identified as $B^0_{(s)} \to \mu^+\mu^-$. The number of $B \to h^+h^-$ decays triggered as TIS is calculated for the full BDT range from the number of $B^0 \to K^+\pi^-$ decays in data corrected for the expected fraction of $B \to h^+h^-$ decays is mode occupies. Apart form the trigger and particle identification requirements the same selection is used for $B^0 \to K^+\pi^-$ decays as $B^0_{(s)} \to \mu^+\mu^-$, therefore only the trigger and particle identification efficiencies are corrected for. The efficiencies are calculated using a combination of data and simulated decays for each BDT bin and the same BDT *pdf* as $B^0_{(s)} \to \mu^+\mu^-$ decays is assumed for $B \to h^+h^-$ decays.

5.3.2 Semi-leptonic decays

The mass *pdfs* of semi-leptonic backgrounds vary across the BDT range therefore these *pdfs* are evaluated using simulated decays separated into each BDT bin. An Argus function [36] convoluted with an Gaussian Function is used to describe the mass distributions. The shapes of $B^0 \to \pi^-\mu{+}\nu_\mu$ and $B_s^0 \to K^-\mu{+}\nu_\mu$ are extremely similar and therefore these backgrounds are modelled with one common *pdf*. Similarly one mass *pdf* is used to model $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ decays.

The expected yields of the semi-leptonic backgrounds in each BDT bin is estimated by normalising to the number of $B^+ \to J/\psi K^+$ decays observed via

$$\mathcal{N}_x^{exp} = \mathcal{N}_{B^+\to J/\psi K+} \cdot \frac{f_x}{f_u} \cdot \frac{\mathcal{B}_x}{\mathcal{B}_{B^+\to J/\psi K^+}} \cdot \frac{\epsilon_x}{\epsilon^{B^+\to J/\psi K^+}} \tag{5.6}$$

where x represents each background decay. The background estimation can be factorised as

$$\mathcal{N}_x^{exp} = \beta \cdot f_x \cdot \epsilon_x \cdot \mathcal{B}_x \tag{5.7}$$

where β combines the background yield, detection and selection efficiency and hadronisation factors of $B^+ \to J/\psi K^+$ decays, it is the same for all backgrounds. The β term is evaluated using the same method as the normalisation of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions described in Section 5.4. The efficiencies and yields are evaluated across the full BDT range whereas the detection and selection efficiency of each background, ϵ_x, are evaluated separately for each BDT bin from information from both data and simulated decays. The hadronisation factors and branching fractions are specific to each background and were possible measured, rather than predicted, branching fractions are used.

5.4 Normalisation

As introduced earlier the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are measured by normalising the number of observed $B^0_{(s)} \to \mu^+\mu^-$ decays to the number of observed $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ decays. The normalisation parameters $\alpha_{d(s)}$, in equation 5.3 for $B^0_{(s)} \to \mu^+\mu^-$ decays depend on the yields of the normalisation decays, the ratio of the detection and selection efficiencies and the hadronisation factors. The evaluation of each of these terms are described in the following sections. In addition to the normalisation decay $B_s^0 \to J/\psi\phi$ decays are used to check to normalisation parameters used to measure the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions therefore the yield of decays and the detection and selection efficiencies must also be evaluated. This is done in the same way as the normalisation channels.

5.4.1 $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ yields

The yields of $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ decays, $\mathcal{N}^{obs}_{norm}$, are calculated from data using maximum likelihood fits to each year of data taking. The $B^+ \to J/\psi K^+$ mass *pdf* is modelled by an Ipathia function [] and the fit includes components for combinatorial background and $B^+ \to J/\psi\pi^+$ decays that are mis-reconstructed as $B^+ \to J/\psi K^+$. The mass *pdf* parameters are determined from both data and simulated decays. The $B^0 \to K^+\pi^-$ yields are calculated in the same way at the BDT calibration and the same trigger requirements are used. However for the normalisation the total number of $B^0 \to K^+\pi^-$ decays across the full BDT range is needed rather the bin-by-bin yields. Figure 5.2 and 5.3 show the mass fits used to calculate the Run 1 and Run 2 $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ yields.

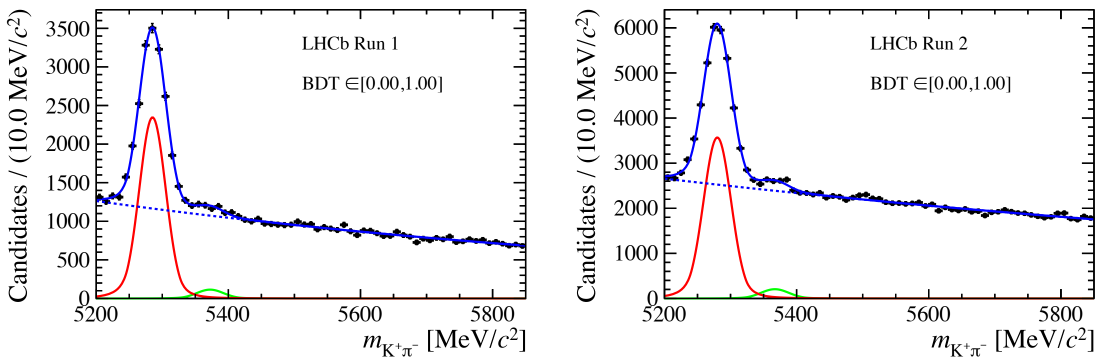


Fig. 5.2 Mass fit to measure $B^0 \to K^+\pi^-$ yield for the normalisation for Run 1 (left) and Run 2 (right) data.

5.4.2 Efficiency ratio

The efficiency ratio in equation 5.3 is split into several separate efficiency terms

$$\frac{\epsilon_{norm}}{\epsilon_{B^0_{(s)}\to\mu^+\mu^-}} = \frac{\epsilon^{Acc}_{norm}}{\epsilon^{Acc}_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{\epsilon^{RecSel|Acc}_{norm}}{\epsilon^{RecSel|Acc}_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{\epsilon^{Trig|RecSel}_{norm}}{\epsilon^{Trig|RecSel}_{B^0_{(s)}\to\mu^+\mu^-}} \tag{5.8}$$

for the detector acceptance, ϵ^{Acc}, reconstruction and selection efficiencies, $epsilon^{RecSel|Acc}$, and the trigger efficiency, $\epsilon^{Trig|RecSel}$.

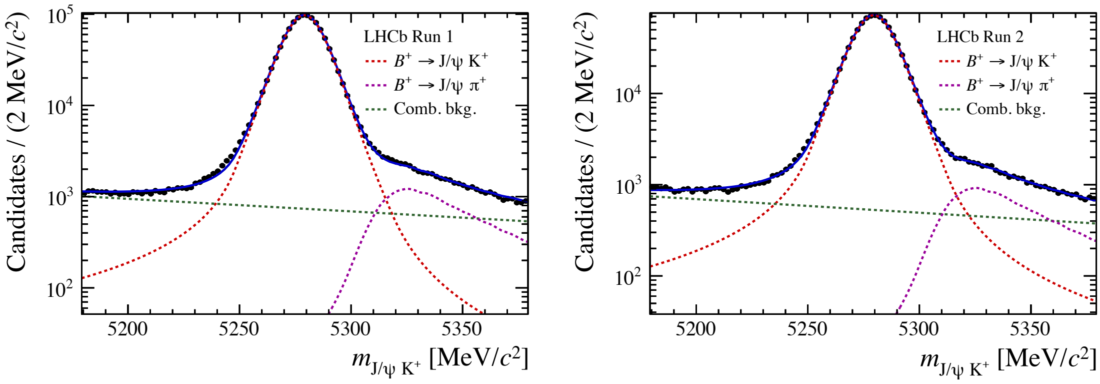


Fig. 5.3 Mass fit to measure $B^+ \to J/\psi K^+$ yield for the normalisation for Run 1 (left) and Run 2 (right) data.

The detector acceptance efficiency gives the efficiency for the decay products to be within the LHCb detector acceptance. This efficiency is evaluated on simulated decays to for decay products that fall within the range [10,400] mrad. The range is chosen to be slightly larger than the detector acceptance so that particles recovered by the magnetic field are included. To keep this efficiency similar for $B^0_{(s)} \to \mu^+\mu^-$ and $B^0 \to K^+\pi^-$ decays, the hadrons from $B^0 \to K^+\pi^-$ are required to be within the muon detector acceptance.

The reconstruction and selection efficiencies are calculated as the reconstruction efficiency of decays that are within the detector acceptance and the selection efficiency of reconstructed decays. The selection and reconstruction efficiencies are evaluated from a combination of information from data and simulated decays to ensure accurate selection efficiency ratios. Similar to the fraction of $B^0_s \to \mu^+\mu^-$ in each BDT bin, a correction is applied for the lifetime used in simulated $B^0_s \to \mu^+\mu^-$ decays assuming $A_{\Delta\Gamma} = +1$.

The trigger efficiencies for decays passing the reconstruction and selection are evaluated for each decay by data driven methods as descried in **??**.

The efficiencies are calculated for $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ separately to account for difference in the decays and kinematics. The ratio of efficiencies between signal and normalisation channels in the normalisation parameters ensures that systematic uncertainties arising from the use of simulated decays cancel out and will not effect the precision of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions.

5.4.3 Hadronisation factors

The normalisation factors depend on the hadronisation factors, f_u, f_s, f_d, that give the probability of a b or $\bar{b}$ quark to form a B^+, B^0_s or B^0, respectively. The hadronisation factors f_d and f_u are equal therefore the $B^0 \to \mu^+\mu^-$ branching fraction does not depend on hadronisation factors. For the $B^0_s \to \mu^+\mu^-$ the ratio f_s/f_d is used in the normalisation, since $f_d = f_u$. This ratio was measured at LHCb for pp collisions at $\sqrt{s} = 7$ TeV, and it is used for the different LHC $\sqrt{s}$ energies. However for Run 2 the f_s/f_d ratio must be modified fora small observed relative production difference. The uncertainty on the hadronisation factor ratio contributes the largest uncertainty to the $B^0_s \to \mu^+\mu^-$ branching fraction. Alternatively the $B^0_s \to \mu^+\mu^-$ decay could be normalised using a different B^0_s decay however the precision of the measured branching fractions and abundance of B^0_s decays, such as $B^0_s \to J/\psi\phi$, are not high enough at present to provide a lower overall uncertainty on the measured branching fraction.

5.4.4 Normalisation parameters

The yields, efficiencies and hadronisation factors are combined to produce separate normalisation factors for each year of data taking and each normalisation channel. The consistency of the efficiencies and yields for each normalisation channel are checked for each year by comparing the branching fraction ratios $\mathcal{B}(B^0 \to K^+\pi^-)/\mathcal{B}(B^+ \to J/\psi K^+)$ and $\mathcal{B}(B^+ \to J/\psi K^+)/\mathcal{B}(B^0_s \to J/\psi\phi)$ with the PDG values. The yearly normalisation factors are combined for each channel using to produce the normalisation factors for Run 1 and for Run 2 taking into account correlations between the parameters. A weighted average of the normalisation factors for $B^0 \to K^+\pi^-$ and is used to produce the overall normalisation factors for Run 1 and Run 2 as shown in Table 5.3.

Normalisation Paramters	Run 1	Run 2
$\alpha_d \times 10^{11}$	2.877 ± 0.101	3.521 ± 0.155
$\alpha_s \times 10^{10}$	1.071 ± 0.072	1.306± 0.095

Table 5.3 Noramlisation parameters for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ for Run 1 and Run 2.

5.5 Results

As described earlier in Section 5.1 the $B_s^0 \to \mu^+\mu^-$ and $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions are measured by a simultaneous maximum likelihood fit to the dimuon invariant mass of the Run 1 and Run 2 data sets, each divided into four BDT bins.

In the fit the mass *pdfs* and fraction of $B_{(s)}^0 \to \mu^+\mu^-$ decays in each BDT bin are constrained within Gaussian limits using the expected values and uncertainties. The yield of the combinatorial background is left free in the fit in each BDT bin and the slope of the mass distribution is constrained to have the same value across all bins for each data set. The yields of the backgrounds from $B \to h^+h^-$, $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$, $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$ and $B_c^+ \to J/\psi\mu^+\nu_\mu$ decays in each BDT bin are constrained around the expected values, similarly to the signal fractions but the mass shapes are fixed in the fit.

The branching fraction results from the fit are;

$$
\begin{aligned}
\mathcal{B}(B_s^0 \to \mu^+\mu^-) &= (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9} \\
\mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.5^{+1.2+0.2}_{-1.0-0.1}) \times 10^{-10}
\end{aligned}
\tag{5.9}
$$

Figure 5.4 shows the fit results for $B_{(s)}^0 \to \mu^+\mu^-$ candidates in the 4 BDT bins for both Run 1 and Run 2 data and Figure **??** the 2-dimensional likelihood profile for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ branching fraction measurements. The statistical significance of the $B_s^0 \to \mu^+\mu^-$ signal is 7.8 σ making this measurement the first single experiment observation of the $B_s^0 \to \mu^+\mu^-$ decay. While the significance of the $B^0 \to \mu^+\mu^-$ signal is less at 1.6σ, therefore the CLs method [] is used to place an upper limit on the branching fraction of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.4 \times 10^{-10}$ at the 95 % confidence level.

The quoted $B_s^0 \to \mu^+\mu^-$ branching fraction assumes the Standard Model value for $A_{\Delta\Gamma}$, applying the corrections detailed in Section 5.2.3 for $A_{\Delta\Gamma}$ values of 0 and -1 shift the central value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ by 4.6 % and 10.9 %, respectively. All results are consistent with the predictions of the Standard Model.

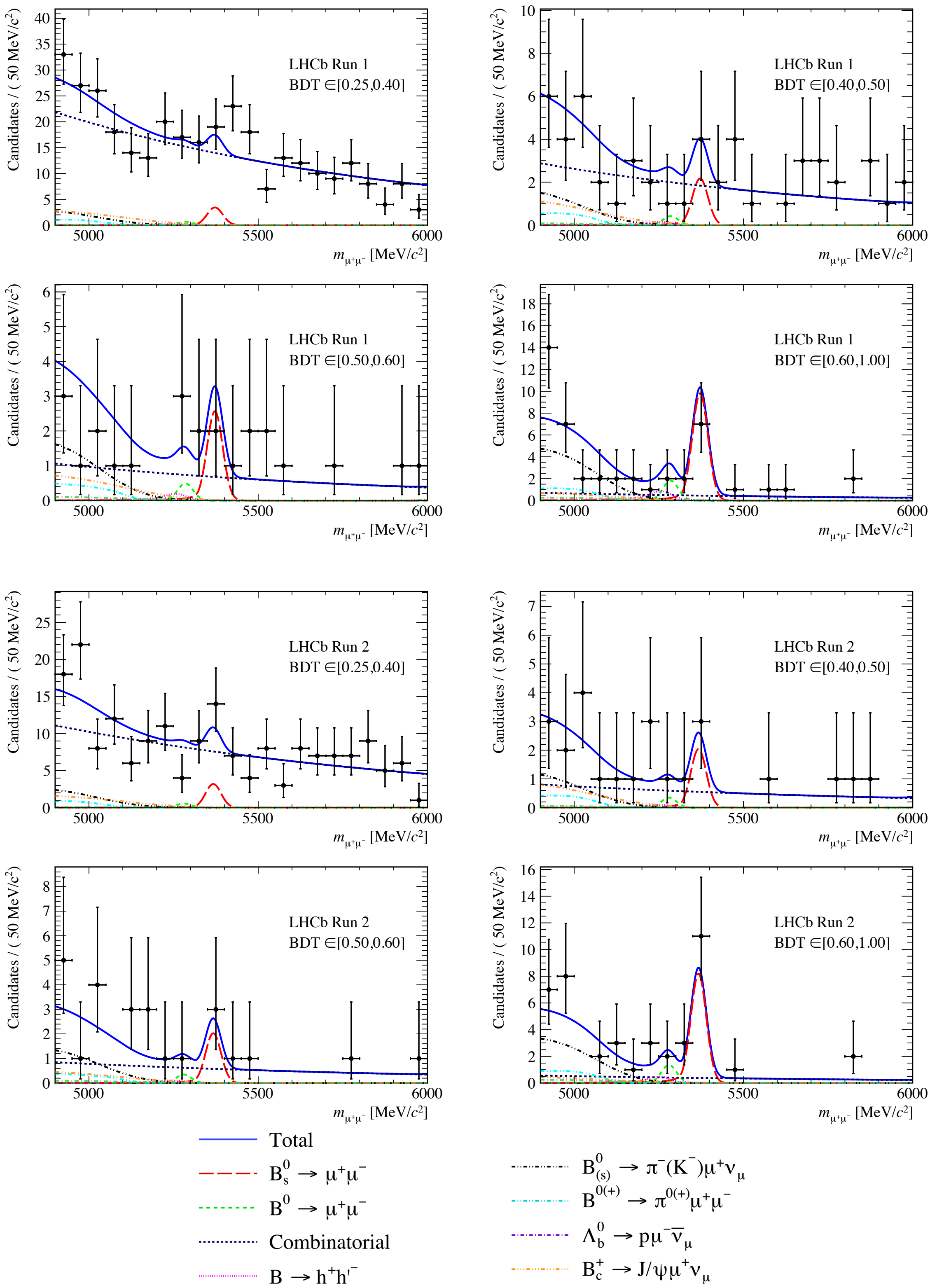


Fig. 5.4 Mass distribution in BDT bins for selected $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ candidates with the fit overlaid for Run 1 and Run 2 data. The fit includes components for $B^0 \to \mu^+\mu^-$, $B^0_s \to \mu^+\mu^-$, combinatorial backgrounds, mis-identified $B \to h^+h^-$ decays and backgrounds from semi-leptonic decays.

Chapter 6

Measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime

This chapter describes the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime. Section 6.1 presents an overview of the analysis strategy used to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime from data, the mass and decay time distributions of $B_s^0 \to \mu^+\mu^-$ decays and backgrounds passing the selection must to known for the optimisation of the analysis strategy and the measurement of the effective lifetime. The *pdfs* of the mass and decay time distributions are described in Sections 6.2 and 6.3 for signal and background decays. Due to the very rare nature of $B_s^0 \to \mu^+\mu^-$ decays, the measurement strategy has been optimised to produce the lowest expected uncertainty on the measured $B_s^0 \to \mu^+\mu^-$ effective lifetime, the optimisation studies are detailed in Section 6.4. Finally the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime is presented in Section 6.5.

The work presented in this chapter was completed for this thesis except the areas where the same method as the branching fraction analysis is used. This includes the background mass *pdf* evaluation and the expected signal and background yields for the data set, as well as the yields from data.

6.1 Analysis strategy

The $B_s^0 \to \mu^+\mu^-$ effective lifetime is measured from the decay time distribution of $B_s^0 \to \mu^+\mu^-$ candidates passing the selection criteria described in Section **??**. However the selection requirements do not completely separate real $B_s^0 \to \mu^+\mu^-$ decays from the backgrounds, therefore to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime either the *pdfs* describing the decay time distributions of signal and backgrounds must be know or the background candidates must be removed from the data set leaving only the signal distribution. Several approaches were investigated to determine which would produce stable results for the measured $B_s^0 \to \mu^+\mu^-$ effective lifetime on 4.4 fb^{-1} and yield the smallest expected statistical uncertainty on the result. The most successful approach uses the sPlot statistical method as described in [34] that provides a way to statistically untangle the signal and background distributions in a data set.

This method produces a two step strategy to measure the effective lifetime. This first step is a unbinned maximum likelihood fit to the dimuon invariant mass spectrum, where components are included in the *pdf* for $B_s^0 \to \mu^+\mu^-$ decays and each background decay. The mass fit measures the yields of the signal and background decays and from the fit sWeights are calculated for each component in the mass fit. The second step is to apply the sWeights of $B_s^0 \to \mu^+\mu^-$ decays to the data set, effectively removing all background decays, and perform an unbinned maximum likelihood fit to the signal weighted decay time distribution to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime. In the final fit only the $B_s^0 \to \mu^+\mu^-$ decay time *pdf* is needed to measure the effective lifetime. Due to the low statistics expected for the data set, Run 1 and Run 2 data are combined and the maximum likelihood fit to the mass and weighted decay time distributions are performed to the combined data.

A requirement of the sPlot procedure is that the variable used to calculate the sWeights and the variable from which the observable is measured must be independent. The correlation of the mass and decay time for $B_s^0 \to \mu^+\mu^-$ decays and combinatorial background decays has been evaluated using simulated decays and data. The correlation is of the order of a few percent, as shown in Table 6.1, therefore mass can be used to accurately determine sWeighted to me sure the $B_s^0 \to \mu^+\mu^-$ effective lifetime.

Year	$B_s^0 \to \mu^+\mu^-$ correlation	$b\bar{b} \to \mu^+\mu^- X$ correlation
2011	-0.008	0.003
2012	-0.006	0.008
2015	-0.006	0.010
2016	0.008	0.002

Table 6.1 Correlation between mass and decay time for candidate from $B_s^0 \to \mu^+\mu^-$ simulated decays and combinatorial background decays from data for 2011, 2012, 2015 and 2016 data taking conditions. The full effective lifetime selection is applied to simulated $B_s^0 \to \mu^+\mu^-$ decays and decays in data must pass the effective lifetime selection requirements apart from the global BDT cut and have a dimuon invariant mass of 5447 MeV/c^2.

