

UNIVERSITY OF CAMBRIDGE

This is the very exciting title of my thesis  
something to do with  $B_s^0 \rightarrow \mu^+ \mu^-$  branching  
fraction and effective lifetime.

*Hannah Mary Evans*

*Selwyn College*

Something about the date and that it is being submitted for a PhD.



## **Declaration**

Here I shall say the appropriate lines that are needed to say that this thesis is actually mine.

Hannah Evans  
February 2017



# **Abstract**



# Acknowledgements



# Preface



# Table of contents

<b>Abstract</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>Preface</b>	<b>ix</b>
<b>1 CERN, the LHC and LHCb</b>	<b>1</b>
1.1 The LHC . . . . .	1
1.2 LHCb Experiment . . . . .	4
1.2.1 Tracking . . . . .	5
1.2.2 Particle Identification . . . . .	13
1.2.3 The Trigger . . . . .	21
1.2.4 LHCb Software and Simulation . . . . .	23
1.2.5 LHCb data collected so far . . . . .	25
<b>2 Theory</b>	<b>27</b>
<b>3 Selection of <math>B_s^0 \rightarrow \mu^+ \mu^-</math> and <math>B \rightarrow h^+ h^-</math> decays</b>	<b>29</b>
3.1 Backgrounds . . . . .	30
3.2 Simulated Particle Decays . . . . .	31
3.3 Trigger . . . . .	34
3.4 Cut Based Selection . . . . .	35
3.4.1 Development of the stripping selection . . . . .	36
3.4.2 Optimisation of $B_s^0 \rightarrow \mu^+ \mu^-$ stripping selection . . . . .	39
3.4.3 Stripping selection and offline cuts . . . . .	47
3.5 Particle Identification . . . . .	47
3.6 Multivariate Classifiers . . . . .	49
3.6.1 Boosted Decision Trees . . . . .	51
3.6.2 The BDTS . . . . .	53

3.6.3	Global BDT . . . . .	54
3.6.4	Global BDT cut optimisation . . . . .	59
<b>Bibliography</b>		<b>67</b>
<b>A Distributions of input variables for the global BDT</b>		<b>71</b>

# Chapter 1

## CERN, the LHC and LHCb

The European Organisation for Nuclear Research (CERN) was founded in 1954, it began with 12 member states as a organisation to encourage European collaboration and the study of nuclear physics. The collaborative nature of CERN allowed large-scale expensive experiments over the years that individual member states would not have been able to afford. The Proton Synchrotron (PS) was CERN's flagship accelerator, operational in 1959 it had a circumference of 628 m and accelerated protons to 25 GeV. The PS was the highest energy particle accelerator at that time. Now 62 years since its foundation, CERN has grown to include 22 member states<sup>1</sup> 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 shall discuss the LHC and the LHC beauty experiment, one of the experiments that studies the products of particle collisions at the LHC.

### 1.1 The LHC

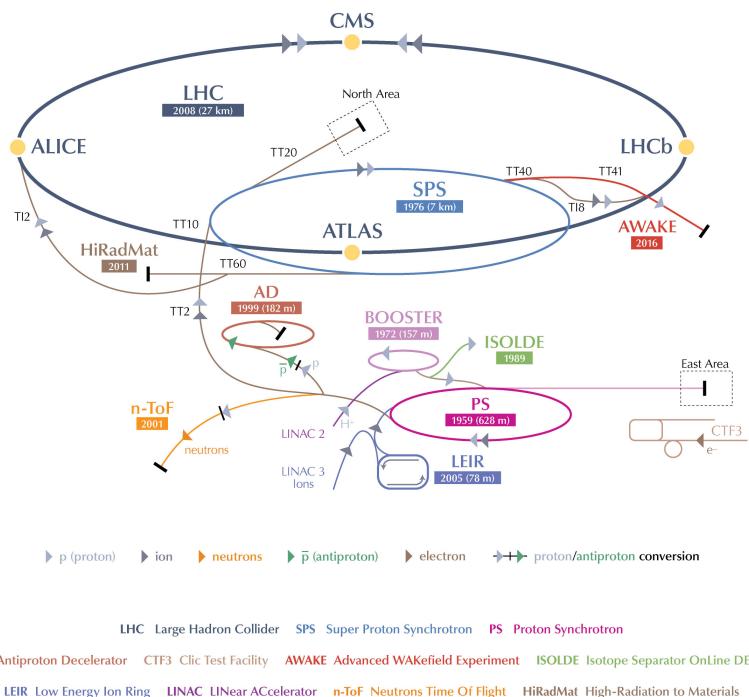
The LHC is a proton synchrotron designed to accelerate and collide two beams of protons with a centre-of-mass energy of 14 TeV. Although operation of the LHC began in 2010 it is yet to reach the design energy. The purpose of the LHC is to provide high energy proton collisions, the products of which are used for precision tests of the Standard Model (SM) and to search for new physics particles that go beyond the scope 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

---

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

of particle collisions. The LHC can also accelerate lead-nuclei up to 2.76 TeV per nucleon, it is only the products from proton collisions that are relevant for the topic of this thesis.

The protons for the LHC come 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. 1.1, 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 for the LHC.



**Fig. 1.1** 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, these are passed to the Proton Synchrotron Booster that accelerates the protons 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.

The protons leave the chain of accelerators with of energy of 450 GeV per proton and in bunches of  $10^{11}$  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.

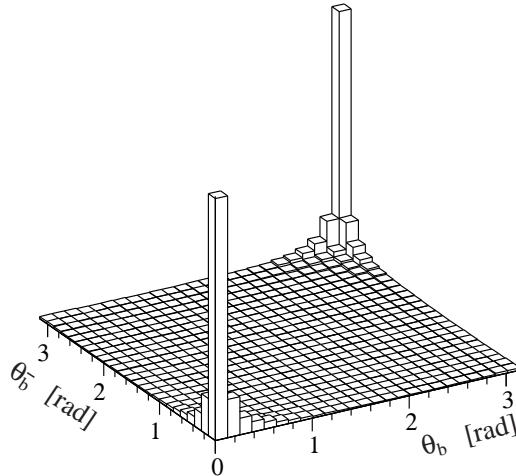
The centre-of-mass energy of a collider is an important measure of its performance as it dictates what particles could be produced in collisions, another important measure of collider performance is the instantaneous luminosity a collider can provide. The instantaneous luminosity,  $\mathcal{L}$ , is a measure of how many collision occur per second, it is given by

$$\mathcal{L} = \frac{N^2 f n_b}{\mathcal{F}}. \quad (1.1)$$

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  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . To reach this luminosity the LHC can have up to 2808 proton bunches per beam with a revolution frequency of 11.245 kHz, therefore the separation between proton bunches can be as short as 25 ns. 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.

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  $\sqrt{s} = 7 \text{ TeV}$  in 2010 and 2011 then this energy was increased to  $\sqrt{s} = 8 \text{ TeV}$  for 2012. After Run 1 came first long shutdown (LS1), during this time 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  $\sqrt{s} = 13 \text{ TeV}$ , 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 particles that are beyond the scope of the SM in proton collisions, these two experiments operate at the full instantaneous luminosity of the LHC. ALICE studies quark-gluon plasma produced in heavy ions 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 is a very forward experiment studying particles that are thrown



**Fig. 1.2** 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].

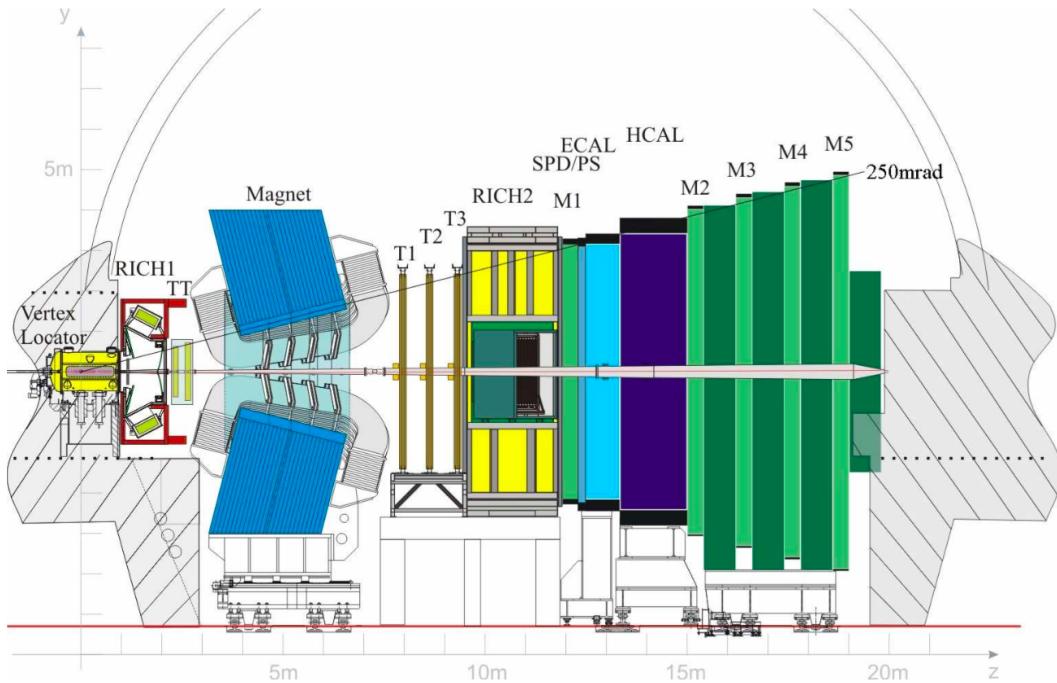
forward in LHC collisions to understand similar processes that occur in cosmic rays. Finally there is the Large Hadron Collider Beauty experiment (LHCb) that will be described in the next section.

## 1.2 LHCb Experiment

The LHCb experiment was built to study CP violation and rare decays of  $b$ -hadrons to search for new physics processes that could be revealed in these decays. At the LHC  $b\bar{b}$  pairs are produced with large transverse momentum, 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 hadronize to form a range of  $b$ -hadrons, including  $B^+$ ,  $B_s^0$  and  $\Lambda_b^0$ , that are studied by LHCb. The large transverse momentum of the  $b\bar{b}$  pairs means that the quarks and subsequent  $b$ -hadrons are boosted along the beam pipe. The angular distribution of  $b\bar{b}$  pairs from  $pp$  collision is shown in Figure 1.2.

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. This angular coverage what chosen to exploit the small angles relative the the beam pipe that  $b\bar{b}$  pairs are produced at. A cross-section of the detector is shown in Figure 1.3, where a right handed coordinate system is used. Protons collide at the interaction point on the left hand side of the diagram, the products of the collisions then travel through the detector leaving information in the different sub detectors along the length of the detector.

The information deposited in the sub detectors is reconstructed to determine what happened in the proton collisions.



**Fig. 1.3** Cross section of the LHCb detector [3].

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. These detectors and their performance are described briefly in the following sections along with the trigger system and software needed to perform physics analyses. Finally the data recorded by the experiment during Run 1 and Run 2 is presented in Section 1.2.5. For a full description of the detector and its performance during Run 1 see references [4, 5].

### 1.2.1 Tracking

The tracking system within the LHCb experiment consists of the VELO, the 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 basic principle that the passage of high energy charged particles through silicon or ionised gas causes 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. Precise particle track and momentum measurements are necessary to obtain the accurate particle mass and decay time measurements that help distinguish between different  $b$ -hadron decaying in the LHCb detector.

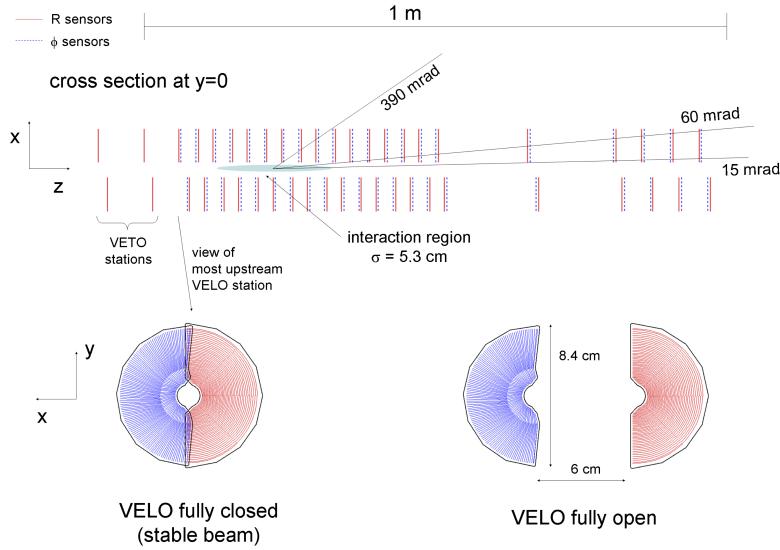
### 1.2.1.1 The VELO

The VELO is a silicon detector surrounding the interaction point. Its main goal is to provide precise information about the interaction vertices and secondary decay vertices of particles produced in proton collisions. Information the VELO provide enables precise measurements of particle lifetimes and impact parameters of particles tracks necessary for physics analyses.

The VELO is made up 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 hole in the centre of the sensors for the beams to pass through. The arrangement of sensors along the  $z$  axis, shown in figure 1.4, it designed such that the sensors cover the full LHCb acceptance and a charged particle within the 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  $\phi$  coordinates as shown in Figure 1.5. The  $r$ ,  $\phi$  coordinates and the  $z$  placement of the sensors are used to reconstruct charged particle trajectories. Cylindrical coordinates were chosen to allow for fast reconstruction for particle trajectories in the VELO.

The momentum resolution achievable for charged tracks 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 achieve this the sensors in the VELO need to be as close as possible to the interaction point. This is achieved by making the VELO out of two retractable halves and including the interaction point within the coverage of the VELO. During data taking, when the VELO is recording particle tracks the sensors are only 8mm from the beam axis. However during the injection phase of the beam the width of the beam is much greater, therefore the 2 halves of the VELO can retract so that they are 3 cm from the nominal beam axis.

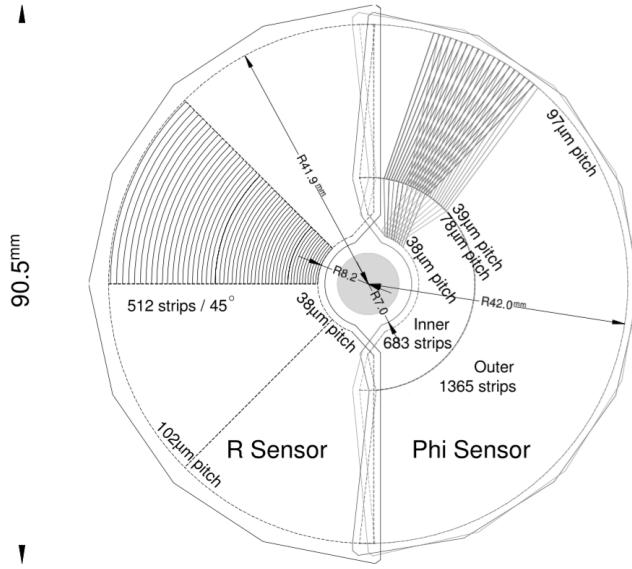


**Fig. 1.4** The VELO layout and position of sensors along the beam axis [3].

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 1.4 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 act as a veto for 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. Events with large numbers of primary vertices are difficult and time consuming to reconstruct and lead to less precise measurements of particle decay properties. Information from the VELO is used to reject events with high numbers of primary vertices to ensure the best use of information from the detector.

