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 **??** 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, described in Section X, it the first step in the selection, it selects events that could contain an interesting physics process and these events are saved to be used in physics analyses. $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ candidates are reconstructed from events that have passed the trigger. For each candidate it is useful to know whether is was a particle 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 information from only the candidate was enough to cause a trigger line to select the event
- TIS, triggered independant of signal - a candidate is identified as TIS if another part of the event independant 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 and also events where a combination of information from the candidate and something else in the event was needed for a trigger line to select the event

$B_s^0 \to \mu^+\mu^-$ decays are very rare decays and therefore trigger requirements used to select these decays are chosen to keep a high efficiency for selecting $B_s^0 \to \mu^+\mu^-$ decays are 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 acheived for selecting $B_s^0 \to \mu^+\mu^-$ decays. The L0Global trigger combines all trigger lines present in the L0 trigger, it selects an event provided at least one L0 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 exclude.

Slightly different trigger decisions are used to select $B \to h^+h^-$ decays but the same trigger lines are used. To be useful as a validation channel the efficiency of the trigger requirements as a function of the decay time needs to be similar to the $B_s^0 \to \mu^+\mu^-$ triggers, this is acheived by requiring $B \to h^+h^-$ decays to be TIS at each level of the trigger.

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

3.4 Cut Based Selection

The $B_s^0 \to \mu^+\mu^-$ 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 $B_s^0 \to \mu^+\mu^-$ 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 $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ decays for the $B_s^0 \to \mu^+\mu^-$ effective lifetime measurement uses the same stripping lines as those in the $B_{(s)}^0 \to \mu^+\mu^-$ 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 $B_s^0 \to \mu^+\mu^-$ 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 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 $B^0_{(s)} \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$ 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 $B^0_{(s)} \to \mu^+\mu^-$ 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 $B^0_{(s)} \to \mu^+\mu^-$, $B \to h^+h^-$ 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. The cuts on muon transverse momentum, track χ^2 and isMuon (when used) are applied in the reconstruction and cannot be changed.

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 3.5. The efficiencies are evaluated using 2012 sim06 simulated events. No trigger requirements have been applied to that only the effect of the stripping selection on the efficiencies can be assessed. Only cuts that are applied to select $B_{(s)}^0 \to \mu^+\mu^-$ decays are evaluated, all candidates have had the isMuon, track $\chi^2/ndof$ and muon transverse mometum cuts already applied during the reconstruction and events that do not pass these cuts are not included because decays that do not pass these requirements are not included in the samples of simulated events. The efficiencies of the stripping selection to select $B^0 \to \mu^+\mu^-$ decays is extremely similar to the efficiency to select $B_s^0 \to \mu^+\mu^-$ decays therefore it is not considered in these studies.

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 3.2 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 $B^+ \to J/\psi K^+$

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

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

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	0.25 GeV/c < p_T < 40 GeV/c	0.25 GeV/c < p_T
		ghost probability < 0.3	
K^+			Track χ^2/ndof < 3
			p_T > 0.25 GeV/c
			Minimum IP χ^2 > 25
			0.25 GeV/c < p_T

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

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 $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ efficiencies is much closer to 1. Also the ratio of efficiencies for $B^0 \to K^+\pi^-$ is shown in Figure fig:ratio$_p lots Bd2KPi$.

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

The set of events removed by each cut in 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 3.3 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. It is important that any increase in $B_s^0 \to \mu^+\mu^-$ selection efficiency from the stripping is not removed when the trigger requirements are applied, Figure 3.4 shows that the trigger efficiencies are relatively flat across a large range of FD χ^2 and daughter IP χ^2 cut values therefore the efficiency gained in the stripping is effected uniformly by the trigger.

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 3.3 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 $B^0_{(s)} \to \mu^+\mu^-$ 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 $B^0_{(s)}$ FD χ^2 in $B_s^0 \to \mu^+\mu^-$ correspond to a change in the $B^0_{(s)}$ 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 $B^0_{(s)} \to \mu^+\mu^-$, $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ stripping lines along side the amount of data retained for the set of cuts . 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. The $B^0_{(s)} \to \mu^+\mu^-$ Branching Fraction analysis also uses stripping lines for $B^0_s \to J/\psi\phi$ and $B \to J/\psi K^*$ as well.