The sWeights are calculated using the RooFit package [37], however the raw sWeights from the mass fit cannot be used directly in the maximum likelihood fit to measure the effective lifetime. The normalisation of the sWeights will not produce the correct statistical uncertainty on the effective lifetime measurement. Therefore the sWeights are re-normalised via

$$\omega_i^{'} = \omega_i \cdot \frac{\sum\limits_j \omega j}{\sum\limits_j \omega j^2} \tag{6.1}$$

where ω_i are the sWeights values for each decay. The re-normalised sWeights will produce the correct statistical uncertainty in a maximum likelihood fit to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime.

The approach outlined here is suited to the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime because the mass *pdfs* are accurately known for the signal and background decays in the data set from the branching fraction analysis. Furthermore no knowledge is needed to the decay time *pdfs* in the final fit, this is advantageous because decay time distribution of combinatorial background decays is challenging to accurately model. However the overall performance of this strategy depends on the maximum likelihood fit to the invariant mass distribution; how many background components are included in the fit and the mass range the fit covers. The determination of the final fit configuration was done using toy studies, described in Section 6.4, that study a range of different mass ranges largest being 4900 - 6000 MeV/c^2. Therefore the development of the fit configuration requires the mass and decay time *pdfs* of all background within the largest mass range need to be known as well as the signal *pdfs*.

6.2 Mass *pdfs*

The selection criteria used to identify $B_s^0 \to \mu^+\mu^-$ candidates for the $B_{(s)}^0 \to \mu^+\mu^-$ branching fraction and $B_s^0 \to \mu^+\mu^-$ effective lifetime measurements are very similar. Therefore the background decays passing the $B_s^0 \to \mu^+\mu^-$ effective lifetime selection and in the mass range 4900 - 6000 MeV/c^2 are the same as those passing the branching fraction selection, although the yields will be different.Therefore the maximum likelihood fit to extract the sWeights is very similar to the fit to the mass distribution used to measure the $B_{(s)}^0 \to \mu^+\mu^-$ yields for the measurement of the $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions. The *pdf* used in the mass fit has the form

$$\mathcal{P}_{tot}(m) = N_{sig}\mathcal{P}_{sig}(m) + \sum_i N_{bkg}^i \mathcal{P}_{bkg}^i(m) \tag{6.2}$$

where i represents a particular background, $N_{sig(bkg)}$ are the signal (background) yields and $P_{sig(bkg)}$ are the signal (background) *pdfs*. The background decays include; $B^0 \to \mu^+\mu^-$, $B \to h^+h^-$, $\Lambda_b^0 \to p\mu^-\nu_\mu$, $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$,$B_c^+ \to J/\psi\mu^+\nu_\mu$ and combinatorial background decays. For the effective lifetime measurement the $B^0 \to \mu^+\mu^-$ decay is included as a background.

The $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ mass *pdfs* are described by the same a Crystal ball functions used in the branching fraction measurements, with the Run 1 parameters given in Table 5.1. The Run 1 and Run 2 data sets are combined for the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime therefore only one mass *pdf* is needed to describe $B_{(s)}^0 \to \mu^+\mu^-$ decays in data. The choice of Run 1 or Run 2 parameters in the *pdf* has a negligible affect on the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime as shown in Section **??**.

Mis-identified semi-leptonic decays, $\Lambda_b^0 \to p\mu^-\nu_\mu$, $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$ and $B_c^+ \to J/\psi\mu^+\nu_\mu$ are each described by an Argus function convoluted with a Gaussian function evaluated from simulated decays using the same method as described in Section 5.3. The particle identification requirements and the cut on the global BDT use in the selection of candidates for the effective lifetime measurement are taken into account in the evaluation of the *pdf* shapes. However unlike the *pdfs* used in the branching fraction analysis a separate *pdf* is used for each background decay.

Backgrounds from mis-identified $B \to h^+h^-$ decays are described by the double Crystal Ball function evaluated using the method described in Section 5.3 with the effective lifetime particle identification requirements applied. Finally the combinatorial background is modelled with a decaying exponential where the slope is not constrained in the final fit.

The mass *pdfs* for the signal and backgrounds are evaluated for the mass range 4900 to 6000 MeV/c^2 to be used in the toy studies described in Section 6.4. The parameters used describing the background shapes are given in Appendix **??**.

6.3 Decay time *pdfs*

The efficiency of the selection criteria to identity candidates for the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime varies as a function of decay time for both signal and background decays, biasing the decay time distribution. The bias arises because variables used in the selection and the global BDT, such as the isolations and the B meson impact parameter and flight distance significance, are correlated with the decay time. Consequently cuts placed on these variables have a non-uniform efficiency across the decay time range. Therefore the *pdf* describing the decay time changes from a decaying exponential to

$$\mathcal{P}(t) = \epsilon(t) \times e^{-t/\tau} \tag{6.3}$$

where $\epsilon(t)$ is the selection efficiency as a function of decay time. The decay time distribution and selection efficiency as a function of decay time are shown in Figure 6.1 for simulated $B_s^0 \to \mu^+\mu^-$ decays at different stages through the selection. The cut on the global BDT causes the biggest decay time bias as expected since it is the hardest selection cut applied.

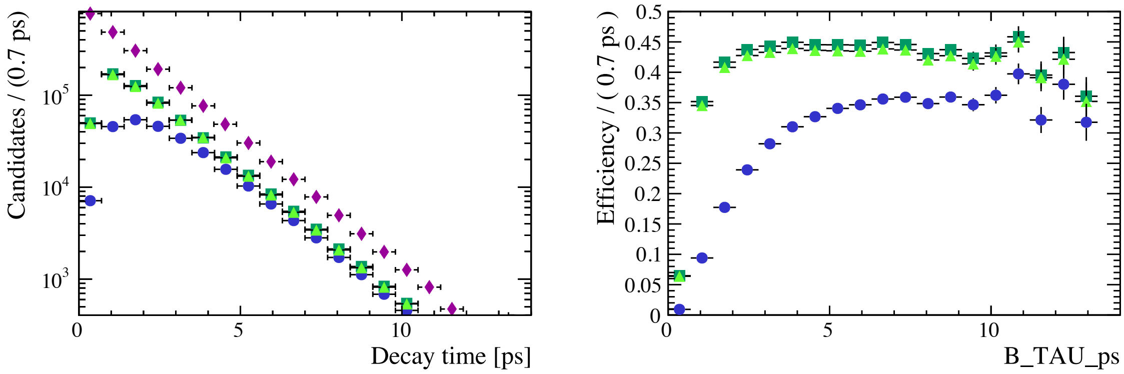


Fig. 6.1 Decay time distribution (left) and selection efficiency as a function of decay time (right) for 2012 $B_s^0 \to \mu^+\mu^-$ simulated decays at different different stages of the selection process. The decay time distributions and efficiencies are shown for reconstructed decays that pass the trigger, stripping and pre-selection cuts (turquoise), the decays that go on to pass PID requirements (green) and decays that pass all selection requirement including the global BDT cut (blue). Also the decay time distribution is shown for all generated simulated decays (purple).

To measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime the efficiency of the selection on $B_s^0 \to \mu^+\mu^-$ decay as a function of decay time must be accurately modelled. The determination of $\epsilon(t)$ for $B_s^0 \to \mu^+\mu^-$ decays is described in Section 6.3.1. Although the sPlot method used to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime means that the decay time *pdfs* of the backgrounds present in the data set are not needed, realistic descriptions of the background decay time *pdfs* are necessary for optimising the mass fit configuration. The background *pdfs* are used described in Section 3.1.

6.3.1 $B_s^0 \to \mu^+\mu^-$

The selection efficiency of $B_s^0 \to \mu^+\mu^-$ decays as a function of decay time is modelled by an 'acceptance' function. A range of different models were investigated for the acceptance function, the parameterised acceptance

$$\epsilon(t) = \frac{[a(t-t_0]^n}{1 + a(t-t_0]^n} \tag{6.4}$$

used in [38] was found to best describe the $B_s^0 \to \mu^+\mu^-$ decay time efficiency. The acceptance function parameters are taken from a fit to simulated $B_s^0 \to \mu^+\mu^-$ decays and are fixed in the fit to data. The parameters could not be determined from data because there are too few $B_s^0 \to \mu^+\mu^-$ decays in data and the efficiency distribution of the more abundant $B \to h^+h^-$ decays after the selection is quite different to that of $B_s^0 \to \mu^+\mu^-$. The decay time efficiency for each year of data taking is slightly different therefore simulated decays from each year of data taking must be used to determine the acceptance parameters.

In general simulated decays model distributions in data reasonably well, however the number of tracks present in an event are not well modelled in the simulation. Although the $B_s^0 \to \mu^+\mu^-$ decay time distribution does not depend on the number of tracks present in the event, the isolations used in the global BDT do. Therefore the selection efficiency as a function of decay time depends on the number of tracks in the event and cannot be accurately described by simulated decays alone. To overcome this the number of tracks in an event for simulated $B_s^0 \to \mu^+\mu^-$ decays are weighted using information from the number of tracks per event for $B^0 \to K^+\pi^-$ decays in both data and simulation.

The selection requirements listed in Table **??** are used to identify $B^0 \to K^+\pi^-$ decays in data and simulated decays but importantly the global BDT cut is not applied. The $\mathrm{DLL}_{K\pi}$ variable is used to separate $B^0 \to K^+\pi^-$ decays from other $B \to h^+h^-$ decays in data and the loose trigger requirements used for the branching fraction analysis are applied to data and simulated decays to keep a high trigger efficiency[1]. The same requirements are applied to simulated decays. The distribution of the number of tracks present in events containing $B^0 \to K^+\pi^-$ decays is obtained from data by performing a maximum likelihood fit to the B^0 mass distribution and extracting sWeights. The distribution of the weighted number of tracks per event in data is compared with the distribution in simulated $B^0 \to K^+\pi^-$ decays. The mass fits to $B^0 \to K^+\pi^-$ decays in data are shown in Figure 6.2 and the normalised distributions

[1] The Hlt2Phys Dec trigger decision was not correctly implemented in 2016 simulated decays, therefore the DEC decisions of a combination of trigger lines designed to select $B \to h^+h^-$ are used to emulate the Hlt2Phys DEC trigger decision. The trigger lines are Hlt2Topo2BodyDecision, Hlt2B2HH_Lb2PPiDecision, Hlt2B2HH_Lb2PKDecision Dec, Hlt2B2HH_B2PiPiDecision, Hlt2B2HH_B2PiKDecision, Hlt2B2HH_B2KKDecision and Hlt2B2HH_B2HHDecision. These trigger lines are applied to both data and simulated decays.

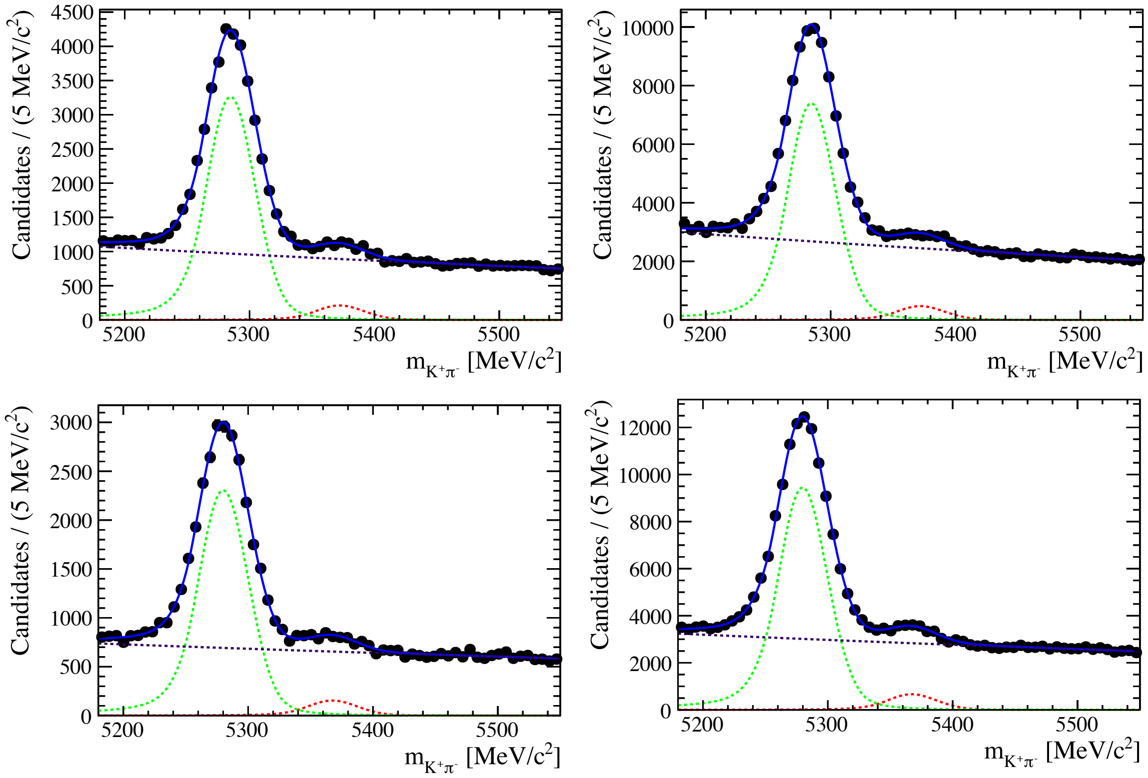


Fig. 6.2 Maximum likelihood fits to the mass distribution of $B^0 \to K^+\pi^-$ candidates in 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data. The mass *pdf* includes components for $B^0 \to K^+\pi^-$ (green), $B_s^0 \to K^+\pi^-$ (red) and combinatorial background (purple).

of the number of tracks per event in weighted data and simulated decays are shown in Figure 6.3. Each year of data taking is kept separate and the same simulation version is used for $B^0 \to K^+\pi^-$ simulated decays as available for $B_s^0 \to \mu^+\mu^-$ decays.

The distributions of the number of tracks per event for $B^0 \to K^+\pi^-$ decays in data and simulated decays are used to weight $B^0 \to K^+\pi^-$ decays so that the distribution in simulation matches that in data. The weights are evaluated by taking the ratio of the normalised histograms in Figure 6.3 for the number of tracks per event in data and simulation for each year. The affect on the decay time distribution of using these weights and then applying the global BDT cut is shown in Figure 6.4 for the simulated $B^0 \to K^+\pi^-$ decays. The difference between the decay time distributions with and without the weights is not large but clearly noticeable at low decay times where the change in selection efficiency is greatest.

The same weights are applied to simulated $B_s^0 \to \mu^+\mu^-$ decays by binning the number of tracks per event for $B_s^0 \to \mu^+\mu^-$ decays in the same way to used for

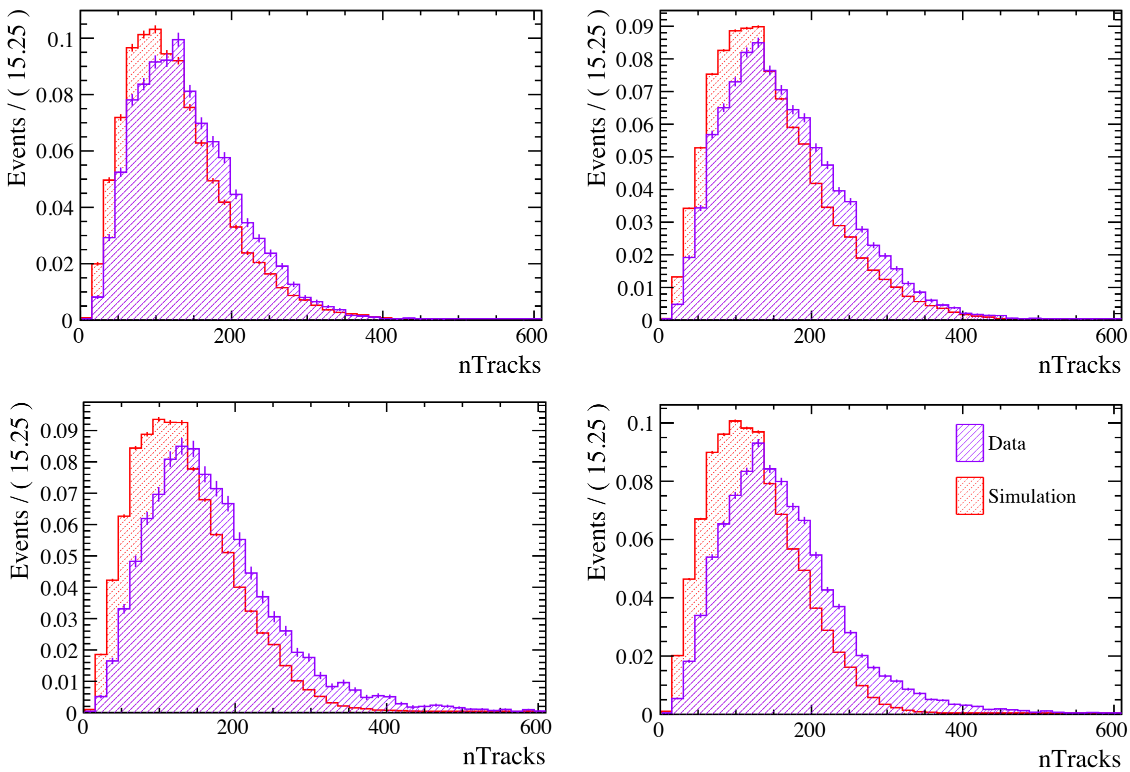


Fig. 6.3 Normalised histograms of the number of tracks per event in simulated $B^0 \to K^+\pi^-$ decays and weighted $B^0 \to K^+\pi^-$ decays in data for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data.

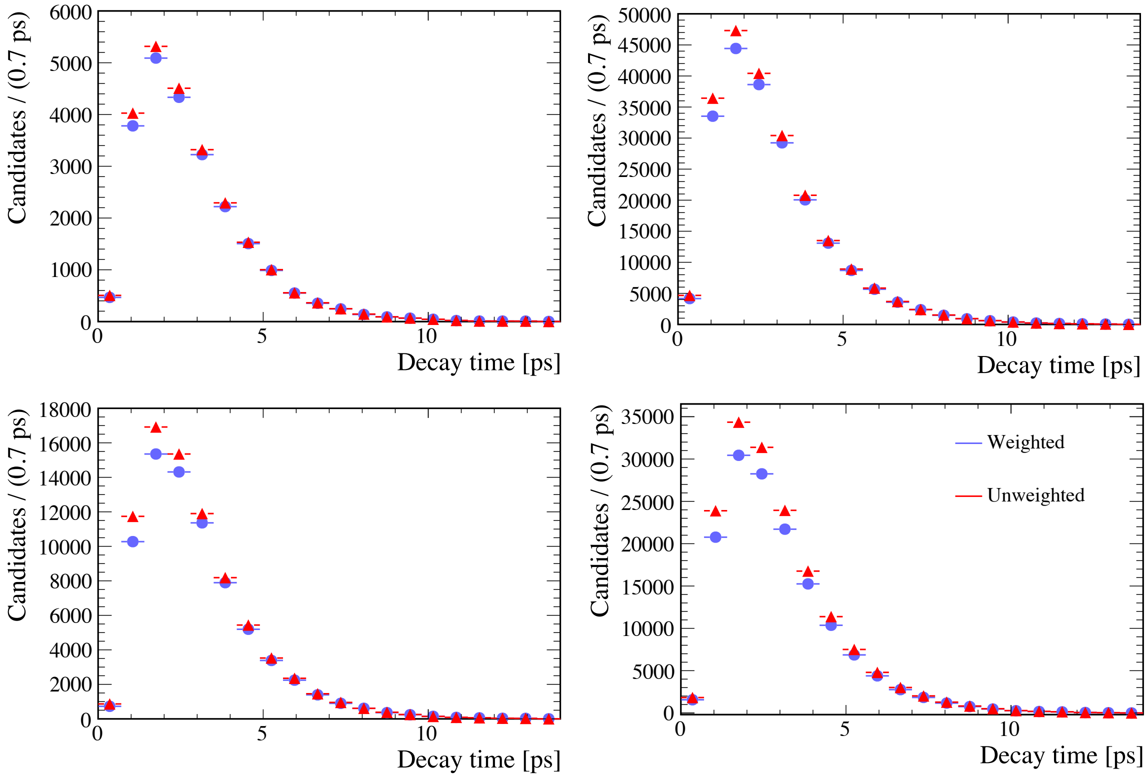


Fig. 6.4 Decay time distributions for weighted and un-weighted $B^0 \to K^+\pi^-$ simulated decays for for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data taking conditions.

$B^0 \to K^+\pi^-$ decays. The weights are applied to decays that pass selection but before the global BDT cut is applied. The change in the decay time distribution for simulated decays after the global BDT cut is shown in Figure 6.6 in the comparison of weighted and un-weighted $B_s^0 \to \mu^+\mu^-$ decay time distributions. Similarly to $B^0 \to K^+\pi^-$ decays the biggest effect is at low decay times where the change in selection efficiency is greatest as seen in Figure 6.1.

The reweighting relies on the number of tracks per event being very similar for $B^0 \to K^+\pi^-$ and $B_s^0 \to \mu^+\mu^-$ decays, this cannot be evaluated in data due to the small number of $B_s^0 \to \mu^+\mu^-$ decays in data. However Figure 6.6 shows a comparison of the number of tracks per event for simulated $B_s^0 \to \mu^+\mu^-$ and $B^0 \to K^+\pi^-$ decays for each year and resulting distributions are rather similar.