The VELO achieves a vertex resolution of 10 - 20  $\mu\text{m}$  transverse to the  $z$  direction and 50 - 100  $\mu\text{m}$  along the  $z$  direction, the resolution of each track depends on the number of tracks in each event as shown in Figure 1.6. 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 1.7 shows the IP resolution for 2012 data, for a track with transverse momentum of 1 GeV/c it has an impact parameter resolution of 35  $\mu\text{m}$ .

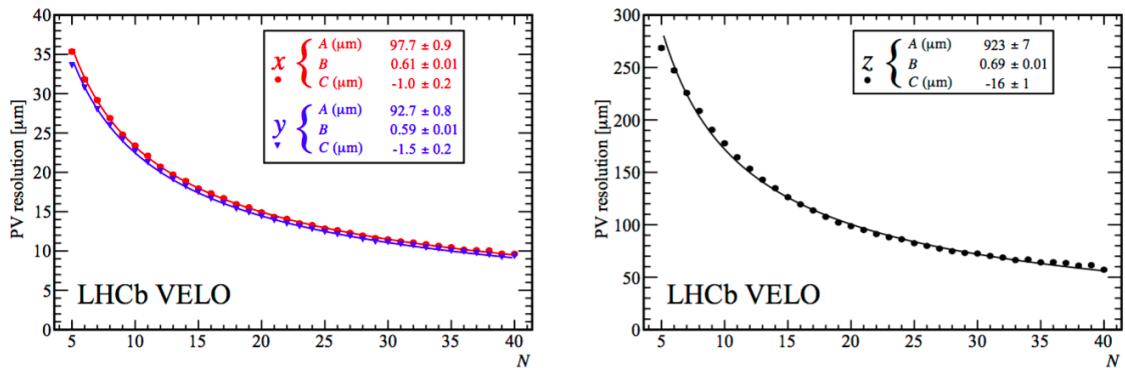


**Fig. 1.5** Diagram of  $r$  and  $\phi$  sensor layouts [3].

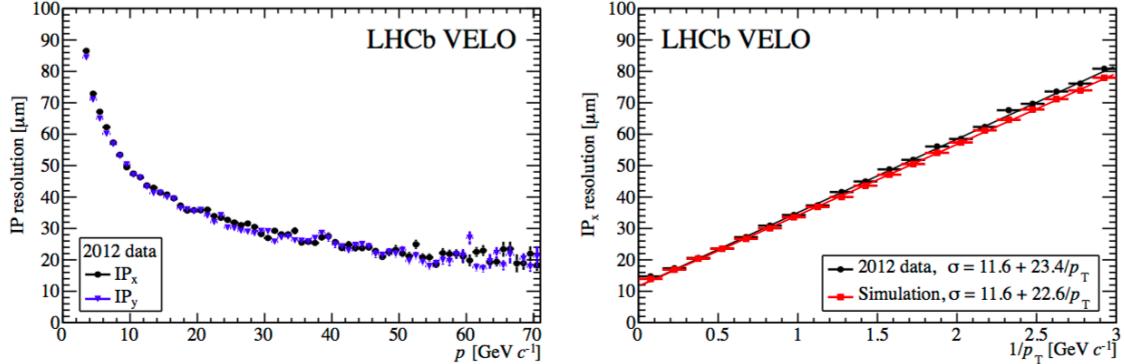
### 1.2.1.2 The Tracking Stations

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 tracking 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



**Fig. 1.6** 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].

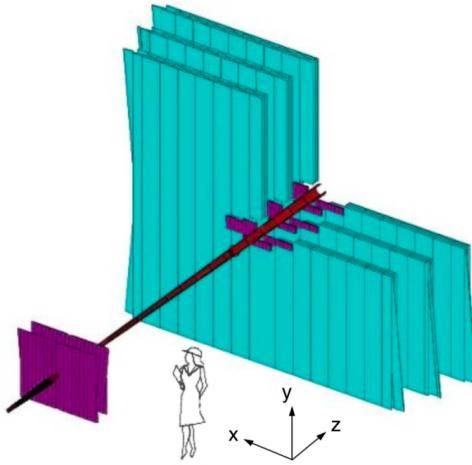


**Fig. 1.7** Velo performance for impact parameter resolution as a function of momentum (left) and inverse transverse momentum (right) for 2012 data [6].

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 for tracks when combined with the other tracking stations. The TT has a resolution of  $50 \mu\text{m}$  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 are 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 600cm by 490 cm, this is illustrated in Figure 1.8. 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.

The IT has very similar in design to the TT, each station is made of 4 layers of silicon trackers and it has with a track resolution of  $50 \mu\text{m}$ . The silicon trackers are arranged in a cross shape around the beam pipe, as shown in Figure 1.8, 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



**Fig. 1.8** Sizes of the TT and T stations [3].

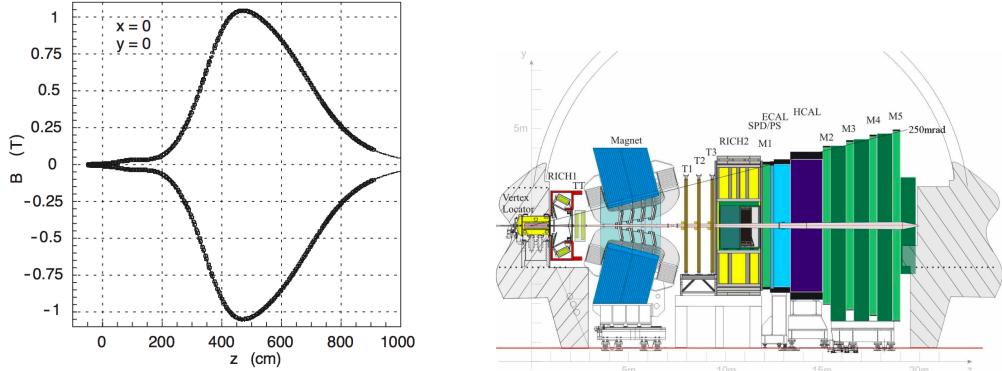
the LHCb full angular acceptance which includes tracks bend by the magnetic field. The tubes have a fast drift time of 50 ns giving a better than  $200 \mu\text{m}$  track resolution.

### 1.2.1.3 The 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 from the radius of curvature of the particle track the particle momentum can be determined.

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 1.9 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 4Tm for track that travels 10m through the detector.

The polarity of the magnetic field is periodically switched so that it 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.



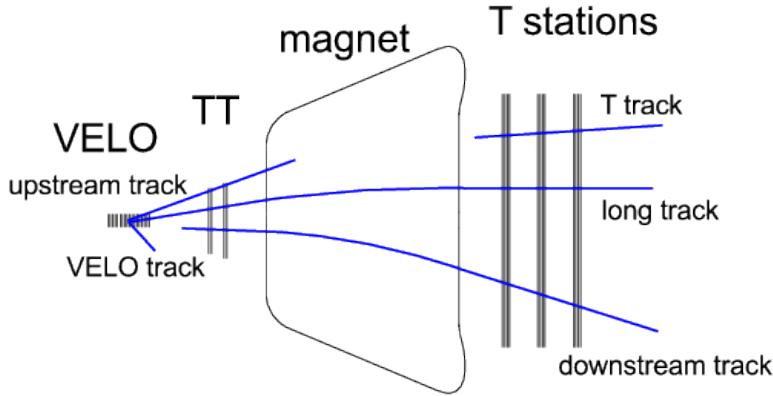
**Fig. 1.9** Magnet field of the dipole magnet along the length of the LHCb detector (left) and the layout to the LHCb detector [3]. The peak strength of the field occurs between the TT and T1-3 station.

#### 1.2.1.4 Track reconstruction and performance

The information left by the passage of charged particles in the VELO, TT and T stations is combined using track reconstruction algorithms to find trajectories of charged particles through 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 as seeds 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 Kalam Fitter which takes into account multiple scattering and energy loss within the detector. For each track the fitter returns the  $\chi^2$  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 1.11.

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 only leave tracks in the VELO. These tracks are useful for reconstructing primary vertices.
- **Upstream tracks** are made by low momentum particles that only leave hits in the tracking stations upstream of the magnet, the VELO and TT. 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



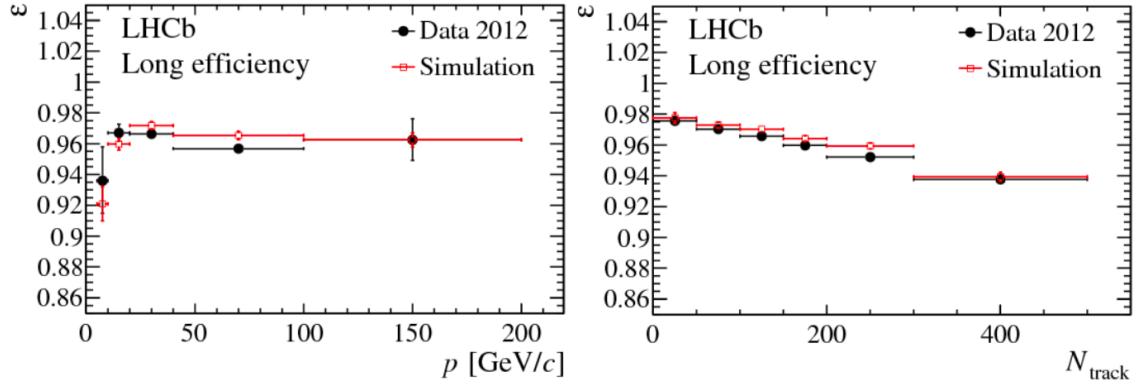
**Fig. 1.10** Different types of tracks that are reconstructed at LHCb [7].

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 decays of long lived neutral particles, that travel out of the VELO before decaying. These particles only leave tracks in 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 different parameters of the events, including the momentum to the particle producing the track and the number of tracks present in the event, as shown in Figure 1.11 for 2012 data. In Run 1 long tracks were correctly reconstructed on average of 96% of the time.

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 the two tracks have many hits in common, when this happens the track with the highest number of total hits is used and the other is discarded. The second type of incorrect tracks are ghost tracks when track segments in different detectors are incorrectly joined



**Fig. 1.11** Long track reconstruction efficiency as a function of momentum (left) and number of track in the event (right) for 2012 data [7].

together. This most often 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.

## 1.2.2 Particle Identification

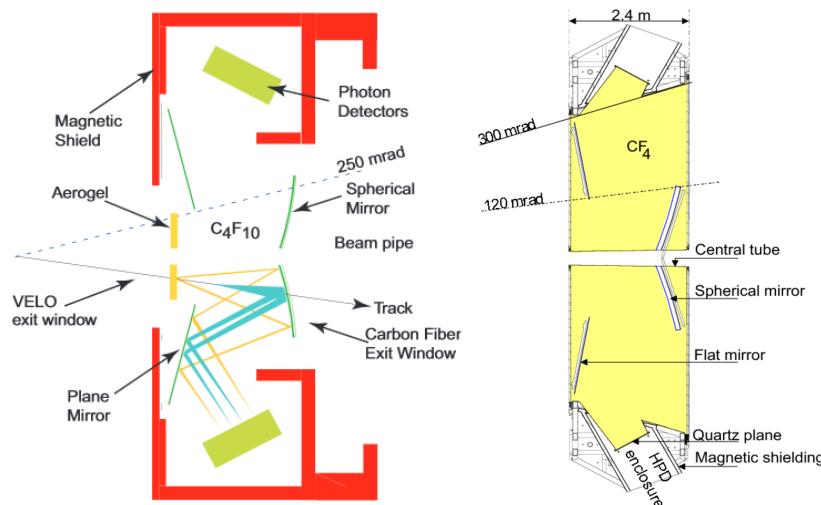
In LHCb the particles identification (PID) detectors consist of two ring imaging Cherenkov (RICH) detectors, electromagnetic and hadronic calorimeters and the muons stations. Together these detectors distinguish between different charged leptons and hadrons and between neutral particles such as photons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

### 1.2.2.1 RICH

RICH detectors are used at LHCb to distinguish between charged hadrons and leptons that have a momentum between 0 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,  $\theta_c$ , 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 produced is called Cherenkov radiation. The RICH detectors measure the angle of light produced as particles pass through them, the angle gives a measurement of the particle's speed which when combined with momentum from the tracking stations gives that particle's mass and therefore its identity.



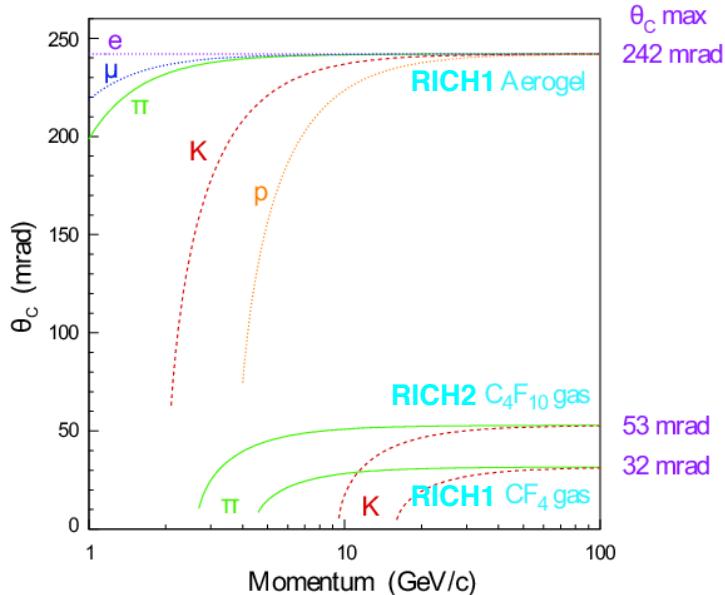
**Fig. 1.12** Diagram of the RICH1 detector (left) and the RICH 2 detector (right) [3]. For Run 2 the aerogel radiator in the RICH 1 detector was removed.

The 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 1-60 GeV/c. The RICH 1, is illustrated in Figure 1.12, it contains two different radiator materials; at the front of the detector is a aerogel sensitive to particles with a momentum up to 10 GeV/c, behind the aerogel is a gas radiator sensitive to particles in the momentum range 10 - 60 GeV/c. The aerogel radiation was removed after Run 1, therefore the RICH 1 is only sensitive to particles in the momentum range 10 - 60 GeV/c in Run 2. As charged particles travel through the RICH 1, the rings of

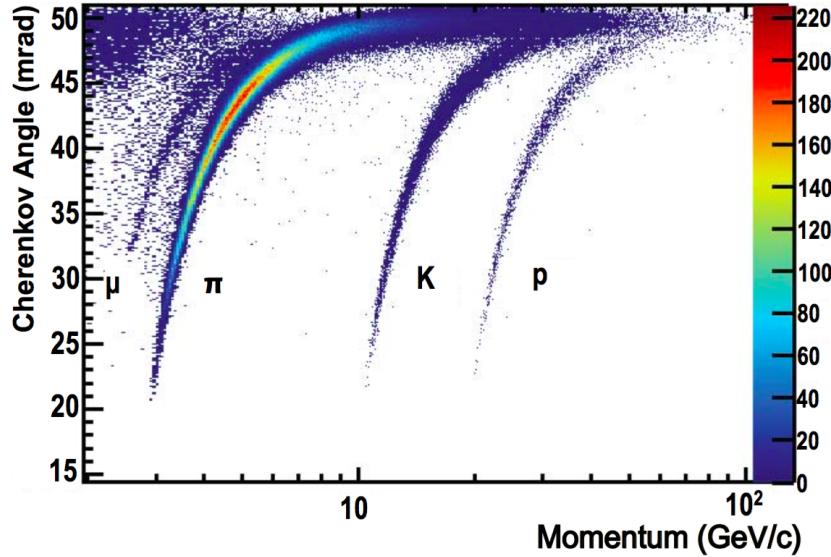
light produced are focused by spherical mirrors onto Hybrid Photon Detectors (HPDs), the radii of the detected rings provides information about how fast the particle was travelling.

