Chapter 3

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

The analysis described in Chapter X for the $B_s^0 \to \mu^+\mu^-$ effective lifetime requires $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays to be identified in the data sets recorded by the LHCb experiment. Although $B_s^0 \to \mu^+\mu^-$ 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 that can mimic a decay in the detector. The background processes are described in Section 3.1. To understand different aspects of the selection and analysis of decays, particle decays with a similar topology to are used. $B \to h^+h^-$ decays, where $h = K, \pi$, are used because they have large branching fractions and are well understood from previous LHCb analyses as well as a similar topology to $B_s^0 \to \mu^+\mu^-$ decays. The measurement of the $B_{(s)}^0 \to \mu^+\mu^-$ Branching Fractions, described in Chapter X, requires the use of $B^+ \to J/\psi K^+$ decays, as well as $B_{(s)}^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays, to be used as a normalisation channel. This Chapter describes the selection of $B_{(s)}^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ decays for the effective lifetime and Branching fraction analyses. The analyses share many of the same selection requirements. The selection occurs in several stages, and the development of the selection relies on simulated events which are detailed in Section 3.2. The first step to select decays is choosing what requirements to place on the trigger which is followed by a set of loose selection requirements to remove obvious background events. These two steps are described in Sections 3.3 and **??**. A tighter selection is applied to the output of the stripping as described in Section 3.5 and particle identification requirements are used in Section 3.4.2 to further reduced background events. Finally a multivariate classifier is described in Section **??** used as the final step in the selection to reduced the backgrounds to a level suitable for the analysis in

Chapter X to be preformed. Throughout this Chapter $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are selected in the same way.

The LHCb collaboration has published a number of papers studying the $B_s^0 \to \mu^+\mu^-$ decay, the selection described in this Chapter has been built up over a number of years by a range of different collaboration members. The studies detailed in sections X, X and Y was done for this thesis.

3.1 Backgrounds

The reconstruction process, outlined in Section X, produces numerous $B_s^0 \to \mu^+\mu^-$ candidates from pairs of muons created by pp collisions and recorded in the detector. Some candidates will have come from real $B_s^0 \to \mu^+\mu^-$ decays but there are other processes that occur during pp collisions that can create two muons that when combined look a lot like a $B_s^0 \to \mu^+\mu^-$ decay.

The main sources of background decays that mimic $B_s^0 \to \mu^+\mu^-$ decays are:

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

- Inelastic proton collisions that create two muons at the primary vertex. The muons for a good vertex and are combined to for a B_s^0 that decays instantaneously. This type of background is prompt combinatorial background.

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

- Random combinations of muons produced by separate semi-leptonic decays. The $B_s^0 \to \mu^+\mu^-$ 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 $B_s^0 \to \mu^+\mu^-$ backgrounds in this way are $B^0 \to \pi^-\mu+\nu_\mu$, $B_s^0 \to K^-\mu+\nu_\mu$, $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$ and $B_c^+ \to J/\psi\mu^+\nu_\mu$ where $J/\psi \to \mu^+\mu^-$.

- $B \to h^+h^-$ decays, where $h = K, \pi$, where both hadrons are mis-identified as muons. This usually occurs when the hadrons decay in flight. Similarly to mis-identified semi-leptonic decays the reconstructed B_s^0 candidate mass is lower than expected.

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

The selection aims to separate real $B_s^0 \to \mu^+\mu^-$ decays from the background to produce a set of $B_s^0 \to \mu^+\mu^-$ candidates with a high signal purity from which the B_s^0 effective lifetime can be measured. This is challenging because $B_s^0 \to \mu^+\mu^-$ decays are highly suppressed decays therefore reconstructed candidates are predominately made from background decays.

I could put a plot showing the mass plot from the previous analysis or I could make a plot something like Siim has to illustrate what i mean but that feels a bit like copying!

3.2 Simulated Particle Decays

Simulated particle decays, as described in Section X, are used to develop the selection and analysis of $B_s^0 \to \mu^+\mu^-$ decays. Large clean samples of simulated decays are needed to separate signal decays from background decays and to understand the impact of selection criteria on decays present in data. 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 generated the decays.