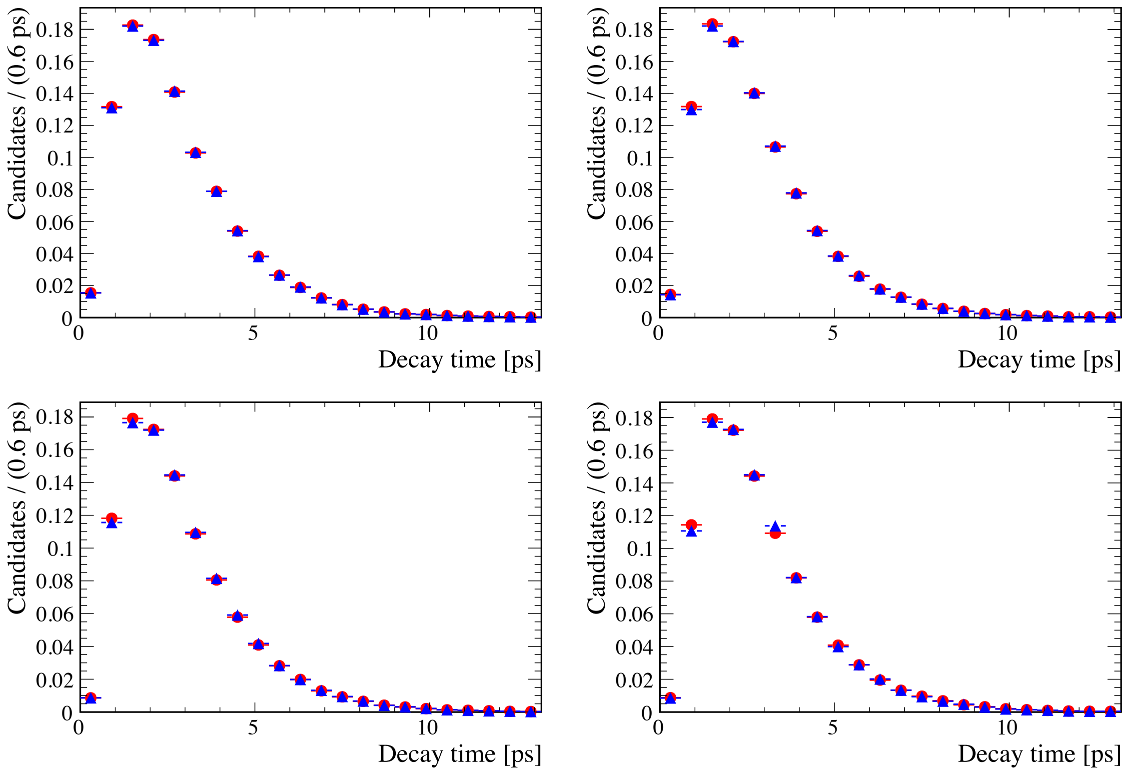


Fig. 6.5 Decay time distributions for weighted and un-weighted $B_s^0 \to \mu^+\mu^-$ simulated decays for for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data taking conditions. Distributions have been normalised to have unit area.

The decay time efficiency will now be accurately modelled in the weighted simulated $B_s^0 \to \mu^+\mu^-$ decays and the parameters in the acceptance function can be evaluated. The number of simulated decays available for each year does not correspond to the proportions of decays present in each year of the data. Therefore weights are used

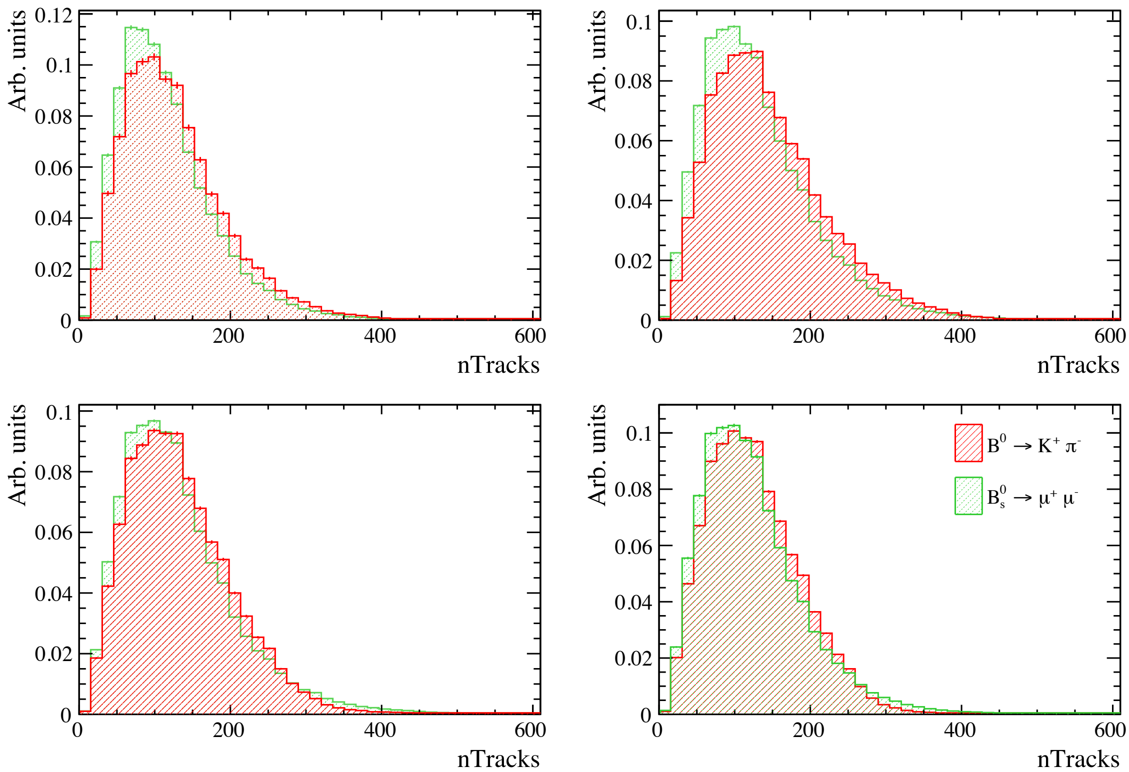


Fig. 6.6 Normalised histograms of the number of tracks per event in simulated $B^0 \to K^+\pi^-$ and $B_s^0 \to \mu^+\mu^-$ decays in data for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right) data.

to combine the simulated decays so that the combined set of decays has the same proportions of decays for each year as the complete data set. The proportion of events of each year is taken from the number of $B_s^0 \to J/\psi\phi$ decays in data for each year corrected for the selection differences for $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ decays. The $B_s^0 \to J/\psi\phi$ yields, $Y^{J/\psi\phi}$, are extracting from maximum likelihood fits to the mass spectrum of candidates in each year of data. The selection applied to identify candidates is the very similar to that applied to $B_s^0 \to \mu^+\mu^-$ decays apart from the particle identification and global BDT requirements. This decay is chosen because the ratio of the efficiencies for the stripping, trigger and pre-selection requirements of $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ decays is uniform across the different years making $B_s^0 \to J/\psi\phi$ decays a good proxy for $B_s^0 \to \mu^+\mu^-$. The weights applied to simulated $B_s^0 \to \mu^+\mu^-$ decays are

$$\omega_i = \frac{Y_i^{J/\psi\phi}\epsilon_i}{\sum_j Y_j^{J/\psi\phi}\epsilon_j} \cdot \frac{\sum_k N_k^{\mu^+\mu^-}}{N_i^{\mu^+\mu^-}} \tag{6.5}$$

where i represents the year and $N^{\mu^+\mu^-}$ the number of simulated $B_s^0 \to \mu^+\mu^-$ decays available for the year passing the full $B_s^0 \to \mu^+\mu^-$ selection and ϵ_i the efficiency of the particle identification and global BDT requirements for $B_s^0 \to \mu^+\mu^-$ decays that have passed all other selection requirement evaluated from simulated decays. The weights applied to simulated $B_s^0 \to \mu^+\mu^-$ decays and values of the different components of the weights are given in Table 6.2.

Year (i)	Y_i	ϵ_i	N_i	ω_i	$\mathcal{N}_i \equiv N_i\omega_i$
2011	19190	0.412	70448	1.72	131364
2012	42103	0.406	254822	1.03	262461
2015	8571	0.410	222820	0.24	53917
2016	37765	0.406	124870	1.88	235218

Table 6.2 Weights used to combine simulated $B_s^0 \to \mu^+\mu^-$ decays for each year to determine the acceptance function. Weights ensure the proportion of simulated events for each year matches what is expected in data.

An unbinned maximum likelihood fit is performed to the combined simulated $B_s^0 \to \mu^+\mu^-$ decays to determine the acceptance parameters in equation 6.4. In the fit the acceptance parameters are free and the $B_s^0 \to \mu^+\mu^-$ lifetime is constrained to the weighted average of lifetimes used to generate each year of simulated decays. The fit results are shown in Figure 6.7 and the acceptance parameters are given in Table **??**.

Figure 6.8 shows the selection efficiency histogram as a function of decay time with the acceptance function overlaid.

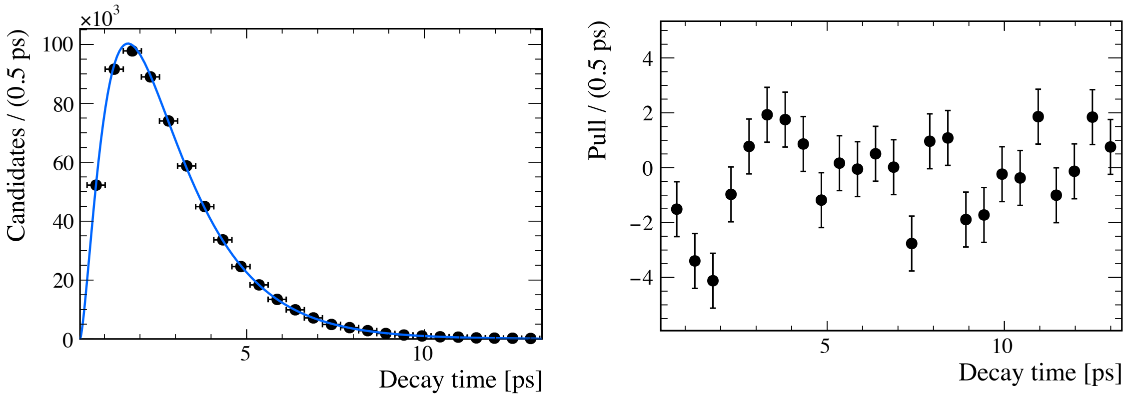


Fig. 6.7 The maximum likelihood fit to the combined decay time distribution (left) of 2011, 2012, 2015 and 2016 simulated $B_s^0 \to \mu^+\mu^-$ decays and the pull distribution of the fit.

Parameter	Value
a	0.574 ± 0.011 ps $^{-1}$
n	1.49 ± 0.03
t_0	0.313 ± 0.007 ps

Table 6.3 Parameters for the $B_s^0 \to \mu^+\mu^-$ acceptance function determined from weighted simulated $B_s^0 \to \mu^+\mu^-$ decays.

6.3.2 Backgrounds

decay time pdf

The final fit to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime does not require knowledge of the decay time *pdfs* of the backgrounds. However the fit configuration is developed using toy studies that use the mass and decay time distributions of both signal and background decays. Therefore realistic models of the decay time *pdfs* are needed to determine the optimal fit configuration.

The selection biases the decay time distributions of the backgrounds in the same way as the $B_s^0 \to \mu^+\mu^-$ decay time. Therefore they are described by the same *pdfs* as in equation 6.3.

The backgrounds from semi-leptonic, $B \to h^+h^-$ and $B^0 \to \mu^+\mu^-$ decays are assigned the same acceptance function as $B_s^0 \to \mu^+\mu^-$ decays because the decay time efficiency of these backgrounds roughly the same as $B_s^0 \to \mu^+\mu^-$ decays. The

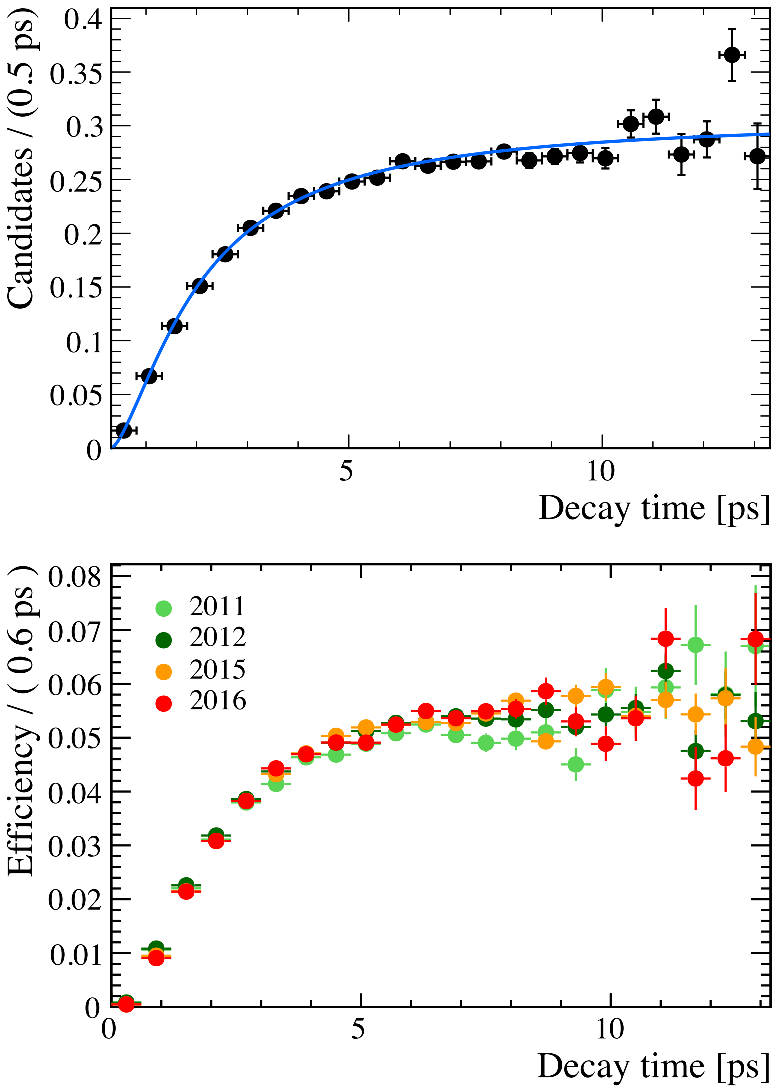


Fig. 6.8 The selection efficiency histogram as a function of decay time with the acceptance *pdf* overlaid for weighted 2011, 2012, 2015 and 2016 simulated $B_s^0 \to \mu^+\mu^-$ decays (left) and the efficiency histograms for each year separately for weighted simulated $B_s^0 \to \mu^+\mu^-$ decays.

acceptances of these backgrounds does not need to be as accurately known as the acceptance function of the signal because very few background decays from these sources will be present in the data set after the selection and the final results does not depend on the acceptance function of the backgrounds. The lifetimes of these background decays are taken from a fit to simulated decays for, for $B \to h^+h^-$ the fit is performed to a combined set of $B \to h^+h^-$ decays representing what is expected in data.

The decay time *pdf* of the combinatorial background is more challenging to determine. This background arises from random combinations of muons in the event and not from one source, therefore there is no single lifetime that describes the background. Furthermore the global BDT which is designed to separate $B^0_s \to \mu^+\mu^-$ decays from combinatorial background decays will have a will have a different efficiency as a function of decay time for the combinatorial background compared to the signal. The decay time *pdf* of the combinatorial background cannot be evaluated from simulated decays or decays in data that pass the $B^0_s \to \mu^+\mu^-$ selection because there are too few candidates left. Therefore the decay time *pdf* of the combinatorial background for $B^0_s \to \mu^+\mu^-$ decays is evaluated from combinational background of $B \to h^+h^-$ decays using candidates in data that pass the $B \to h^+h^-$ selection and have a reconstructed mass greater than 5447 MeV/c^2, above the B^0_s signal region. The decay time *pdf* for combinatorial background decays is modelled by

$$P_{cbg}(t) = \epsilon(t) \times \left(f \cdot e^{-\Gamma_1 t} + (1-f) \cdot e^{-\Gamma_2 t}\right) \tag{6.6}$$

where Γ_1 and Γ_2 are two independent lifetimes used to describe the background, f describes the fraction of candidates with each lifetime and the same acceptance shape as in equation 6.4 is used for describe the decay time efficiency. The lifetimes are different, one describes a long lived component and the other a short lived component that are evident in the data. The decay time acceptance is flat at large decay times, therefore the lifetimes of the combinatorial background decays are determined from a maximum likelihood fit of equation 6.6, setting $\epsilon(t) = 1$, to candidates with a decay time above 2.5 ps. The acceptance function parameters are then determined from a maximum likelihood fit to the full decay time range using equation 6.6 where the lifetimes and the fraction of candidates with each lifetime are fixed. The results are shown in Figure 6.9 and the *pdf* parameters in Table 6.4, the t_0 parameter is fixed in the fit to improve fit stability.

This model for the background assumes that the decay time distribution of $B \to h^+h^-$ candidates formed by random combinations of kaons and pions is the same as that

Parameter	Value
a	1.45 ± 0.12 ps $^{-1}$
n	1.92 ± 0.17
t_0	0.290 ps
Γ_1	0.06 ± 0.05 ps $^{-1}$
Γ_2	0.77 ± 0.17 ps $^{-1}$
f	0.032 ± 0.027

Table 6.4 Parameters to described the background decay time distribution from combinatorial background decays in data passing the $B \to h^+h^-$ selection.

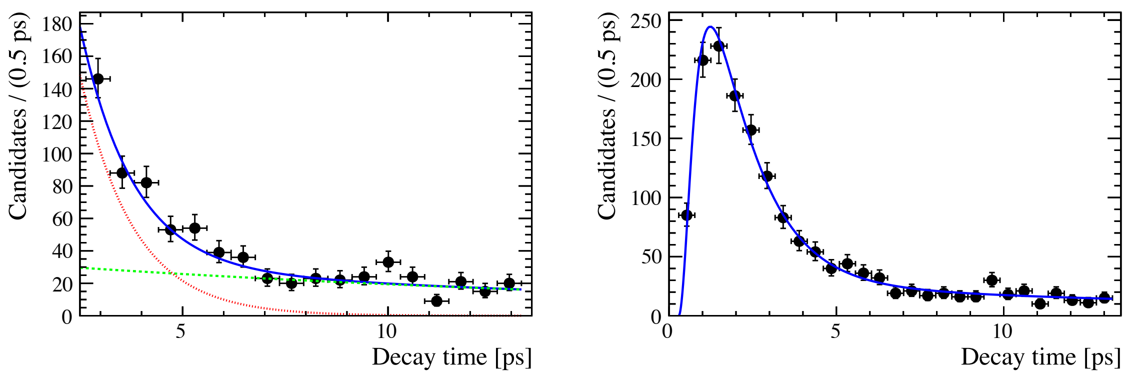


Fig. 6.9 The maximum likelihood to determine the lifetimes of the background (left), with the long live component (green) and the short lived component (red) shown and the acceptance parameters (right) of combinatorial background decays in data passing the $B \to h^+h^-$ selection requirements.

of $B_s^0 \to \mu^+\mu^-$ candidates formed by randomly combining muons in the event. There are too few candidates passing the $B_s^0 \to \mu^+\mu^-$ selection to verify this assumption, the validity of this model and the impact of the toy studies is investigated in Section **??**.

6.4 Toy Studies for fit optimisation

The strategy to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime was described earlier in Section 6.1, however given the extremely rare nature of $B_s^0 \to \mu^+\mu^-$ decays, the stability and performance of the final fit will be highly dependant on maximum likelihood fit to the invariant mass distribution. Toy studies were performed to determine the mass range and background components included in the maximum likelihood fit would produce the smallest expected uncertainty on the measured effective lifetime for the data set.

The expected number of signal and background decays in the data set passing the $B_s^0 \to \mu^+\mu^-$ selection in the mass range 4600 - 6000 MeV/c^2 were used as the basis for the toy studies. The expected background yields were calculated using the same methods described in Section 5.3 but taking into account the looser particle identification requirement and the cut placed on the global BDT. The number of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays are calculated using the normalisation factors in Section 5.4 and assuming the branching fraction values predicted by the Standard Model. The expected yields are shown in Table 6.5.

Decay	Expected yield
$B_s^0 \to \mu^+\mu^-$	30.94
$B^0 \to \mu^+\mu^-$	3.27
$B \to h^+h^-$	9.68
$\Lambda_b^0 \to p\mu^-\nu_\mu$	13.34
$B^0 \to \pi^-\mu{+}\nu_\mu$	40.50
$B_s^0 \to K^-\mu{+}\nu_\mu$	9.13
$B^+ \to \pi^+\mu^+\mu^-$	6.01
$B^0 \to \pi^0\mu^+\mu^-$	4.86
$B_c^+ \to J/\psi\mu^+\nu_\mu$	9.79
Combinatorial background	66.23

Table 6.5 Number of expected decays in data passing the $B_s^0 \to \mu^+\mu^-$ effective lifetime selection.

The toy studies are performed by generating the mass and decay time distributions for the expected number of signal and background decays using the *pdfs* described in Section 6.2 and 6.3 and assuming the Standard Model prediction for $\tau_{\mu\mu}$ and taking the slope of the combinatorial background mass *pdf* from simulated decays. Then sWeights are computed from an unbinned maximum likelihood fit to the invariant mass distribution and the lifetime and its inverse are measured by a unbinned maximum likelihood fit to the signal weighted decay time distribution. A series of different mass ranges and background components included in the mass fit were tested. For each possible configuration 10,000 toy studies were performed and the performance of each configuration was evaluated using using a couple of different metrics. The first, is the median expected uncertainty of the $B_s^0 \to \mu^+\mu^-$ lifetime and inverse lifetime, the median rather than the mean uncertainty is used due to the asymmetric spread of uncertainties observed for the expected statistics. The second measure, is the pull distributions of any free parameters in the fit, where the pull is defined as $(x-\mu)/\sigma$ with x the measured parameter value, μ the value used in the generation and σ the uncertainty on the measured parameter value. Ideally the pull distributions will be Gaussian in shape with a mean at 0 and a width of 1.