The RICH 2 detector is located upstream of the RICH 1, between the last tracking station and before the first muon station. The RICH 2 consists of a gas radiator sensitive to particles with a momentum range 50 - 100 GeV/c and the detection of the light produced is similar to the RICH 1 as illustrated in Figure [? ]. 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, they are shielded from the magnet field using iron sheets ensuring the field is less than 2mT across the HPDs. 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 it's momentum from the tracking stations realise the particles mass and therefore it's identity. Figure 1.14 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 1.13 shows what is expected for the different radiators.



**Fig. 1.13** Expected Cherenkov angles produced by different particles travelling through the different radiators in the RICH detectors [3].



**Fig. 1.14** Cherenkov angles for isolated tracks as a function of momentum in the RICH1 detector for 2011 data [8].

### 1.2.2.2 Calorimeters

The calorimeter system consists of the Scintillating Pad Detector (SDP), Pre-Shower (PS), electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The main purposes of the calorimeters are to identify electrons, photons and hadrons with high transverse momentum to be used in the first level of the trigger the L0, and also 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 it is 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.

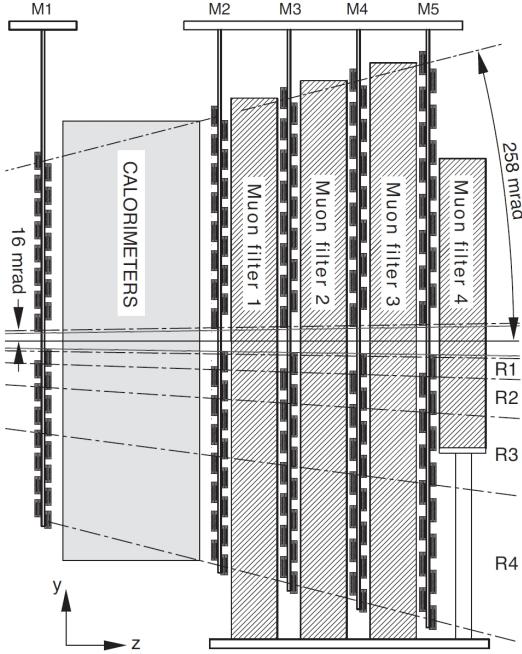
The SPD, PS and ECAL identify electrons, positrons and photons. The SPD is a layer of scintillating material at the start of the calorimeter system, it separates electron and photon showers created later in the calorimeter because only charged particles will produce light in the SPD. Next in the calorimeter 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  $\delta E/E = 69\%/\sqrt{(E)} \oplus 9\%$ .

### 1.2.2.3 Muon Stations

The muons stations are designed to identify highly penetrating muons, for use in the trigger and offline analyses. Muons are produced in made  $b$ -hadron decays, good muon identification is necessary trigger events containing muons and to distinguish topologically similar decays such as  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ,  $B^0 \rightarrow K^+ \pi^-$ ,  $B_s^0 \rightarrow K^+ K^-$  and  $B_s^0 \rightarrow K^+ \pi^-$  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 1.15 that track and identify highly penetrating 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 MWPC and between each station is 80cm of lead absorber ensuring only highly penetrating muons pass through the muon detector. A muon must have a momentum of at least 3 GeV/c to pass through the calorimeters and M2 and M3, 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



**Fig. 1.15** Layout of the muon stations [3].

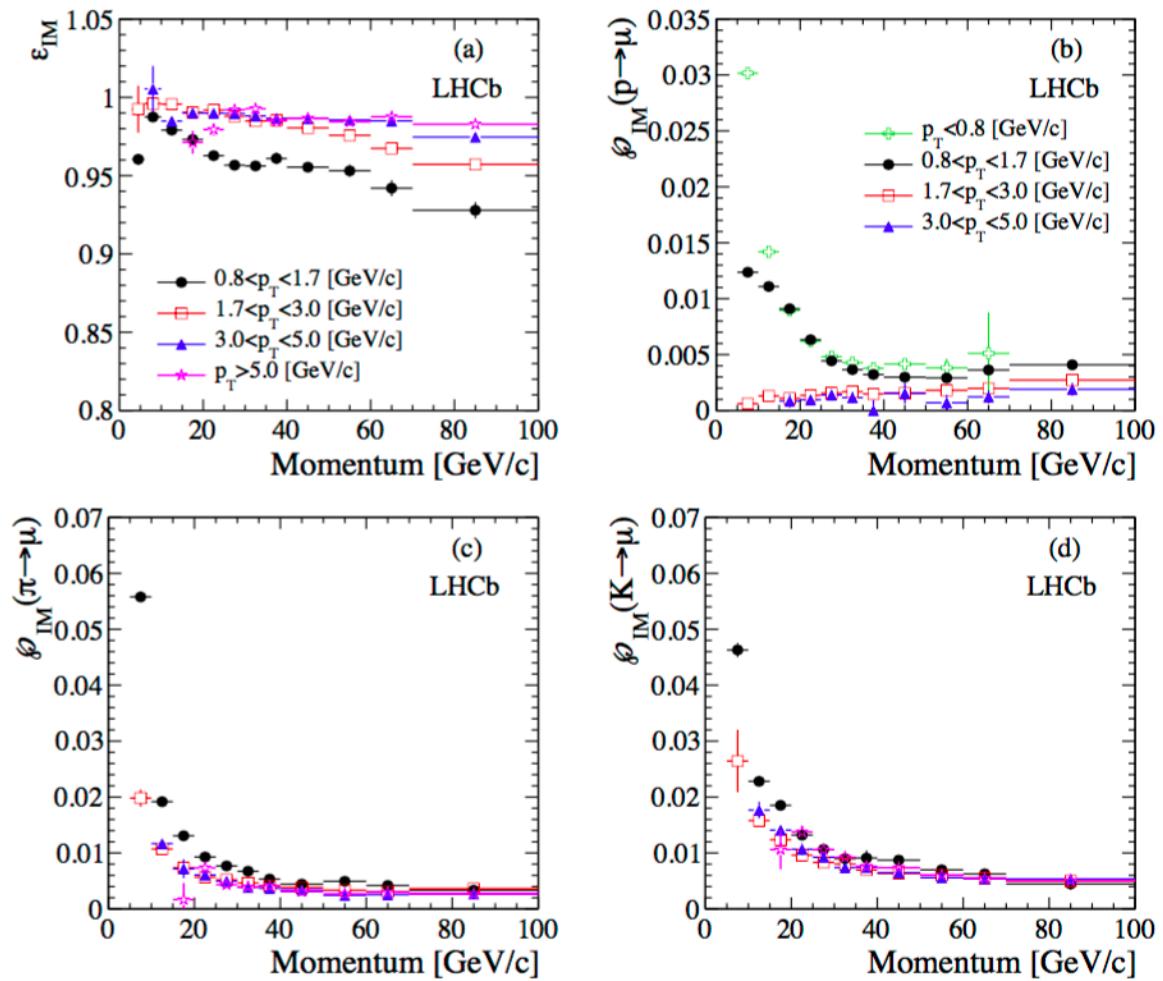
trigger. M1 is located before the calorimeters to improve the transverse momentum measurement of the muons. The last two stations have lower spatial resolution and are designed to identify highly penetrating muons. After the muon stations there is an iron wall to stop any particles from travelling downstream 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 station lie outside the magnetic field which allows for fast reconstruction of the tracks and a muons.

#### 1.2.2.4 PID information 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 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 1.16 shows the efficiency for the isMuon selection at selecting muons and hadron mis-identification probabilities. The mis-ID rate is higher for lower momentum particles, which is expected given there are less hits in the muons detectors. The main contribution is the misidentifying particles as muons comes from the kaons and pions that decay in flight, the muons from these decays are then detected in the muon stations.



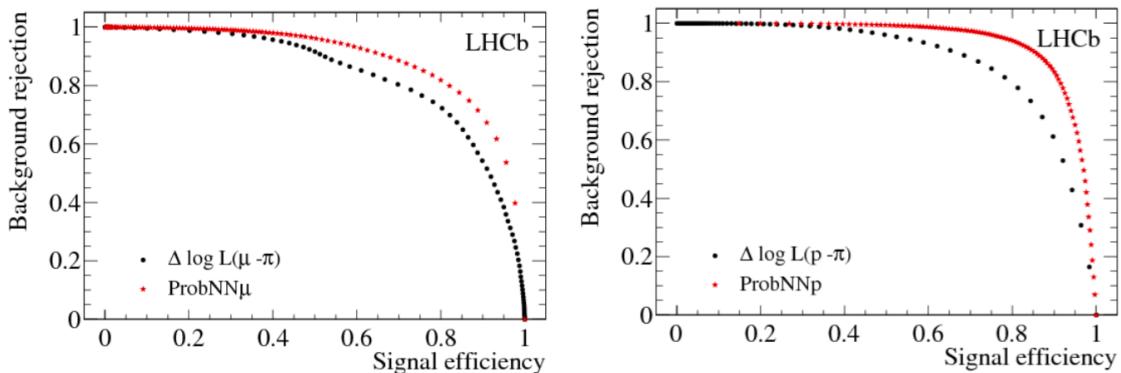
**Fig. 1.16** Muon efficiency (top left) and misidentification probabilities for protons (top right), pions (bottom left) and kaons (bottom right) for isMuon criteria [9].

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 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), it gives a measure of how likely each particle hypothesis is compared to 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  $DLL_{K\pi}$  etc.

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 probabiltiy of a particle being a muon is given by  $\text{ProbNN}\mu$  and the probabiltiy of a particle being a pion is given by  $\text{ProbNN}\pi$  etc.

Figure 1.17 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. 1.17** Muon (left) and proton (right) signal efficiency vs background rejection for DLL and ProbNN PID variables [3].

### 1.2.3 The 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 that selects potentially interesting events to be saved for later physics analysis. 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 known as the level-zero (L0) trigger and reduces the 40 MHz event rate to 1 MHz at which the full detector can be read out. The software trigger is known 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 accepted by a trigger line at both the L0 and HLT are available for use in physics analyses.

#### 1.2.3.1 The L0 trigger

The L0 trigger runs synchronously to the LHC bunch crossing. Its purpose is to reduce the event 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 identify the number of collisions in an event and are predominantly used for luminosity measurements [10].

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 efficient 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.

### 1.2.3.2 The 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.

The HLT1 is the first level of the HLT. It runs on the output of the L0 checking the decisions made by the L0 trigger lines and reducing the event rate. Time constraints in the HLT1 do not allow for full event reconstruction using all LHCb sub-detectors, instead the HLT1 runs reconstruction and selection algorithms on event information from the VELO and tracking stations. These 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, HLT2, runs on the output of the HLT1 which provides an event rate that is low enough to allow event reconstruction that includes all detector subsystems. The trigger lines in the HLT2 are designed to select decays relevant to specific physics analyses or particle decay topologies, this is made possible by detailed information from the 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 HLT2 lines in 2011, 200 in 2012 and 450 in 2015. Furthermore, 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 [11]. The majority of the changes to the HLT for Run 2 are not relevant for the analysis discussed in this thesis, the overall impact is that a more

detailed reconstruction is used in the decisions of the Run 2 HLT lines compared to Run 1.

### 1.2.3.3 Trigger Decisions

The trigger lines in the L0 and HLT return three different types of decisions that are used to classify events. The choice of which type of trigger decision to use depends on the particular physics analysis and the signal decay of interest, the decisions can either be used line by line or as global decisions taking all lines together. The different decisions are:

- **TOS**, ‘triggered on signal’, tracks and hits that make up signal candidate of a physics analysis are sufficient for the event to pass the trigger line.
- **TIS**, ‘triggered independant of signal’, if the tracks and hits associated with the signal candidate of a physics analysis are removed from the event, other tracks and hits would still cause the event to pass the trigger line.
- **Dec**, refers to whether the event was accepted by the trigger line.

### 1.2.4 LHCb 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 [12] is a C++ framework that is the basis for the software applications needed to process the data at LHCb [13]. 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 [14].

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 the analysts

therefore a ‘stripping’ procedure is used to break up the data into a manageable size for physics analyses. 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 as possible but reduce the number background events. Only events that pass a stripping line selection are available 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 separately 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 [15] files from the output of their stripping lines, these files display the data in histogram 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 [16, 17], this package uses PYTHIA [18, 19] to model proton-proton collisions and the production of particles, then the EVTGEN [20] application to calculate the decays of these particles. Final state radiation is modelled using PHOTOS [21]. Both PYTHIA and EVTGEN have been tuned for the production and decay of particles within the LHCb 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 detector response to the simulated events is processed by the BOOLE application which uses GEANT4 [22, 23] to model the detector. The output is a digitised response of the detector which 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 [24, 25], 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 [26] 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

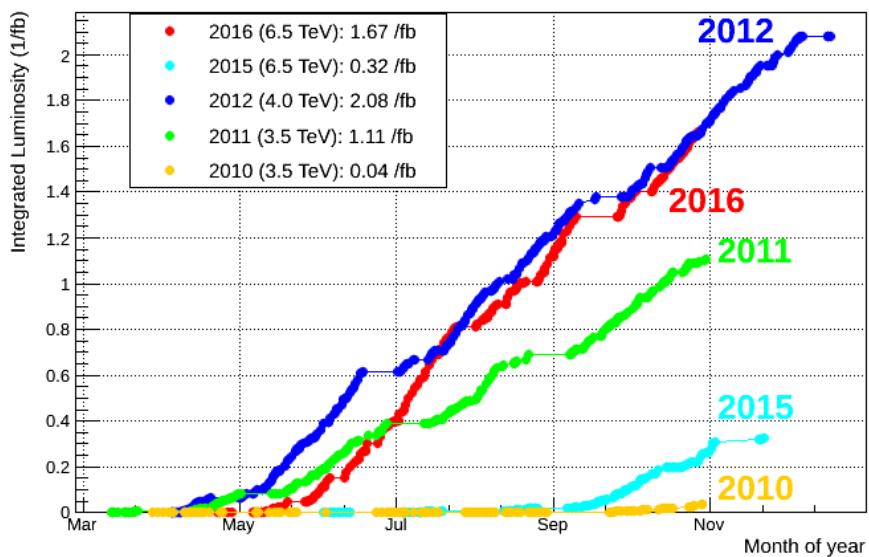
the large amounts of data produced by LHCb without having to store the data where the analyst is.

### 1.2.5 LHCb data collected so far

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 at a centre-of-mass energy of 7 TeV, this was increased to 8 TeV in 2012. The luminosity recorded by LHCb in each was;  $0.04 \text{ fb}^{-1}$  in 2010,  $1.10 \text{ fb}^{-1}$  in 2011 and  $2.08 \text{ fb}^{-1}$  in 2012. After Run 1 the LHC and experiment 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  $0.32 \text{ fb}^{-1}$  in 2015 and  $1.67 \text{ fb}^{-1}$  in 2016 both at a centre-of-mass energy of 13 TeV. Figure 1.18 shows the integrated luminosity collected by LHCb in each year of data taking. The recorded luminosity of Run 2 is currently less than 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  $5 \text{ fb}^{-1}$  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  $23 \text{ fb}^{-1}$  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  $1.1 \text{ fb}^{-1}$ .