There exist multiple versions of the simulation because it is updated as understanding of the detector increases and to incorporate differences in data taking conditions, such as the trigger lines or center-of-mass energy, present in 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.

Simulated $b\bar{b} \to \mu^+\mu^- X$ decays are used to understand the combinatorial background of $B_s^0 \to \mu^+\mu^-$ decays, however producing a large enough sample of these decays to be useful is computational expensive and produces large output files to save generated decays. Therefore cuts are applied at the generation level for $b\bar{b} \to \mu^+\mu^- X$ decays to reduce the size of the samples that are saved and to speed production. The cuts, listed in Table 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 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 X.

3.3 Trigger

The trigger is the first step in the selection process and the structure of the trigger is described in Section X. Since $B_s^0 \to \mu^+\mu^-$ decays are very rare a broad set of trigger requirements is used in order to keep a high proportion of $B_s^0 \to \mu^+\mu^-$ decay at this step of the selection. Specific trigger lines are not used in the selection but rather the combined results of a large selection of trigger lines at each level of the trigger. The combinations of trigger lines used are the L0Global, Hlt1Phys and Hlt2Phys triggers. The L0Global trigger combines all trigger lines present in the L0 trigger, it selects an event provided at least one L0 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.

Different trigger decisions on these lines are used to select decays for the Branching Fraction and effective lifetime analyses. The Branching fraction selection imposed the loosest trigger requirements by requiring a event to pass the 'Dec' decision at each trigger level as illustrated in set 'A' of Table X. Trigger decisions are defined in Section X. The effective lifetime analysis has slightly more constrained trigger requirement, requiring an event passes either the 'TIS' or 'TOS' decision at each level of the trigger

as illustrated in set 'B' of Table X. The trigger choice for the effective lifetime is motivated by the determination of the acceptance function in Section X.

Events are required to be either TIS, triggered independent of signal or TOS, triggered on signal, on the trigger lines used at each level of the trigger. Slightly different trigger requirements are used to select $B \to h^+h^-$ decays used to develop and validate the effective lifetime analysis, the same broad trigger lines are used but the requirement on the output varies depending on the use of the $B \to h^+h^-$ events. The are two sets of trigger requirements, set 'A' and 'C', in Table 3.2 are used to select $B \to h^+h^-$ decays, it will be made clear in later sections where $B \to h^+h^-$ decays are used which trigger requirements are imposed.

There was a problem with the implementation of the Hlt2Phys Dec decision in 2016 simulated events.This only affect the selection of $B \to h^+h^-$ decays. In order to emulate this trigger a combination of Hlt2 lines that select $B \to h^+h^-$ events, listed in Table 3.3, is used instead of HLT2Phys when the Dec decision is required.

3.4 Cut Based Selection

The $B^0_s \to \mu^+\mu^-$ and $B \to h^+h^-$ 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 $B^0_s \to \mu^+\mu^-$ decays and the backgrounds that mimic them. 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 by exploiting differences between the decays and the backgrounds that mimic them. The selection of $B^0_s \to \mu^+\mu^-$ and $B \to h^+h^-$ decays for the $B^0_s \to \mu^+\mu^-$ effective lifetime measurement uses the same stripping lines as those in the $B^0_{(s)} \to \mu^+\mu^-$ Branching Fraction measurements. These lines were designed at the start of Run 1 by studying the efficiencies of different selection cuts from simulated events []. 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 $B^0_s \to \mu^+\mu^-$ events. These studies are detailed in Sections 3.4.0.1 and 3.4.0.2.

The offline selection cuts are applied to the output of the stripping selection(, only candidates that pass the stripping selection can be used to develop the analysis). The

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

Table 3.1 Simulated decays used for developing the selection and the analysis method listed according to the studies the decays are used in. Requirements are imposed on $b\bar{b} \to \mu^+\mu^- X$ decays as they decays are generated, the cuts are included alongside the decay type.