The details of the toy studies performed are given in Section 6.4.2, however first is a discussion of whether the $B_s^0 \to \mu^+\mu^-$ lifetime or inverse lifetime should be measured given the expected number of decays present in the data set.

6.4.1 To fit for τ or τ^{-1}

During the development of the fit strategy the toy studies produced biased pull distributions for the measured $B_s^0 \to \mu^+\mu^-$ effective lifetime no matter what mass fit configuration or acceptance function was used. The pull distribution for the effective lifetime, $\tau_{\mu\mu}$, is shown in Figure 6.10 for a simplified configuration where no acceptance function is used and only signal and combinatorial background decays are generated in the mass range 4900 - 6000 MeV/c^2. The distribution is clearly not Gaussian in shape. The bias was more pronounced in early stages of the analysis development which was done assuming the expect signal and background yields of only the Run 1 data set.

The log-likelihood profile of the fit at a function of $\tau_{\mu\mu}$ reveals the cause of the biased pull distribution. For the simplified studies illustrated in Figure **??** the decay time is modelled by

$$N(t, \tau^{\mu\mu}) = N_0 e^{-t/\tau_{\mu\mu}} \tag{6.7}$$

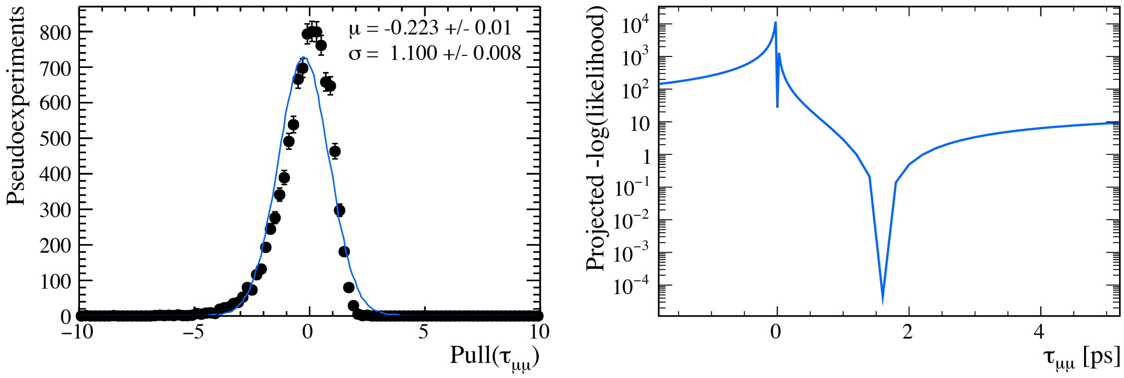


Fig. 6.10 Pull distribution (left) for $\tau_{\mu\mu}$ using a simplified configuration where no acceptance function is used and only signal and combinatorial background decays are generated in the mass range 4900 - 6000 MeV/c^2 with the expected statistics for 4.4 fb^{-1} of data. Likelihood profile for $\tau_{\mu\mu}$ (right).

The likelihood profile as a function of decay time for this model is shown in Figure **??** and there is a clear discontinuity at the zero. The discontinuity arises because the value of $N(t, \tau)$ approaches zero as τ reduces in value until at the origin when $\tau = 0$ and $N(t, \tau)$ jumps to infinity. The jump in value is reflected as the discontinuity in the log-likelihood profile. At the low statistics expected for the data set, particularly when only Run 1 data was considered, the fitted value for $\tau_{\mu\mu}$ is only a few standard deviations of the discontinuity, therefore leading the bias in the statistical uncertainty and hence the pull distributions. However as the number of expected signal and background decays are increased the $\tau_{\mu\mu}$ pull distributions become Gaussian in shape as shown in Figure 6.11. This is as expected for when the statistical uncertainty deceases and the discontinuity of Figure **??** is no longer within a few standard deviations of the measured $\tau_{\mu\mu}$.

The bias in the $\tau_{\mu\mu}$ pull distribution shows that the distribution cannot be interpreted in the usually way and also that the statistical uncertainties from the maximum likelihood to the weighted decay time distribution may not be correct.

Another way to assess the accuracy of the statistical uncertainties returned by the maximum likelihood fit is the coverage of the uncertainties; the percentage of fitted $\tau_{\mu\mu}$ values from the toy studies that fall within 1, 2 and 3 standard deviations of the lifetime used is the input value for the toy studies. Table 6.6 shows the coverage of the statistical uncertainties for $\tau_{\mu\mu}$ for 10,000 toy studies for the expected $B_s^0 \to \mu^+\mu^-$ and combinatorial background yields with 4.4 fb^{-1} alongside the intervals expected for a Gaussian distribution. The simple toy configuration used to produce the log-likelihood function is used. A comparison between the coverage of $\tau_{\mu\mu}$ and the Gaussian intervals

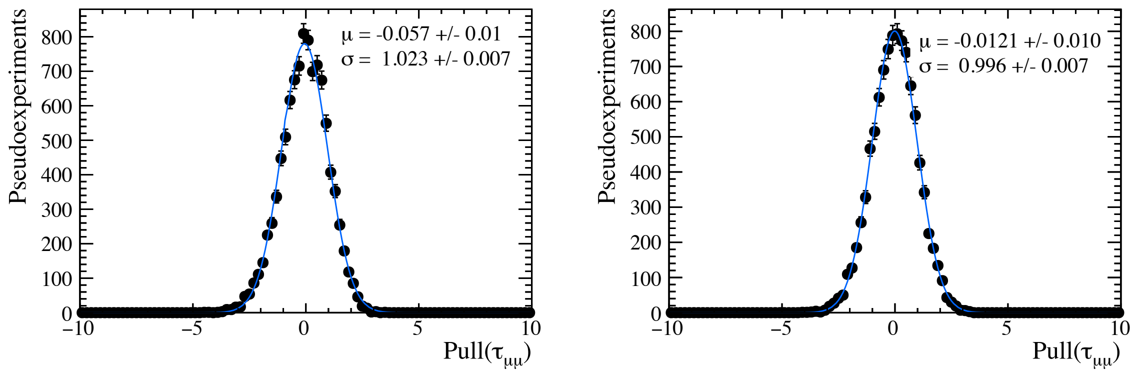


Fig. 6.11 Pull distribution for $\tau_{\mu\mu}$ using simplified toys for 50 fb^{-1} (left) and 300 fb^{-1} (right) using a simplified configuration where no acceptance function is used and only signal and combinatorial background decays are generated in the mass range 4900 - 6000 MeV/c^2.

shows that the coverage of the statistical uncertainties is very close to the expected values.

	$\tau_{\mu\mu}$	$\Gamma_{\mu\mu}$	Gaussian
1σ	68.50%	67.92%	68.27%
2σ	93.44%	95.91%	95.45 %
3σ	98.06%	99.55%	99.73 %

Table 6.6 Coverage.

Alternatively, a way to get around having a biased pull distribution is to measure the inverse of the effective lifetime, $\tau_{\mu\mu}^{-1} \equiv \Gamma_{\mu\mu}$. The pull distributions for $\Gamma_{\mu\mu}$ are shown in Figure 6.12 and produce unbiased pull values regardless of the amount of data. This is unsurprising given the smooth log-likelihood profile as a function of $\Gamma_{\mu\mu}$ also shown in Figure 6.12. Furthermore the statistical coverage of $\Gamma_{\mu\mu}$ is closer to the expected Gaussian coverage than the coverage of $\tau_{\mu\mu}$. However the lifetime is a more interesting variable from a physics point of view due to its relationship with $A_{\Delta\Gamma}$.

Ideally the fit strategy would be performed to extract the lifetime not the inverse lifetime, however for the moment the maximum likelihood fit for both $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ will be used in the toy studies. The statistical coverage for both parameters is good and using either is reasonable. The final decision will be made based on the statistical coverage for the observed number of decays in the data set.

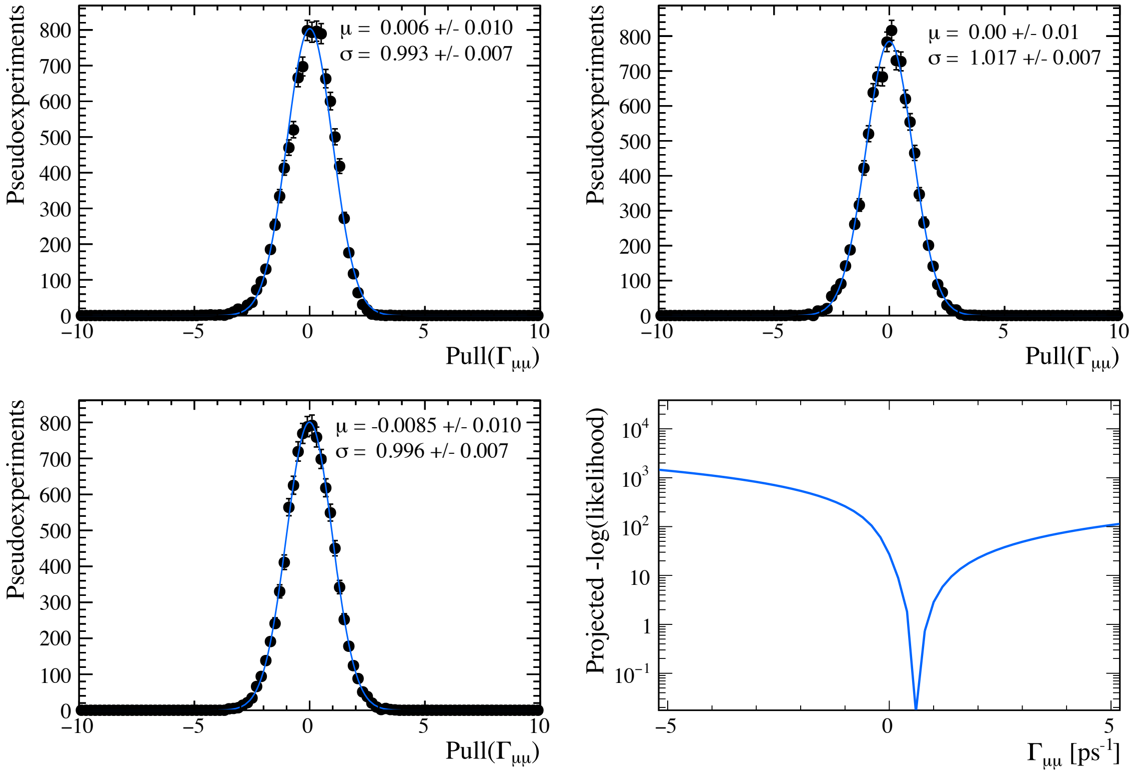


Fig. 6.12 Pull distribution for $\Gamma_{\mu\mu}$ using simplified toy studies for 4.4 (top left), 50 (top right) and 300 (bottom left) fb^{-1} and the likelihood profile as a function of $\Gamma_{\mu\mu}$ (bottom right).

6.4.2 Toy results

The mass distribution of expected number $B_s^0 \to \mu^+\mu^-$ candidates passing the effective lifetime selection is shown in Figure 6.13 alongside the corresponding decay time distribution. The contributions from the different signal and background sources are shown and the backgrounds beneath the B_s^0 mass peak are the combinatorial background and the tails of the $B \to h^+h^-$, and $\Lambda_b^0 \to p\mu^-\nu_\mu$ backgrounds. The expected mass distribution is used to determine a range of mass fit configurations to be tested using toy studies to find the configuration that produces the smallest expected uncertainty on the measurement of $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$.

In each mass fit configuration the mass *pdf* in equation 6.2 is used and the mass ranges and backgrounds included in the *pdf* for the different configurations are given in Table 6.7.

For each possible mass fit configuration the $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and combinatorial background yields are left free in the fit whereas the yields of any other backgrounds are constrained to their expected values. The mass shapes of all components are fixed in the maximum likelihood fit except the slope of the combinatorial background because this is not accurately known in data. The Standard Model prediction, $\tau_{\mu\mu} = \mathrm{tH}$, for the $B_s^0 \to \mu^+\mu^-$ effective lifetime is used to generate events for the toy studies where τ_H is taken from the PDG value and regardless of which background components are included in the mass fit all backgrounds are generated for each mass range. The *pdfs* used in the toy studies are detailed in Appendix **??**.

A total of 10,000 toy studies are performed for each mass configuration and the results are given in Table 6.8. The mean and widths of $\Gamma_{\mu\mu}$, the $B_s^0 \to \mu^+\mu^-$ yield and combinatorial background yield and slope as well as the median expected uncertainty on $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are used to measure the performance of each mass fit configuration. The pull distribution of the fit for $\tau_{\mu\mu}$ is not used to assess the performance of each mass fit configuration given the discussion in Section 6.4.1.

The expected statistical uncertainties for $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are smallest for mass of the simplest mass fit configuration, number 11, where the mass range is restricted to 5320 - 6000 MeV/c^2 and only the $B_s^0 \to \mu^+\mu^-$ and combinatorial background components are used in the total mass *pdf*. The mean and widths for the different pull distributions are consistent with the expected mean of 0 and width of 1 for this fit configuration. The larger mass ranges with more background components included in the mass *pdf* have larger expected uncertainties for $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ as well as clearly biased pull distributions. It is not surprising that the simplest fit performs the best given the very low expected number of events.

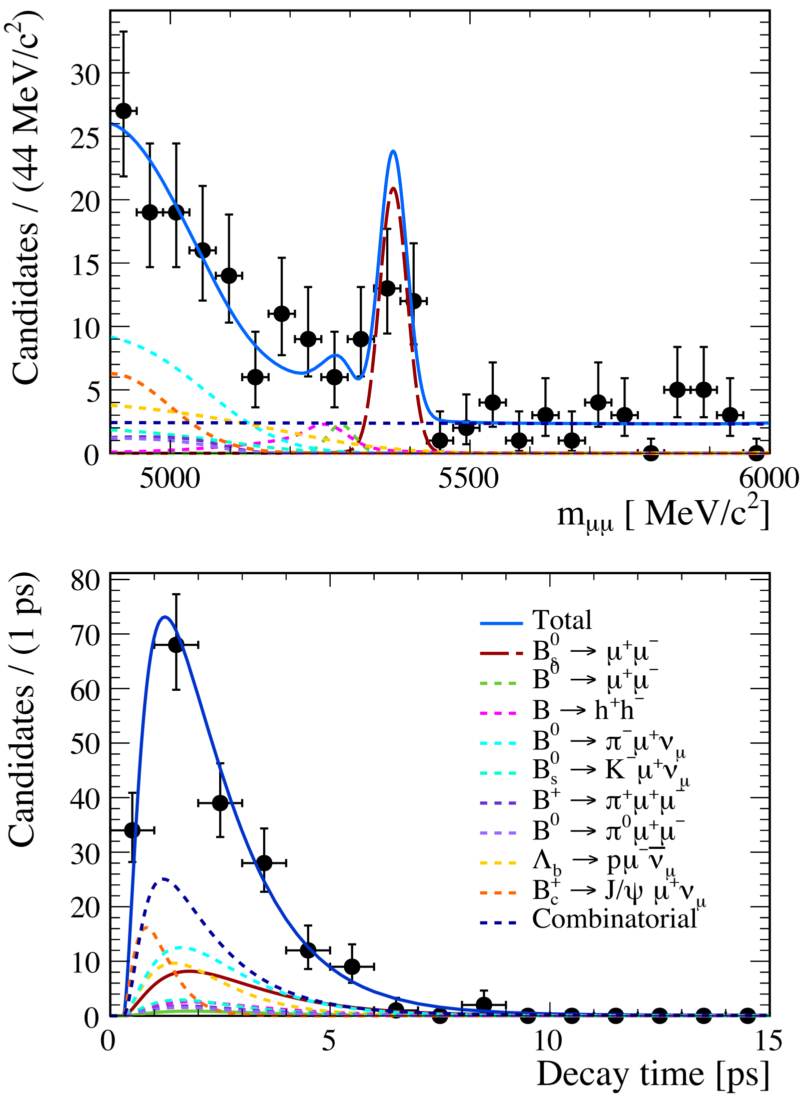


Fig. 6.13 Mass and decay time distributions for the generated decays in the mass range 4900 - 6000.

Fit no.	Mass Range / MeV/c^2	Components included in the mass *pdf*	
		Free yields free	Fixed yields
1.	4900 - 6000	$B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, comb. bkg.	$B \to h^+h^-$, $\Lambda_b^0 \to p\mu^-\nu_\mu$, $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$, $B_c^+ \to J/\psi\mu^+\nu_\mu$
2.	4900 - 6000	$B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, comb. bkg.	$B \to h^+h^-$, $\Lambda_b^0 \to p\mu^-\nu_\mu$, $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$, $B_c^+ \to J/\psi\mu^+\nu_\mu$ Mass *pdfs* for $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$ are combined and so are $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$.
3.	5150 - 6000	$B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, comb. bkg.	$B \to h^+h^-$, $\Lambda_b^0 \to p\mu^-\nu_\mu$, $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$,
			$B^0 \to \pi^0\mu^+\mu^-$, $B_c^+ \to J/\psi\mu^+\nu_\mu$
4.	5150 - 6000	$B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, comb. bkg.	$B \to h^+h^-$, $\Lambda_b^0 \to p\mu^-\nu_\mu$, $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$, $B_c^+ \to J/\psi\mu^+\nu_\mu$ Mass *pdfs* for $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$ are combined and so are $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$.
5.	5200 - 6000	$B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, comb. bkg.	$B \to h^+h^-$, $\Lambda_b^0 \to p\mu^-\nu_\mu$, $B_c^+ \to J/\psi\mu^+\nu_\mu$
6.	5200 - 6000	$B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, comb. bkg.	$B \to h^+h^-$, $\Lambda_b^0 \to p\mu^-\nu_\mu$
7.	5200 - 6000	$B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, comb. bkg.	-
8.	5200 - 6000	$B_s^0 \to \mu^+\mu^-$, comb. bkg.	-
9.	5250 - 6000	$B_s^0 \to \mu^+\mu^-$,comb. bkg.	-
10.	5300 - 6000	$B_s^0 \to \mu^+\mu^-$, comb. bkg.	-
11.	5320 - 6000	$B_s^0 \to \mu^+\mu^-$, comb. bkg.	-

Table 6.7 Mass ranges and components included in the different mass fit configurations tested using to toy studies. In the mass fit the slopes of the combinatorial background decays are not fixed but the shapes of all other mass *pdfs* are fixed.

Fit	$\mathcal{N}(B_s^0 \to \mu^+\mu^-)$		$\mathcal{N}$(Comb.)		λ		$\tau_{\mu\mu}$	$\Gamma_{\mu\mu}$		
	Mean	Width	Mean	Width	Mean	Width	σ/ps	Mean	Width	σ/ps $^{-1}$
1.	-0.043	1.002	-0.048	1.050	-0.071	1.005	0.40	0.021	0.946	0.15
2.	-0.064	1.015	-0.018	1.048	-0.092	1.006	0.35	0.013	0.970	0.13
3.	-0.034	0.973	-0.067	1.023	-0.019	0.999	0.42	0.031	0.935	0.12
4.	-0.042	0.981	-0.066	1.024	-0.018	0.999	0.41	0.028	0.942	0.15
5.	-0.094	0.997	0.100	1.007	-0.228	1.018	0.40	0.017	0.933	0.40
6.	-0.124	1.024	0.110	1.009	-0.242	1.021	0.32	-0.008	0.973	0.12
7.	-0.367	1.045	1.248	0.923	-1.823	1.104	0.33	-0.091	0.975	0.12
8.	-0.521	1.049	1.770	0.983	-2.425	1.075	0.34	-0.114	0.969	0.12
9.	-0.473	1.044	1.296	0.918	-1.883	1.126	0.34	-0.126	-.993	0.12
10.	-0.101	1.013	0.396	0.989	-0.571	1.068	0.31	-0.043	0.985	0.11
11.	0.050	1.006	0.060	1.013	-0.123	1l013	0.29	0.024	0.982	0.11

Table 6.8 Results for the toy studies testing the mass fit configuration. The mean and width of the pull distributions for the $B_s^0 \to \mu^+\mu^-$ and combinatorial background yields and the slope of the combinatorial background mass *pdf*, λ, are shown along with the expected statistical uncertainty on $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$. The uncertainties on the means are 0.010 and the widths are 0.007 for all configurations.

Therefore the fit configuration number 11 is chosen to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime. Figure 6.14 gives an example of the mass and decay time maximum likelihood fits for the chosen configuration, it is clear from Figure 6.13 that the number of background decays from $B^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $\Lambda_b^0 \to p\mu^-\nu_\mu$ is extremely mass above 5320 MeV/c^2 therefore these backgrounds do not need to be modelled in the mass *pdf*. The affect on the final result of not modelling these backgrounds is estimated in Section **??**.

