Chapter 5

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

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

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

5.1 Analysis Strategy

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

selection. The $B^0_{(s)} \to \mu^+\mu^-$ branching fractions can be given by

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

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

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

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

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

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

where *norm* inidicates the normalisation channel. The normalisation parameters can be combined into one normalisation parameter $\alpha_{d(s)}$ for each of the B^0_s and B^0 decays. The normalisation proceedure removes the uncertainty from $\sigma_{b\bar{b}}$.

Therefore the number of observed $B^0_{(s)} \to \mu^+\mu^-$ decays and the normalisation paramters, $\alpha_{d(s)}$, need to be evaluated to measure the branching fractions. The selection described in Chapter **??** allows $B^0_{(s)} \to \mu^+\mu^-$ candidates to be classified by their dimuon invariant mass and global BDT output. As illustrated in Figure **??**. A simultaneous unbinned maximum likelihood fit is performed to the dimuon invariant mass distribution in 4 BDT bins to measure the observed number of $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ decays. The Run 1 and Run 2 data are kept seperate and the fit is applied simulatanrously to both data sets. To measure the number of $B^0_{(s)} \to \mu^+\mu^-$ decays the fit requires knowledge of the mass shapes and the fraction of $B^0_{(s)} \to \mu^+\mu^-$ decays in

each BDT bin and the number of background decays and their mass shapes in each binl. The mass shapes and fraction of $B^0_{(s)} \to \mu^+\mu^-$ decay in each BDT bin are described by probability density functions (*pdfs*). The evalutaion of the $B^0_{(s)} \to \mu^+\mu^-$ mass *pdfs* and the fraction of decays in each BDT bin are described in Section 5.2. The expected number of background decays and their mass *pdfs* in each BDT bin are described in Section 5.3.

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

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

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

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

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

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

5.2.1 Mass *pdfs*

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

The parameters are evaluated using different methods:

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

All parameters are evaluted seperatly for the B^0 and B^0_s for each year of data taking. The resulting parameter vales are in good agreement across each year in the Run 1 and Run 2 data sets. The weighted average of the yearly parameters therefore is used to produced the mass *pdfs* for Run 1 and Run 2 and are given in Table **??**.

5.2.2 BDT *pdfs*

The global BDT distribution for $B^0_{(s)} \to \mu^+\mu^-$ decays is expected to be uniform between 0 and 1 as designed by the flattening proceedure descriebd in Section **??**. The fraction of $B^0_{(s)} \to \mu^+\mu^-$ decays in a BDT bin should simply be proportional to the bin with. However the global BDT was trained and flattened using simulated decays therefore to avoid differences between simulated decays and data affecting the expected fraction of $B^0_{(s)} \to \mu^+\mu^-$ decays in each BDT bin, the BDT *pdf* is evaluated from data. This process is know as the BDT callibration. The global BDT is designed to used only kinematic and goemetric information to classify candidates and include

no PID information. Therefore the BDT distributions of $B \to h^+h^-$ decays, that are topologically simimlar to $B^0_{(s)} \to \mu^+\mu^-$ decays, will be the same as $B^0_{(s)} \to \mu^+\mu^-$ decays. $B^0 \to K^+\pi^-$ decays are used to callibrate the BDT response because it is the most abundant $B \to h^+h^-$ decay.