Trigger Line	Trigger decision
set A	
L0Global	Dec
Hlt1Phys	Dec
Hlt2Phys	Dec
set B	
L0Global	TIS or TOS
Hlt1Phys	TIS or TOS
Hlt2Phys	TIS or TOS
set C	
L0Global	TIS
Hlt1Phys	TIS
Hlt2Phys	TIS

Table 3.2 Trigger lines used to select $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays. Set 'A' is used to select decays for the Branching Fraction analysis. Set 'B' is used to select $B_s^0 \to \mu^+\mu^-$ decays for the effective lifetime analysis. Sets 'A' and 'C' are used to select $B \to h^+h^-$ decays used to develop the $B_s^0 \to \mu^+\mu^-$ effective lifetime analysis.

$B \to h^+h^-$ trigger lines
Hlt2Topo2BodyDecision Dec
Hlt2B2HH Lb2PPiDecision Dec
Hlt2B2HH Lb2PKDecision Dec
Hlt2B2HH B2PiPiDecision Dec
Hlt2B2HH B2PiKDecision Dec
Hlt2B2HH B2KKDecision Dec
Hlt2B2HH B2HHDecision Dec

Table 3.3 Trigger lines used to emulate the Hlt2Phys_Dec decision for $B \to h^+h^-$ data and simulated events.

Trigger decision				
Trigger Line	Branching Fraction analysis	$B_s^0 \to \mu^+\mu^-$ effective lifetime	$B \to h^+h^-$ effective lifetime	
			set A	set B
L0Global	Dec	TIS or TOS	TIS	Dec
Hlt1Phys	Dec	TIS or TOS	TIS	Dec
Hlt2Phys	Dec	TIS or TOS	TIS	Dec

Table 3.4 Trigger lines used to select $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays. Set 'A' is used to select decays for the Branching Fraction analysis. Set 'B' is used to select $B_s^0 \to \mu^+\mu^-$ decays for the effective lifetime analysis. Sets 'A' and 'C' are used to select $B \to h^+h^-$ decays used to develop the $B_s^0 \to \mu^+\mu^-$ effective lifetime analysis.

stripping selection imposes loose selection requirements onto $B_s^0 \to \mu^+\mu^-$ candidates so that as much information as possible is still available to develop the analysis and understand background events after the stripping selection. The offline selection further refined the data, removing background candidates. The full set of cuts applied in the stripping and offline selection to select $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays from Run 1 and Run 2 data are presented in Section X.

3.4.0.1 Published Run 1 Stripping Selection

The measurement of the $B_s^0 \to \mu^+\mu^-$ Branching Fraction, described in Chapter X, uses $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$ decays to normalise the number of observed $B_s^0 \to \mu^+\mu^-$ decays to the number created in proton-proton collisions. There are three stripping lines that select $B_{(s)}^0 \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$ and $B \to h^+h^-$ candidates, where $h = K, \pi$, 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. However, the selection of $B^+ \to J/\psi K^+$ decays must diverge from $B_s^0 \to \mu^+\mu^-$ due to additional particles in the final state of the decay. Any changes made to the $B_{(s)}^0 \to \mu^+\mu^-$ 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 [] to select $B_{(s)}^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ decays are listed in Table **??**.

The variables used in the stripping selection are:

- the reconstructed mass, M - the mass and momenta of the decay products of the B meson (or J/ψ) are combined to provide its reconstructed mass. Cuts on the mass remove candidates with a reconstructed mass far from the expected particle mass that are clearly to background. Loose mass requirements are made on for the $B_s^0 \to \mu^+\mu^-$ selection to allow for the offline study of semi-leptonic backgrounds that have a mass less than the B_s^0 mass when mis-identified as $B_s^0 \to \mu^+\mu^-$ decays;

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

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

- 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 $B_s^0 \to \mu^+\mu^-$, this distance would ideally be zero because they originate from the same vertex;

- decay time, τ - this 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 **??** 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 values due to the heavy B meson masses however an upper limit removes unphysical backgrounds

- momentum, p - an upper limit on the momentum of a particle removes unphysical backgrounds

- ghost probability - defined in Section **??** provides the probability of a tracking being composed on random hits in the detector.