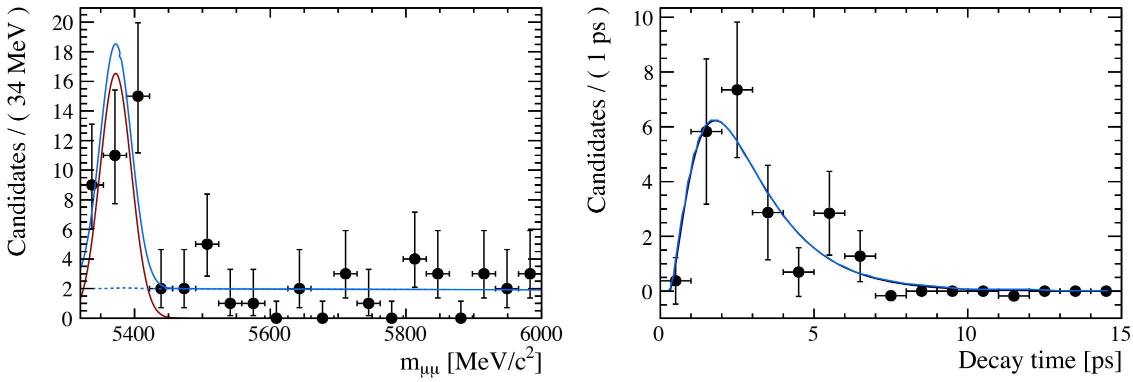


Fig. 6.14 Toy example of the mass and decay time maximum likelihood fits for the chosen fit configuration where only components for $B_s^0 \to \mu^+\mu^-$ and combinatorial background are modelled in the mass *pdf*.

The expected uncertainties for the chosen fit configuration for $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are $\sigma(\tau_{\mu\mu}) = 0.28$ ps and $\sigma(\Gamma_{\mu\mu}) = 0.11$ ps $^{-1}$. However due to the low expected number of decays there is a large spread in the expected uncertainties as shown in Figure 6.15. Therefore the uncertainties on the measurements would range between 0.1 - 0.8 ps for $\tau_{\mu\mu}$ and 0.07 - 0.2 ps $^{-1}$ for $\Gamma_{\mu\mu}$.

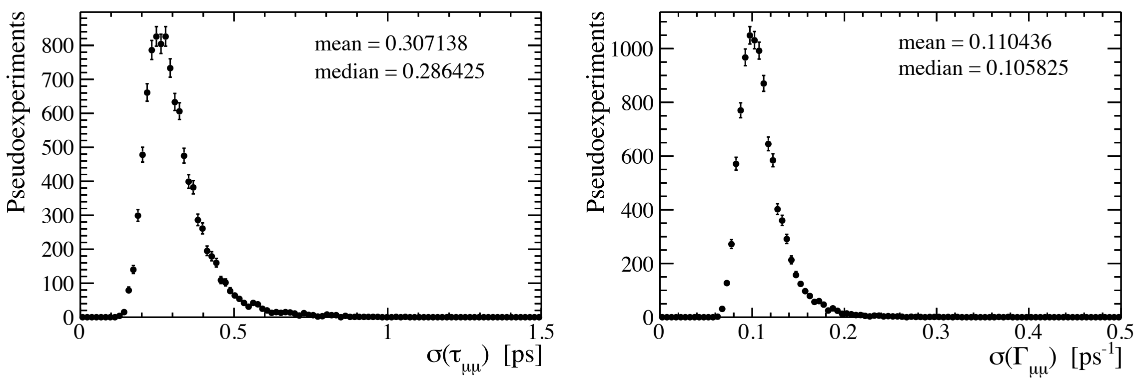


Fig. 6.15 Expected statistical uncertainties for $\tau_{\mu\mu}$ (left) and $\Gamma_{\mu\mu}$ (right) using fit configuration number 11.

6.5 Results

The results of the unbinned maximum likelihood fit to the dimuon mass distribution and the sWeighted decay time of $B_s^0 \to \mu^+\mu^-$ candidates for 4.4 fb^{-1} of Run 1 and Run 2 data are shown in Figure 6.16. The number of observed decays was 22 ± 6 $B_s^0 \to \mu^+\mu^-$ decays and 20 ± 6 combinatorial background decays. The measured values of $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are

$$\tau_{\mu\mu} = 2.04 \pm 0.44 \text{ ps} \tag{6.8}$$

$$\Gamma_{\mu\mu} = 0.489 \pm 0.117 \text{ ps}^{-1} \tag{6.9}$$

where the uncertainties are only statistical. The results are consistent with the Standard Model prediction of $\tau_{\mu\mu} = \tau_H$ within 1 σ and within 1.5 σ of $\tau_{\mu\mu} = \tau_L$.

The observed number of decays is lower than expected and the statistical coverage of the uncertainties has been checked using toy studies generated with the observed number of decays. In the toy studies the all background decays are generated at the expected level. The coverage of both $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ statistical uncertainties is good, as shown in Table 6.9, therefore the result of the more interesting $\tau_{\mu\mu}$ and its statistical uncertainty can be trusted as accurate.

	$\tau_{\mu\mu}$	$\Gamma_{\mu\mu}$	Gaussian
1σ	68.83%	67.76%	68.27%
2σ	93.11%	95.55%	95.45 %
3σ	97.92%	99.67%	99.73 %

Table 6.9 Coverage of observed decays.

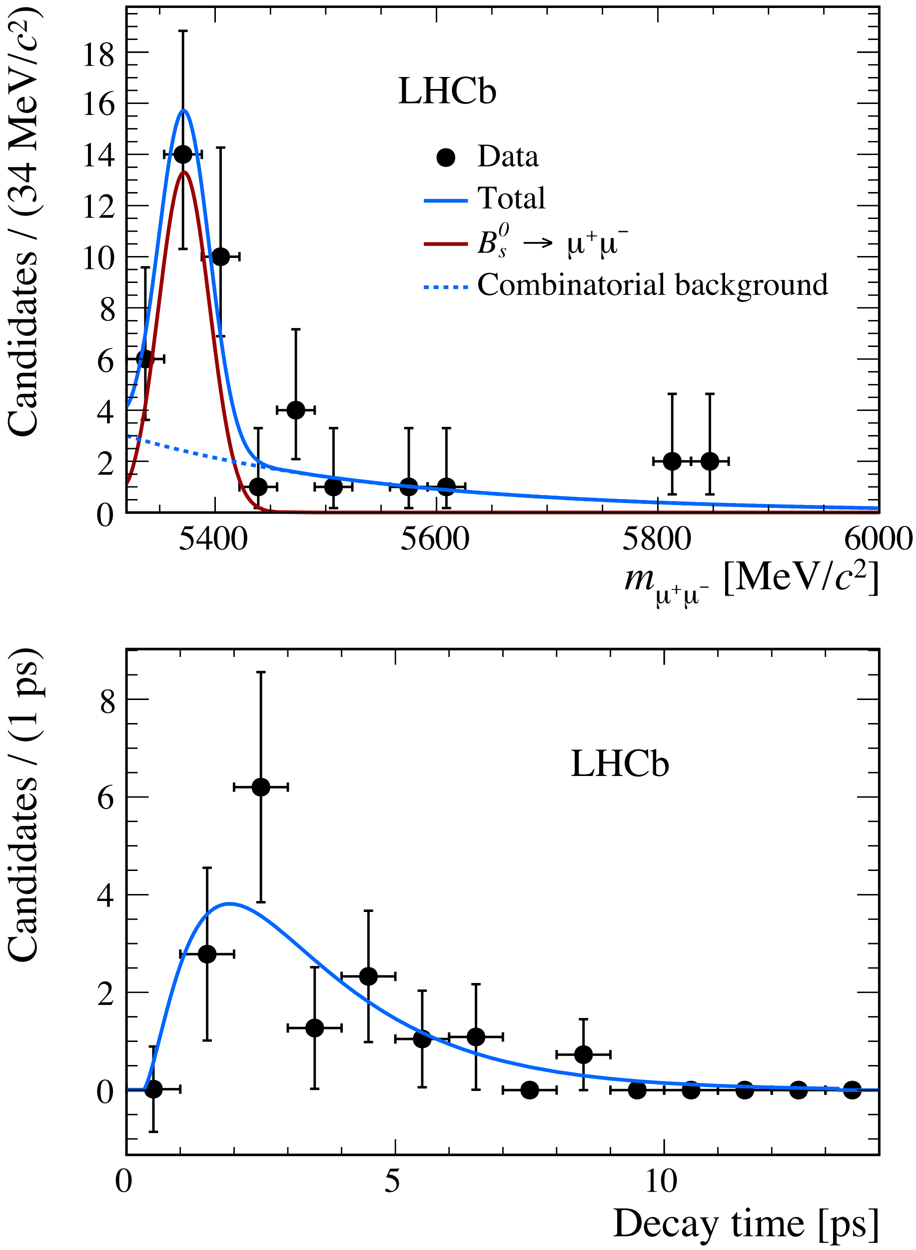


Fig. 6.16 Maximum likelihood fit to the invariant mass distribution (left) and weighted decay time distribution (right) of $B_s^0 \to \mu^+\mu^-$ candidates in 4.4 fb^{-1} of data collected by the LHCb experiment. $B_s^0 \to \mu^+\mu^-$ candidates are described by the red peak in the mass plot and combinatorial background by the blue dashed line, the total *pdf* is given by the solid blue line.

Chapter 7

Systematic Uncertainties and Cross Checks

The measured $B^0_s \to \mu^+\mu^-$ effective lifetime presented in Chapter **??** is influenced by various systematic biases arising from different areas of the analysis procedure. In this Chapter the size of systematic biases is estimated and several cross checks are made on the measurement strategy to ensure the uncertainty quoted on the final result is correct. The total systematic uncertainty for measuring $\tau_{\mu\mu}$ is give at the end of the Chapter.

7.1 Fit Accuracy

The fit strategy used to measure the $B^0_s \to \mu^+\mu^-$ effective lifetime was presented in Chapter 6, the final fit configuration was chosen by optimising two different figures of merit; the mean and width of the pull distributions of free parameters in the fit and the expected uncertainties on $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$. The values for these parameters from 10,000 toy studies using the final fit configuration and assuming the Standard Model $B^0_s \to \mu^+\mu^-$ branching fraction and effective lifetime are given in Table **??**. The same toy set up as described in Section 6.4 but in these toy studies only $B^0_s \to \mu^+\mu^-$ and combinatorial background decays are generated. Based on these results several aspects of the fit deserve further investigation, this includes the stability of the fit with different $\tau_{\mu\mu}$ values, the biased pull distribution for $B^0_s \to \mu^+\mu^-$ yields and the overall bias in the measured values of $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$. The pull distributions for $\tau_{\mu\mu}$ are biased, as discussed in Section 6.4.1, and are therefore not used to evaluate the fit performance and the pull distribution for $\Gamma_{\mu\mu}$ can be used as a measure of the fit performance instead. Furthermore the uncertainty on $\Gamma_{\mu\mu}$ is not longer needed to evaluate the fit performance since the final measured results will be quoted in terms of $\tau_{\mu\mu}$.

7.1.1 Fit stability with $\tau_{\mu\mu}$ values

The value of $\tau_{\mu\mu}$ is accurately returned by the fit for the expected number of decays. However, it is necessary to understand if this is due to accurate background subtraction by the sPlot method or it it could be resulting from similarities between the decay time distributions of $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays. As a test, toy studies were performed for a range of generated $B_s^0 \to \mu^+\mu^-$ lifetimes different to the Standard Model prediction. Only $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays were generated in the toy studies so that the small contamination from mis-identified backgrounds does not mask the effects of using different lifetimes. The bias arising from the small contribution of mis-identified backgrounds in the mass range of the fit is evaluated in Section 7.2. The results of 10,000 toy studies are shown in Table 7.1 and the different lifetimes all return accurate pull distributions for the fitted $\Gamma_{\mu\mu}$ values with means and widths consistent with 0 and 1, respectively and the expected uncertainties for $\tau_{\mu\mu}$ are similar. Therefore the fit returns an accurate measured value for a range of $B_s^0 \to \mu^+\mu^-$ lifetimes.

τ	$\mathcal{N}(B_s^0 \to \mu^+\mu^-)$		$\mathcal{N}$(Comb.)		λ		$\Gamma_{\mu\mu}$		$\sigma(\tau_{\mu\mu})$
	Mean	Width	Mean	Width	Mean	Width	Mean	Width	ps
τ_H	-0.097	1.020	-0.062	1.030	-0.013	0.992	-0.000	0.993	X
τ_L	-0.098	1.019	-0.062	1.018	-0.003	0.987	-0.010	1.018	X
$\tau_{B_s^0}$	-0.102	1.017	-0.059	1.032	-0.027	0.996	0.001	0.989	X

Table 7.1 Results from 10,000 toy studies using the final fit configuration for the expected number of decays and taking at $\tau_{\mu\mu}$ the Standard Model predication (τ_H), the lifetime of the light B_s^0 mass eigenstate (τ_L) and the mean lifetime of the B_s^0 ($\tau_{B_s^0}$). The mean and width of the pull distributions for the $B_s^0 \to \mu^+\mu^-$ and combinatorial background yields and the slope of the combinatorial background mass *pdf*, λ, are shown along with the median statistical uncertainty on $\tau_{\mu\mu}$. The uncertainties on the means are 0.010 and widths are 0.007 for both configurations.

7.1.2 $B_s^0 \to \mu^+\mu^-$ yields

The pull distributions for the $B_s^0 \to \mu^+\mu^-$ yields in the mass fit have biased mean values of ~ 0.010 ps as shown in Tables 7.1, implying that the mass fit does not accurately estimate the $B_s^0 \to \mu^+\mu^-$ yield. However the pull distribution for $\Gamma_{\mu\mu}$ is accurate therefore this bias in the $B_s^0 \to \mu^+\mu^-$ yield could originate from a different source.

In the toy studies the number of expected decays, N_{expt}, in the mass range 4900 - 6000 MeV/c^2 are given in Table 6.5. These numbers are used as the basis for the toy studies however the number of decays generated is fluctuated for each study about the expected value using a Poisson distribution, therefore the number of decays generated, N_{gen}, is different to the expected number. This enables an extended maximum likelihood fit, where the total number of event is a free parameter, to be used to fit the mass distribution. To achieve an accurate pull distribution of the measured $B_s^0 \to \mu^+\mu^-$ yields the uncertainties on the measured yields must be distributed according to a Gaussian function. This will be true in the high statistics limit where a Poisson distribution is a good approximation of a Gaussian distribution. However the current data available does not contain high number of $B_s^0 \to \mu^+\mu^-$ decays therefore the uncertainty on the measured yields is proportional to $\sqrt{N_{gen}}$ and does not have a Gaussian distribution. This effect would shift the mean value of the pull distribution but not lead to an incorrect estimation of the $B_s^0 \to \mu^+\mu^-$ yield. The 'fractional bias', $(N_{meas} - N_{gen})/N_{gen}$, where N_{meas} is the measured $B_s^0 \to \mu^+\mu^-$ yield is shown in Figure 7.1 and supports this explanation by producing a mean consistent at zero, with a bias of 0.8 %.

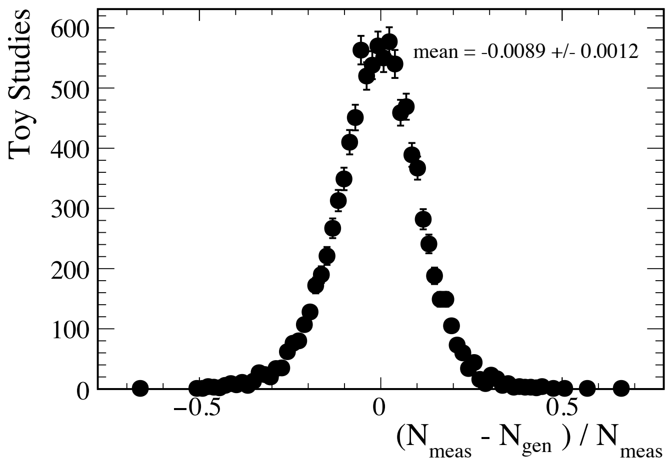


Fig. 7.1 Fractional bias of the measured $B_s^0 \to \mu^+\mu^-$ yield from toy studies for the expected number of decays with 4.4 fb^{-1}. Only $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays are included in the toy studies.

Furthermore, the pull distributions for $B_s^0 \to \mu^+\mu^-$ yields for toy studies with higher statistics have means that move towards 0 as shown in Figure 7.3 for the expected number of decays with 50 and 300 fb^{-1}. Therefore the mass fit returns accurate yields for $B_s^0 \to \mu^+\mu^-$ and the biased pull distribution arises from the low statistics of the data set. The same reasoning can be applied to the pull distribution of the yields of combinatorial background decays that have a sightly less bias mean value of ~ 0.006 compared to the $B_s^0 \to \mu^+\mu^-$ yields.

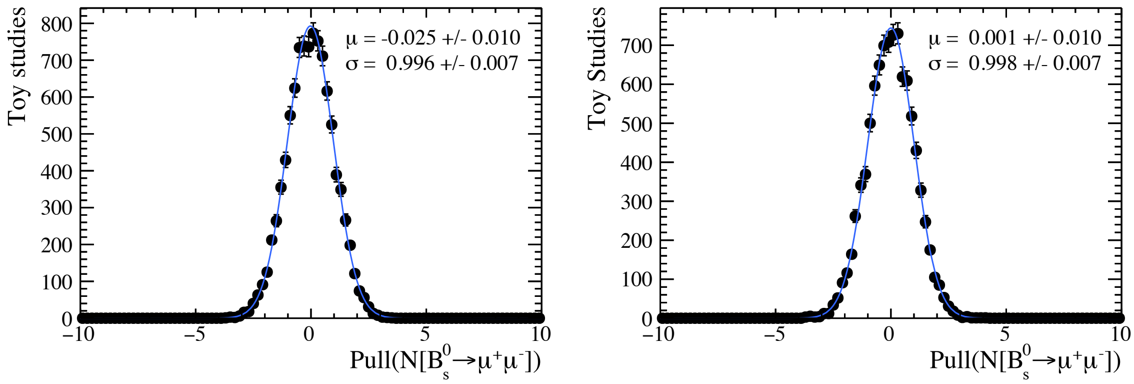


Fig. 7.2 Pull distribution for $B_s^0 \to \mu^+\mu^-$ measured yields from 10,000 toy studies for the expected number of decays in 50 and 300 fb^{-1}. Only $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays are included in the toy studies.

7.1.3 Overall bias on $\tau_{\mu\mu}$

The remaining area of the fit to investigate is any underlying bias in the fit on the measured value of $\tau_{\mu\mu}$. As discussed in Section **??** the pull distribution for the measured effective lifetime is biased for the expected number of statistics but the pull distribution for $\Gamma_{\mu\mu}$ produces a mean and width consistent with 0 and 1, respectively. However the coverage of the uncertainties of both $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ are reasonable and the biased $\tau_{\mu\mu}$ pull arises from the likelihood function as discussed in Section 6.4.1.

The overall bias in the fit for measuring $\tau_{\mu\mu}$ is evaluated from the difference between the measured and generated values from toy studies. A total of 10,000 toy studies are performed generating only $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays so the fit bias is not masked by contamination from mis-identified backgrounds. The difference between the measured and generated $\tau_{\mu\mu}$ values is evaluated for toy studies with the expected and also the observed number of $B_s^0 \to \mu^+\mu^-$ decays. The fit bias is evaluated for the observed number of decays because these are fewer than expected. The resulting distributions are shown in Figure 7.3. The mean of the $\tau_{\mu\mu}$ difference distribution is 0.02 ps for the expected number of decays and 0.03 ps for the observed number of decays. Therefore the larger uncertainty of 0.03 ps is used as the measured of the systematic uncertainty cause by the fit on the final result for $\tau_{\mu\mu}$.

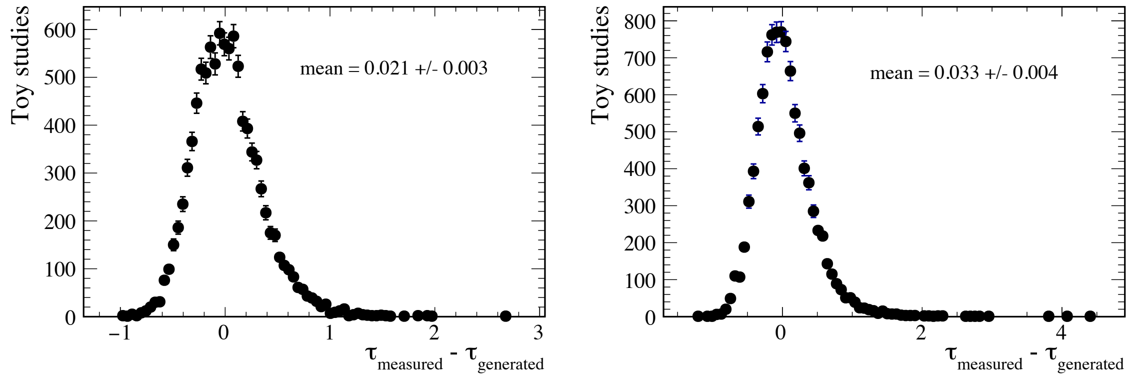


Fig. 7.3 Overall bias in $\tau_{\mu\mu}$, evaluated as the difference between the measured, $\tau_{measured}$, and the generated, $\tau_{generated}$, lifetimes for toy studies using the expected (left) and observed (right) $B_s^0 \to \mu^+\mu^-$ and combinatorial background yields.

7.2 Background contamination

The mass fit configuration used to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime only includes components for $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays and only candidates with a B_s^0 mass between 5320 - 6000 MeV/c^2 are used in the mass fit. Although the majority of background decays from mis-identified semi-leptonic and $B \to h^+h^-$ decays and $B^0 \to \mu^+\mu^-$ decays fall outside this mass window, as shown in Figure 6.13, the tails of some backgrounds still end up in the mass window. The backgrounds that need to be considered are $B^0 \to \mu^+\mu^-$, $B \to h^+h^-$, $\Lambda_b^0 \to p\mu^-\nu_\mu$, $B^0 \to \pi^-\mu{+}\nu_\mu$, $B_s^0 \to K^-\mu{+}\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$, $B_c^+ \to J/\psi\mu^+\nu_\mu$, with $B \to h^+h^-$, $B^0 \to \mu^+\mu^-$ and $\Lambda_b^0 \to p\mu^-\nu_\mu$ being of particular importance for the chosen mass range.