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 $B^0_{(s)} \to \mu^+\mu^-$ stripping selection by a factor of 7 and the $B \to h^+h^-$ stripping selection by a factor of 4.5. Table 3.4 shows the number of events passing the original stripping selection listed in Table **??** for the last published analysis, the $B \to h^+h^-$ stripping lines lets through the most candidates therefore the retention of this line is most important. The final set of cuts used in the stripping selection must be a compromise between the selection efficiency and the amount of data that passes the selection. The selection cuts of B^0_s FD $\chi^2 > 121$ and minimum muon IP $\chi^2 > 9$ would increase the $B^0_{(s)} \to \mu^+\mu^-$ selection efficiency by from 71 % to 82 % and the amount of data retained would be doubled. The increase of the data retained by the $B \to h^+h^-$ and $B^+ \to J/\psi K^+$ lines is smaller and the efficiencies are similar to the $B^0_{(s)} \to \mu^+\mu^-$ selection efficiencies. Therefore these cuts are applied in the stripping selection for this analysis.

Stripping Lines	Events	Retention/%
$B^0_{(s)} \to \mu^+\mu^-$	898880	0.0022
$B \to h^+h^-$	14502295	0.0831
$B^+ \to J/\psi K^+$	3344568	0.0087
$B^0_s \to J/\psi\phi$	456787	0.0011
$B \to J/\psi K^*$	12574956	0.018
Total	31779975	-

Table 3.4 The number of events passing stripping lines used for the $B^0_s \to \mu^+\mu^-$ analysis from the selection listed in Table **??**. Does not include correlation between lines, which is expected to be 42 % between $B^0_{(s)} \to \mu^+\mu^-$ and $B \to h^+h^-$ lines.

3.4.1 Stripping selection and offline cuts

The complete list of selection cuts applied in the cut based selection to select $B^0_s \to \mu^+\mu^-$ 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

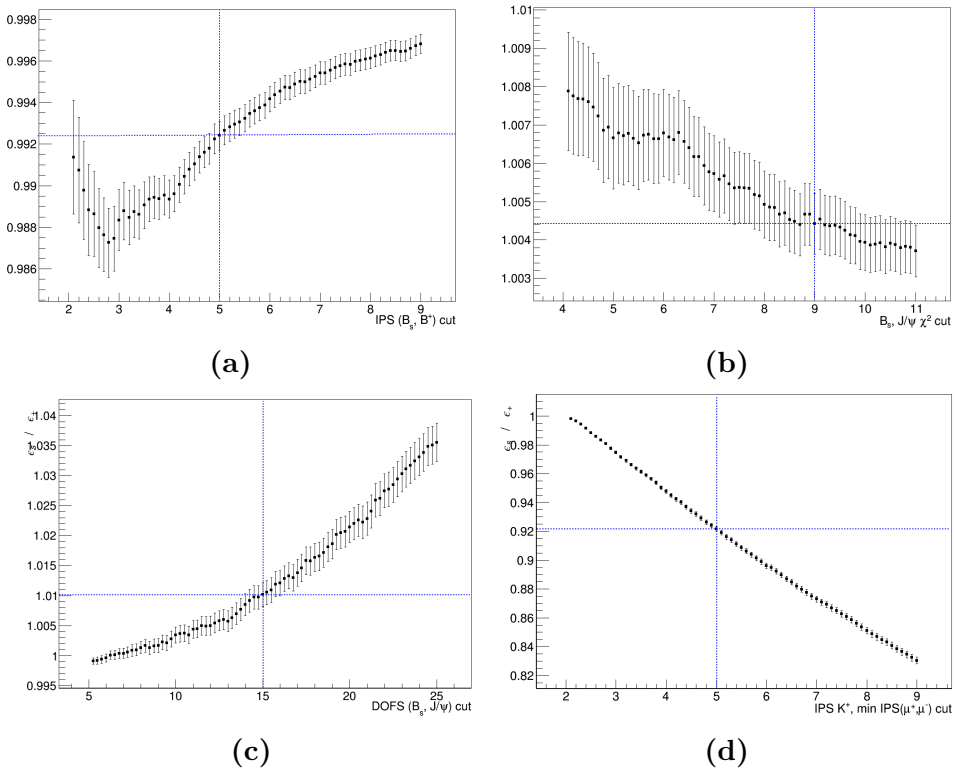
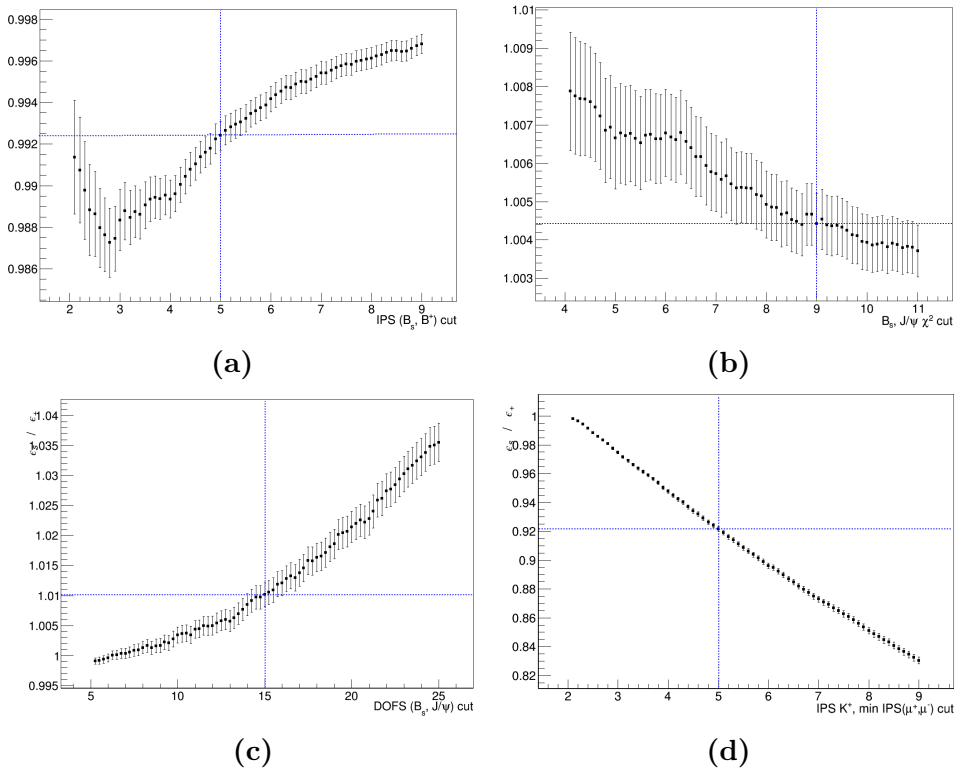