- impact parameter (IP) χ^2 - this is the change in the fit for a primary vertex (PV) caused by removing one track in the fit. In a $B_s^0 \to \mu^+\mu^-$ decay, the B_s^0

is produced at the PV therefore it should have a small IP χ^2 value whereas the muons will be displace from the PV because of the relatively long lifetime of the B_s^0 and therefore will have a large IP χ^2;

- minimum muon impact parameter (IP) χ^2 - this is the IP χ^2 of the muons with respect to all PVs in the event, this is to remove prompt muons created at any PV in the event and therefore reduce the prompt combinatorial background.

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

3.4.0.2 Efficiency of $B_{(s)}^0 \to \mu^+\mu^-$ stripping line

The efficiencies of the stripping lines for selecting $B_{(s)}^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ decays are shown in Table 3.6. The efficiencies are evaluated using 2012 sim06 simulated events and a loose set of trigger requirements (set A from Table 3.2) have been imposed. This set of trigger requirements is used in the Branching Fraction measurement in Chapter X. Only cuts that are applied to select $B_{(s)}^0 \to \mu^+\mu^-$ decays are evaluated, the efficiencies for the isMuon and track $\chi^2/ndof$ are not included because decays that do not pass these requirements are not included in the samples of simulated events.

The selection efficiencies are similar across the different decays for the selection cuts that are shared with the $B_{(s)}^0 \to \mu^+\mu^-$ selection, and the total efficiency is around 70 % for all decays. The similarity of selection efficiencies across the different decays is further illustrated in Figure **??** which shows the ratio of the efficiencies $B_{(s)}^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ where each cut has been applied independently. With the exception of the IP χ^2 cuts on the daughter particles, the ratio of efficiencies is well within 2% of 1 for the range of cuts values shown. The ratio of the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ efficiencies for the daughter particle IP χ^2 markedly deviates from unity, showing that the IP χ^2 distribution of the muons and kaon are very different as seen previous in [**?**]. If the other selection cuts are applied to the simulated events before the daughter IP χ^2 requirement the ratio of $B_{(s)}^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ efficiencies is much closer to 1.

The efficiencies for most of the stripping cuts is $\sim 97\%$, however, the efficiencies of the cuts on the FD χ^2 of the $B_{(s)}^0$ or J/ψ and the daughter IP χ^2 of the muon or hadron

Particle	$B_s^0 \to \mu^+\mu^-$	$B \to h^+h^-$	$B^+ \to J/\psi K^+$
B_s^0 or B^+	|M - M_{PDG}| < 1200 MeV/c^2	|M - M_{PDG}| < 500 MeV/c^2	|M - M_{PDG}| < 500 MeV/c^2
	DIRA > 0	DIRA > 0	Vertex χ^2/ndof < 45
	FD χ^2 > 121	FD χ^2 > 121	IP χ^2 < 25
	IP χ^2 < 25	IP χ^2 < 25	
	Vertex χ^2/ndof < 9	Vertex χ^2/ndof < 9	
	DOCA < 0.3 mm	DOCA < 0.3 mm	
		τ < 13.248 ps	
		p_T > 500 MeV/c	
J/ψ			|M - M_{PDG}| < 60 MeV/c^2
			DIRA > 0
			FD χ^2 > 225
			Vertex χ^2/ndof < 9
			DOCA < 0.3 mm
Daughter μ or h	Track χ^2/ndof < 3	Track χ^2/ndof < 3	Track χ^2/ndof < 3
	isMuon = True		isMuon = True
	Minimum IP χ^2 > 9	Minimum IP χ^2 > 9	Minimum IP χ^2 > 25
		0.25 GeV/c < p_T < 40 GeV/c	
		p < 350 GeV/c	
		ghost probability < 0.3	
K^+			Track χ^2/ndof < 3
			p_T > 0.25 GeV/c
			Minimum IP χ^2 > 25

Table 3.5 Selection requirements applied during the stripping selection for Run 1 data used in the $B_{(s)}^0 \to \mu^+\mu^-$ Branching Fraction analysis [] to select $B_{(s)}^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ decays. M_{PDG} corresponds to the Particle Data Group[] mass of each particle.