The number of expected background decays and their mass *pdfs* in the mass range 4900 - 6000 MeV/c^2 were computed using the methods described in Chapter 5. The number of decays from each background type expected in the smaller mass range 5320 - 6000 MeV/c^2 are computed by integrating the mass *pdfs*. The expected yields for each background in both mass ranges are given in Table 7.2. The expected yields are all < 1 for each background source in the mass range 5320 - 6000 MeV/c^2.

The impact of backgrounds not modelled in the mass fit on the measured $\tau_{\mu\mu}$ value is evaluated using two sets of toy studies. The toy studies have the same general

Decay	Expected yield in mass range	
	4900 - 6000 MeV/c^2	5320 - 6000 MeV/c^2
$B_s^0 \to \mu^+\mu^-$	30.9	30.5
$B^0 \to \mu^+\mu^-$	3.3	0.2
$B \to h^+h^-$	9.7	0.9
$\Lambda_b^0 \to p\mu^-\nu_\mu$	13.3	0.6
$B^0 \to \pi^-\mu{+}\nu_\mu$	40.5	0.1
$B_s^0 \to K^-\mu{+}\nu_\mu$	9.1	0.0
$B^+ \to \pi^+\mu^+\mu^-$	6.0	0.0
$B^0 \to \pi^0\mu^+\mu^-$	4.9	0.6
$B_c^+ \to J/\psi\mu^+\nu_\mu$	9.8	0.0
Combinatorial background	66.2	40.6

Table 7.2 Number of expected decays in data passing the $B_s^0 \to \mu^+\mu^-$ effective lifetime selection in the mass ranges 4900 - 6000 MeV/c^2 and 5320 - 6000 MeV/c^2.

set up as described in Section 6.4. One set of toy studies assumes there are no backgrounds other than the combinatorial background and therefore only $B_s^0 \to \mu^+\mu^-$ and combinatorial background candidates are generated. The second set of toy studies generates all possible background decays. The expected yields are fluctuated using a Poisson distribution around their expected values to 2 decimal places. For each configuration 10,000 toy studies were performed and the pull distributions for $\Gamma_{\mu\mu}$ of each toy set up is compared. The pull distributions for $\tau_{\mu\mu}$ are not used due to their non-Gaussian distribution disused in Section 6.4.1.

The inclusion of all the background decays causes a shift in the mean of the $\Gamma_{\mu\mu}$ pull distribution of 0.024 ps $^{-1}$ as shown in Figure 7.4. Therefore, assuming the expected uncertainty in Section 6.4.2 of 0.28 ps for $\tau_{\mu\mu}$ the systematic shift from not including all backgrounds in the fit configuration is 0.007 ps.

The expected number of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays assumes the Standard Model branching fractions, however the current world averages (excluding the results presented in Chapter 5) are different and the expected number of $B_s^0 \to \mu^+\mu^-$ decays decreases and the expected number of $B^0 \to \mu^+\mu^-$ decays increases, the changes in the yields are given in Table 7.3. The toy studies were repeated with the world average but the shift in the mean of the pull distribution was smaller, therefore the larger value from the SM predictions are used.

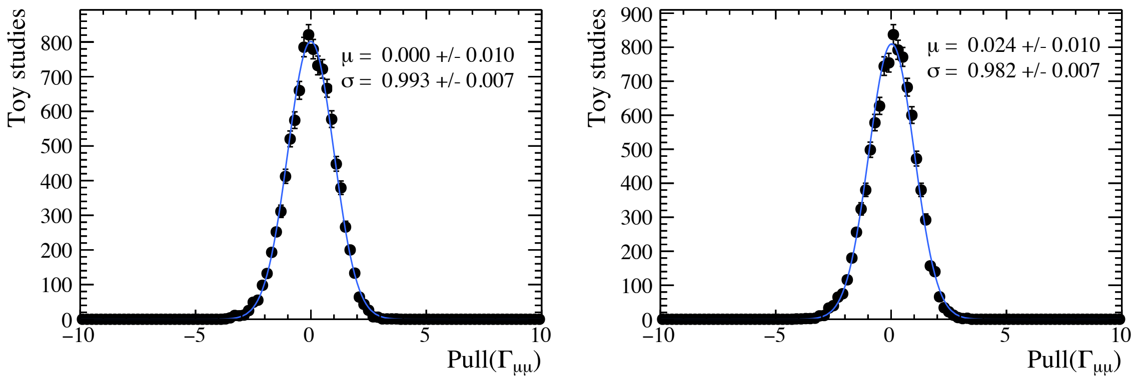


Fig. 7.4 Pull distribution for $\Gamma_{\mu\mu}$ from 10,000 toy studies, using background sources of combinatorial background only (left) and combinatorial background, mis-identified semi-leptonic and $B \to h^+h^-$ decays and $B^0 \to \mu^+\mu^-$ decays (right).

Decay	Expected yield in 5320 - 6000 MeV/c^2	
	Standard Model	World average
$B_s^0 \to \mu^+\mu^-$	30.5	22.5
$B^0 \to \mu^+\mu^-$	0.2	0.7

Table 7.3 Number of expected decays in data passing the $B_s^0 \to \mu^+\mu^-$ effective lifetime selection in the mass range 5320 - 6000 MeV/c^2, using the Standard Model and world average branching fraction values.

7.3 Mass *pdf* parameters

The data collected in Run 1 and Run 2 are combined for the measurement of the $B_s^0 \to \mu^+\mu^-$ effective lifetime and the mass and decay time fits are applied to the combined data. However the parameters used in the mass *pdf* in Table 5.1 were evaluated specifically for Run 1 data and different parameters are available for Run 2 data. Therefore the influence of the choice of mass *pdf* parameters on the measured $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$ values must be evaluated.

Toy studies are performed to understand the size of the impact of the mass fit choice on the effective lifetime measurement. Only $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays are generated to separate mass *pdf* effects from the contamination of mis-identified backgrounds and $B^0 \to \mu^+\mu^-$ decays in the mass window. $B_s^0 \to \mu^+\mu^-$ candidates are generated using the Run 1 parameters in Table 5.1 but the mass fit is performed using the Run 2 parameters in Table 5.2. The pull distribution for 10,000 toy studies of $\Gamma_{\mu\mu}$ from this configuration are compared with those from toy studies where Run 1 parameters are used to generate and fit the mass distribution. The change in the measured lifetime with the mass *pdf* parameters is negligible as shown in Figure 7.5 that overlays the pull distributions for the two toy configurations studied. Therefore no systematic uncertainty is assigned.

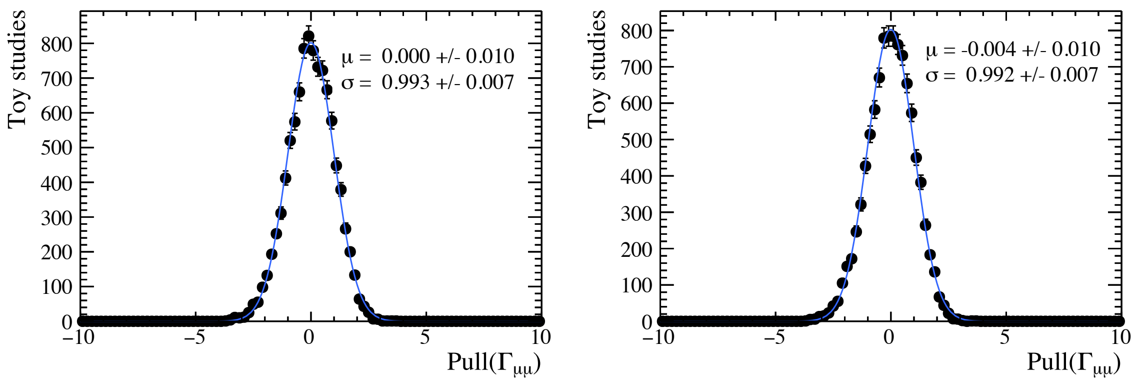


Fig. 7.5 Pull distribution for $\Gamma_{\mu\mu}$ from 10,000 toy studies where the $B_s^0 \to \mu^+\mu^-$ mass distribution is generated using the Run 1 parameters and the mass fit is performed using Run 1 parameters (left) and Run 2 parameters (right).

7.4 Acceptance function accuracy

The decay time acceptance function is determined from weighted simulated decays, as discussed in Section 6.3.1. It relies on the assumption that the weighted simulated decays model the data reasonable well. To test this assumption, as well as the strategy used to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime, the lifetimes of the more abundant $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ decays are measured.

$B^0 \to K^+\pi^-$ decays have a much larger branching fraction the $B_s^0 \to \mu^+\mu^-$ decays at XXX and are therefore more abundant in data. The selection requirements used to identify these decays in 2011, 2012, 2015 and 2016 data are detailed in Chapter 4 and are kept very similar to the $B_s^0 \to \mu^+\mu^-$ selection. All candidates are required to be TIS, although this considerably reduces the statistics, in order to keep the $B \to h^+h^-$ trigger efficiency similar to that of decays. The trigger lines that identify $B_s^0 \to \mu^+\mu^-$ decays in data are relatively unbiased with respect to the B_s^0 decay time distribution and the same is true for trigger lines that identify $B \to h^+h^-$ decays as TIS, whereas TOS triggers that identify $B \to h^+h^-$ decays create a large bias on the decay time distribution due to the dependence on the trigger lines on IP, IP χ^2 and flight distance variables. The $\mathrm{DLL}_{K\pi}$ variable is used to separate $B^0 \to K^+\pi^-$ decays from other $B \to h^+h^-$ decays, as detailed in Section 4.4.3 and candidates are reconstructed with the daughter with the highest $\mathrm{DLL}_{K\pi}$ value assigned the kaon mass hypothesis and the daughter particle with the lowest $\mathrm{DLL}_{K\pi}$ value the pion mass hypothesis.

The measurement of the lifetime is performed in the same way as the $B_s^0 \to \mu^+\mu^-$ effective lifetime measurement and all years of data are combined into one data set. An extended unbinned maximum likelihood fit is performed to the $B^0 \to K^+\pi^-$ mass distribution. Components for $B^0 \to K^+\pi^-$, $B_s^0 \to K^+\pi^-$ and combinatorial background decays are included in the mass fit. Both B meson decays are modelled by Crystal Ball functions with parameters taken from some source. Combinatorial background decays are modelled by an exponential function with the slope left free in the mass fit. The mass fit, shown in Figure 7.6, is used to calculate sWeighted that are re-normalised using equation **??**. The lifetime, $\tau_{K\pi}$, and the inverse lifetime $\Gamma_{K\pi}$, are measured from an unbinned maximum likelihood fir to the sWeighted decay time distribution.

The decay time *pdf* has the same form as the one used for the $B_s^0 \to \mu^+\mu^-$ decays in equation 6.3 and the acceptance parameters are found from a fit to weighted $B^0 \to K^+\pi^-$ simulated decays using the same method described in Section 6.3.1. In the same way as $B_s^0 \to \mu^+\mu^-$ decays the number of tracks in the event are weighted using the same weights as for $B_s^0 \to \mu^+\mu^-$ decays. However the yearly weights applied to combined simulated $B^0 \to K^+\pi^-$ decays from each year are not dependant on $B_s^0 \to J/\psi\phi$ decays. Since $B^0 \to K^+\pi^-$ decays have a high yield in data, mass fits to each year are used to find the yields and combined the simulated decays from each

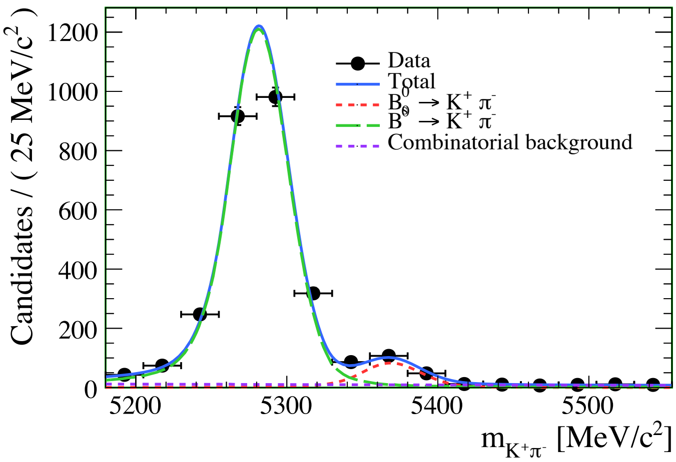


Fig. 7.6 Maximum likelihood fit to the mass distribution of $B^0 \to K^+\pi^-$ decays for data taken in 2011, 2012, 2015 and 2016. Components for $B^0 \to K^+\pi^-$, $B_s^0 \to K^+\pi^-$ and combinatorial background decays are included.

year. The same mass *pdf* used to fit the combined mass distribution is applied to each year. This simplifies the efficiency terms and the weights combining simulated decays in different years are then computed as

$$\omega_i = \frac{Y_i^{data}}{\sum\limits_j Y_j^{data}} \cdot \frac{\sum\limits_k N_k^{K\pi}}{N_i^{K\pi}} \tag{7.1}$$

where Y_i^{data} are the $B^0 \to K^+\pi^-$ yields in each year of data and $N_i^{K\pi}$ the number of simulated decays available for each year. The acceptance function fit is shown in Figure 7.7, for consistency the same simulated versions are used for simulated $B^0 \to K^+\pi^-$ decays as for $B_s^0 \to \mu^+\mu^-$ decays for each year.

The measured $B^0 \to K^+\pi^-$ lifetime is

$$\tau_{K\pi} = 1.52 \pm 0.03 \text{ ps} \tag{7.2}$$

where only the statistical uncertainty is given and the decay time fit is shown in Figure 7.8. The measured results are consistent with the PDF value of 1.520 ± 0.004 ps []. Therefore the measurement strategy used to find the $B_s^0 \to \mu^+\mu^-$ effective lifetime has been shown to work. The statistical uncertainty of the $B^0 \to K^+\pi^-$ decay time fit are assigned as systematic uncertainty to provide a measure of how well the acceptance function can be determined from weighted simulated decays for measuring $\tau_{\mu\mu}$.

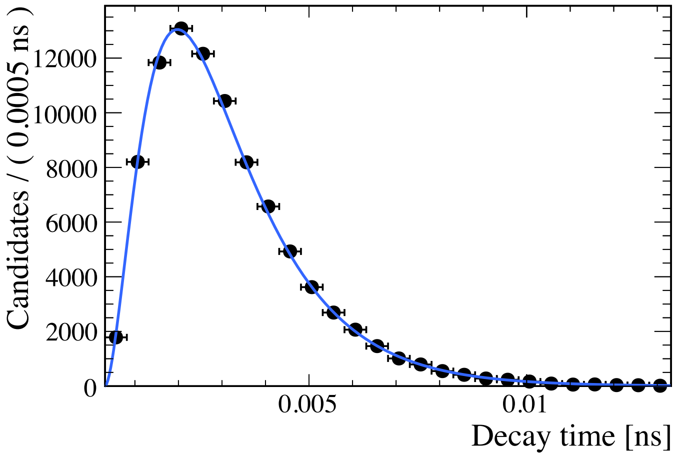


Fig. 7.7 Decay time distribution in weighted 2011, 2012, 2015 and 2016 simulated decays and the maximum likelihood fit results to determine the acceptance function parameters.

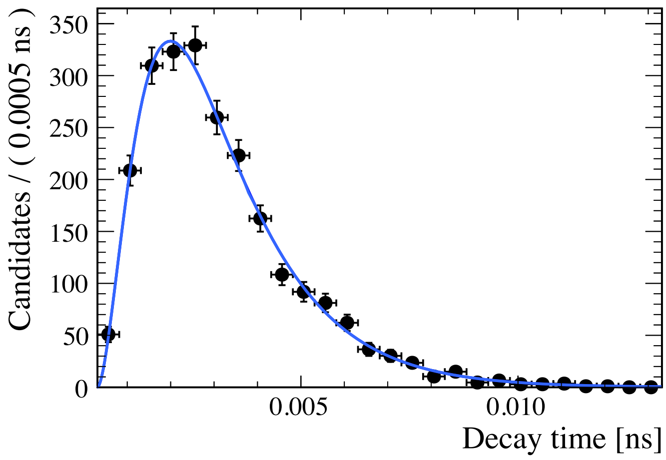


Fig. 7.8 Maximum likelihood fit to the signal weighted decay time distribution of $B^0 \to K^+\pi^-$ decays for data taken in 2011, 2012, 2015 and 2016.

However, the determination of the $B_s^0 \to \mu^+\mu^-$ acceptance function relies on weights taken from the number of tracks in an event from $B^0 \to K^+\pi^-$ decays in data and simulation. Although the measurement of the $B^0 \to K^+\pi^-$ lifetime shows the procedure and weighting method works for $B^0 \to K^+\pi^-$ decays it does not show that the weights taken from $B^0 \to K^+\pi^-$ decays for the number of tracks in an event can be applied to other decays. Therefore as a cross check, the lifetime of $B_s^0 \to K^+K^-$ decays is also measured.

The same measurement strategy is used for $B_s^0 \to K^+K^-$ decays as $B_s^0 \to \mu^+\mu^-$ and $B^0 \to K^+\pi^-$ and candidates in 2012 and 2015 data are identified using the selection requirements in Chapter 4. Only 2012 and 2015 data are used due to the available simulation versions of simulated $B_s^0 \to K^+K^-$ decays compared to $B_s^0 \to \mu^+\mu^-$. Once again TIS triggers are used to keep a lifetime unbiased trigger efficiency and candidates are reconstructed assuming both daughters are kaons. The mass *pdf* includes $B_s^0 \to K^+K^-$ and combinatorial background decays only and the same *pdf* is used for $B_s^0 \to K^+K^-$ as for $B_s^0 \to K^+\pi^-$. An unbinned maximum likelihood fit used to extract the sWeights is shown in Figure 7.9.

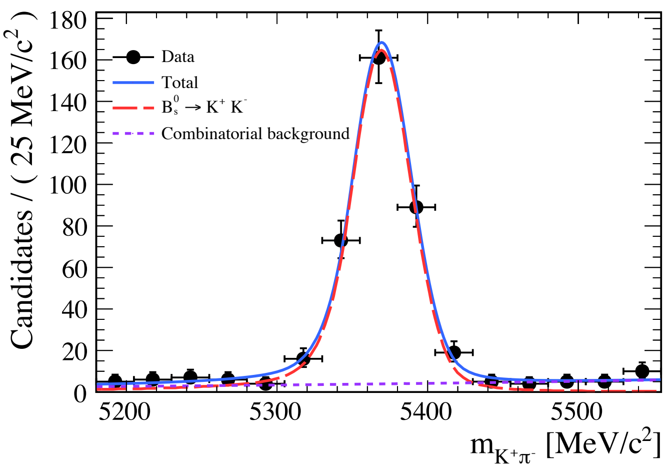


Fig. 7.9 Maximum likelihood fit to the mass distribution of $B_s^0 \to K^+K^-$ decays for data taken in 2012 and 2015. Components for $B_s^0 \to K^+K^-$ and combinatorial background decays are included in the mass fit.

The $B_s^0 \to K^+K^-$ acceptance is found using the same method as $B_s^0 \to \mu^+\mu^-$ with decays used to determine the relative proportions of decays in each year of data. Figure 7.10 shows the acceptance function fit and the decay time fit to data. The measured values for the lifetime, τ_{KK}, is

$$\tau_{KK} = 1.39 \pm 0.06 \text{ ps} \tag{7.3}$$

where only the statistical uncertainty is given and the fit results are shown in Figure 7.11. The measured values are consistent with the predicted value of 1.395 ± 0.020 ps [39] and therefore shows that $B^0 \to K^+\pi^-$ weights can be used for other decays as well as $B^0 \to K^+\pi^-$.

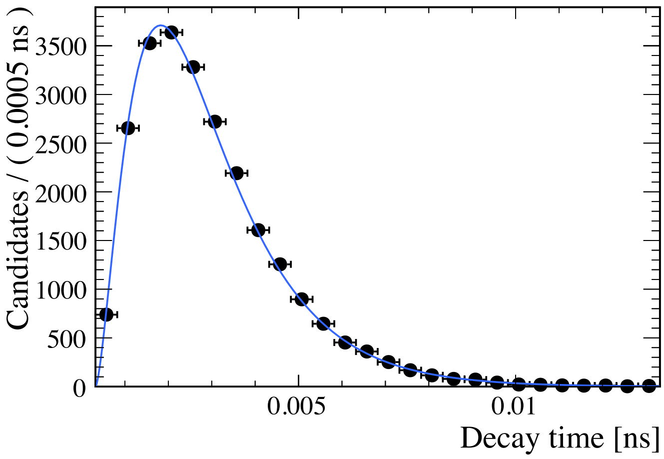


Fig. 7.10 Decay time distribution in weighted 2012 and 2015 simulated decays and the maximum likelihood fit results to determine the acceptance function parameters.

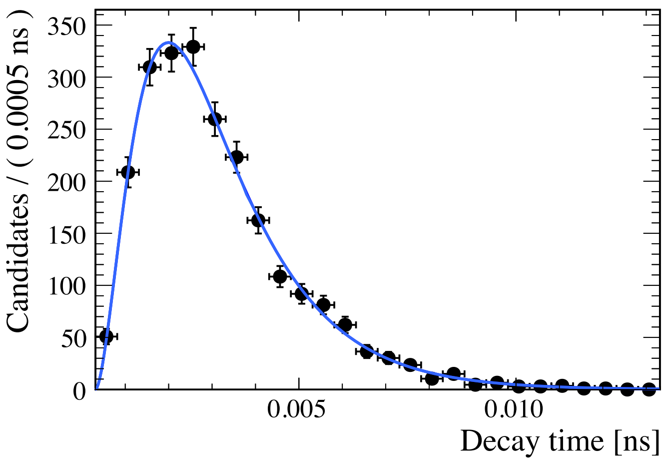


Fig. 7.11 Maximum likelihood fit to the signal weighted decay time distribution of $B_s^0 \to K^+K^-$ decays for data taken in 2012 and 2015 data.