**Fig. 1.18** Integrated luminosity collected by the LHCb experiment in each year of data taking. Source: LHCb.

## **Chapter 2**

### **Theory**



# Chapter 3

## Selection of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B \rightarrow h^+ h^-$ decays

The analysis described in Chapter X for the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime requires  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B \rightarrow h^+ h^-$  decays to be identified in the data sets recorded by the LHCb experiment.  $B \rightarrow h^+ h^-$  decays, where  $h = K, \pi$ , are used To understand different aspects of the selection and analysis of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays because they have large branching fractions, a similar topology to  $B_s^0 \rightarrow \mu^+ \mu^-$  decays and are well understood from previous LHCb analyses. Although  $B_s^0 \rightarrow \mu^+ \mu^-$  decays leave a clear 2 muon signature in the detector, the selection of these decays is challenging because it is a very rare process and there are many other processes (Sect. 3.1) that can mimic a  $B_s^0 \rightarrow \mu^+ \mu^-$  decay in the detector. This Chapter describes the selection of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  and  $B \rightarrow h^+ h^-$  decays for the measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime (Chapter ??), much of the selection is shared with the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Branching Fraction analysis(Chapter ??). The selection occurs in several stages and the development of the selection relies on simulated events (Sect. 3.2). The first step to select  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B \rightarrow h^+ h^-$  decays is choosing what requirements to place on the trigger (Sect. 3.3) which is followed by a cut based selection to remove obvious background events (Sect. 3.4. Then particle identification variables (Sect. 3.5) to reduced background events from mis-identified semi-leptonic and  $B \rightarrow h^+ h^-$  decays and finally multivariate classifiers (Sect. 3.6.1) are used as the last step in the selection to reduced the backgrounds to a low enough level so that the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime can be measured from the data.

The LHCb collaboration has published a number of papers studying the  $B_s^0 \rightarrow \mu^+ \mu^-$  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

3.4.2, 3.5 and 3.6.4 were completed for this thesis as well as all Figures and quoted efficiencies.

### 3.1 Backgrounds

The reconstruction process (Sect. ??) produces numerous  $B_s^0 \rightarrow \mu^+\mu^-$  candidates from pairs of muons created during  $pp$  collisions. Some candidates will have come from real  $B_s^0 \rightarrow \mu^+\mu^-$  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  $B_s^0 \rightarrow \mu^+\mu^-$  decay. The selection aims to separate real  $B_s^0 \rightarrow \mu^+\mu^-$  decays from the background to produce a set of  $B_s^0 \rightarrow \mu^+\mu^-$  candidates with a high signal purity from which the  $B_s^0$  effective lifetime can be measured. The main background sources that mimic  $B_s^0 \rightarrow \mu^+\mu^-$  decays are:

- Elastic collisions of protons that produce a pair of muons via the exchange of a photon,  $pp \rightarrow p\mu^+\mu^-p$ . The protons travel down the beam pipe and are undetected leaving the muons to be reconstructed as  $B_s^0 \rightarrow \mu^+\mu^-$ . 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  $B_s^0$  that decays instantaneously. This type of background is prompt combinatorial background.
- $B^0 \rightarrow \mu^+\mu^-\gamma$  decays where the photon is not reconstructed. The presence of the photon in the decay means that  $B_s^0 \rightarrow \mu^+\mu^-\gamma$  decays are not helicity suppressed and could therefore be a sizable background, however the photon gains a large transverse momentum resulting in the reconstructed  $B_s^0$  mass being much lower than expected.
- Random combinations of muons produced by separate semi-leptonic decays. The  $B_s^0 \rightarrow \mu^+\mu^-$  candidates formed in this way are long lived combinatorial background because the reconstructed  $B_s^0$  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  $B_s^0$  candidate is lower than expected due to the missing particle information. The semi-leptonic decays that contribute to  $B_s^0 \rightarrow \mu^+\mu^-$  backgrounds in this way are  $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ ,  $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ ,  $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ ,  $B^0 \rightarrow \pi^0\mu^+\mu^-$  and  $B_c^+ \rightarrow J/\psi\mu^+\nu_\mu$  where  $J/\psi \rightarrow \mu^+\mu^-$ .

- $B \rightarrow h^+h^-$  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  $B_s^0$  candidate mass is lower than expected.
- $B^0 \rightarrow \mu^+\mu^-$  decays that are identical to  $B_s^0 \rightarrow \mu^+\mu^-$  decays apart from the difference in the  $B$  meson masses. The decay is irrelevant for the measurement of the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime and is therefore a background for this measurement.

The separation of  $B_s^0 \rightarrow \mu^+\mu^-$  decays from the backgrounds is challenging because  $B_s^0 \rightarrow \mu^+\mu^-$  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  $B_s^0 \rightarrow \mu^+\mu^-$  and the backgrounds, however background from mis-identified semi-leptonic and  $B \rightarrow h^+h^-$  decays and long lived combinatorial background are difficult to remove. The dimuon invariant mass distribution from the last published  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  Branching Fraction analysis by LHCb is shown in Figure 3.1, components for background from mis-identified semi-leptonic and  $B \rightarrow h^+h^-$  decays are present below the  $B_s^0$  mass and the long lived combinatorial background has an almost flat distribution across the entire mass range.

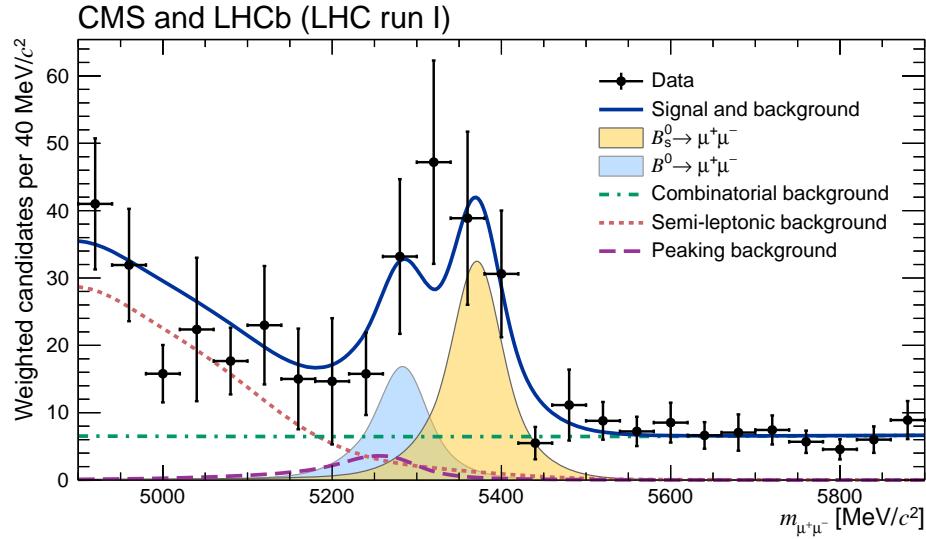
## 3.2 Simulated Particle Decays

Simulated particle decays, as described in Section ??, are used to develop the selection and analysis of  $B_s^0 \rightarrow \mu^+\mu^-$  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. The simulated decays used for studies documented in this thesis are listed in Table 3.1 along with the data taking conditions and simulation versions used to generate 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 of 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. The simulated decays in Table 3.1 listed under the studies they are used in.

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

**Table 3.1** Simulated decays used for developing the selection and the analysis of  $B_s^0 \rightarrow \mu^+ \mu^-$  listed according to the studies the decays are used in. Cuts are applied to  $b\bar{b} \rightarrow \mu^+ \mu^- X$  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.



**Fig. 3.1** Weighted dimuon invariant mass spectrum from combined analysis of CMS and LHCb Run 1 data for  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  Branching Fraction measurements [27]. Backgrounds included in the mass fit are mis-identified semi-leptonic decays in red, mis-identified  $B \rightarrow h^+h^-$  decays in purple and long lived combinatorial background in green.

Simulated  $b\bar{b} \rightarrow \mu^+\mu^-X$  decays are used to understand the combinatorial background of  $B_s^0 \rightarrow \mu^+\mu^-$  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  $b\bar{b} \rightarrow \mu^+\mu^-X$  decays to reduce the size of the samples that are saved and to speed production. The cuts, listed in Table 3.1, 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 area where the simulation falls short of reality. The distributions of 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 [] 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 ??.

### 3.3 Trigger

The trigger (Sect. 1.2.3) 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.  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  candidates are reconstructed from events that have passed the trigger. For each candidate it is useful to know whether it 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

$B_s^0 \rightarrow \mu^+\mu^-$  decays are very rare decays and therefore trigger requirements used to select these decays are chosen to keep a high efficiency is kept at this step of the selection. The trigger lines L0Global, Hlt1Phys and Hlt2Phys are used and candidates are required to be TOS or TIS at each level of the trigger. These trigger lines combine the decisions of many individual lines used in the trigger which allows a high efficiency to be achieved for selecting  $B_s^0 \rightarrow \mu^+\mu^-$  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.

Slightly different trigger decisions are used to select  $B \rightarrow h^+h^-$  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 needs to be similar to the  $B_s^0 \rightarrow \mu^+\mu^-$  triggers, this is achieved by requiring  $B \rightarrow h^+h^-$  decays to be TIS at each level of the trigger.

The requirements imposed on the trigger to select  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  decays is shown in Table 3.2.

Trigger Line	Trigger decision
<i>Select <math>B_s^0 \rightarrow \mu^+\mu^-</math> decays</i>	
L0Global	TIS or TOS
Hlt1Phys	TIS or TOS
Hlt2Phys	TIS or TOS
<i>Select <math>B \rightarrow h^+h^-</math> decays</i>	
L0Global	TIS
Hlt1Phys	TIS
Hlt2Phys	TIS

**Table 3.2** Trigger lines used to select  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  decays.

## 3.4 Cut Based Selection

The  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  candidates that pass 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  $B_s^0 \rightarrow \mu^+\mu^-$  decays and the backgrounds that mimic them. The selection of  $B \rightarrow h^+h^-$  decays is kept as close as possible to that of  $B_s^0 \rightarrow \mu^+\mu^-$  decays. The cuts based selection is compared of two parts; the stripping selection and the offline selection.

The stripping selection, as described in Section 1.2.4, 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 for  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  decays is described in Sections 3.4.1 and 3.4.2.

The offline selection cuts are applied to the output of the stripping selection. Overall stripping selection imposes loose selection requirements onto  $B_s^0 \rightarrow \mu^+\mu^-$  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 full set of cuts applied in the stripping and offline selection to select  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  decays from Run 1 and Run 2 data are presented in Section 3.4.3.

### 3.4.1 Development of the stripping selection

The stripping selection used to select  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  decays for the  $B_s^0 \rightarrow \mu^+\mu^-$  effective lifetime measurement uses the same stripping lines as the selection of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays for the Branching Fraction measurement. The selection were designed at the start of Run 1 by studying the efficiencies of different selection cuts from simulated events [28]. 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 investigate where improvements can be made to the efficiency of the stripping selection used to select  $B_s^0 \rightarrow \mu^+\mu^-$  events.

In addition to  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  decays the measurement of the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  Branching Fractions requires  $B^+ \rightarrow J/\psi K^+$  decays.  $B^0 \rightarrow K^+\pi^-$  and  $B^+ \rightarrow J/\psi K^+$  decays are used to normalise the number of observed  $B_s^0 \rightarrow \mu^+\mu^-$  decays to the number created in  $pp$  collisions.

There are three stripping lines that select  $B_{(s)}^0 \rightarrow \mu^+\mu^-$ ,  $B^+ \rightarrow J/\psi K^+$  and  $B \rightarrow h^+h^-$  candidates, the selection of the normalisation channels is kept as similar as possible to the signal selection to avoid introducing systematic uncertainties in the normalisation procedure of the branching fraction. However, the selection of  $B^+ \rightarrow J/\psi K^+$  decays must diverge from  $B_s^0 \rightarrow \mu^+\mu^-$  due to additional particles in the final state of the decay. Any changes made to the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  stripping selection to improve the selection efficiency must be included in the selection to the normalisation channels to keep the systematic uncertainties under control, therefore all three stripping lines must be studied together. The stripping selection cuts applied for the Run 1 Branching Fraction analysis [27, 29] to select  $B_{(s)}^0 \rightarrow \mu^+\mu^-$ ,  $B \rightarrow h^+h^-$  and  $B^+ \rightarrow J/\psi K^+$  decays are listed in Table 3.3.

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  $J/\psi$ ) are combined to provide its reconstructed mass. Cuts on 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  $B_s^0 \rightarrow \mu^+\mu^-$  selection to allow for the study of semi-leptonic backgrounds that have a mass less than the  $B_s^0$  mass when mis-identified as  $B_s^0 \rightarrow \mu^+\mu^-$  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

Particle	$B_s^0 \rightarrow \mu^+ \mu^-$	$B \rightarrow h^+ h^-$	$B^+ \rightarrow J/\psi K^+$
$B_s^0$ or $B^+$	$ M - M_{PDG}  < 1200 \text{ MeV}/c^2$ DIRA > 0 FD $\chi^2 > 225$ IP $\chi^2 < 25$ Vertex $\chi^2/\text{ndof} < 9$ DOCA < 0.3 mm $\tau < 13.248 \text{ ps}$ $p_T > 500 \text{ MeV}/c$	$ M - M_{PDG}  < 500 \text{ MeV}/c^2$ DIRA > 0 FD $\chi^2 > 225$ IP $\chi^2 < 25$ Vertex $\chi^2/\text{ndof} < 9$ DOCA < 0.3 mm $\tau < 13.248 \text{ ps}$ $p_T > 500 \text{ MeV}/c$	$ M - M_{PDG}  < 500 \text{ MeV}/c^2$ Vertex $\chi^2/\text{ndof} < 45$ IP $\chi^2 < 25$ DIRA > 0 FD $\chi^2 > 225$ Vertex $\chi^2/\text{ndof} < 9$ DOCA < 0.3 mm $\tau < 13.248 \text{ ps}$ $p_T > 500 \text{ MeV}/c$
$J/\psi$			
Daughter $\mu$ or $h$	Track $\chi^2/\text{ndof} < 3$ isMuon = True Minimum IP $\chi^2 > 25$ $p_T > 0.25 \text{ GeV}/c$ ghost probability < 0.3	Track $\chi^2/\text{ndof} < 3$ Track $\chi^2/\text{ndof} < 3$ isMuon = True Minimum IP $\chi^2 > 25$ $0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$ 0.25 GeV/ $c$ < $p_T$	Track $\chi^2/\text{ndof} < 3$ isMuon = True Minimum IP $\chi^2 > 25$ $0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$ 0.25 GeV/ $c$ < $p_T$
$K^+$			

**Table 3.3** Selection requirements applied during the stripping selection for Run 1 data used in the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Branching Fraction analysis [27, 29] to select  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ,  $B \rightarrow h^+ h^-$  and  $B^+ \rightarrow J/\psi K^+$  decays.  $M_{PDG}$  corresponds to the Particle Data Group [30] mass of each particle.

and decay vertices<sup>1</sup> 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)  $\chi^2$  - 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  $\chi^2$  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  $B_s^0 \rightarrow \mu^+ \mu^-$ , this distance would ideally be zero because the muons originate from the same vertex;
- decay time,  $\tau$  - 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 1.2.2 that returns True for muons and False for other particles;
- 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;

---

<sup>1</sup>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.

- 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)  $\chi^2$  - this is the change in the fit for a primary vertex (PV) caused by removing one track in the fit. In a  $B_s^0 \rightarrow \mu^+\mu^-$  decay, the  $B_s^0$  is produced at the PV therefore it should have a small IP  $\chi^2$  value whereas the muons will be displaced from the PV because of the relatively long lifetime of the  $B_s^0$  and therefore will have a large IP  $\chi^2$ ;
- minimum impact parameter (IP)  $\chi^2$  - this is the IP  $\chi^2$  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  $B \rightarrow h^+h^-$  decays compared to  $B_s^0 \rightarrow \mu^+\mu^-$  because  $B \rightarrow h^+h^-$  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  $B \rightarrow h^+h^-$  decays in the stripping are the later applied to  $B_s^0 \rightarrow \mu^+\mu^-$  candidates in the offline selection.