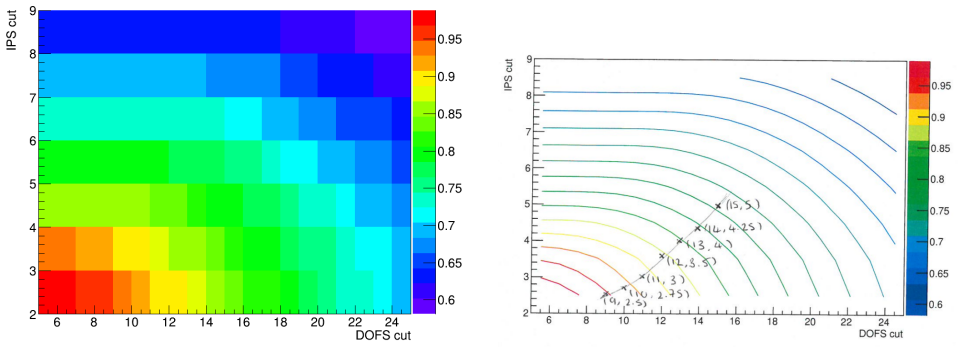
pair are lower at 83% and 80%, respectively. Therefore improvements to the stripping selection efficiencies could be achieved by altering these two selection requirements.

The set of events removed by each cut in a stripping selection is not independent. Therefore the effect of changing on cut on the total efficiency of a stripping selection must be considered. Figure 3.2 shows the total efficiency of the $B_s^0 \to \mu^+\mu^-$ stripping line on simulated $B_s^0 \to \mu^+\mu^-$ events that have passed the loose trigger requirement for a range of cut values for the FD χ^2 and daughter IP χ^2 requirements. As expected the lower the cut values are the more efficient the stripping line becomes. However 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. The curve on Figure 3.2 is used as a guide to study a set of FD χ^2 and daughter IP χ^2 cut values in more depth, taking in to account both the signal efficiency and the amount of data retained by the selection, the set of chosen cuts aims to keep both cuts as tight as possible for a certain efficiency. Any changed applied to the $B_{(s)}^0 \to \mu^+\mu^-$ stripping line must be propagated through into the stripping lines for $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ decays in order to keep the selection as similar as possible for across all the decays. Changes to the $B_{(s)}^0$ FD χ^2 in $B_s^0 \to \mu^+\mu^-$ correspond to a change in the $B_{(s)}^0$ FD χ^2 in $B \to h^+h^-$ and the J/ψ FD χ^2 for $B^+ \to J/\psi K^+$, similarly changes to the muon IP χ^2 in $B_s^0 \to \mu^+\mu^-$ correspond to changes in the hadron IP χ^2 in $B \to h^+h^-$ and the muon IP χ^2 in $B^+ \to J/\psi K^+$. Table **??** shows the total efficiencies of the $B_{(s)}^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ stripping lines along side the amount of data retained for the set of cuts illustrated in Figure 3.2b. The data retention is computed by applying the stripping selection to a sub-set of 2012 data, with the trigger requirements applied, to find the number of events that pass the stripping lines for each pair of FD χ^2 and daughter IP χ^2 cuts. 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 16% can be gained in the stripping selection efficiencies by using the loosest cuts in Table **??** however the loosest cuts increases the amount of data passing the $B_{(s)}^0 \to \mu^+\mu^-$ stripping selection by a factor of 7. 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 $B_{(s)}^0 \to \mu^+\mu^-$ selection efficiency by from 71 % to 82 % and the amount of data retained would be doubled. The increase of the data retained by the $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ lines is smaller and the

	Efficiency		
Requirement	$B_s^0 \to \mu^+\mu^-$	$B \to h^+h^-$	$B^+ \to J/\psi K^+$
B |M - M_{PDG}|	(100.00 ± 0.00)%	(±)%	(±)%
$B^0_{(s)}$ or J/ψ DIRA	(99.43 ± 0.01) %	(±)%	(±)%
$B^0_{(s)}$ or J/ψ FD χ^2	(83.89 ± 0.06) %	(±)%	(±)%
$B^0_{(s)}$ or J/ψ IP χ^2	(96.88 ± 0.03) %	(±)%	(±)%
$B^0_{(s)}$ or J/ψ vertex χ^2/ndof	(97.36 ± 0.03) %	(±)%	(±)%
$B^0_{(s)}$ or J/ψ DOCA	(99.86 ± 0.01) %	(±)%	(±)%
μ or h minimum IP χ^2	(80.47 ± 0.06) %	(±)%	(±)%
Efficiency after above cuts	(73.75 ± 0.07) %	(±)%	(±)%
Efficiency after all striping line cuts	(73.75 ± 0.07) %	(±)%	(±)%

