

Search for the rare decay

$$\Lambda_b^0 \rightarrow p \pi^- \mu^+ \mu^-$$

Eluned Anne Smith

High Energy Physics

Blackett Laboratory

Imperial College London

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

This thesis reports the branching fraction measurement of the rare Cabibbo-suppressed decay $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$. The decay is observed for the first time with a 5.5σ deviation from the background-only hypothesis. This is the first observation of a $b \rightarrow d$ quark transition in the baryon sector. The dataset used for the measurement corresponds to 3 fb^{-1} of pp collisions collected at the LHCb experiment at CERN. The branching fraction is measured using $\Lambda_b^0 \rightarrow J/\psi (\rightarrow \mu^+\mu^-) p\pi^-$ as a normalisation channel and is measured as

$$\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-) = (6.9 \pm 1.9 \pm 1.1^{+1.3}_{-1.0}) \times 10^{-8},$$

where the first error is the statistical uncertainty, the second is the systematic uncertainty and the third is the uncertainty on $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-)$. The measurement of $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)$ can be combined with the branching fraction measurement for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ to give constraints on the ratio of CKM matrix elements $|\frac{V_{td}}{V_{ts}}|$. Such a determination of $|\frac{V_{td}}{V_{ts}}|$ requires a theory prediction for the ratio of the relevant form factors.

This thesis also reports the ratio of tracking efficiencies, ϵ_{rel} , between data and simulation for $K_s^0 \rightarrow \pi^+\pi^-$ decays occurring within the LHCb detector acceptance. As K_s^0 particles are long-lived, their associated tracking efficiencies are less precisely determined compared to those of shorter-lived particles. The average value of ϵ_{rel} for $K_s^0 \rightarrow \pi^+\pi^-$ decays, where the K_s^0 has a flight distance of $\gtrsim 1 \text{ m}$, is found to be

$$\epsilon_{\text{rel}} = 0.70 \pm 0.02.$$

To perform this calibration measurement a novel technique was developed which has the potential to be used in measuring the value of ϵ_{rel} for other decays involving long-lived particles.

Contents

Declaration of originality	7
Copyright declaration	8
Acknowledgements	9
List of Figures	10
List of Tables	21
List of abbreviations and definitions	24
1 Introduction	30
2 Theory	33
2.1 Local gauge invariance	34
2.2 Generation of fermion masses	35
2.2.1 Introducing weak hypercharge	41
2.3 The generation of gauge boson masses	42
2.3.1 Illustrating the unitary gauge using a U(1) symmetry	42
2.3.2 Giving mass to the W^\pm and Z^0 bosons	44
2.4 Form factors for hadronic transitions	46
2.5 The flavour problem	47
2.6 Using $b \rightarrow d l l$ decays to search for new physics	49
2.6.1 Measuring V_{ts} and V_{td} using tree-level process	49

2.6.2	Discrepancies between tree-level and higher-order decays	50
2.7	Measuring V_{td} using semileptonic mesonic decays	53
2.8	Measuring V_{td} using semileptonic baryonic decays	53
3	The LHCb detector	56
3.1	Trigger	59
3.1.1	L0 trigger	60
3.2	The HLT trigger	61
3.2.1	The TISTOS method for efficiency estimations	63
3.3	The VELO	65
3.3.1	Measuring the impact parameter	65
3.3.2	Measuring the PV position	65
3.3.3	Reconstructing tracks in the VELO	67
3.4	The TT station, T station and track reconstruction	69
3.5	The RICH detectors	72
3.5.1	The particle identification algorithm	73
3.5.2	The PID performance	74
3.6	Muon identification	78
3.6.1	The efficiency of the muon stations	78
3.6.2	The muon identification procedure	79
3.6.3	Efficiency of the muon ID procedure	79
3.7	Simulation	82
4	Measuring the efficiency of long-lived particles	83
4.1	Current reconstruction efficiency studies carried out at LHCb	84
4.1.1	Long track efficiency measurements	84
4.1.2	Downstream tracking efficiency measurements	86
4.1.3	Measuring the reconstruction efficiency of $K_s^0 \rightarrow \pi^+\pi^-$ as a function of z	90
4.2	Data samples	92