Fig. 3.1 The ratio of $B_s^0 \to \mu^+\mu^-$ to $B^0 \to K^+\pi^-$ stripping efficiencies on MC events when each cut has been applied independe ntly of all other cuts. The current cut values are marked by the blue lines.



(a) **(b)** **(c)** **(d)**

Fig. 3.2 The ratio of $B^0_{(s)} \to \mu^+\mu^-$ to $B^+ \to J/\psi K^+$ stripping efficiencies on MC events when each cut has been applied independently of all other cuts. The current cut values are marked by the blue lines.

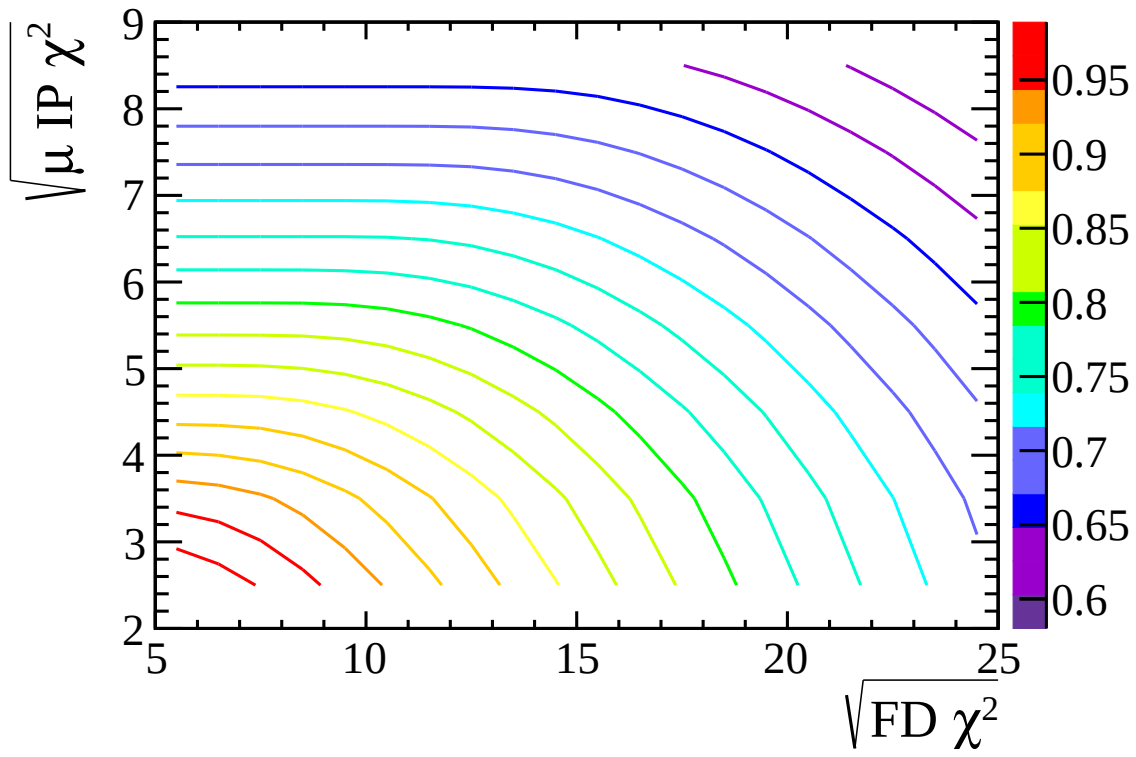
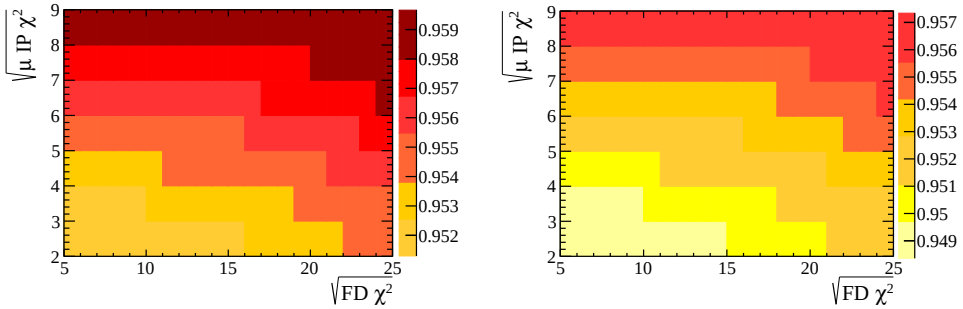


Fig. 3.3 Efficiency figures.



(a) **(b)**

Fig. 3.4 Trigger efficiency figures.

	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 3.5 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 3.3.

		Stripping line efficiency			Stripping line retention		
$B^0_{(s)}$, J/ψ FD χ^2	μ,h$IP\chi^2$	$B^0_s \to \mu^+\mu^-$	$B^0 \to K^+\pi^-$	$B^+ \to J/\psi K^+$	$B^0_s \to \mu^+\mu^-$	$B^0 \to K^+\pi^-$	$B^+ \to J/\psi K^+$
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 3.6 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.

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^0_s mass window. A 'J/ψ veto' can be used to remove background events from $B^+_c \to J/\psi\mu^+\nu_\mu$ decays. The veto works by removing events where one muon from the $B^0_{(s)} \to \mu^+\mu^-$ candidate combined with any other oppositely charged muon in the event has $|m_{\mu\mu} - m\mathrm{J}/\psi| < 30\ MeV/\mathrm{c}^2$. 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 $B^0_s \to \mu^+\mu^-$ and $B \to h^+h^-$ decays is narrower than the range in the stripping selection in Section 3.4.0.1. $B^0_s \to \mu^+\mu^-$ 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 to remove $B^0 \to \mu^+\mu^-$ decays, $B^0_s \to \mu^+\mu^-\gamma$ 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.

3.4.2 Particle Identification

Particle identification (PID) variables are used to refine the selection of $B^0_{(s)} \to \mu^+\mu^-$ 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

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)
	$\lvert m_{\mu\mu} - mJ/\psi \rvert$<30 MeV/$c^2$	$\lvert m_{\mu\mu} - mJ/\psi \rvert$<30 MeV/$c^2$
	isMuon = True	-

Table 3.7 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.

semi-leptonic decays that contribute to $B_s^0 \to \mu^+\mu^-$ backgrounds 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 $J/\psi \to \mu^+\mu^-$.