### 3.4.2 Optimisation of $B_s^0 \rightarrow \mu^+\mu^-$ stripping selection

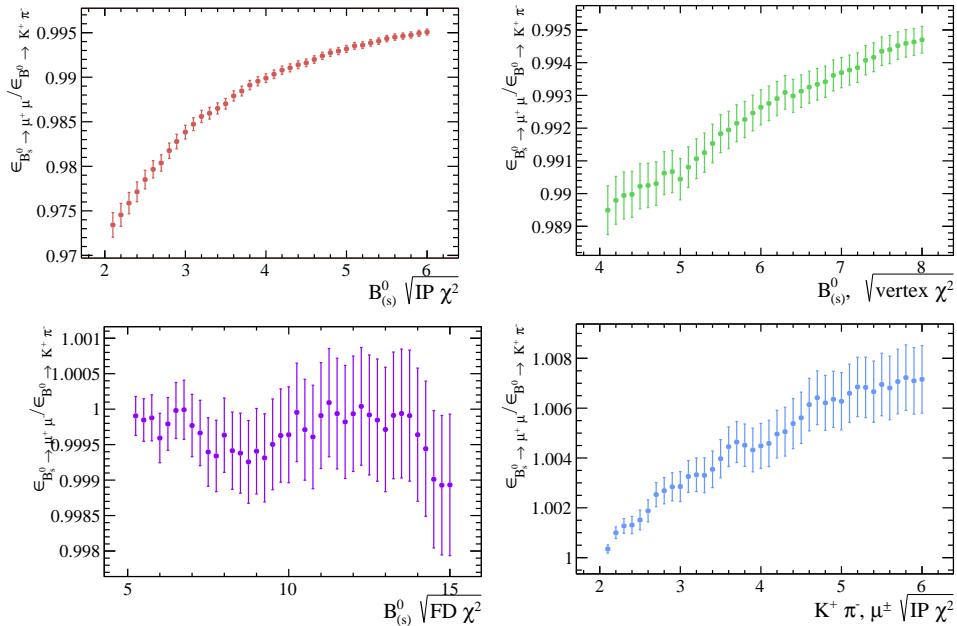
The efficiency of the cuts used in the stripping lines to selecting  $B_{(s)}^0 \rightarrow \mu^+\mu^-$ ,  $B \rightarrow h^+h^-$  and  $B^+ \rightarrow J/\psi K^+$  decays are shown in Table 3.4, only cuts that are in common with the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  stripping lines are listed. The efficiencies are evaluated using 2012 sim06 simulated events that have the minimum track  $p_T$ , track  $\chi^2$  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 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 /it 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 3.3 and 3.2 which show the ratio of selection efficiencies for  $B_s^0 \rightarrow \mu^+\mu^-$  decays to  $B^+ \rightarrow J/\psi K^+$  and  $B^0 \rightarrow K^+\pi^-$  decays for a range of selection cuts. With the

Requirement	Efficiency				$B^+ \rightarrow J/\psi K^+$
	$B_s^0 \rightarrow \mu^+ \mu^-$	$B^0 \rightarrow \mu^+ \mu^-$	$B \rightarrow h^+ h^-$	$B \rightarrow h^+ h^-$	
$B  \mathcal{M} - M_{PDG} $	(100.00 $\pm$ 0.00)%	(100.00 $\pm$ 0.00)%	(98.25 $\pm$ 0.02)%	(99.73 $\pm$ 0.02)%	(99.73 $\pm$ 0.02)%
$B_{(s)}^0$ or $J/\psi$ DIRA	(99.41 $\pm$ 0.01)%	(99.47 $\pm$ 0.01)%	(99.47 $\pm$ 0.01)%	(95.83 $\pm$ 0.08)%	(95.83 $\pm$ 0.08)%
$B_{(s)}^0$ or $J/\psi$ FD $\chi^2$	(83.74 $\pm$ 0.06)%	(83.96 $\pm$ 0.06)%	(83.83 $\pm$ 0.06)%	(82.90 $\pm$ 0.15)%	(82.90 $\pm$ 0.15)%
$B_{(s)}^0$ or $J/\psi$ IP $\chi^2$	(96.78 $\pm$ 0.03)%	(96.93 $\pm$ 0.03)%	(97.44 $\pm$ 0.03)%	(97.52 $\pm$ 0.06)%	(97.52 $\pm$ 0.06)%
$B_{(s)}^0$ or $J/\psi$ vertex $\chi^2$ /ndof	(97.21 $\pm$ 0.03)%	(97.18 $\pm$ 0.03)%	(97.68 $\pm$ 0.02)%	(96.78 $\pm$ 0.07)%	(96.78 $\pm$ 0.07)%
$B_{(s)}^0$ or $J/\psi$ DOCA	(99.82 $\pm$ 0.01)%	(99.80 $\pm$ 0.01)%	(99.83 $\pm$ 0.01)%	(99.58 $\pm$ 0.03)%	(99.58 $\pm$ 0.03)%
$\mu, h, K^+$ minimum IP $\chi^2$	(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 3.4** Stripping line cut efficiencies for  $B_s^0 \rightarrow \mu^+ \mu^-$ ,  $B \rightarrow h^+ h^-$  and  $B^+ \rightarrow J/\psi K^+$  2012 simulated decays. Selection cuts applied are listed in Table 3.3. Efficiencies have been calculated only for cuts that are present in the  $B_s^0 \rightarrow \mu^+ \mu^-$  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.

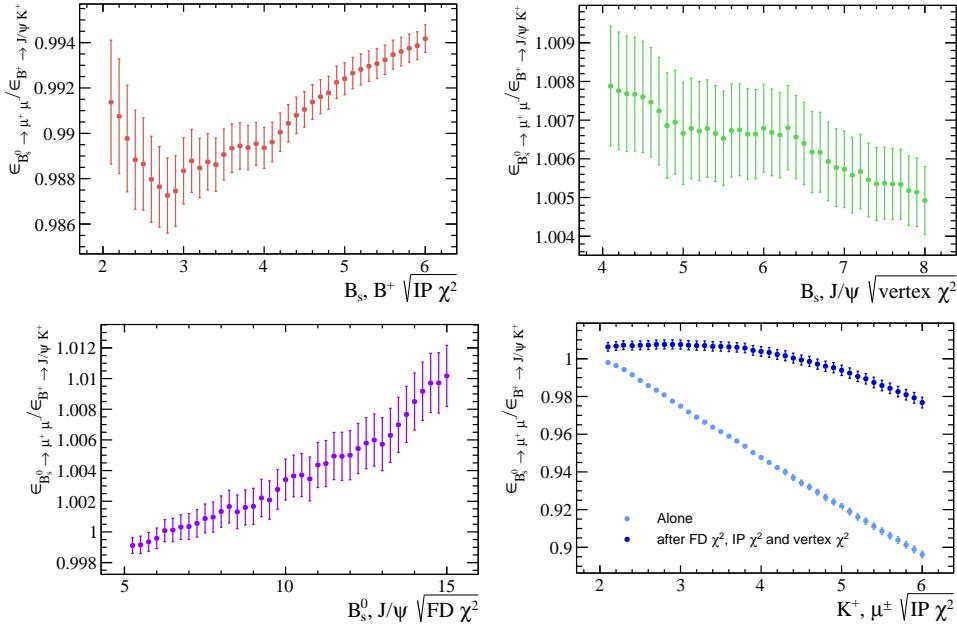
exception of the  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B^+ \rightarrow J/\psi K^+$  IP  $\chi^2$  cuts on the daughter particles, the ratio of efficiencies is well within 3% of 1 for the range of cuts values shown. The ratio of the  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B^+ \rightarrow J/\psi K^+$  efficiencies for the daughter particle IP  $\chi^2$  markedly deviates from unity, showing that the IP  $\chi^2$  distribution of the muons and kaon are very different as seen previous in [28]. If the FD  $\chi^2$ ,  $B_s^0$  or  $J/\psi$  IP  $\chi^2$  and vertex  $\chi^2$  selection cuts are applied to the simulated events before the daughter IP  $\chi^2$  requirement the ratio of  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  and  $B^+ \rightarrow J/\psi K^+$  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  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  selection will have a similar impact on the efficiencies of the normalisation decays.



**Fig. 3.2** The ratio of  $B_s^0 \rightarrow \mu^+\mu^-$  to  $B^0 \rightarrow K^+\pi^-$  stripping efficiencies when each cut has been applied independently of all other cuts. The current cut values are marked by the blue lines.

The efficiencies for most of the stripping cuts as  $\sim 97\%$  or higher, however, the efficiencies of the cuts on the FD  $\chi^2$  of the  $B_{(s)}^0$  or  $J/\psi$  and the daughter IP  $\chi^2$  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 3.4 shows the total efficiency of the  $B_s^0 \rightarrow \mu^+\mu^-$  stripping



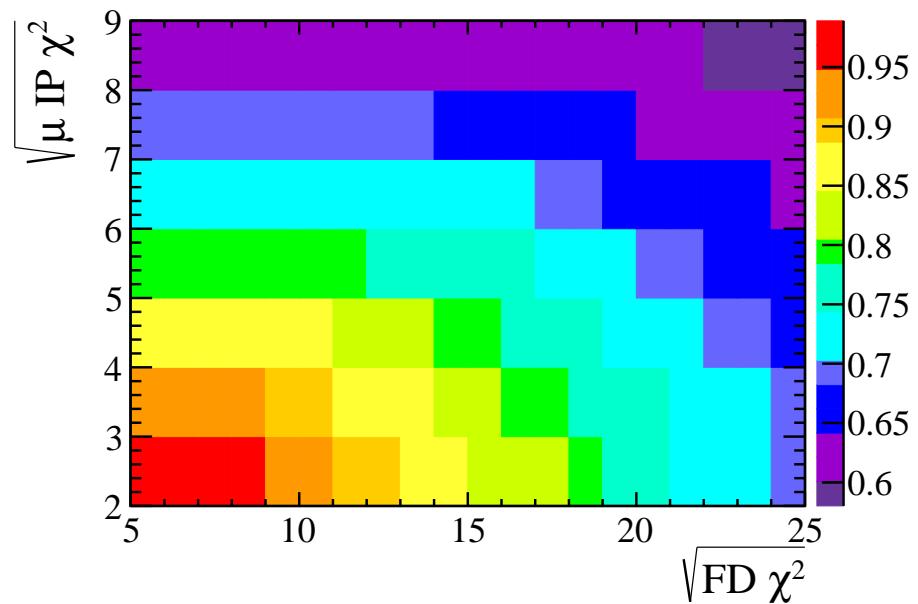
**Fig. 3.3** The ratio of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  to  $B^+ \rightarrow J/\psi K^+$  stripping efficiencies when each cut has been applied independently of all other cuts. The current cut values are marked by the blue lines.

line on simulated  $B_s^0 \rightarrow \mu^+ \mu^-$  decays for a range of FD  $\chi^2$  and daughter IP  $\chi^2$  cut values. As expected the lower the cut values the more efficient the stripping line becomes. It is important that any increase in  $B_s^0 \rightarrow \mu^+ \mu^-$  selection efficiency from the stripping is not removed when the trigger requirements are applied, Figure 3.5 shows that the trigger efficiencies are relatively flat across a large range of FD  $\chi^2$  and daughter IP  $\chi^2$  cut values therefore the efficiency gained by a change in the stripping selection is not lost when trigger requirements are imposed. The selection efficiency for  $B^0 \rightarrow \mu^+ \mu^-$  is very similar to  $B_s^0 \rightarrow \mu^+ \mu^-$  as seen in Table 3.4, therefore only  $B_s^0 \rightarrow \mu^+ \mu^-$  have been studied for different stripping selection cut values.

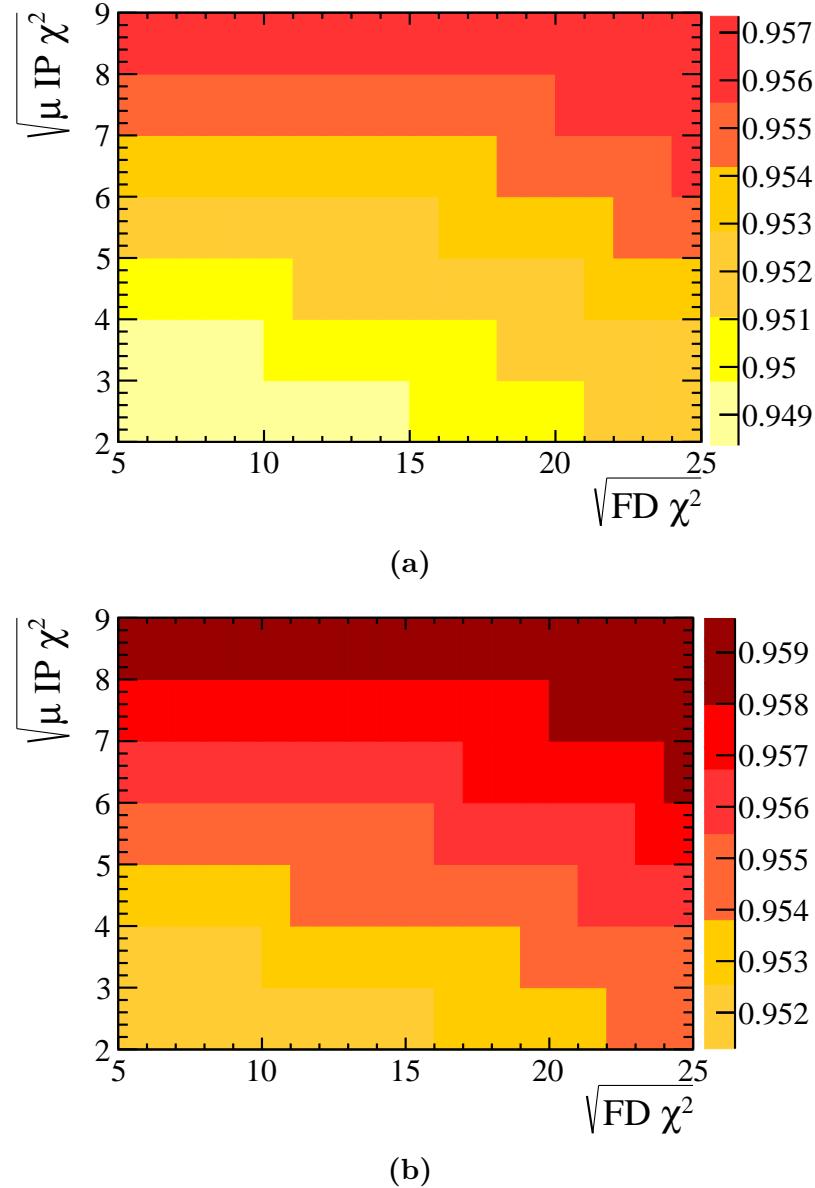
One of the main purposes of the stripping selection, as described in Section ??, is to reduce the size of the data set, therefore the cuts cannot be set as loose as possible.

Any change applied to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  stripping line must be propagated through into the stripping lines for  $B \rightarrow h^+ h^-$  and  $B^+ \rightarrow J/\psi K^+$  decays therefore the retention of all stripping lines must be evaluated.

Table 3.5 shows the total efficiency of the  $B_s^0 \rightarrow \mu^+ \mu^-$  stripping line along side the amount of data retained for the set of cuts on the FD  $\chi^2$  and daughter IP  $\chi^2$  for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ,  $B \rightarrow h^+ h^-$  and  $B^+ \rightarrow J/\psi K^+$  stripping lines. The set of chosen cuts aims to keep both cuts as high as possible for a certain  $B_s^0 \rightarrow \mu^+ \mu^-$  efficiency.