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

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

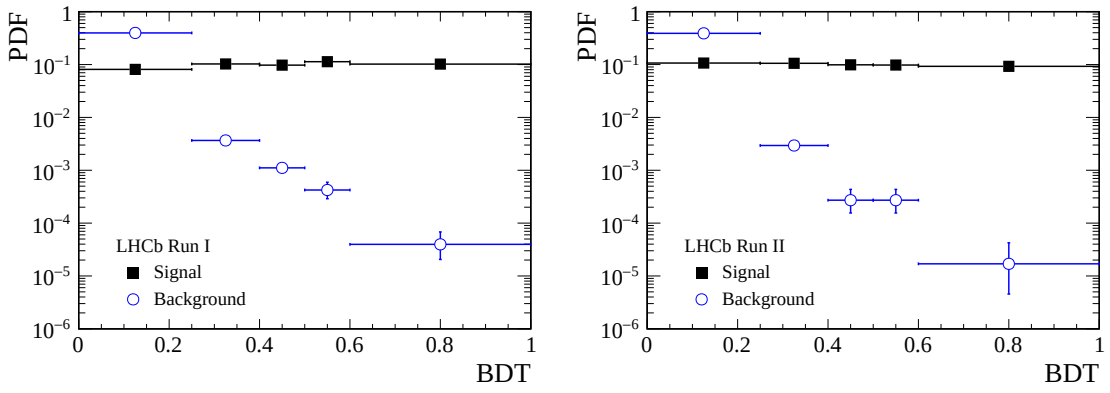

Fig. 5.1 $B^0_{(s)} \to \mu^+\mu^-$ BDT *pdfs* (black squares) for Run 1 and Run 2 data callibrated on $B^0 \to K^+\pi^-$ decay and the combinatorail background decays (blue circles) for $B^0_{(s)} \to \mu^+\mu^-$ candidates in data with a dimuon mass above 5477 MeV/c^2.

5.2.3 Decay time dependance of the $B^0_s \to \mu^+\mu^-$ BDT *pdf*

The output of the gloabl BDT for $B^0_{(s)} \to \mu^+\mu^-$ is correlated with the $B^0_{(s)} \to \mu^+\mu^-$ decay time due the the choice of input variables used in the BDT as listed in Section **??**. This correlation will lead of slightly incorrect estimations fo the $B^0_s \to \mu^+\mu^-$ BDT

pdf. In the SM the $B^0_s \to \mu^+\mu^-$ effective lifetime, $\tau_{\mu\mu}$, is equal to the lifetime of the heavy B^0_s mass eigenstate, τ_H, however in reality $\tau_{\mu\mu}$ could be somewhere in between the lifetimes of the heavy and light mass eignenstates. As described in Chapter (*the Theory Chapter*) the $B^0_s \to \mu^+\mu^-$ effective lifetime is related to the parameter $A_{\Delta\Gamma}$, where $A_{\Delta\Gamma} = +1$ for $\tau_{\mu\mu} = \tau_H$ and $A_{\Delta\Gamma}$ = -1 for $\tau_{\mu\mu} = \tau_L$, where τ_L is the lifetime of the light $B^0_s \to \mu^+\mu^-$ mass eigenstate.

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

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

5.3 Background mass *pdfs* and expected yields

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

- $B \to h^+h^-$ decays (where h = K, π) when both hadrons are mis-identified as muons because the hadrons decay during their flight through the detetor after leaving the VELO. This background falls within the B^0 mass window but not the B^0_s mass window[1] due to the missing energy from the undetected neutrino.

- semi-leptonic decays where one hadron is mis-identified as a muon that include;

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

  - $B^0 \to \pi^-\mu{+}\nu_\mu$ and $B_s^0 \to K^-\mu{+}\nu_\mu$ decays where the final state hadrons are mis-identified as muons. The mass of these backgrounds falls below the B^0 mass window in the left mass sideband
  - $\Lambda_b^0 \to p\mu^-\nu_\mu$ decays when the proton is mis-identified as a muon. The large mass of the Λ_d meases that this background pollutes the B_s^0 and B^0 mass windows and below these windows

- semi-leptonic decays where muon muons in the decay form a good vertex that include;
  - $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ decays where the pion is not detected. The missing hadron means that these backgrounds fall well below the B^0 mass window.
  - $B_c^+ \to J/\psi\mu^+\nu_\mu$ decays where $J/\psi \to \mu^+\mu^-$. The large mass of the B_c^+ causes this background of cover the full mass range 4900 - 6000 MeV/c^2

- combinatorial background formed by the random combination of any two muons in the event, this background is distributed accross the full mass range

To measure the number of $B_s^0 \to \mu^+\mu^-$ decays these backgrounds must be modelled in the invariant mass fit in each BDT bin, therefore the mass *pdfs* and fraction of events present in each BDT bin must be determined. Backgrounds that fall below the B^0 and B_s^0 mass windows still need to be accruatly modelled so that the number of combinatorial background decays that cover the full mass range can be accuratley described/measured. The proceedure is slightly different for $B \to h^+h^-$ decays compared to semi-leptonic decays.