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

A multivariate classifier is an algorithm that learns differences between signal and background candidates in the following manner, firstly 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 events. Then the classifier uses the distributions of the input variables with its knowledge of which events are signal and background to learn the difference between the two types of events. Finally the algorithm is then be applied to a data set containing an unknown mixture of signal and background events and distinguish between them. For each event the algorithm produces a number, typically between -1 and +1, where high numbers indicate signal-like candidates and low numbers

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^+	$\text{DLL}_{K\pi} > 10$
	$\pi-$	$\text{DLL}_{K\pi} <$ -10
$B_s^0 \to K^+K^-$	K^+ and K^-	$\text{DLL}_{K\pi} > 10$

Table 3.8 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 $B_s^0 \to K^+K^-$.

indicating background-like candidates. A cut is placed on the output of the classifier to remove background events with a classifier response less than a particular value and the remaining data set has a higher purity for signal events.

Two multivariate classifiers are used to select $B_s^0 \to \mu^+\mu^-$ decays. Both classifiers are a type called Boosted Decision Trees (BDT) that are described in Section 3.5.1. A range of different classifiers were investigated but BDTs preformed the best at separating signal and background candidates.

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

Both classifiers were developed for the measurement of the $B_{(s)}^0 \to \mu^+\mu^-$ Branching Fractions. In the selection $B_s^0 \to \mu^+\mu^-$ decays for this analysis the second classifier is used to classify candidates into 8 bins containing increasing proportions of signal candidates, not 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 $B_s^0 \to \mu^+\mu^-$ effective lifetime however it was found that the two classifiers developed for the Branching Fraction measurement performed best.

3.5.1 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 candidate is know to be signal or background and a set of variables describing these events. 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 on 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 then uses the knowledge of which candidates are signal or background to assign a value of +1 or -1 to every candidate. A candidate 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 candidates are background. The

Fig. 3.5 Illustration of a decision tree.

final decisions made by the tree are not prefect, some signal (background) events will be mis-classified as background events and given the value of -1 (+1). The decision making process of a decision tree is illustrated in Figure 3.5.

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 dependance of the final decisions on statistical fluctuations. A BDT starts with one decision tree and then assigns weights to events in the sample 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-like candidates and low numbers indicating background-like candidates.

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

$$w = \frac{1-f}{f}, \text{ where } f = \frac{\text{Total misclassified events}}{\text{Total events}}. \tag{3.1}$$

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 β where $w \to w^{\beta}$, this can be specified in the training of the decision tree, 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);, the 'speed' at which the boosting occurs 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 but together they can be used to prevent overtraining of the BDT. Overtraining is when the BDT is very sensitive to the statistical fluctuations in the training sample and 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.

3.5.2 The BDTS

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

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

The signal and background samples used to train the BDTS are simulated $B_s^0 \to \mu^+\mu^-$ decays and $B_s^0 \to \mu^+\mu^-$ candidates in 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 **??**. 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 $B_{(s)}^0 \to \mu^+\mu^-$ and $B \to h^+h^-$ 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. This is illustrated in Figure 3.7.

3.5.3 Global BDT

The global BDT is the final step in selecting $B_s^0 \to \mu^+\mu^-$ decays and it is ver effective at seperating them from long lived combinatorial background decays. The discrimintaing power schieved by the global BDT is mostly dependant on isolation variables. Isolation variables, or just isolations, provide a measure of how isolated or far away each muon from a $B_s^0 \to \mu^+\mu^-$ candidate is from other track in the event. The tracks of the

(a) **(b)**

Fig. 3.6 Signal MC and and background (from somewhere) showing the reponses for the BDTS and how is it flattened.

Fig. 3.7 Something to illustrate how good the BDTS is at working. I think prehaps the BsMuMu stripping line and the B2hh stripping line with all selection applied, but a different mass range to make it clear, and then with and without the BDTS cut. Prehaps no PID requirements to seperate b2hh decays as well.

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

Table 3.9 Selection cuts applied to select candidates for signal and background samples used to train the BDTS.

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

Table 3.10 Training parameters used to specify the train of the BDTS.

(a) **(b)**

Fig. 3.8 Long track and Velo track isolation distributions of signal and background passing cuts in Table X.