**Fig. 3.4** Efficiency of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  stripping selection for  $B_s^0 \rightarrow \mu^+ \mu^-$  simulated decays for a range of cuts on the  $B_s^0$  FD  $\chi^2$  and the minimum muon IP  $\chi^2$ .



**Fig. 3.5** The trigger efficiencies of  $B_s^0 \rightarrow \mu^+\mu^-$  simulated decays across a range of  $B_s^0$  FD  $\chi^2$  and the minimum muon IP  $\chi^2$  cut values. a) shows the efficiencies of the trigger requirements used to select  $B_s^0 \rightarrow \mu^+\mu^-$  decays for the effective lifetime measurement listed in Table 3.2 and b) shows the efficiencies of the trigger requirements used to select  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays for the branching fraction measurement the same trigger lines in Tables 3.2 are used but candidates must pass the DEC trigger decision at each level.

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

**Table 3.5** The efficiency of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  stripping line to select  $B_s^0 \rightarrow \mu^+ \mu^-$  decays and the changing in the date retention for  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ,  $B \rightarrow h^+ h^-$  and  $B^+ \rightarrow J/\psi K^+$  stripping lines for a range of FD  $\chi^2$  and daughter IP  $\chi^2$  cut values. Fractional uncertainty on the retention is less than 1 %.

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  $\chi^2$  and daughter IP  $\chi^2$  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 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 3.5 however the loosest cuts increases the amount of data passing the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  stripping selection by a factor of 7 and the  $B \rightarrow h^+ h^-$  stripping selection by a factor of 4. Table 3.6 shows the number of Run 1 candidates passing the original stripping selection listed in Table 3.3 for the last published analysis. The  $B \rightarrow h^+ h^-$  stripping line lets through the most candidates where as the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  stripping line saves far fewer candidates, therefore a chance in the retention of the  $B \rightarrow h^+ h^-$  line is more significant than the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  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  $B_s^0$  FD  $\chi^2 > 121$  and minimum muon IP  $\chi^2 > 9$  would increase the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  selection efficiency by from 71 % to 82 % and the amount of data retained would be doubled. The increase of the data retained by the  $B \rightarrow h^+ h^-$  and  $B^+ \rightarrow J/\psi K^+$  lines is smaller and the efficiencies are similar to the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  selection efficiencies. Therefore these cuts are applied in the stripping selection for this analysis.

Stripping Lines	Events	Retention / %
$B_{(s)}^0 \rightarrow \mu^+ \mu^-$	898880	0.0022
$B \rightarrow h^+ h^-$	14502295	0.0831
$B^+ \rightarrow J/\psi K^+$	3344568	0.0087
Total	18745743	-

**Table 3.6** The number of events passing stripping lines used for the  $B_s^0 \rightarrow \mu^+ \mu^-$  analysis from the selection listed in Table 3.3 and the percentage of the total LHCb data set that they correspond to. The total does not include correlation between lines, which is expected to be 42 % between  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  and  $B \rightarrow h^+ h^-$  lines.

### 3.4.3 Stripping selection and offline cuts

The complete list of selection cuts applied in the cut based selection to select  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B \rightarrow h^+ h^-$  decays in Run 1 and Run 2 data are listed in Tables 3.7. The stripping selection cuts from Table 3.3 are included with the  $B$  mesons FD  $\chi^2$  and daughter IP  $\chi^2$  requirements updated to the looser values and the selection of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays includes the momentum, ghost track probability and decay time cuts made in the  $B \rightarrow h^+ h^-$  stripping line, but were absent in the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  stripping line.

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  $pp \rightarrow p\mu^+\mu^-p$  decays and a  $J/\psi$  veto is used to remove backgrounds from  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decays. The semi-leptonic  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decays, where  $J/\psi \rightarrow \mu^+ \mu^-$ , contribute to the background of  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  decays when a muon from the  $J/\psi$  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 ‘ $J/\psi$  veto’ can be used to remove background events from  $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$  decays. The veto works by removing events where one muon from the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  candidate combined with any other oppositely charged muon in the event has  $m_{\mu^+ \mu^-} - m_{J/\psi} < 30 \text{ MeV}/c^2$ .

The  $B$  meson mass range for both  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B \rightarrow h^+ h^-$  decays is narrower than the range in the stripping selection in Section 3.4.1.  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates are required to have a dimuon invariant mass greater than  $5320 \text{ MeV}/c^2$ . The motivation comes from mass fit studies that are detailed in Section X. The consequence of this cut is to remove  $B^0 \rightarrow \mu^+ \mu^-$  decays,  $B_s^0 \rightarrow \mu^+ \mu^- \gamma$  backgrounds and most backgrounds from mis-identified semi-leptonic and  $B \rightarrow h^+ h^-$  decays. This can be seen from the mass distribution in Figure 3.1. The expect number of  $B^0 \rightarrow \mu^+ \mu^-$  and mis-identified decays after the full selection can be found in Section X. Similarly the  $B \rightarrow h^+ h^-$  mass window is reduced to remove contributions from mis-identified backgrounds.

The selection applied to Run 1 and Run 2 is the same for all variables except the track ghost probability and track  $\chi^2/ndof$ . Slightly looser cuts are used for Run 2 to take advantage to changes in the reconstruction that were introduced for Run 2.

## 3.5 Particle Identification

Particle identification (PID) variables are used to refine the selection of  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates and to separate different  $B \rightarrow h^+ h^-$  decays.

Particle	$B_s^0 \rightarrow \mu^+\mu^-$	$B \rightarrow h^+h^-$
$B_s^0$ or $B^+$	$5320 \text{ MeV}/c^2 < M < 6000 \text{ MeV}/c^2$	$5100 \text{ MeV}/c^2 < M < 5500 \text{ MeV}/c^2$
DIRA	$> 0$	$> 0$
FD $\chi^2$	$> 121$	$> 121$
IP $\chi^2$	$< 25$	$< 25$
Vertex $\chi^2/\text{ndof}$	$< 9$	$< 9$
DOCA	$< 0.3 \text{ mm}$	$< 0.3 \text{ mm}$
$\tau$	$< 13.248 \text{ ps}$	$< 13.248 \text{ ps}$
$p_T$	$> 500 \text{ MeV}/c$	$> 500 \text{ MeV}/c$
Daughter $\mu$ or $h$	Track $\chi^2/\text{ndof} < 3$ (4) Minimum IP $\chi^2 > 9$ $0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$ $p < 500 \text{ GeV}/c$ ghost probability $< 0.3$ (0.4) $ m_{\mu\mu} - mJ/\psi  < 30 \text{ MeV}/c^2$ isMuon = True	Track $\chi^2/\text{ndof} < 3$ (4) Minimum IP $\chi^2 > 9$ $0.25 \text{ GeV}/c < p_T < 40 \text{ GeV}/c$ $p < 500 \text{ GeV}/c$ ghost probability $< 0.3$ (0.4) $ m_{\mu\mu} - mJ/\psi  < 30 \text{ MeV}/c^2$ -

**Table 3.7** Selection cuts applied to select  $B_s^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  decays, where selection is different between Run 1 and Run 2 the Run 2 values are shown in parenthesis.

In the selection of  $B_s^0 \rightarrow \mu^+\mu^-$  decays PID variables are particularly useful to reduce the backgrounds coming from mis-identified semi-leptonic decays and  $B \rightarrow h^+h^-$  decays and also help to reduce the number of combinatorial background events. However most backgrounds from mis-identified semi-leptonic and  $B \rightarrow h^+h^-$  decays are below the mass cut applied to  $B_s^0 \rightarrow \mu^+\mu^-$  candidates at 5320 MeV/ $c^2$  therefore loose PID requirements can be used to select  $B_s^0 \rightarrow \mu^+\mu^-$  decays ensuring a higher signal selection efficiency.

The PID requirements to select  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  decays are shown Table ?? alongside requirements to separate different  $B \rightarrow h^+h^-$  decays. Two types of PID variables, defined in Section 1.2.2.4, are used; DLL variables and ProbNN variables.

A linear combination of ProbNN variables is used to select  $B_s^0 \rightarrow \mu^+\mu^-$  decays and remove semi-leptonic backgrounds, in addition to the isMuon requirement applied in the stripping selection. The classifiers used in ProbNN variables are tuned to give the best performance depending on the different data taking conditions in the detector for each year. Since different tunes are used to select  $B_s^0 \rightarrow \mu^+\mu^-$  decays in 2016 data compared to Run 1 and 2015 data, the requirement on the linear combination of ProbNN variables varies with the year of data taking. The cuts are chosen to give similar efficiencies for each data sets at selecting signal and removing background across the different years.

The separation of different  $B^0 \rightarrow K^+\pi^-$  and  $B_s^0 \rightarrow K^+K^-$  decays is done via DLL variables. These are useful to separate  $B \rightarrow h^+h^-$  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 the same for each year of data taking.

## 3.6 Multivariate Classifiers

The selection described so far remove a large number of background candidates however because  $B_s^0 \rightarrow \mu^+\mu^-$  decays occur very rarely the data is still dominated by long lived combinatorial background from  $b\bar{b} \rightarrow \mu^+\mu^- X$  decays. To increase the signal purity of the data multivariate classifiers are used to separate  $B_s^0 \rightarrow \mu^+\mu^-$  from the backgrounds.

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.

Decay	Particle	PID requirements
$B_s^0 \rightarrow \mu^+ \mu^-$ (Run 1 and 2015)	$m\mu^+$ and $\mu^-$	$\text{ProbNN}\mu * (1 - \text{ProbNN}\pi) * (1 - \text{ProbNN}p) > 0.2$
$B_s^0 \rightarrow \mu^+ \mu^-$ (2016)	$m\mu^+$ and $\mu^-$	$\text{ProbNN}\mu * (1 - \text{ProbNN}\pi) * (1 - \text{ProbNN}p) > 0.4$
$B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow K^+ \pi^-$	$K^+$	$\text{DLL}_{K\pi} > 10$
	$\pi^-$	$\text{DLL}_{K\pi} < -10$
$B_s^0 \rightarrow K^+ K^-$	$K^+$ and $K^-$	$\text{DLL}_{K\pi} > 10$

**Table 3.8** Particle identification requirements to select  $B_s^0 \rightarrow \mu^+ \mu^-$  decays and to separate the  $B \rightarrow h^+ h^-$  decays  $B^0 \rightarrow K^+ \pi^-$  and  $B_s^0 \rightarrow K^+ \pi^-$  from  $B_s^0 \rightarrow K^+ K^-$ .

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. A cut is placed on the output of the classifier to remove background so that the remaining data set has a higher purity for signal events.

Two multivariate classifiers are used to select  $B_s^0 \rightarrow \mu^+ \mu^-$  decays. Both classifiers are a type called a Boosted Decision Tree (BDT) that are described in Section 3.6.1. A range of different classifiers were investigated but BDTs preformed the best at separating signal from background.

The first classifier (Sect. 3.6.2), called the BDTS, is used to remove candidates that are very unlikely to be signal and has a high efficiency to select  $B_s^0 \rightarrow \mu^+ \mu^-$  decays. The second classifier (Sect. 3.6.3), called the global BDT, is the final step in the selection process, the output is used to remove most of the remaining backgrounds and has much lower efficiency to select signal events compared to the BDTS. The cut applied to the output of this classifier is optimised to give the lowest expected uncertainty on the measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime (Sect. 3.6.4).

Both classifiers were developed for the measurement of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Branching Fractions. In the selection  $B_s^0 \rightarrow \mu^+ \mu^-$  decays for this analysis the second classifier is used to classify candidates into 8 bins containing increasing proportions of signal candidates, no candidates are removed based on the output of the second BDT. Therefore the BDTS is necessary to reduce the number of background events to a more manageable level. The use of a single classifier was investigated to select candidates to measure the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime however it was found that the two classifiers developed for the Branching Fraction measurement performed best.

### 3.6.1 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 known 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 perfect, 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 [31] is used 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

$$w = \frac{1-f}{f}, \text{ where } f = \frac{\text{total misclassified events}}{\text{total events}}. \quad (3.1)$$

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  $\beta$  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<sup>2</sup>);, the ‘speed’ at which the boosting occurs ( $\beta$ ) 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 by preforms poorly at classifying candidates in a statistically independent sample. Although this is less common in BDT 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.

### 3.6.2 The BDTS

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

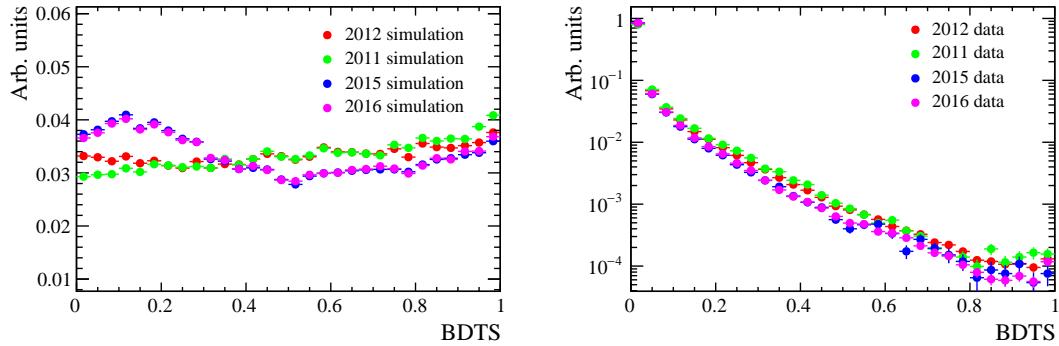
- impact parameter  $\chi^2$  of the  $B_s^0$
- vertex  $\chi^2$  of the  $B_s^0$
- direction cosine of  $B_s^0$
- distance of closest approach of the muons

---

<sup>2</sup>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.

- minimum impact parameter  $\chi^2$  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

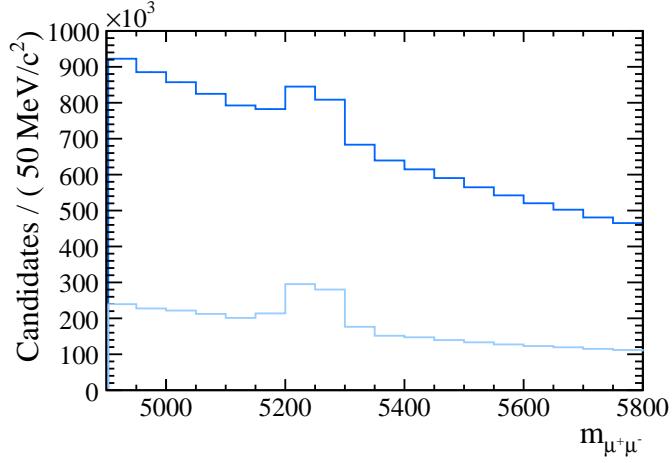
The signal and background samples used to train the BDTS are simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays and  $B_s^0 \rightarrow \mu^+\mu^-$  candidates in a sample of Run 1 data from the mass ranges 4800 - 5000 MeV/ $c^2$  and 5500 - 6000 MeV/ $c^2$ . The selection cuts listed in Table 3.9 are applied to the training samples and the training parameters used in the BDT are listed in Table 3.10. 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 3.6. The BDTS is applied to all candidates passing the  $B_{(s)}^0 \rightarrow \mu^+\mu^-$  and  $B \rightarrow h^+h^-$  stripping lines, and candidates are required to have a BDTS value above 0.05. The performance of the BDTS at removing backgrounds is illustrated in Figure 3.7.