5.3.1 $B \to h^+h^-$ mass and BDT *pdfs*

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

The number of mis-identified $B \to h^+h^-$ decays in each BDT bin is computed from $B^0 \to K^+\pi^-$ decays in data passing the same selection as the signal BDT callibration sample using the equation

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

The probability of a $B \to h^+h^-$ decay passing the $B^0_{(s)} \to \mu^+\mu^-$ particle identification requirements is given by and is the number of $B \to h^+h^-$ decays in data determined from $B^0 \to K^+\pi^-$ decays and scaled for the relative fraction of the total $B \to h^+h^-$ decays. The efficiency of the $B^0_{(s)} \to \mu^+\mu^-$ triggers is given by and is the efficiency TIS triggers used to select $B^0 \to K^+\pi^-$ decays. The trigger and particle identification efficiencies are computed per BDT bin and the BDT pdf before the PID efficiencies is assumed to be the same as the signal *pdf*.

5.3.2 Semi-leptonic mass and BDT *pdfs*

The mass *pdfs* of semi-leptonic backgrounds vary across the BDT range therefore these *pdfs* are evalutaed using simulated decays seperated into each BDT bin. An Argus function [] is used to describe the mass distributions. The shapes of $B^0 \to \pi^-\mu{+}\nu_\mu$ and $B^0_s \to K^-\mu{+}\nu_\mu$ are extremely similar and therefore these backgrounds are modelled with one commmon *pdf*. Similarly one mass *pdf* is used to model $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ decays.

The expected yield of the semi-leptonic backgrounds in each BDT bin is estimated by normalising to the number of decays observed via

$$\begin{aligned}\mathcal{N}^{exp}_x &= \mathcal{N}_{B^+\to J/\psi K+} \cdot \frac{f_x}{f_u} \cdot \frac{\mathcal{B}_x}{\mathcal{B}_{B^+\to J/\psi K^+}} \cdot \frac{\epsilon^{tot}_x}{\epsilon^{B^+\to J/\psi K^+}_{tot}} \\ &= \beta \cdot f_x \cdot \epsilon^{tot}_x \cdot \mathcal{B}_x\end{aligned} \tag{5.6}$$

The noramalisation can be factorised, the β parameter combines the noramlisation information from the decay and is the same for all backgrounds. The evalutation of these parameters is detail in Section X, since they are the same as used to normalise the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions. The hadronisation factor f_x depends on the background and where possible the measured branching fractions of the background are used, else the predicted values are used. The total efficiencies, ϵ^{tot}, include the detector acceptance, trigger, selection, reconstruction and particle identification efficiencies for each decay. The efficiencies are calucated for each BDT seperately using information from a combination of data and simulated decays to enable the expected number of background decays in each bin to be found.

Prehaps I could put the pdf that Flavio showed in the seminar to illustrate the backgrounds and their pdfs? But CBG is there too ...

5.4 Normalisation

As introduced earlier the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions are measured by normalising the number of observed$B^0_{(s)} \to \mu^+\mu^-$ decays the number of observed and $B^0 \to K^+\pi^-$ decays. The normaliation paramters, $_{d(s)}$ in equation 5.3 can be re-written in more detail as

$$\alpha_{d(s)} = \frac{1}{\mathcal{B}_{norm}} \cdot \frac{f_{norm}}{f_{d(s)}} \cdot \frac{\epsilon^{ACC}_{norm}}{\epsilon^{ACC}_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{\epsilon^{RECSEL|ACC}_{norm}}{\epsilon^{RECSEL|ACC}_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{\epsilon^{TRIG|RECSEL}_{norm}}{\epsilon^{TRIG|RECSEL}_{B^0_{(s)}\to\mu^+\mu^-}} \cdot \frac{1}{\mathcal{N}^{obs}_{norm}} \quad (5.7)$$

the efficiency term has been split up into several components indicting the efficiency of different stages of event selection. The evaluations of different terms in equation 5.4 are described in the following sections.

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

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

5.4.2 Detector, reconstruction, selection and trigger efficiencies

The normalisation factors, in equation 5.4, require various efficiencies for identifying $B^0_{(s)} \to \mu^+\mu^-$, $B^0 \to K^+\pi^-$ and decays to be evalutated for each year of data taking.

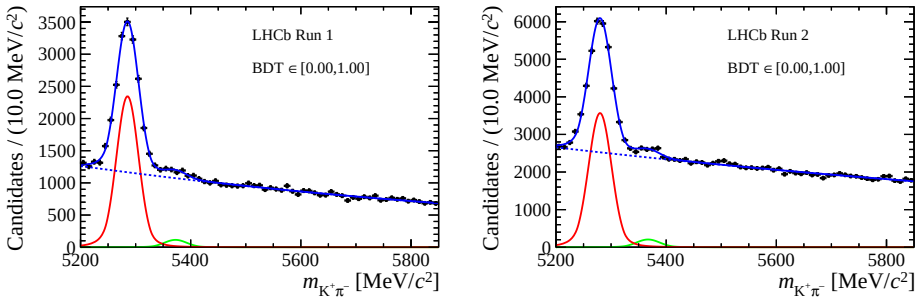

Fig. 5.2 Mass fit to measure $B^0 \to K^+\pi^-$ yield for the noramlisation for Run 1 (left) and Run 2 (right) data.