muons from a rel $B_s^0 \to \mu^+\mu^-$ decays will be, in general, far from other trakcs in the event because the $B_s^0 \to \mu^+\mu^-$ decays tree contains no other tracks apart from the muons. However 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. Various different definitions of isolations have been used acorss different experiments from D0 and CDF to ATLAS, CMS and LHCb. These variables are very useful in the selection of very rare decays like $B_s^0 \to \mu^+\mu^-$ because they enable background to bve 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 muons in $B_s^0 \to \mu^+\mu^-$ candidates and the other compares VELO tracks in the event to the muons. The definition of the track types can be found in Section X. The isolation variables are built from the output of BDTs. For each type of track a BDT is trained on simulated $B_s^0 \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ decays using a set of input variables that describe track and vertex properties. The BDT compares the μ^+ from a $B_s^0 \to \mu^+\mu^-$ candidate with all other tracks in the event, excluding the track of the μ^-, and gives an output, iso_{μ^+}*(track)*, for each possible μ^+ - track pairing. The sample process is repeated for the μ^-. The BDT is designed to produce high output values for muons from $b\bar{b} \to \mu^+\mu^- X$ decays and a low value for muons from $B_s^0 \to \mu^+\mu^-$ decays. The isolation variable of a $B_s^0 \to \mu^+\mu^-$ candidates is then composed of the sum of the highest values of iso_{μ^+}*(track)* and iso_{μ^-}*(track)* for any tracks in the event. The input cariables used to the Long track and VELO track isolations are listes in Appendix **??**. The seperation power of these isolations are shown in Figure 3.8. Full details of the isolation development can be found in **??**.

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

- Long tracks isolation
- VELO tracks 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
- the smallest IP χ^2 with respect to the primary vertex of the $B_s^0 \to \mu^+\mu^-$ of the muons
- vertex χ^2 of the B_s^0
- IP χ^2 of the B_s^0 with respect to the primary vertex
- angle between the momentum vector of the B_s^0 and the vector connecting the production and decay vertices of the B_s^0

A comparison of the signal and background distributions of the input variables are shown in Figures 3.8 and 3.9. These variables were chosen by training a BDT beginign with the most discrimintaing variable, the Long track isolation, and adding variables to determine which improved the performance to the classifer. 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 chosing those that gave the best performance.

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

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

Simulated $B_s^0 \to \mu^+\mu^-$ and $b\bar{b} \to \mu^+\mu^- X$ decays are used to provide large signal and background training samples for the global BDT. In data $B_s^0 \to \mu^+\mu^-$ candidates in the mass range 5431 to 6550 MeV/c^2 consist almost entirely of $b\bar{b} \to \mu^+\mu^- X$ decays, however the number of candidates in this mass range is too small to be a useful as sample of background candidates to train a BDT with comparable preformance to