**Fig. 3.6** BDTS response for simulated  $B_s^0 \rightarrow \mu^+\mu^-$  decays (left) and data with a mass above 5447 MeV/ $c^2$  consisting on  $b\bar{b} \rightarrow \mu^+\mu^-X$  decays.

### 3.6.3 Global BDT

The global BDT is the final step in selecting  $B_s^0 \rightarrow \mu^+\mu^-$  decays and it is very effective at separating them from long lived combinatorial background decays. The 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  $B_s^0 \rightarrow \mu^+\mu^-$  candidate is from other tracks in the event. The tracks of the muons from a real  $B_s^0 \rightarrow \mu^+\mu^-$  decays will be, in general, far from other tracks in the event because the  $B_s^0 \rightarrow \mu^+\mu^-$  decays tree contains no other tracks apart from the



**Fig. 3.7** Invariant mass spectrum for  $B \rightarrow h^+ h^-$  decays in 2016 data passing the selection requirements in Table 3.9 before and after the BDTS cut is applied.

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

**Table 3.9** 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 to that the BDTS can be used on  $B \rightarrow h^+ h^-$  decays. The trigger requirement used are the same as those used in the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Branching Fraction analysis.

Parameter	Value
nTrees	250
nEventsMin	400
MaxDepth	3
$\beta$	1.0
nCuts	20

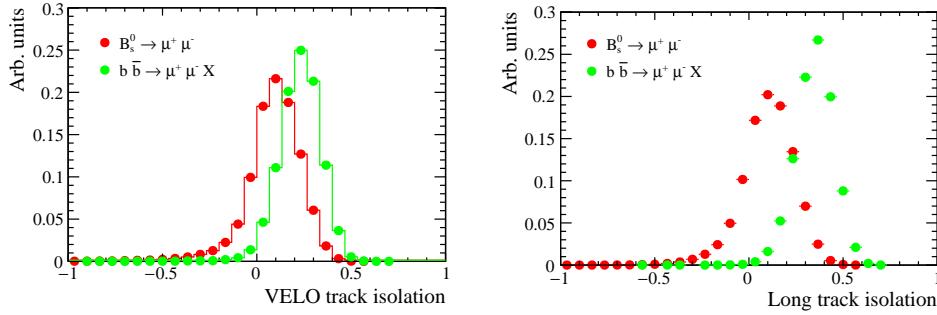
**Table 3.10** Training parameters used to specify the training of the BDTS.

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  $B_s^0 \rightarrow \mu^+ \mu^-$  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  $B_s^0 \rightarrow \mu^+ \mu^-$  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  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays using a set of input variables that describe track and vertex properties. The BDT compares the  $\mu^+$  from a  $B_s^0 \rightarrow \mu^+ \mu^-$  candidate with all other tracks in the event, excluding the track of the  $\mu^-$ , and gives an output,  $iso_{\mu^+}(track)$ , for each possible  $\mu^+$  and track pairing. The process is repeated for the  $\mu^-$ . The BDT is designed to produce high output values for muons from  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays and a low value for muons from  $B_s^0 \rightarrow \mu^+ \mu^-$  decays. The isolation variable of a  $B_s^0 \rightarrow \mu^+ \mu^-$  candidate is then composed of the sum of the highest values of  $iso_{\mu^+}(track)$  and  $iso_{\mu^-}(track)$  for any tracks in the event. The separation power of these isolations are shown in Figure 3.8.

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
- $\sqrt{\Delta\phi^2 + \Delta\eta^2}$ , where  $\Delta\phi$  is the difference in azimuthal angles of the muons and  $\Delta\eta$  the difference in the pseudorapidity of the muons



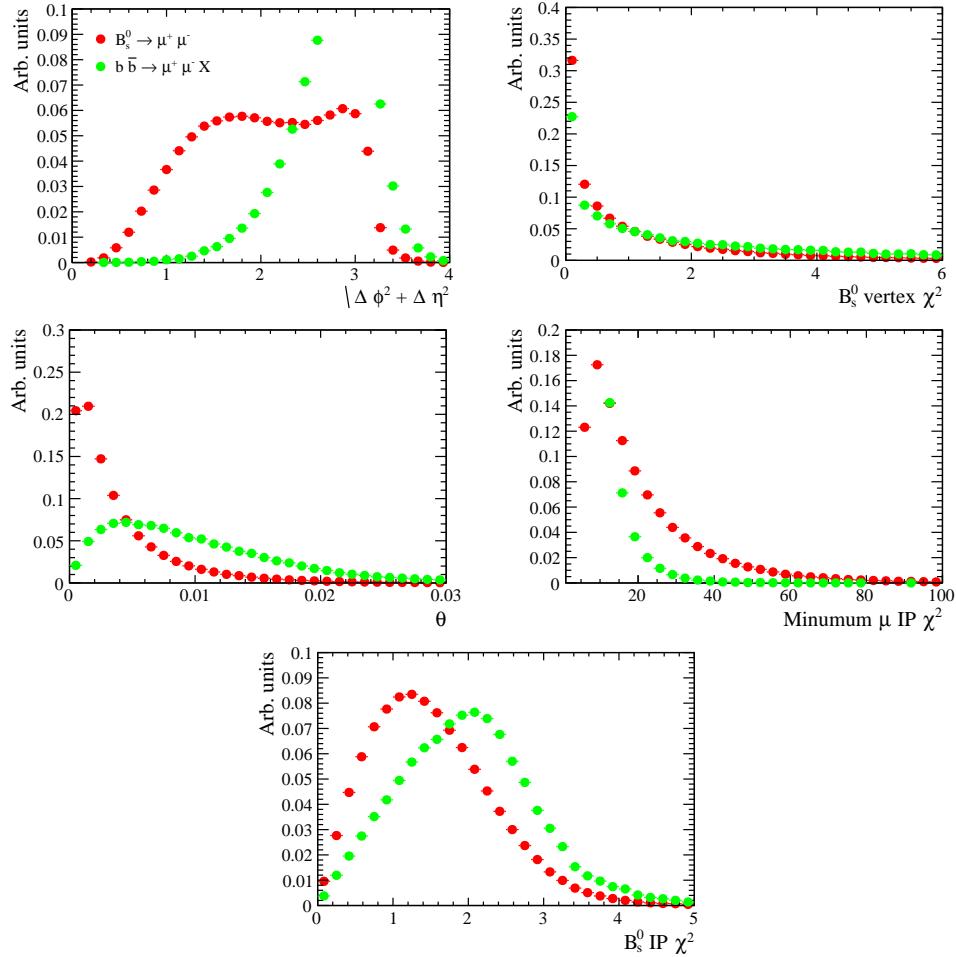
**Fig. 3.8** Long track and VELO track isolation distributions of simulated  $B_s^0 \rightarrow \mu^+\mu^-$  and  $b\bar{b} \rightarrow \mu^+\mu^-X$  decays used to train the global BDT passing cuts in Table 3.12.

- the smallest IP  $\chi^2$  with respect to the primary vertex of the  $B_s^0 \rightarrow \mu^+\mu^-$  of the muons
- vertex  $\chi^2$  of the  $B_s^0$
- IP  $\chi^2$  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 3.8 and 3.9. 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 3.11. These parameters were chosen by scanning across a range of variables and choosing those that gave the best performance.

Parameter	Value
nTrees	1000
MinNodeSize	1%
MaxDepth	3
$\beta$	0.75
nCuts	30

**Table 3.11** Training parameters used to specify the training of the global BDT.



**Fig. 3.9** Distributions of input variables of the global BDT from simulated  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $b \bar{b} \rightarrow \mu^+ \mu^- X$  decays used to train the global BDT passing cuts in Table 3.12.

Simulated  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays are used to provide large signal and background training samples for the global BDT. In data  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates in the mass range 5431 to 6550 MeV/ $c^2$  consist almost entirely of  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays, however the number of candidates in this mass range is too small to be a useful sample of background candidates to train a BDT with comparable performance to one trained entirely on simulated decays. The simulated sample  $b\bar{b} \rightarrow \mu^+ \mu^- X$  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  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays. The cuts, listed in Table 3.1, were applied to the simulated decays as they were generated to reduce the number of events saved on disk. Also the stripping selection cuts in Table 3.3 were applied and candidates that did not pass the stripping selection were not saved. Unfortunately the  $b\bar{b} \rightarrow \mu^+ \mu^- X$  sample therefore does not include candidates that are selected by the looser stripping selection described in Section ???. 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  $\chi^2$  and daughter IP  $\chi^2$  listed in Table 3.3 must be used to select  $B_s^0 \rightarrow \mu^+ \mu^-$  candidates. The complete list of selection requirements applied to the training samples used to develop global BDT are listed in Table 3.12, the same selection is applied to  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays.

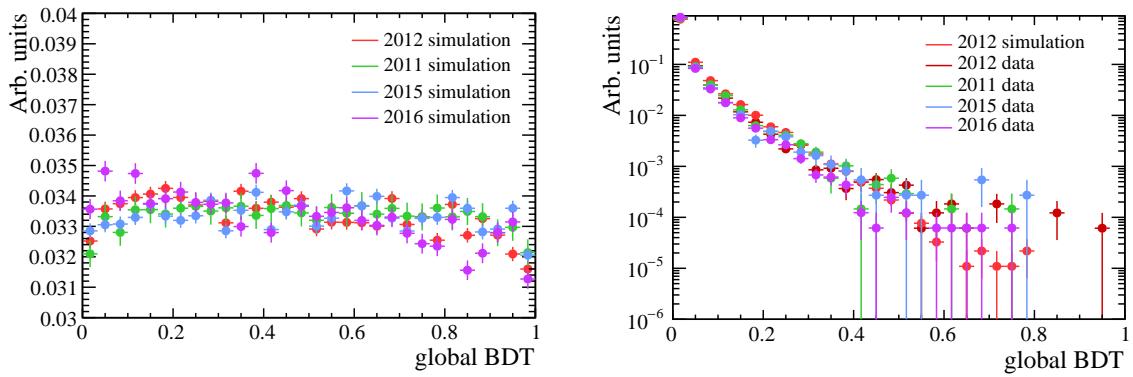
The global BDT is used for data taken in all years and in the same way as the BDTS the final output of the global BDT is flattened to have a response between 0 and 1 that is uniform for signal and the background peaks at zero, as shown in Figure 3.10 for each year of data taking. The flattening is important for the measurement of the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Branching Fractions where 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 and background rejection of the global BDT is shown in Figure 3.11 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.

### 3.6.4 Global BDT cut optimisation

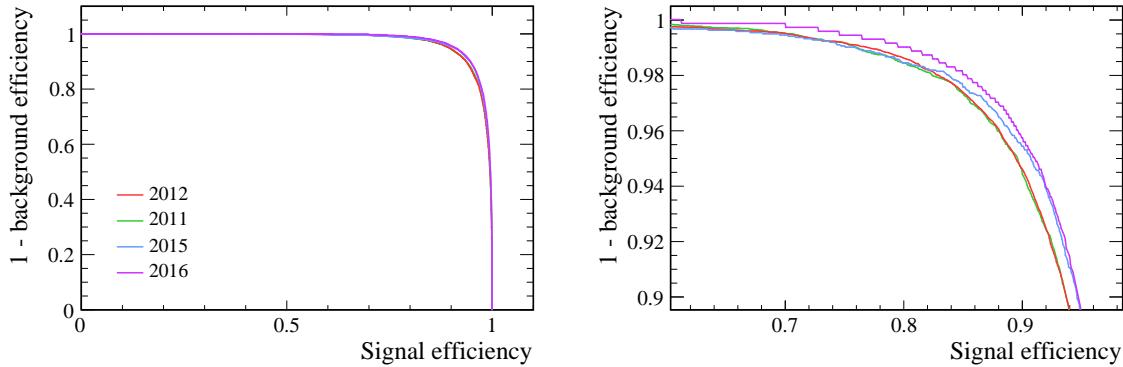
A cut is placed on the output of the global BDT to select  $B_s^0 \rightarrow \mu^+ \mu^-$  and  $B \rightarrow h^+ h^-$  decays. The cut value has been optimised to give the smallest expected uncertainty on the measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  effective lifetime,  $\tau_{\mu\mu}$ , and its inverse,  $\tau_{\mu\mu}^{-1}$ . This is

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

**Table 3.12** Selection cuts applied to select candidates for signal and background samples used to train the BDT. The trigger requirements imposed are those used to select decays for the  $B_{(s)}^0 \rightarrow \mu^+ \mu^-$  Branching Fraction measurement.



**Fig. 3.10** Global BDT output distributions for  $B_s^0 \rightarrow \mu^+ \mu^-$  simulated decays (left) and  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays from simulation and data.



**Fig. 3.11** Global BDT performance for 2011, 2012, 2015 and 2016 data taking conditions. Signal efficiency is calculated from  $B_s^0 \rightarrow \mu^+ \mu^-$  simulated decays and background rejection from data passing the  $B_s^0 \rightarrow \mu^+ \mu^-$  selection with  $m_{\mu^+ \mu^-} > 5447 \text{ MeV}/c^2$ . The full BDT range is shown in the left plot and only the most sensitive region is shown in the right.

done by using toy experiments for the expected number of  $B_s^0 \rightarrow \mu^+ \mu^-$  combinatorial background decays for different cut on the global BDT output.

The fit procedure to extract  $\tau_{\mu\mu}$  from the data is described in depth in Chapter X 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  $B_s^0 \rightarrow \mu^+ \mu^-$  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  $B_s^0 \rightarrow \mu^+ \mu^-$  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 [32]
- 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 X.