Table 3.6 Stripping line efficiencies for $B_s^0 \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ 2012 simulated decays after broad trigger requirements. Selection cuts applied are listed in Table 3.5.



(a) (b)

Fig. 3.2 Efficiency figures.

3.4.1 Stripping selection and offline cuts

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

Additional selection requirements are applied after the stripping to remove specific backgrounds. A lower bound is placed on the B meson transverse momentum to remove pairs of muons originating from $pp \to p\mu\mu p$ decays and a J/ψ veto is used to remove backgrounds from $B_c^+ \to J/\psi\mu^+\nu_\mu$ decays. The semi-leptonic $B_c^+ \to J/\psi\mu^+\nu_\mu$ decays, where $J/\psi \to \mu^+\mu^-$, contribute to the background of $B^0_{(s)} \to \mu^+\mu^-$ decays when a muon from the J/ψ forms a good vertex with the muon from the B_c^+ decay. Due to the high mass of the B_c^+ this could place mis-reconstructed candidates within the B_s^0 mass window. A 'jpsi veto' can be used to remove background events from $B_c^+ \to J/\psi\mu^+\nu_\mu$ decays. The veto works by removing events where one muon from the $B^0_{(s)} \to \mu^+\mu^-$ candidate combined with any other oppositely charged muon in the event has $|m_{\mu\mu} - m\mathrm{J}/\psi| < 30\ MeV/\mathrm{c}^2$. The veto has a rejection power of X % on $B_c^+ \to J/\psi\mu^+\nu_\mu$ events that have passed $B^0_{(s)} \to \mu^+\mu^-$ selection cuts in Table **??** and rejects only % of $B^0_{(s)} \to \mu^+\mu^-$ signal events. The expected number of $B_c^+ \to J/\psi\mu^+\nu_\mu$ events after the full selection can be found in Section X.

The B meson mass range for both $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays is narrower than the range in the stripping selection in Section 3.4.0.1. $B_s^0 \to \mu^+\mu^-$ candidates are required to have a dimuon invariant mass greater than 5320 MeV/c^2. The motivation comes from mass fit studies that are detailed in Section X. The consequence of this cut is

$B^0_{(s)}$, J/ψ FD χ^2	μ,h$IP\chi^2$	Stripping line efficiency			Stripping line retention		
		$B_s^0 \to \mu^+\mu^-$	$B^0 \to K^+\pi^-$	$B^+ \to J/\psi K^+$	$B_s^0 \to \mu^+\mu^-$	$B^0 \to K^+\pi^-$	$B^+ \to J/\psi K^+$
15	5.00	XX %	XX %	XX %	XX	XX	XX
14	4.25	XX %	XX %	XX %	XX	XX	XX
13	4.00	XX %	XX %	XX %	XX	XX	XX
12	3.50	XX %	XX %	XX %	XX	XX	XX
11	3.00	XX %	XX %	XX %	XX	XX	XX
10	2.75	XX %	XX %	XX %	XX	XX	XX
9	2.50	XX %	XX %	XX %	XX	XX	XX

Table 3.7 Retention of data and stripping line efficiencies. Efficiencies are $B^0 \to K^+\pi^-$ but retention is $B \to h^+h^-$ because the stripping line selects all $B \to h^+h^-$ decays, h is K or π.

to remove $B^0 \to \mu^+\mu^-$ decays, $B_s^0 \to \mu^+\mu^-\gamma$ backgrounds and most backgrounds from mis-identified semi-leptonic and $B \to h^+h^-$ decays. The expect number of $B^0 \to \mu^+\mu^-$ and mis-identified decays after the full selection can be found in Section X. Similarly the $B \to h^+h^-$ 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 expect the track ghost probability and track χ^2/ndof. Slightly looser cuts are used for Run 2 to take advantage to changes in the reconstruction that were introduced for Run 2.