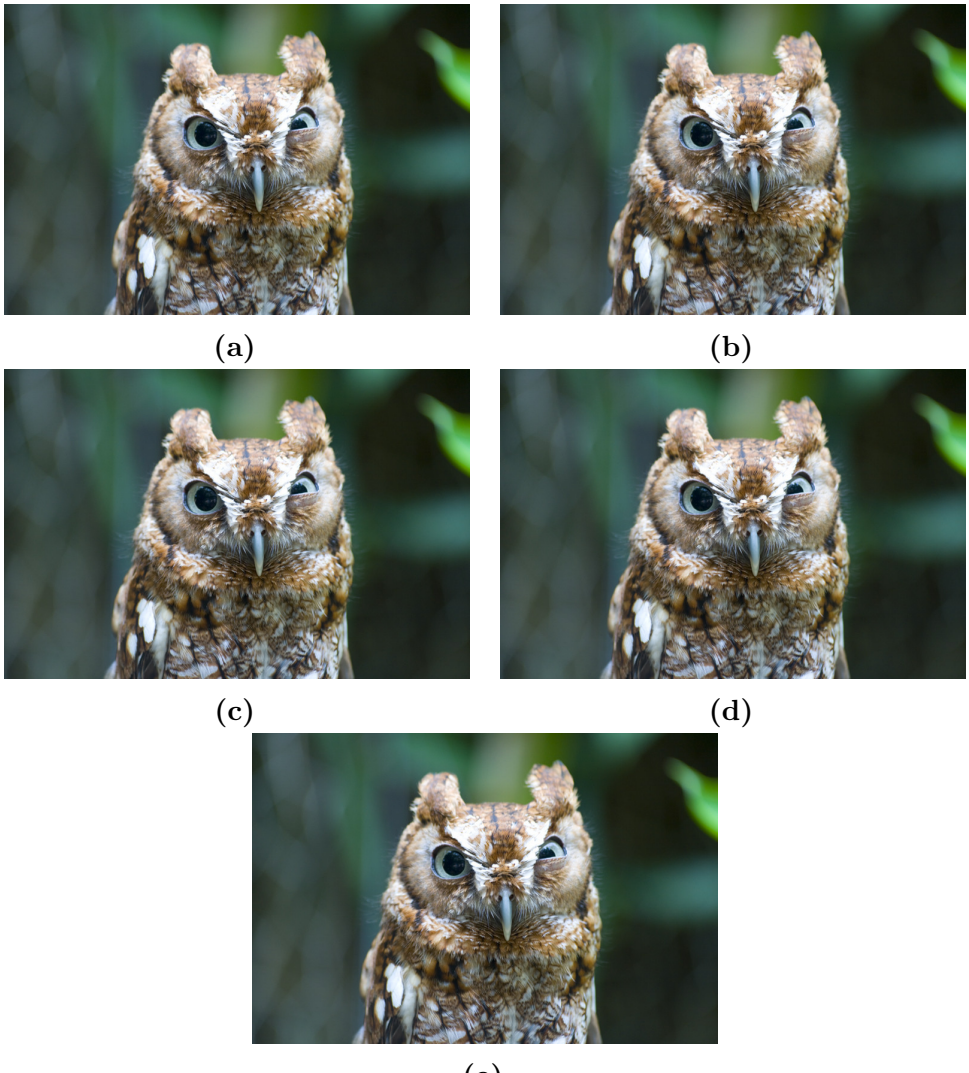


Fig. 3.9 Distributions of variables that go into the BDT after some set of selection cuts. Do I want another one that is for the BDTS?

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

Selection applied to BDT training samples.	
B_s^0 FD $\chi^2 > 225$	$\mu^\pm$ $p_T > 500$ MeV/c
B_s^0 IP $\chi^2 < 25$	$\mu^\pm$ track χ^2/ndof < 3
B_s^0 Vertex χ^2/ndof < 9	$\mu^\pm$ minimum IP $\chi^2 > 25$
B_s^0 DOCA < 0.3 mm	$\mu^\pm$ 0.25 GeV/c $< p_T <$ 40 GeV/c
B_s^0 $\tau < 13.248$ ps	$\mu^\pm$ $p < 500$ GeV/c
B_s^0 $p_T > 500$ MeV/c	
isMuon == 1	B_Hlt1Phys_Dec ==1
B_L0Global_Dec ==1	B_Hlt2Phys_Dec ==1
$4900 < \mathrm{M}_{\mu^+\mu^-} < 6000$ MeV/c^2	

Table 3.12 Selection cuts applied to select candidates for signal and background samples used to train the BDT.

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. The flattening is important for the measurement of the $B_{(s)}^0 \to \mu^+\mu^-$ Branching Fractions where a simultaneous fit is applied to the dimuon

(a) **(b)**

Fig. 3.10 Signal MC and and background (from somewhere) showing the reponses for the BDT and how is it flattened.

invariant mass in bins of BDT, flattening the BDT output enavke bins containing equal proportions of signal decays to be easily created.

3.5.4 Global BDT cut optimisation

A cut is place on the output of the global BDT to select $B_s^0 \to \mu^+\mu^-$ decays to measure the $B_s^0 \to \mu^+\mu^-$ effective lifetime, $\tau_{\mu\mu}$. The motivation comes for the fit choice of how to extract the lifetime from the data set detailed