7.5 Incorrectly assigned primary vertices and additional detector resolution effects

Measuring the $B_s^0 \to \mu^+\mu^-$ effective lifetime accurately relies on the B_s^0 candidate being assigned to the correct primary vertex in the event, incorrect assignment would lead to the wrong value for the decay time of an event. In the papers [40, 41] that study the lifetime of $B \to J/\psi X$ decays at LHCb a component is included into the decay time fit to model the number of incorrectly assigned primary vertices (PVs) as well as the resolution of the detector. The decay time fit consists of a *pdf* describing X convoluted by the sum of three Gaussian functions; two narrow Gaussian functions model the decay time resolution of the detector and a third wider Gaussian function corresponds to $< 1\%$ of decays assigned incorrect PVs. The decay time fit to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime does not explicitly model incorrectly assigned PVs or the detector resolution, however these effects will to some degree be included into the acceptance function.

A similar model is the papers [40, 41] is used to check the affect of decays with incorrectly assigned PVs and detector resolution effects that are not included in the acceptance function on the measured lifetime. A set of 1 million decays are generated using the decay time model

$$\epsilon(t)[\mathcal{R}(t) \otimes e^{-t/\tau}] \tag{7.4}$$

where $\epsilon(t)$ is the acceptance function with parameters given in Table 6.3 and $\mathcal{R}(t)$ is the resolution function composed of 3 Gaussian functions. Decays are generated assuming the $B_s^0 \to \mu^+\mu^-$ effective lifetime is equal to the lifetime of the heavy B_s^0 mass eigenstate of XX $\pm$ YY ps. A fit is then performed to the generated decay time distribution but only the resolution term is not included in the fit *pdf*. The measured $\tau_{\mu\mu}$ value is then compared to the value used to generate the decays.

The resolution function is determined from weighted simulated $B_s^0 \to \mu^+\mu^-$ decays that were used to compute the acceptance function in Section 6.3.1. The difference between the reconstructed decay time and the 'true' decay time with which decays were generated is computed for each decay that passes the full selection, the resulting distribution is fitted with a resolution function composed of the sum of three Gaussian functions. Each Gaussian function has the same mean value, which is left free in the fit, but different widths that are also free in the final fit. The fit parameters are shown in Table 7.4.The resulting distribution has a similar form to those used in [40, 41], where the detector resolution is modelled with two narrow Gaussian and the Gaussian for incorrectly assigned PVs is broader and describes a small fraction of decays.

Parameter	Fit value
μ (ps)	0.00063 $\pm$ 0.00005
$\sigma_1(ps)$	5.62$\pm$0.07
f_1	0.006
$\sigma_2(ps)$	0.0573$\pm$ 0.0003
f_2	0.313
$\sigma_3(ps)$	0.0294 $\pm$ 0.0001
f_3	0.681

Table 7.4 Parameters from fit to the difference between the reconstructed decay time and the true decay time for simulated decays that pass the $B_s^0 \to \mu^+\mu^-$ effective lifetime selection. The mean used for all Gaussian is μ and σ_i are the widths of each Gaussian which make up a fraction f_i for the total sum.

The result from the fit to the generated decays without the resolution function included is $\tau\,() = 1.6098 \pm 0.0014 ps$, which is consistent with the lifetime of generate events. The difference between the lifetime used to generate events and the fitted value is 0.0002 ps, a factor of 10 smaller than the lowest systematic uncertainty. This cross check shows that the presence of incorrectly assigned PVs or detector resolution effects that are not included in the acceptance function have a negligible effect on the $B_s^0 \to \mu^+\mu^-$ effective lifetime.

However this check assumed that simulated decays provide a good estimate of the number of incorrectly assigned PVs, as a check Figure 7.12 shows the number of $B^0 \to K^+\pi^-$ events passing the selection outlined in Section **??** for simulated $B^0 \to K^+\pi^-$ decays and sWeighted decays data for each year. On average there are more PVs per event in simulation compared to data therefore using simualteion would give an overestimation of the number of incorrectly assigned PVs expected in data.

7.6 Combinatorial background decay time model

The decay time distribution of combinatorial background decays is largely unknown due to the nature of the background, the model used for this distribution in the toy studies is described in Section 3.1. The decay time distribution of combinatorial background decays consists of mostly a short lived component with a lifetime of ps and a long lived conpoment with a lifetime of Y ps.

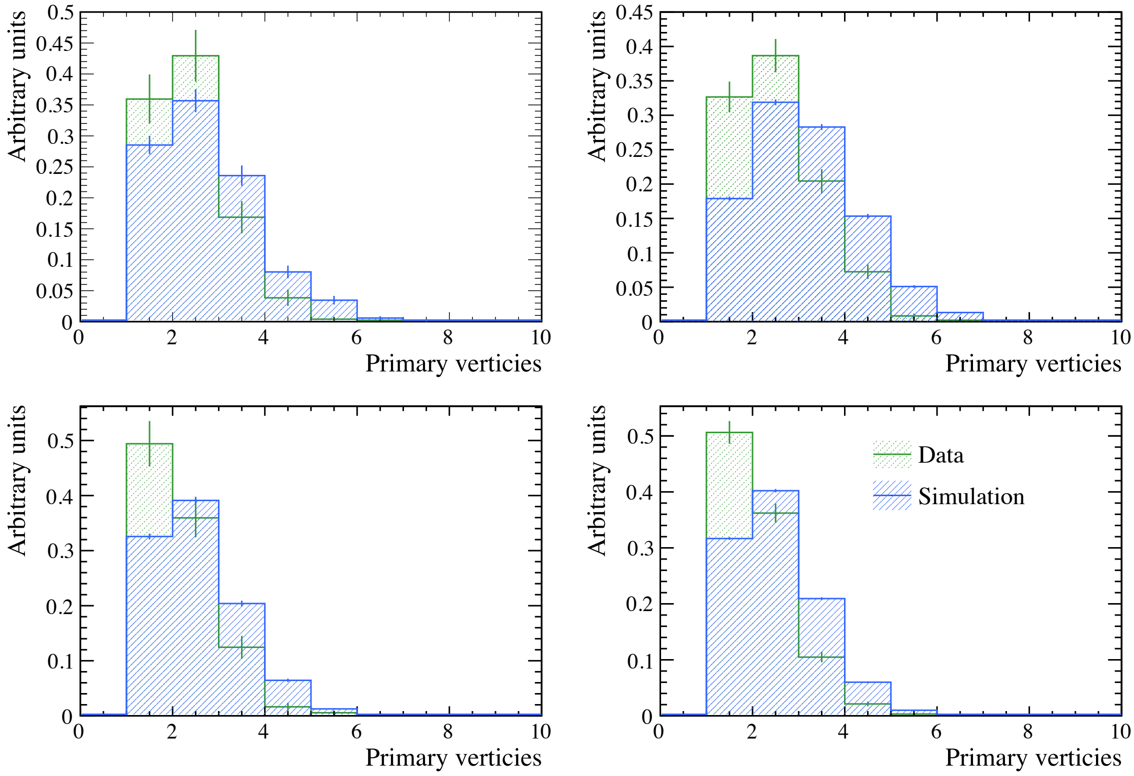


Fig. 7.12 The distributions for the number of primary vertices in an event $B^0 \to K^+\pi^-$ data and simulated decays for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right), for events that pass the selection in Section **??**.

The sWeighting method is sensitive to the background decay time components that are significantly longer lived than the signal lifetimes and can lead to biased estimated of $\tau_{\mu\mu}$ and $\Gamma_{\mu\mu}$. During the selection an upper decay time cut is applied to remove long-lived backgrounds, this cut is very effective at stopping the bias entering the final results. However the remaining effect must be evaluated.

As discussed in Section 3.1 determining the decay time distribution of combinatorial background decays is challenging because there are too few decays left in either $B_s^0 \to \mu^+\mu^-$ decay or simulated decays after the selection requirements to determine the decay time *pdf*. Therefore the combinatorial background of $B \to h^+h^-$ decays in the mass range 5600 - 6000 MeV/c^2 is used. The validity of using $B \to h^+h^-$ combinatorial background to model $B_s^0 \to \mu^+\mu^-$ combinatorial background is studied by comparing the average lifetime of combinatorial background decays in bins of global BDT. The average lifetimes are shown in Table 7.5 for decays in data passing the selection requirements for $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays and in the mass ranges 5447 - 6000 MeV/c^2 and 5600 - 6000 MeV/c^2, respectively. At low values of the global BDT the average lifetimes are similar and for both $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ backgrounds and the lifetime increases with the output of the global BDT. Overall $B \to h^+h^-$ combinatorial background decays are longer lived than $B_s^0 \to \mu^+\mu^-$ combinatorial background decays therefore making the $B \to h^+h^-$ combinatorial background decay time model a conservative estimate for $B_s^0 \to \mu^+\mu^-$ combinatorial background as far as the affect of long lived components in concerned. The model used for the combinatorial background decay time currently introduces no significant bias into the pull distribution of $\Gamma_{\mu\mu}$ for toy studies as shown in Table 6.8. However the size of a systematic bias from the choice of the decay time distribution is estimated by two sets of toy studies. The first uses the background decay time distribution in Table 6.4 and the second set uses the same decay time distribution of expect the lifetime of each component are made longer by 1 standard deviation and the fraction of the long lived component is increase by one standard deviation as well. For both sets of toy studies only combinatorial background and $B_s^0 \to \mu^+\mu^-$ decays are generated and 10,000 studies are performed for each configuration.

The resulting pull distributions for $\Gamma_{\mu\mu}$ are shown in Figure 7.13, the difference in the mean value of the distributions for the two studies is negligible and the width changes by 0.008 ps $^{-1}$ between the two studies. The change in the width is the largest and therefore is taken as the systematic uncertainty, assuming the median expected uncertainties for $\tau_{\mu\mu}$. Therefore the systematic uncertainty due to the combinatorial background decay time model is 0.002 ps for $\tau_{\mu\mu}$.

	$B_s^0 \to \mu^+\mu^-$		$B \to h^+h^-$	
BDT bin	mean decay time / ps	Number of candidates	mean decay time / ps	Number of candidates
1	1.178 ± 0.005	50,695	1.124 ± 0.001	964,502
2	1.936 ± 0.098	244	2.394 ± 0.022	8,838
3	2.570 ± 0.327	46	2.781 ± 0.051	2,373
4	2.210 ± 0.361	17	3.023 ± 0.076	1,125
5	2.582 ± 1.103	4	3.417 ± 0.112	655
6	2.540 ± 0.390	3	3.978 ± 0.187	313
7	2.868 ± 1.048	2	4.626 ± 0.363	109
8	-	0	5.706 ± 0.683	35

Table 7.5 The mean decay time of $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ candidates in 2011, 2012, 2015 and 2016 data in bins of the global BDT output. The mass ranges 5447 - 6000 MeV/c^2 and 5600 - 6000 MeV/c^2 are used for $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays respectively.

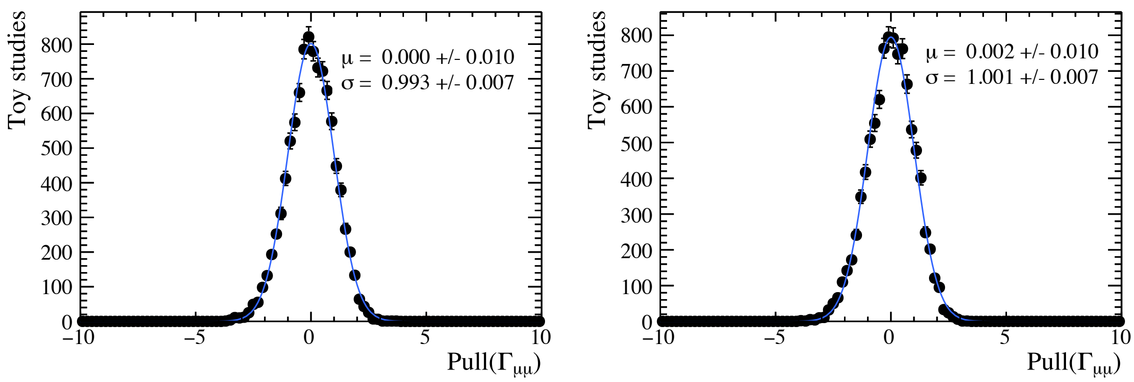


Fig. 7.13 Pull distributions for $\Gamma_{\mu\mu}$ from 10,000 toy studies using the nominal combinatorial background decay time model (left) and the nominal model with the lifetimes and fraction of longer lived decays increased by one standard deviation (right).

7.7 Mix of B_s^0 mass eigenstates

In the Standard Model the $B_s^0 \to \mu^+\mu^-$ effective lifetime is equal to the lifetime of the heavy B_s^0 mass eigenstate. However the real $B_s^0 \to \mu^+\mu^-$ effective lifetime could be in between the lifetimes of the light and heavy mass eigenstates.

As shown in Section 6.3 the selection efficiency used to identify $B_s^0 \to \mu^+\mu^-$ decays in data is not uniform across the decay time range. The selection rejects a greater proportion of candidates with short lifetimes compared to candidates with longer lifetimes. Therefore the presenece of any B_s^0 light mass eigenstates decaying as $B_s^0 \to \mu^+\mu^-$ in data could be masked by the bias in the decay time distribution because the efficiency to selection light B_s^0 mass eigenstates is lower than the efficiency to select heavy B_s^0 mass eigenstates.

The size of this effect has been estimated using two simple toy studies. The first assumes that the selection has no bias on the decay time distribution and 1 million candidates are generated with equal contributions from the heavy and light B_s^0 mass eigenstates.A second set of 1 million candidates are generated with the same mix of eigenstates but with a more realistic model, using the $B_s^0 \to \mu^+\mu^-$ acceptance function. A fit is performed to the first set of candidates with a single exponential function and the second set with the acceptance function and exponential function in order to find $\tau_{\mu\mu}$ for each distribution. The acceptance parameters are fixed in the second fit.

The values of $\tau_{\mu\mu}$ are compared for the two studies and a systematic uncertainty is assigned for the change in $\tau_{\mu\mu}$ caused by the inclusion of the acceptance function. The maximum likelihood fits used to measure $\tau_{\mu\mu}$ in the two toy studies are shown in Figure 7.14 and the differences between the measured lifetimes for the two studies is 0.018 ps.

7.8 Production asymmetry of B_s^0 and $\bar{B}_s^0$

The $B_s^0 \to \mu^+\mu^-$ effective lifetime is the mean lifetime of an unbiased sample of $B_s^0 \to \mu^+\mu^-$ decays, as discussed in Chapter **??**, and it is given by

$$\tau_{\mu\mu} \equiv \frac{\int_0^\infty t \left\langle \Gamma \left(B_s^0(t) \to \mu^+\mu^- \right) \right\rangle dt}{\int_0^\infty \left\langle \Gamma \left(B_s^0(t) \to \mu^+\mu - \right) \right\rangle dt} \tag{7.5}$$

where the untagged decay rate is

$$\left\langle \Gamma \left(B_s^0(t) \to \mu^+\mu^- \right) \right\rangle = \Gamma \left(B_s^0(t) \to \mu^+\mu^- \right) + \Gamma \left(\bar{B}_s^0(t) \to \mu^+\mu^- \right). \tag{7.6}$$

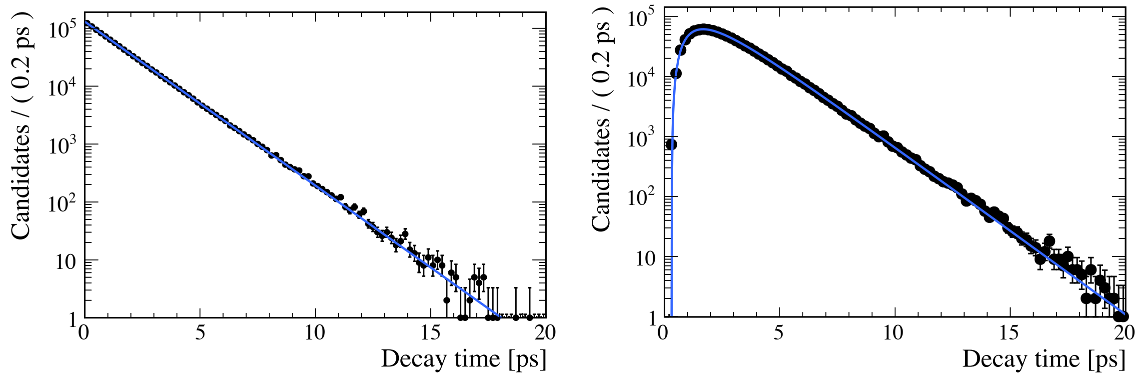


Fig. 7.14 Maximum likelihood fits to the decay time distribution to measure $\tau_{\mu\mu}$ for $B^0_s \to \mu^+\mu^-$ decays that are composed of an equal mix from the heavy and light mass eigenstates. Decay time distributions are generated assuming a flat acceptance function (left) and the acceptance function used to describe $B^0_s \to \mu^+\mu^-$ decays in data (right).

and assumes that B^0_s and $\bar{B}^0_s$ mesons are produced at equal rates. This assumption has been made in the measured value of $\tau_{\mu\mu}$ presented in Chapter **??**. However the LHC is a *pp* collider therefore B^0_s and $\bar{B}^0_s$ mesons are not produced at equal rates, the affect of the production asymmetry of B^0_s and

The production asymmetry is given by

$$A_p \equiv \frac{\sigma\left(B^0_s\right) - \sigma\left(\bar{B}^0_s\right)}{\sigma\left(B^0_s\right) + \sigma\left(\bar{B}^0_s\right)} \tag{7.7}$$

where $\sigma\left(B^0_s\right)$ and $\sigma\left(\bar{B}^0_s\right)$ are the production cross-sections for B^0_s and $\bar{B}^0_s$ mesons respectively. The production asymmetry was measured by LHCb in 2011 at a centre of mass energy of 7 TeV as $A_p = (1.09 \pm 2.61 \pm 0.66)\%$ []. The presence of the production asymmetry modifies the $B^0_s \to \mu^+\mu^-$ decay rate to be

$$\Gamma\left(B^0_s(t) \to \mu^+\mu^-\right) = \left(\frac{1+A_p}{2}\right)\Gamma(B^0_s(t) \to \mu^+\mu^-) + \left(\frac{1-A_p}{2}\right)\Gamma(\bar{B}^0_s(t) \to \mu^+ \tag{7.8}$$

The affect of the production asymmetry on the measured lifetime can be determined from equations 7.5 and 7.8 using the decay rates of $B^0_s \to \mu^+\mu^-$ and $B^0_s \to \bar{}\mu^+\mu^-$ given in Section **??** as

$$\Gamma(B_s^0(t) \to \mu^+\mu^-) = \frac{1}{2} N_{\mu^+\mu^-} |A_{\mu^+\mu^-}|^2 (1+|\epsilon|^2) e^{-\Gamma_s t} \Big\{ \cosh\frac{\Delta\Gamma_s t}{2} + C_\lambda \cos(\Delta m_s t) + A_{\Delta\Gamma}^{\mu^+\mu^-} \sinh\frac{\Delta\Gamma_s t}{2} + S_\lambda \sin(\Delta m_s t) \Big\}, \quad (7.9)$$

$$\Gamma(\bar{B}_s^0(t) \to \mu^+\mu^-) = \frac{1}{2} N_{\mu^+\mu^-} |A_{\mu^+\mu^-}|^2 (1+a)(1+||^2) e^{-\Gamma_s t} \Big\{ \cosh\frac{\Delta\Gamma_s t}{2} - C_\lambda \cos(\Delta m_s t) + A_{\Delta\Gamma}^{\mu^+\mu^-} \sinh\frac{\Delta\Gamma_s t}{2} - S_\lambda \sin(\Delta m_s t) \Big\} \quad (7.10)$$

The total decay rate then becomes

$$\Gamma\left(B_s^0(t) \to \mu^+\mu^-\right) = \frac{1}{2} N_{\mu^+\mu^-} |A_{\mu^+\mu^-}|^2 (1+|\epsilon|^2) e^{-\Gamma_s t} \Big\{ \cosh\frac{\Delta\Gamma_s t}{2} + A_{\Delta\Gamma}^{\mu^+\mu^-} \sinh\frac{\Delta\Gamma_s t}{2} + A_p \left[C_\lambda \cos(\Delta m_s t) + S_\lambda \sin(\Delta m_s t)\right] \Big\} + \mathcal{O}(a). \quad (7.11)$$

where the production asymmetry has introduced an additional oscillatory term which disappears when $A_p = 0$.Using the relationships $\Delta\Gamma_s = \Gamma_L - \Gamma_H$ and $\Gamma_s = (\Gamma_L + \Gamma_H)/2$ and ignoring terms $\mathcal{O}(a)$, the decay rate becomes

$$\Gamma\left(B_s^0(t) \to \mu^+\mu^-\right) \simeq N' \Big\{ (1 - A_{\Delta\Gamma)}^{\mu^+\mu^-} e^{-\Gamma_L t} + (1 + A_{\Delta\Gamma}^{\mu^+\mu^-}) e^{-\Gamma_H t} + 2A_p e^{-\Gamma_s t} \left[C_\lambda \cos(\Delta m_s t) + S_\lambda \sin(\Delta m_s t)\right] \Big\}. \quad (7.12)$$

where $N' \equiv \frac{1}{4} N_{\mu^+\mu^-} |A_{\mu^+\mu^-}|^2 (1+||^2)$. This decay rate can be used to calculate the $B_s^0 \to \mu^+\mu^-$ effective lifetime in the presence of a production asymmetry. Using integration by parts the contributing terms to the effective lifetime become

$$\int_0^\infty t \left\langle \Gamma\left(B_s^0(t) \to \mu^+\mu^-\right) \right\rangle dt = N' \left\{ \frac{1 - A_{\Delta\Gamma}^{\mu^+\mu^-}}{\Gamma_L^2} + \frac{1 + A_{\Delta\Gamma}^{\mu^+\mu^-}}{\Gamma_H^2} + 2A_p \left[C_\lambda \frac{\Gamma_S^2 - \Delta m_s^2}{(\Delta m_s^2 + \Gamma_s^2)^2} + S_\lambda \frac{2\Gamma_S \Delta m_s}{(\Delta m_s^2 + \Gamma_S^2)^2} \right] \right\} \quad (7.13)$$

and

$$\int_0^\infty \left\langle \Gamma\left(B_s(t) \to \mu^+\mu^-\right) \right\rangle dt = N' \left\{ \frac{1 - A_{\Delta\Gamma}^{\mu^+\mu^-}}{\Gamma_L} + \frac{1 + A_{\Delta\Gamma}^{\mu^+\mu^-}}{\Gamma_H} + 2A_p \left[C_\lambda \frac{\Gamma_s}{\Delta m_s^2 + \Gamma_s^2} + S_\lambda \frac{\Delta m_s}{\Delta m_s^2 + \Gamma_s^2} \right] \right\} \quad (7.14)$$

The affect of the production asymmetry on the effective lifetime can now be calculated using the PDG values of $\Delta m_s = 17.717$ ps^{-1}, $\Gamma_s = 0.662$ ps^{-1}, $\Gamma_L = 0.703$ ps^{-1} and $\Gamma_H = 0.621$ ps^{-1} []. A value of $A_p = 0.40$ is used, which is 1 standard deviation greater than the value measured by LHCb, and $A^{\mu^+\mu^-}_{\Delta\Gamma} = 0.0$ is chosen. Since $(A^{\mu^+\mu^-}_{\Delta\Gamma})^2 + (C_\lambda)^2 + (S_\lambda)^2 = 1$ the it is assumed that $C_\lambda = S_\lambda = \sqrt{0.5}$.