4.2.1	Selecting K_s^0 particles from data	92
4.2.2	Simulation	92
4.3	Calculating the genuine number of K_s^0 decays as a function of z	94
4.4	The removal of background	97
4.4.1	Unfolding the reconstruction efficiency in the z_{ref} bins	98
4.5	Comparison of simulated events and data distributions	101
4.6	Calculation of $N(z, p)$	102
4.6.1	Verifying the calculation using generator-level simulated events . .	102
4.7	Systematic checks	104
4.7.1	Finite binning of p and z	104
4.7.2	Assuming that the background shape is the same across all p and η .	104
4.7.3	The modelling of material interactions in simulation	105
4.8	Results	107
4.9	Comparison between the results from this analysis and previous studies .	111
5	Selection and backgrounds in the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ analysis	113
5.1	Overview	114
5.1.1	Analysis strategy	117
5.2	Data samples and simulation	118
5.2.1	Data and simulation samples	118
5.3	Preselection	118
5.3.1	Stripping cuts	119
5.3.2	Trigger requirements	120
5.3.3	Preselection criteria	120
5.3.4	Selections in q^2	121
5.4	Data and simulation agreement	122
5.4.1	Simulation reweighting	122
5.4.2	Resampling the simulation	126
5.5	Reflection backgrounds	128

5.5.1	<i>B</i> reflection vetoes	129
5.5.2	Background from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ decays	134
5.5.3	Double mis-identification of muons as pions	136
5.5.4	Mis-identification from swaps between muons and hadrons	137
5.6	Partially reconstructed backgrounds in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel	138
6	Multivariate methods used in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis	145
6.1	Isolation of the final state tracks	145
6.1.1	Isolation BDT response	147
6.2	The combinatorial Boosted Decision Tree	147
6.3	The BDT performance and optimisation	152
7	Mass fits and efficiency calculations for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis	157
7.1	Mass fits	157
7.1.1	Signal fits	158
7.1.2	Complete fit to the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel	161
7.1.3	Complete fit to the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel	165
7.1.4	Expected significance for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel	168
7.2	Selection efficiency	171
7.2.1	Detector acceptance cuts	172
7.2.2	Stripping and reconstruction efficiency	172
7.2.3	Trigger efficiency, preselection and PID cuts	173
7.2.4	BDT efficiency	174
7.2.5	Total relative efficiency	174
7.2.6	Efficiency as a function of $m_{p\pi}$	175
8	Results	178
8.1	The $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction measurement	178
8.2	Systematic uncertainties on $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)$	180
8.2.1	The TISTOS method	181

8.2.2	PID	182
8.2.3	The effect of the choice of the q^2 distribution on the total relative efficiency	186
8.2.4	The effect of the choice of the BDT efficiency proxy on the total relative efficiency	187
8.2.5	Effect of weighting the simulation and the simulation statistics on the total integrated efficiency value	188
8.2.6	Effect of reweighting the $p\pi$ mass spectrum in simulation on the total integrated efficiency	189
8.2.7	Effect of the fit shape on the fitted signal yield	191
8.2.8	Verifying the fit coverage and bias for the signal channel	195
8.2.9	Verifying the fit coverage and bias for the normalisation channel	196
8.2.10	Errors on the $\Lambda_b^0 \rightarrow p\pi^- J/\psi (\rightarrow \mu^+\mu^-)$ branching fraction measurement	197
8.2.11	Summary of systematic uncertainties	197
9	Conclusions and outlook	199
Bibliography		201
Appendices		210
A Boosted Decision Trees		210
B The <i>sPlot</i> technique		213

Declaration of originality

The work presented in this thesis is the result of collaborative work between members of the LHCb collaboration and myself. All the analysis work (chapters 4–8) presented in this thesis was performed by myself, with the exception of producing the simulation and data samples used in the analysis in [chapter 4](#). All work and plots presented in this thesis that were not the product of my own work are appropriately referenced.

This thesis has not been submitted for any other qualification.

Eluned Smith
September 2016

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

Acknowledgements

I would like to thank Mitesh for his continued support throughout my PhD. I am well aware that the amount of time he gives to his students goes above and beyond the norm and as such count myself lucky to have had him as a supervisor¹. A big thank you for everything.

I am also indebted to Patrick Owen for the invaluable input he made to both analyses presented in this thesis. Thank you. I received a huge amount of support and feedback from all members of the Imperial College LHCb group, particularly from Ulrik Egede and Paula Álvarez Cartelle. In addition, from Sam Hall and William Sutcliffe, who were great company both in and out of work.

I would also like to thank the STFC funding body, who financed my PhD and the LHCb collaboration as a whole, who made the work in this thesis possible.

To my Genevan friends, who made a foreign city feel like home. À Céline, pour toutes nos aventures sportives, à Lorena, por todas las noches de juerga, and to Hannah, who I was lucky enough to have as continued company throughout the last eight years. A big thank you goes to Ellie, because nothing says friendship like letting me loose on your home for a month.