in Section X. The cuts value is chosen to give the smallest expected statistical uncertainty on the measurement of $\tau_{\mu\mu}$ or $\tau_{\mu\mu}^{-1}$. This is done by using toy experiments for the expected number of $B_s^0 \to \mu^+\mu^-$ events and combinatorial background for different cuts on the global BDT output.

The fit proceedure to extract the $\tau_{\mu\mu}$ from the data is described in depth in Chapter X. Briefly the strategy is firsly to perform and unbinned maxiumum likelihood fit to the dimuon invaraiant mass spectrum, where components for $B_s^0 \to \mu^+\mu^-$ and the combinatorial background are included in the fit. The mass fit is used to compute sWeights using the sPlot method []. Then a maximum likelihood fit is prefmoned to the sWeighted decay time distribution to extract $\tau_{\mu\mu}$ and $\tau_{\mu\mu}^{-1}$.

The toy experiments used to chose the cut on the global BDT have the following structure, firstly the number of expected $B_s^0 \to \mu^+\mu^-$ and combinatorial background events are generated using the expected mass and decay time probability density functions. Then the mass fit is performed to extract the sWeights so that the decay time fit to sWeighted decays can be done to extract $\tau_{\mu\mu}$ and $\tau_{\mu\mu}^{-1}$. For each possible cut on the global BDT 10,000 toy experiments were performed to give the median expected uncertainty on $\tau_{\mu\mu}$ and $\tau_{\mu\mu}^{-1}$ for each BDT cut value. Full details of the toy experiment set up and the probability density functions used are given in Appendix X.