In the presence of the production asymmetry $B^0_s \to \mu^+\mu^-$ effective lifetime is 1.520 ps, whereas when there is no production asymmetry and $A_p = 0$ the effective lifetime is 1.522 ps. Therefore the production asymmetry introduces a bias on 0.002 ps into the measurement of $\tau_{\mu\mu}$, this value is assigned as a systematic uncertainty.

7.9 Summary

The complete list of systematic uncertainties for $\tau_{\mu\mu}$ are summarised in Table 7.6. Adding the uncertainties in quadrature leads to a total uncertainty of 0.05 ps for $\tau_{\mu\mu}$. The uncertainties correspond to 11 % of the observed statistical uncertainty for $\tau_{\mu\mu}$. The small size of the total systematic uncertainties compared to the statistical uncertainties is expected given the observed number of decays.

Sources	Uncertainty
Fit accuracy	0.03
Background contamination	0.007
Acceptance function	0.03
Combinatorial background decay time model	0.008
Mix of B^0_s eigenstates	0.018
Production asymmetry	0.002
Total	0.05

Table 7.6 Summary of the systematic uncertainties on $\tau_{\mu\mu}$, the total uncertainty is achieved by adding the separate uncertainties in quadrature.

Chapter 8

Summary and Future Outlook

Hmm what to put in here?

Summary of the BF analysis with a contour plot of the BFs? And the summary plot that goes into the paper.

Conclusions about the lifetime, discussion about the size of the uncertainties and the future prospects. Therefore I need nice eversions of the plots that I showed to Patrick. Also I should download the paper and the plots.

lumi mean median std dev 4.4 0.348 0.322 0.134 8.0 0.2238 0.2157 0.0494 50 0.078969 0.0787 0.0057 300 0.03172 0.03171 0.00091

Right, so what do I want to put into my thesis.

The search/testing on the SM has been going on for many years, this has included the search for bs2Mumu and bd2mumu decays which has gone on for over 30 years. LHCb was designed to measure b decays, Bs2MuMu was outlined in the roadmap of the experiment. First evidence came from LHCb in X with Y, and the first eviddence and first obervation came in the combination paper. With the observations allows measurements to be made of properties such as the EL. Bs2MuMu decays are important to This thesis decouments the first single experiment observation of bs2mumu and the first measurement of the Bs2MuMu effective lifeimte. I should say what the measurements are and also perhaps talk about the ratio of Bs/Bd? Although I'm not sure that is relevant for this but I could inlucde the contour plot as well. First give any/all BF details then move on the the EL. The measured BF allow contraints to be places on NP models, here's a paper that uses the rresults already! EL, it's pointless to give $\mathrm{A}_D Gatthemomentandthereforeacomparisonispointless, butthisisnonthelessaveryimportantmeasu$

So what does the future hold for the EL measureent? Mention the BF measurement that is in the EOL, say the senstivity is supposed to be X in the future as reported in Y at the end of the chapter The EL, based on the obeerved number of decays, predictions can be makde about the future. However first let us consider the unceratinty we got for the measured number of decays, look it's worse that expected but perfectly reasonable. So for the future, we expect with 8fb, 50 fn and 300 fb. Therefore it will be a long time before it's interesting but with 30fb it will start to be good. However these will be conservative estimates since the analysis stratgu will change as more data is avalible. These are the expected statistical unceratinties and the corresponding systmeatics must decrease too, the largest could are limited by data but some are not, these must be address in the future. An alternative method would be lifetime unbiased which would remove the problems from the acceptance function. I would say to put the plots for the current uncertaintues and those at the end of Run 2 and just the numbers for the 50 and 300 fb prediction. Some nice sentence to finish off. Bs2MuMu has been a decay to watch for years and it still is for the future, the precision of the BF makes it imprnat since NP could still be there and also the Bd is syet to be observed! The ratio is well important. And with more data we get better measurements of A$_DGwhichwillffutherconstrainNP$.

Bibliography

[1]

[2] S. Amato *et al.*, “LHCb technical proposal,” 1998.

[3] A. A. Alves, Jr. *et al.*, “The LHCb Detector at the LHC,” *JINST*, vol. 3, p. S08005, 2008.

[4] “LHCb technical design report: Reoptimized detector design and performance,” 2003.

[5] R. Aaij *et al.*, “LHCb Detector Performance,” *Int. J. Mod. Phys.*, vol. A30, no. 07, p. 1530022, 2015.

[6] R. Aaij *et al.*, “Performance of the LHCb Vertex Locator,” *JINST*, vol. 9, p. 09007, 2014.

[7] R. Aaij *et al.*, “Measurement of the track reconstruction efficiency at LHCb,” *JINST*, vol. 10, no. 02, p. P02007, 2015.

[8] R. W. Forty and O. Schneider, “RICH pattern recognition,” Tech. Rep. LHCb-98-040, CERN, Geneva, Apr 1998.

[9] M. Adinolfi *et al.*, “Performance of the LHCb RICH detector at the LHC,” *Eur. Phys. J.*, vol. C73, p. 2431, 2013.

[10] F. Archilli *et al.*, “Performance of the Muon Identification at LHCb,” *JINST*, vol. 8, p. P10020, 2013.

[11] R. Aaij *et al.*, “Absolute luminosity measurements with the LHCb detector at the LHC,” *JINST*, vol. 7, p. P01010, 2012.

[12] O. Lupton and G. Wilkinson, *Studies of* $\mathrm{D}^0 \to \mathrm{K}^0_\mathrm{S}\mathrm{h}^+\mathrm{h}'^-$ *decays at the LHCb experiment*. PhD thesis, Oxford U., Jul 2016. Presented 14 Sep 2016.

[13] P. Mato, “GAUDI-Architecture design document,” 1998.

[14] R. Antunes-Nobrega *et al.*, *LHCb computing: Technical Design Report*. Technical Design Report LHCb, Geneva: CERN, 2005. Submitted on 11 May 2005.

[15] F. Stagni *et al.*, “LHCbDirac: Distributed computing in LHCb,” *J. Phys. Conf. Ser.*, vol. 396, p. 032104, 2012.

[16] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth.*, vol. A389, pp. 81–86, 1997.

[17] I. Belyaev, T. Brambach, N. H. Brook, N. Gauvin, G. Corti, K. Harrison, P. F. Harrison, J. He, C. R. Jones, M. Lieng, G. Manca, S. Miglioranzi, P. Robbe, V. Vagnoni, M. Whitehead, J. Wishahi, and the LHCb Collaboration, "Handling of the generation of primary events in Gauss, the LHCb simulation framework," *Journal of Physics: Conference Series*, vol. 331, p. 032047, 2011.

[18] M. Clemencic, G. Corti, S. Easo, C. R. Jones, S. Miglioranzi, M. Pappagallo, and P. Robbe, "The LHCb simulation application, Gauss: Design, evolution and experience," *J. Phys. Conf. Ser.*, vol. 331, p. 032023, 2011.

[19] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[20] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.

[21] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth.*, vol. A462, pp. 152–155, 2001.

[22] P. Golonka and Z. Was, "PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays," *Eur. Phys. J.*, vol. C45, pp. 97–107, 2006.

[23] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.

[24] J. Allison *et al.*, "Geant4 developments and applications," *IEEE Trans. Nucl. Sci.*, vol. 53, p. 270, 2006.

[25] I. Bird, "Computing for the Large Hadron Collider," *Ann. Rev. Nucl. Part. Sci.*, vol. 61, pp. 99–118, 2011.

[26] W. L. C. Grid. http://www.cern.ch/LHCgrid.

[27] S. Paterson and A. Tsaregorodtsev, "DIRAC Infrastructure for Distributed Analysis," Feb 2006.

[28] L. Anderlini, A. Contu, C. R. Jones, S. S. Malde, D. Muller, S. Ogilvy, J. M. Otalora Goicochea, A. Pearce, I. Polyakov, W. Qian, B. Sciascia, R. Vazquez Gomez, and Y. Zhang, "The PIDCalib package," Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul 2016.

[29] D. Martinez Santos, *Study of the very rare decay* $B_s \to \mu^+\mu^-$ *in LHCb*. PhD thesis, Santiago de Compostela, Universidade de Santiago de Compostela, Santiago de Compostela, 2010. Presented on 05 May 2010.

[30] V. Khachatryan *et al.*, "Observation of the rare $B^0_s \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data," *Nature*, vol. 522, pp. 68–72, 2015.

[31] R. Aaij *et al.*, “Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment,” *Phys. Rev. Lett.*, vol. 111, p. 101805, 2013.

[32] C. Patrignani *et al.*, “Review of Particle Physics,” *Chin. Phys.*, vol. C40, no. 10, p. 100001, 2016.

[33] A. Hoecker *et al.*, “TMVA: Toolkit for Multivariate Data Analysis,” *PoS*, vol. ACAT, p. 040, 2007.

[34] M. Pivk and F. R. Le Diberder, “SPlot: A Statistical tool to unfold data distributions,” *Nucl. Instrum. Meth.*, vol. A555, pp. 356–369, 2005.

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

[36] H. Albrecht *et al.*, “Search for hadronic b→u decays,” *Phys. Lett.*, vol. B241, pp. 278–282, 1990.

[37] W. Verkerke and D. P. Kirkby, “The RooFit toolkit for data modeling,” *eConf*, vol. C0303241, p. MOLT007, 2003. [,186(2003)].

[38] R. Aaij *et al.*, “Measurement of the CP violating phase ϕ_s in $\bar{B}_s^0 \to J/\psi f_0(980)$,” *Phys. Lett.*, vol. B707, pp. 497–505, 2012.

[39] R. Aaij *et al.*, “Effective lifetime measurements in the B_s^0ß$K^+K^{,B^0ßK^+}$ and B_s^0ß^+K decays,” *Phys. Lett.*, vol. B736, pp. 446–454, 2014.

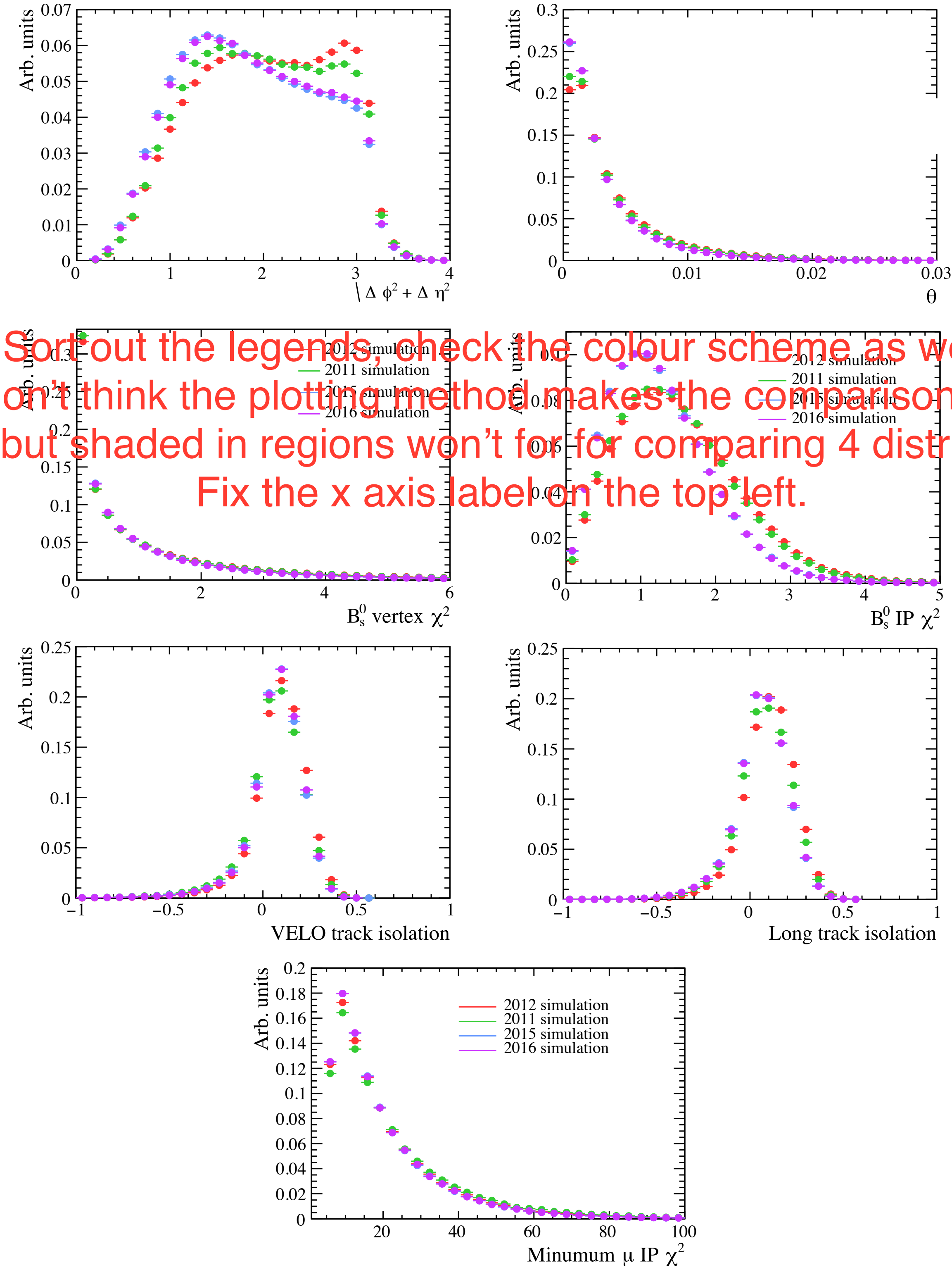
[40] R. Aaij *et al.*, “First study of the CP -violating phase and decay-width difference in $B_s^0 \to \psi(2S)\phi$ decays,” *Phys. Lett.*, vol. B762, pp. 253–262, 2016.

[41] R. Aaij *et al.*, “Measurement of CP violation in $B^0 \to J/\psi K_S^0$ decays,” *Phys. Rev. Lett.*, vol. 115, no. 3, p. 031601, 2015.

Appendix A

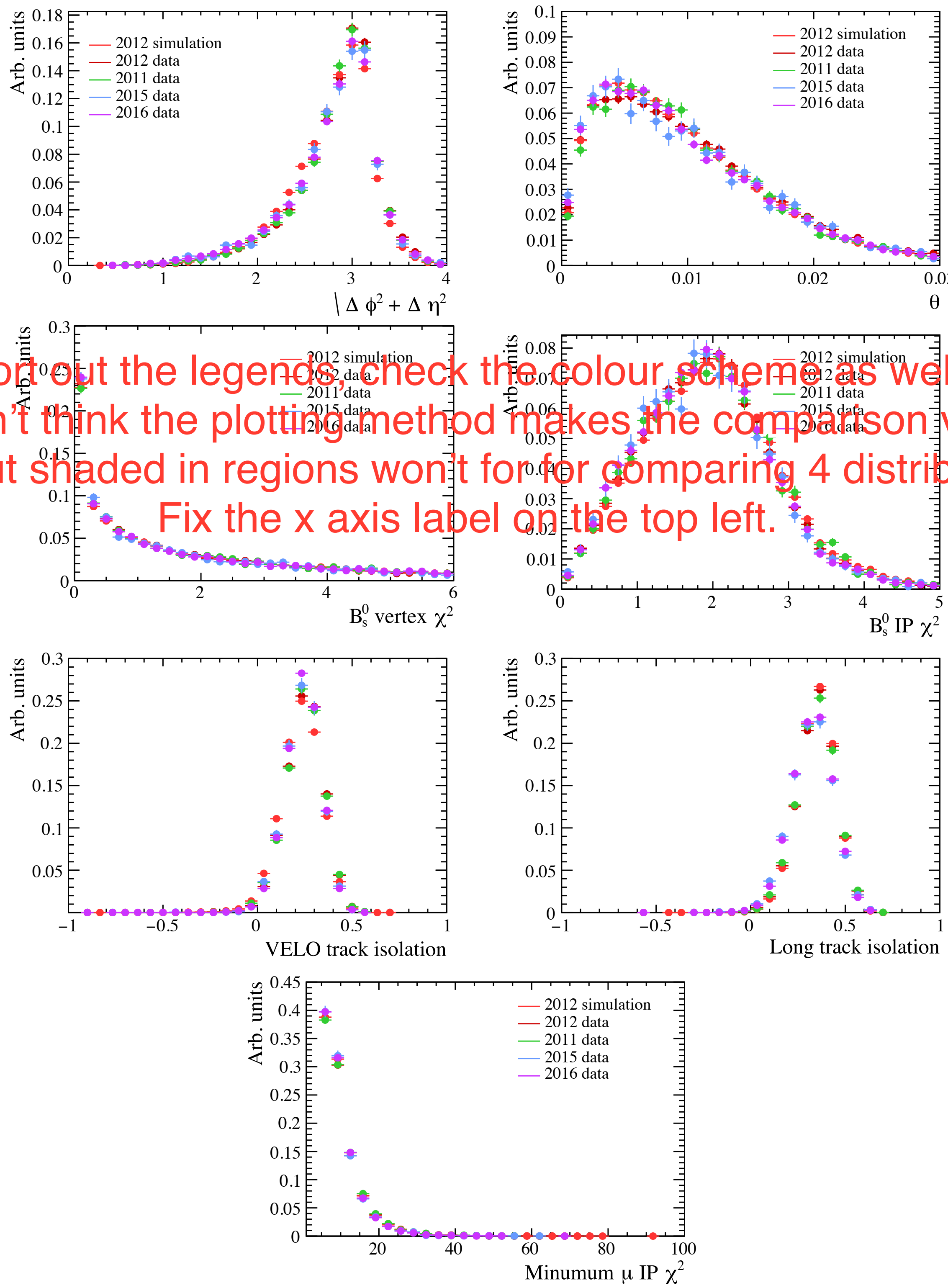
Distributions of input variables for the global BDT

Comparison of the signal and background distributions of the input variables used in the global BDT for 2011, 2012, 2015 and 2016 data taking conditions. Signal distributions are from simulated $B_s^0 \to \mu^+\mu^-$ decays for each year that have passed the selection cuts in Table 4.11. The background distributions are from $b\bar{b} \to \mu^+\mu^- X$ decays in 2011, 2012, 2015 and 2016 data with $m_{\mu\mu} > 5447$ evcc and 2012 simulated $b\bar{b} \to \mu^+\mu^- X$ decays passing the selection cuts in Table 4.11.



Sort out the legends, check the colour scheme as well.
I don't think the plotting method makes the comparison very clear but shaded in regions won't for for comparing 4 distributions.
Fix the x axis label on the top left.

Fig. A.1 Signal distribution for input variables for the global BDT for $B_s^0 \to \mu^+\mu^-$ simulated decays in 2011, 2012, 2015 and 2016.



Sort out the legends, check the colour scheme as well.
I don't think the plotting method makes the comparison very clear but shaded in regions won't for for comparing 4 distributions.
Fix the x axis label on the top left.

(a)

Fig. A.2 Background distribution for input variables from $b\bar{b} \to \mu^+\mu^- X$ decays in 2011, 2012, 2015 and 2016 data with $m_{\mu\mu} > 5447$ MeV$/c^2$ and 2012 simulated $b\bar{b} \to \mu^+\mu^- X$ decays.

Appendix B

Development of multivariate classifiers for the $B_s^0 \to \mu^+\mu^-$ effective lifetime measurement

Details of the input variables used and the development of the classifiers.