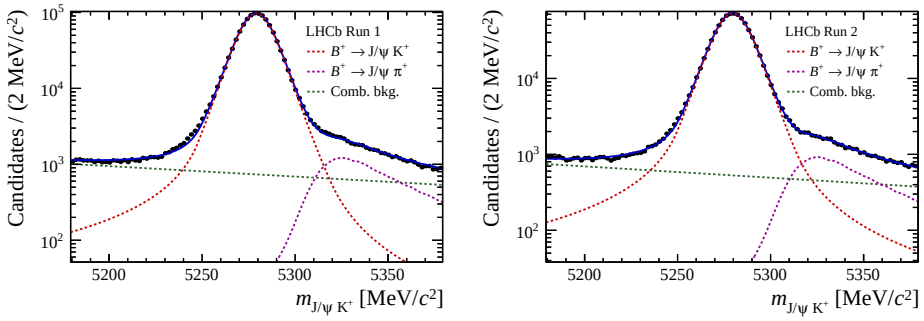

Fig. 5.3 Mass fit to measure yield for the noramlisation for Run 1 (left) and Run 2 (right) data.

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

The reconstruction efficiency of decays that are within the detector acceptance and the selection efficiency of reconstructed decays is given by $\epsilon^{RECSEL|ACC}$. The selection and reconstruction efficiencies are evaluated from a combination of information from data and simulated decays to ensure accurate selection efficiency ratios. Similar to the $B^0_s \to \mu^+\mu^-$ BDT *pdf* a correction is applied for the lifetime used in simulated $B^0_s \to \mu^+\mu^-$ decays assuming $A_{\Delta\Gamma} = +1$.

The trigger efficiencies for decays passing the reconstruction and selection, $\epsilon^{TRIG|RECSEL}$, are evalutaed for each decay by data driven methods as descrided in **??**.

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

5.4.3 Hadronisation factors

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

The various yields, efficiencies and hadronisation factors are combined to produce seperate normalisation factors for each year of data taking and each noramlisation channel. The factors are combined to produced one set of normalisation factors for Run 1 and for Run 2. *I could put the yearly equation in and say that the weighted average is taken of the two modes?.* The final noramlisation parameters are given in Table **??**. The noramlisation factors are also used in determining the number of expected background decays in each BDT bin as detail in the earlier Section X.

5.5 Results

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

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

The branching fraction results from the fit are;

$$\begin{aligned} \mathcal{B}(B^0_s \to \mu^+\mu^-) &= (2.8 \pm 0.6) \times 10^{-9} \\ \mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.6^{+1.1}_{-0.9}) \times 10^{-10} \end{aligned} \tag{5.8}$$

Figure **??** shows the fit results for $B^0_{(s)} \to \mu^+\mu^-$ candidates in the 4 BDT bins for both Run 1 and Run 2 data and Figure 5.4 the 2-dimensioanly likelihood profile for the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ branching fraction measurements. The statistics significancal of the $B^0_s \to \mu^+\mu^-$ signal is 7.9 σ which is the first single experiment obseravtion of the $B^0_s \to \mu^+\mu^-$ decay. While the significance of the $B^0 \to \mu^+\mu^-$ signal is less at 1.9σ, therefore the CLs method is used to place an upper limit on the branching fraction of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.4 \times 10^{-10}$ at the 95 % confidence level.

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

Fig. 5.4 2D likelihood for B^0 and B_s^0 BFs.

some things I should learn about the BF anyslsis or I think I should mention. The correlaction, orlack ok, between the mass and BDT output. The cascade B decays that are removed by the lower 4900 mass cut (Alessio) and also the decays that contribute to CBG (Siim).

prehaps put plots and numbers in the appendix??