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

Table 3.8 Selection cuts applied to select $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays, where selection is different between Run 1 and Run 2 the Run 2 values are shown in parenthesis next to the Run 1 values.

3.4.2 Particle Identification

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

In the selection of $B^0_{(s)} \to \mu^+\mu^-$ decays PID variables are particularly useful to reduce the backgrounds coming from mis-identified semi-leptonic decays and $B \to h^+h^-$ decays and also help to reduce the number of combinatorial background events. The semi-leptonic decays that contribute to $B^0_s \to \mu^+\mu^-$ backgrounds are $B^0 \to \pi^-\mu{+}\nu_\mu$, $B^0_s \to K^-\mu{+}\nu_\mu$, $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$, $B^0 \to \pi^0\mu^+\mu^-$ and $B^+_c \to J/\psi\mu^+\nu_\mu$ where $J/\psi \to \mu^+\mu^-$.

The PID requirements to select $B^0_{(s)} \to \mu^+\mu^-$ decays are shown Table **??** alongside requirements to separate different $B \to h^+h^-$ 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 $B^0_s \to \mu^+\mu^-$ 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 $B^0_s \to \mu^+\mu^-$ 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 $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ decays is done via DLL variables. These are useful to separate $B \to h^+h^-$ decays where h is either a pion or kaon because the variables compare different particle hypotheses with the pion hypotheses. The selection requirements used are the same for each year of data taking.

3.5 Multivariate Classifiers

The output of the stripping selection still includes many background decays, further cuts shown in Table **??** reduce the background decays. Some selection cuts are designed to remove specific background decays and the selection for $B^0_s \to \mu^+\mu^-$ decays used in the Branching Fraction and effective lifetime analyses starts to diverge slightly.

The BDTS is a multivariate classifier that is designed to reduce the number of combinatorial background events. It is a Boosted Decision Tree (BDT) (see Section **??** for a detailed description) that is trained on $B^0_s \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^-X$ simulated decays that have passed the $B^0_{(s)} \to \mu^+\mu^-$ selection requirements in Table **??**. The BDTS uses variables similar to those in the stripping selection to classify events;

- impact parameter χ^2 of the

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

Table 3.9 Particle identification requirements to select $B_s^0 \to \mu^+\mu^-$ decays and to separate the $B \to h^+h^-$ decays $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+\pi^-$ from $B_s^0 \to K^+K^-$.

- vertex χ^2 of the $B^0_{(s)}$
- direction cosine of
- distance of closest approach of the muons
- minimum impact parameter χ^2 of the muons with respect to all primary vertices in the event
- impact parameter of the $B^0_{(s)}$, this is the distance of closest approach of the B to the primary vertex

The BDTS is applied to all candidates passing the $B^0_{(s)} \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ stripping lines, and candidates are required to have a BDTS value above 0.05. The chosen cut value has a efficiency of X % on $B_s^0 \to \mu^+\mu^-$ decays and reject X % of $b\bar{b} \to \mu^+\mu^- X$ decays.

The simulated $b\bar{b} \to \mu^+\mu^- X$ decays are used to train the multivariate classifier (Sect. **??**) used in the Branching Fraction and effective lifetime analyses and variables included within the classifier (Sect. **??**) as well as the BDTS. The simulated $b\bar{b} \to \mu^+\mu^- X$ decays had tighter cuts on the FD χ^2 of the $B^0_{(s)}$ and minimum IP χ^2 of the muons than those listed Table **??** applied when it was created. Events that did not pass those cuts are not available for use in the classifier training, therefore these must be applied to all data and simulated decays to ensure the most effective and reliable performance of the multivariate classifiers used in the analyses.

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* $\mathrm{D}^0 \to \mathrm{K}^0_\mathrm{S}\mathrm{h}^+\mathrm{h}'^-$ *decays at the LHCb experiment*. PhD thesis, Oxford U., Jul 2016. Presented 14 Sep 2016.

[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.