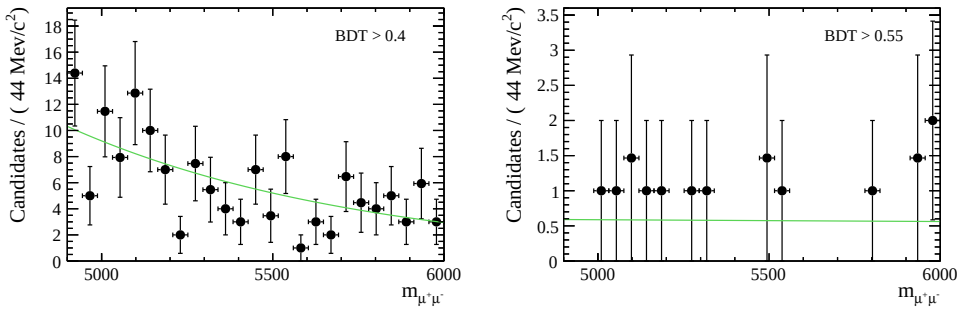
The number of expected $B_s^0 \to \mu^+\mu^-$ and combinatorial background events for diferent BDT cut values is derived from the expected number of decays passing in the all the selection cuts and BDT > 0.55 for Run 1 and Run 2 data as described in Section X. These predictions assume the SM branching fraction for $B_s^0 \to \mu^+\mu^-$ and are given in Table **??**. Since the output of teh BDT is flattened the number of $B_s^0 \to \mu^+\mu^-$ decays is evenly distributed across the BDT range, therefore the expected number of $B_s^0 \to \mu^+\mu^-$ decays is straight forward to calculated for each BDT cut value. The numerb of combinatirual background decays expected after each BDT is is computed from simulated $b\bar{b} \to \mu^+\mu^- X$ decays using the ratio

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

where $\epsilon(BDT > X)$ is the efficiency of the cuts BDT > X. The full selection requirement are applied to the simulated decays before taking the effieicny. The ratios for the different cuts values are shown in Table 3.13. 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.

The mass distribution of the combinatorial background is a decaying exponential, it was observed from the simualted $b\bar{b} \to \mu^+\mu^- X$ decays that the slope of the mass dustrubution changed with the BDT cut value as illustrated in Figure 3.11. The change in slope is accounted for when generating events for the toy experiment by changing the slope parameter (λ) for each BDT cut. Table 3.14 shows the slope of the mass distibtuion for different BDT cuts values evealutaed from $b\bar{b} \to \mu^+\mu^- X$ simulated decays.

The results from 10,000 toy experiments for BDT cut values every 0.05 in the range 0.4 - 0.65 are shown in Table 3.15 along with the expected number of $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays for each BDT cut value. The median uncertainty of the fit for $\tau_{\mu\mu}$ and $\tau_{\mu\mu}^{-1}$ are given along with the signal significance ($\mathcal{S} = S/\sqrt{S+B}$) for each BDT cut. The highest signal significance and lowest expected uncertainties occur for a BDT cut of 0.55, therefore this cut value is used to select $B_s^0 \to \mu^+\mu^-$ decays.



(a) **(b)**

Fig. 3.11 Mass distribution of decays after BDT cuts of 0.4 and 0.55.

BDT1flat cut	R_ϵ
0.40	8.69
0.45	3.91
0.50	1.91
0.55	1.00
0.60	0.55
0.65	0.32

Table 3.13 The ratio of efficiencies of cuts on the global BDT to select $b\bar{b} \to \mu^+\mu^- X$ decays relative to a cut of 0.55 on the global BDT.

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

Table 3.14 The slope of the combinatorial background mass distribution for different cut value on the global BDT evaluated from $b\bar{b} \to \mu^+\mu^- X$ simulated decays.

BDT cut	$B_s^0 \to \mu^+\mu^-$	Combinatorial background	$S/\sqrt{(S+B)}$ /ps	σ () /ps^{-1}	σ ()
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 3.15 The number of expected $B_s^0 \to \mu^+\mu^-$ and combinatorial background decays and 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* $\mathrm{D}^0 \to \mathrm{K}^0_\mathrm{S}\mathrm{h}^+\mathrm{h}'^-$ *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.