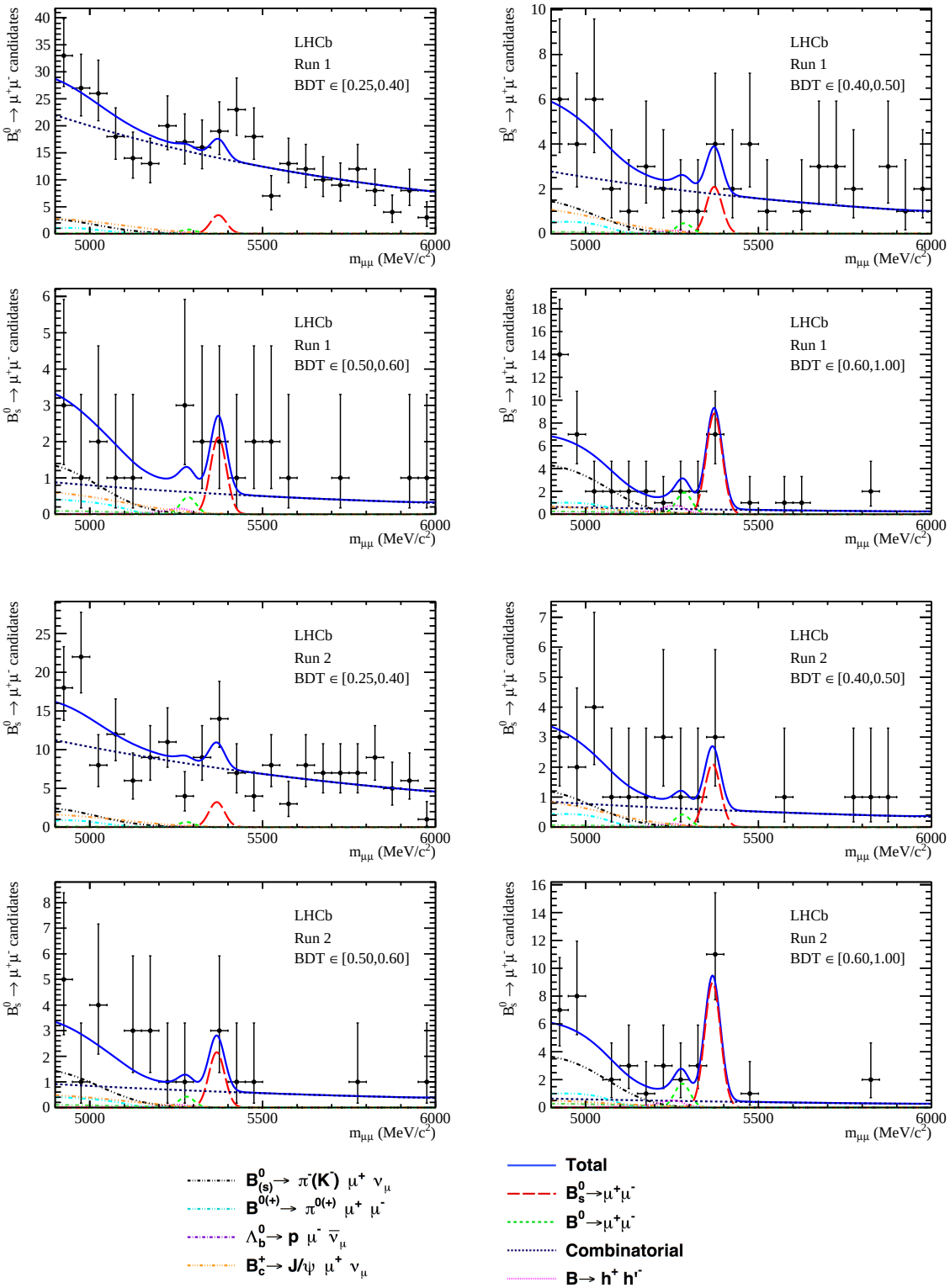

Fig. 5.5 Mass distribution in BDT bins for selected $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ candidates with the fit overlaid for Run 1 and Run 2 data. The components in the total mass pdf (blue) are; $B^0_s \to \mu^+\mu^-$ (red), $B^0 \to \mu^+\mu^-$ (green), $B \to h^+h^-$ (magenta), $B^0 \to \pi^-\mu{+}\nu_\mu$ and $B^0_s \to K^-\mu{+}\nu_\mu$ (black), $B^+_c \to J/\psi\mu^+\nu_\mu$ (orange), $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ (cyan), $\Lambda^0_b \to p\mu^-\nu_\mu$ (green) and combinatorial background (blue). *Think about the key, ideally consistent throughout all images!*

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.