The number of expected  $B_s^0 \rightarrow \mu^+ \mu^-$  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  $\text{BDT} > 0.55$  for Run 1 and Run 2 data but in the mass range  $4900 < m_{\mu^+ \mu^-} < 6000 \text{ MeV}/c^2$ . These predictions assume the SM branching fraction

for  $B_s^0 \rightarrow \mu^+ \mu^-$  and are given in Table 3.13. Since the output of the global BDT is flattened the number of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays is evenly distributed across the BDT range, therefore the expected number of  $B_s^0 \rightarrow \mu^+ \mu^-$  decays is straight forward to calculate for each BDT cut value. The number of combinatorial background decays expected after each BDT is computed from simulated  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays using the ratio

$$R = \frac{\epsilon(BDT > X)}{\epsilon(BDT > 0.55)} \quad (3.2)$$

where  $\epsilon(BDT > X)$  is the efficiency of the cuts  $BDT > X$ . The  $B_s^0 \rightarrow \mu^+ \mu^-$  selection requirements are applied to the simulated decays before taking the efficiency. The ratios for the different cuts values are shown in Table 3.14. 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
$B_s^0 \rightarrow \mu^+ \mu^-$	30.94
Combinatorial background	66.23
Total	97.17

**Table 3.13** Expected number of  $B_s^0 \rightarrow \mu^+ \mu^-$  and combinatorial background candidates after the  $B_s^0 \rightarrow \mu^+ \mu^-$  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	$R_\epsilon$
0.40	8.69
0.45	3.91
0.50	1.91
0.55	1.00
0.60	0.55
0.65	0.32

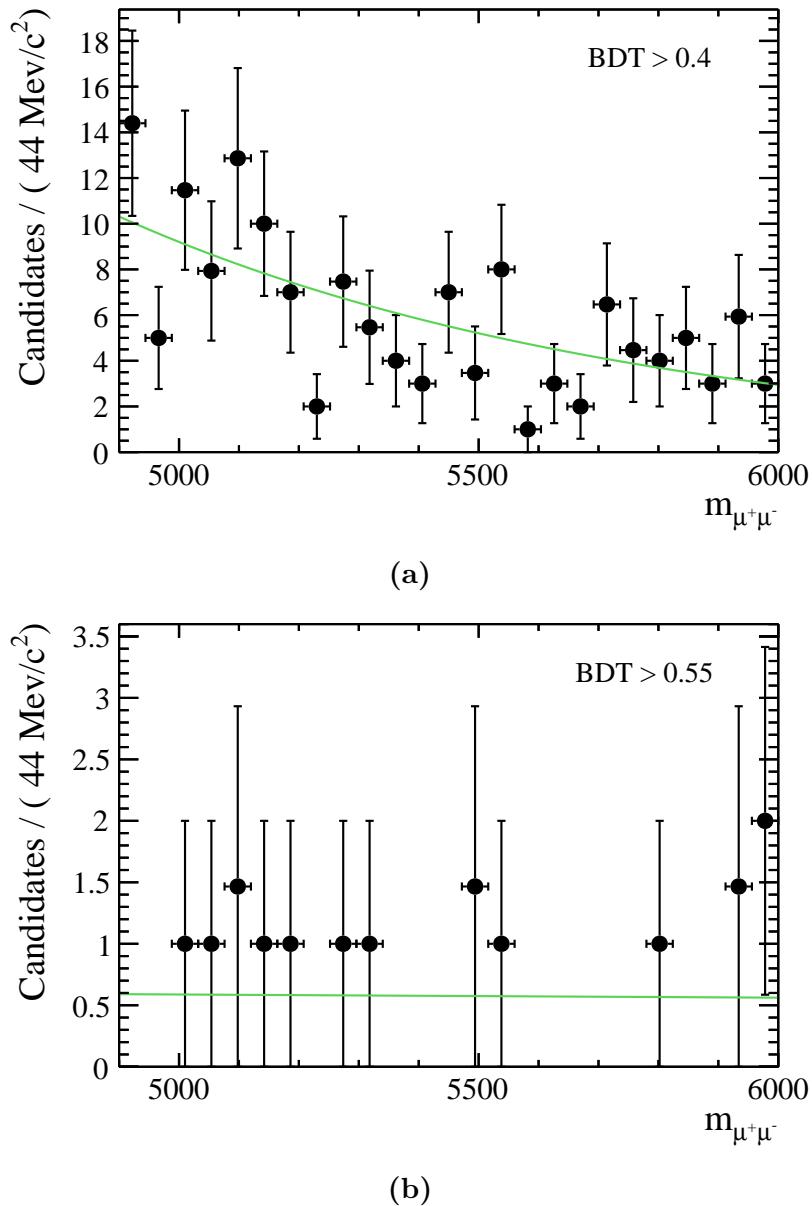
**Table 3.14** The ratio of efficiencies of cuts on the global BDT to select  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays relative to a cut of 0.55 on the global BDT.

The mass distribution of the combinatorial background is a decaying exponential, it was observed from the simulated  $b\bar{b} \rightarrow \mu^+\mu^-X$  decays that the slope of the mass distribution changed with the BDT cut value as illustrated in Figure 3.12. The change in slope is accounted for when generating events for the toy experiment by changing the slope parameter ( $\lambda$ ) for each BDT cut. Table 3.15 shows the slope of the mass distribution for different BDT cuts values evaluated from  $b\bar{b} \rightarrow \mu^+\mu^-X$  simulated decays.

BDT cut	$\lambda / c^2 \text{MeV}^{-1}$
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 3.15** The slope of the combinatorial background mass distribution for different cut value on the global BDT evaluated from  $b\bar{b} \rightarrow \mu^+\mu^-X$  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 3.16 along with the expected number of  $B_s^0 \rightarrow \mu^+\mu^-$  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 for a BDT cut of 0.55, therefore this cut value is used to select  $B_s^0 \rightarrow \mu^+\mu^-$  decays. The same cut is applied to the global BDT to select  $B \rightarrow h^+h^-$  decays.



**Fig. 3.12** Mass distribution of simulated decays after global BDT cuts of 0.4 and 0.55 and the  $B_s^0 \rightarrow \mu^+ \mu^-$  selection.

Global BDT cut	$\frac{S}{\sqrt{S+B}}$	$\sigma(\tau_{\mu\mu}) / \text{ps}$	$\sigma(\tau_{\mu\mu}^-) / \text{ps}^{-1}$
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 3.16** 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.



# Bibliography

- [1] C. member states. <http://home.cern/about/member-states>.
- [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] M. Adinolfi *et al.*, “Performance of the LHCb RICH detector at the LHC,” *Eur. Phys. J.*, vol. C73, p. 2431, 2013.
- [9] F. Archilli *et al.*, “Performance of the Muon Identification at LHCb,” *JINST*, vol. 8, p. P10020, 2013.
- [10] R. Aaij *et al.*, “Absolute luminosity measurements with the LHCb detector at the LHC,” *JINST*, vol. 7, p. P01010, 2012.
- [11] O. Lupton and G. Wilkinson, *Studies of  $D^0 \rightarrow K_s^0 h^+ h^-$  decays at the LHCb experiment*. PhD thesis, Oxford U., Jul 2016. Presented 14 Sep 2016.
- [12] P. Mato, “GAUDI-Architecture design document,” 1998.
- [13] R. Antunes-Nobrega *et al.*, *LHCb computing: Technical Design Report*. Technical Design Report LHCb, Geneva: CERN, 2005. Submitted on 11 May 2005.
- [14] F. Stagni *et al.*, “LHCbDirac: Distributed computing in LHCb,” *J. Phys. Conf. Ser.*, vol. 396, p. 032104, 2012.
- [15] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework,” *Nucl. Instrum. Meth.*, vol. A389, pp. 81–86, 1997.

- [16] 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.
- [17] 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.
- [18] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP*, vol. 05, p. 026, 2006.
- [19] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.
- [20] D. J. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth.*, vol. A462, pp. 152–155, 2001.
- [21] 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.
- [22] S. Agostinelli *et al.*, “GEANT4: A Simulation toolkit,” *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.
- [23] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Trans. Nucl. Sci.*, vol. 53, p. 270, 2006.
- [24] I. Bird, “Computing for the Large Hadron Collider,” *Ann. Rev. Nucl. Part. Sci.*, vol. 61, pp. 99–118, 2011.
- [25] W. L. C. Grid. <http://www.cern.ch/LHCgrid>.
- [26] S. Paterson and A. Tsaregorodtsev, “DIRAC Infrastructure for Distributed Analysis,” Feb 2006.
- [27] V. Khachatryan *et al.*, “Observation of the rare  $B_s^0 \rightarrow \mu^+ \mu^-$  decay from the combined analysis of CMS and LHCb data,” *Nature*, vol. 522, pp. 68–72, 2015.
- [28] D. Martinez Santos, *Study of the very rare decay  $B_s \rightarrow \mu^+ \mu^-$  in LHCb*. PhD thesis, Santiago de Compostela, Universidade de Santiago de Compostela, Santiago de Compostela, 2010. Presented on 05 May 2010.
- [29] R. Aaij *et al.*, “Measurement of the  $B_s^0 \rightarrow \mu^+ \mu^-$  branching fraction and search for  $B^0 \rightarrow \mu^+ \mu^-$  decays at the LHCb experiment,” *Phys. Rev. Lett.*, vol. 111, p. 101805, 2013.
- [30] C. Patrignani *et al.*, “Review of Particle Physics,” *Chin. Phys.*, vol. C40, no. 10, p. 100001, 2016.
- [31] A. Hoecker *et al.*, “TMVA: Toolkit for Multivariate Data Analysis,” *PoS*, vol. ACAT, p. 040, 2007.

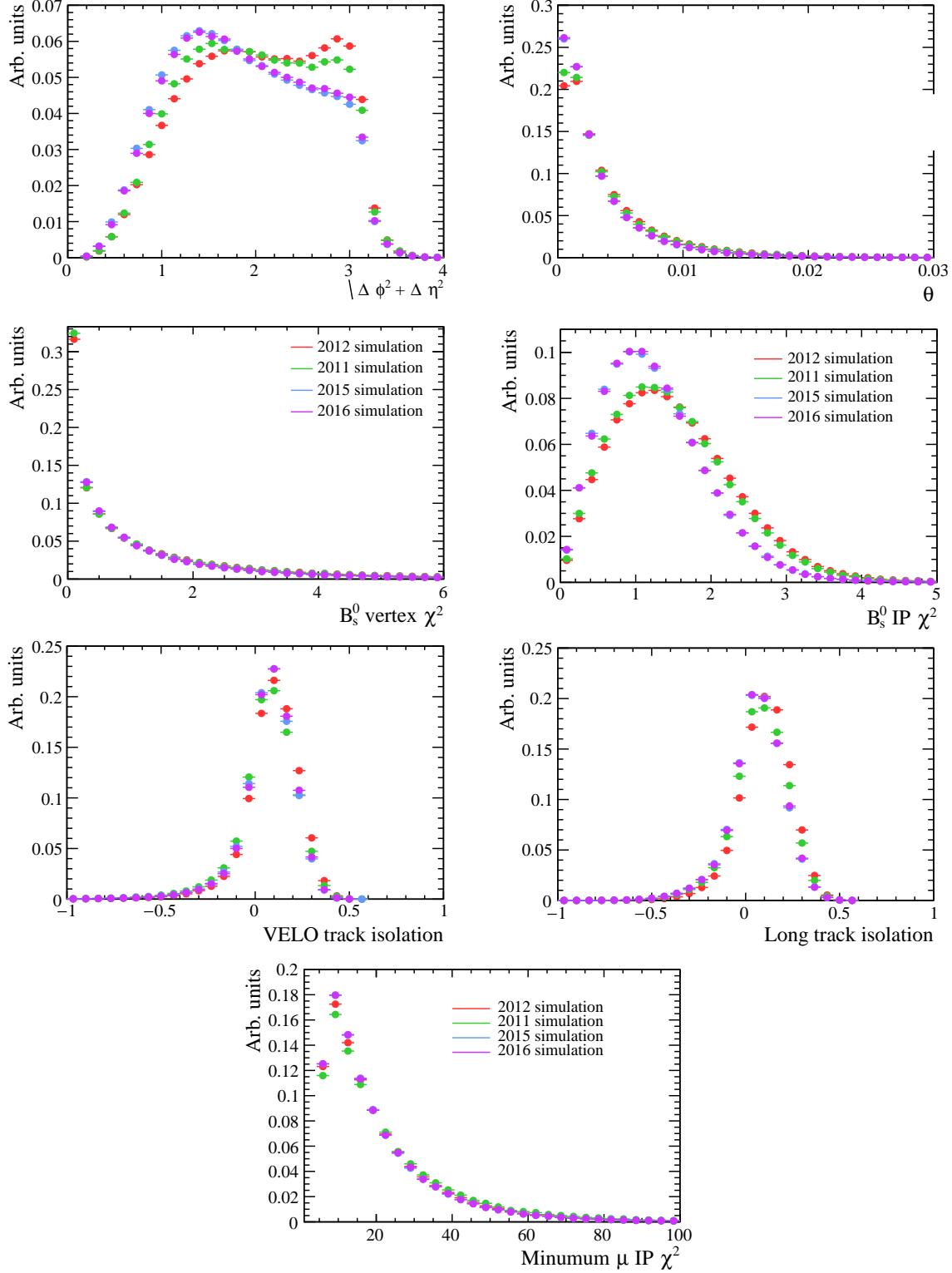
- [32] M. Pivk and F. R. Le Diberder, “SPlot: A Statistical tool to unfold data distributions,” *Nucl. Instrum. Meth.*, vol. A555, pp. 356–369, 2005.



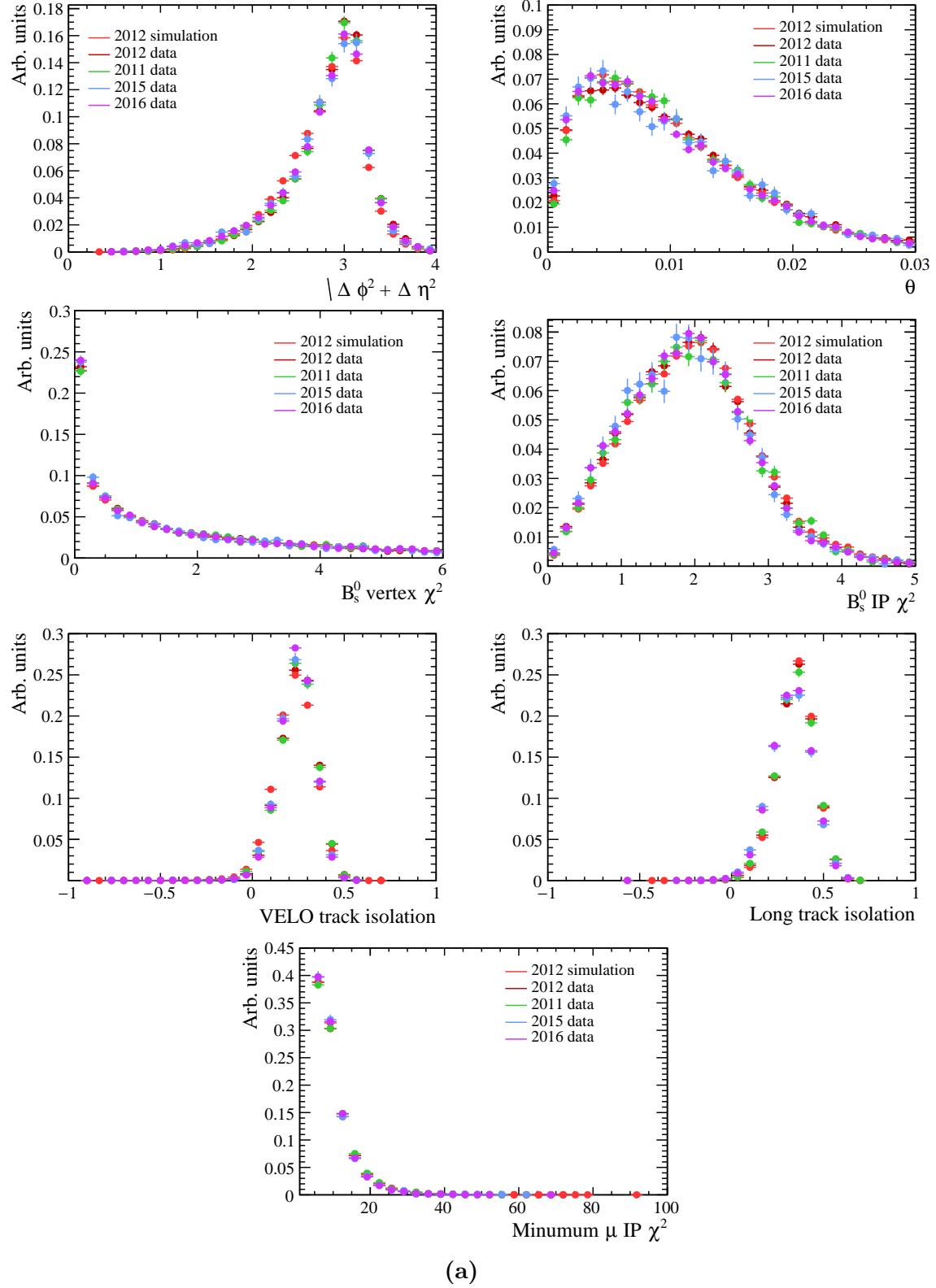
# Appendix A

## 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  $B_s^0 \rightarrow \mu^+ \mu^-$  decays for each year that have passed the selection cuts in Table 3.12. The background distributions are from  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays in 2011, 2012, 2015 and 2016 data with  $m_{\mu\mu} > 5447$  evcc and 2012 simulated  $b\bar{b} \rightarrow \mu^+ \mu^- X$  decays passing the selection cuts in Table 3.12.



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



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