To Mark, who can take much credit for the retention of most of my sanity over the last year.

And finally to my family, for consistently being fabulous, throughout my PhD and always.

¹He is also now one of a select group of people who can understand me even when I break out into full-speaking-speed (or at least I think he can), which in itself deserves some kind of recognition

List of Figures

2.1	The resulting $V(\phi)$ potential for the case of $\mu^2 < 0$, (a) and the shift of fields to the minima, (b), [9].	43
2.2	The lowest order Feynman diagram for B^0 and B_s^0 mixing.	47
2.3	Feynman diagrams for $\Lambda_b^0 \rightarrow p \pi^-(K^-)\mu^+\mu^-$ via a loop, (a), $B^- \rightarrow \pi^-(K^-)\mu^+\mu^-$ via a loop, (b), $\Lambda_b^0 \rightarrow p \pi^-(K^-)\mu^+\mu^-$ via a box diagram, (c), $B^- \rightarrow \pi^-(K^-)\mu^+\mu^-$ via a box diagram, (d).	50
2.4	Comparing measurements of V_{td} , V_{ts} , $ V_{td}/V_{ts} $ using (top to bottom) current experimental measurements of $\Delta M_{q=d/s}$ with the latest theory calculations, as given in the PDG [7], using semileptonic decays, and the combined value for tree and loop level processes, using unitary constraints (CKM fitter [13]). Plot taken from Ref. [15].	52
2.5	Branching fraction as a function of q^2 for $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ decays [24]. . . .	54
3.1	A profile of the LHCb detector [25]. Any reference to x, y, z directions in this thesis refers to this diagram. The interaction point is located inside the Vertex Locator.	57
3.2	The production of $b\bar{b}$ quarks as a function of η . The acceptance for LHCb and General Purpose (GP) detectors are overlaid on top. Figure taken from the LHCb public page.	58
3.3	Overview of the LHCb trigger rates. Figure taken from the LHCb public page.	60
3.4	The 2011 LOMuon and LODiMuon lines efficiency, ϵ_{TOS} , on $B^+ \rightarrow J/\psi (\mu^+\mu^-)K^+$ as a function of P_T (J/ψ) [29].	61

3.5	The <code>Hlt1TrackAllL0</code> TOS efficiency (ϵ_{TOS}) of the decays indicated in the legend as a function of the P_T of the relevant B or D meson, (a), the <code>Hlt1TrackMuon</code> (and <code>Hlt1DiMuonHighMass</code> , <code>Hlt1DiMuonLowMass</code>) ϵ_{TOS} as a function of the P_T of the J/ψ (b). Both plots show 2012 data [30].	62
3.6	The Flight Distance (FD), Impact Parameter (IP) and Primary and Secondary Vertices (PV, SV) for an event. The black lines labelled p represent the colliding proton beams. Image altered from Ref. [35].	64
3.7	Above: the VELO stations as placed along the z direction, below: as shown in the $x - y$ plane in both the open (for injection) and closed (for stable beams) configuration [25].	66
3.8	The resolution of IP_x and IP_y as a function of the momentum of the track, (a), the resolution of IP_x and IP_y as a function of $1/P_T$, (b) [26].	67
3.9	The PV resolution in the x, y directions as a function of track multiplicity, N , (a). The PV resolution in z direction as a function of N , (b). The values in the legends refer to the constants in Equation 3.4 [26].	67
3.10	The tracking efficiency of the VELO for data and simulation as a function of momentum (left) and η (right).	68
3.11	The Outer Tracker in the T stations (light blue) made up of straw-modules and the Inner Tracker and TT station (purple) made up of silicon micro-strips [36].	69
3.12	A display of the reconstructed tracks for an event along with the assigned hits in blue. The insert shows a zoom of the VELO region in the $x-y$ plane [25].	70
3.13	Different types of track in the LHCb detector along with the corresponding magnetic field strength in the y direction [25]	70
3.14	The optical set-up for the RICH1 detector [42].	73
3.15	Fitted invariant mass distributions for $\Lambda^0 \rightarrow p\pi^-$, (a), and $D^0 \rightarrow K^-\pi^+$, (b) [43].	75

3.16	The kaon efficiency and the misidentification rate for pions as kaons for different PID cuts (filled and hollow markers) (a), the proton efficiency and the misidentification rate for pions as protons for different PID cuts (filled and hollow markers), (b) and the proton efficiency and the misidentification rate for kaons as protons for different PID cuts (filled and hollow markers), (c) [43].	76
3.17	Reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator [43].	77
3.18	The five muon stations (a) (the muon filters refer to iron absorbers - see text) and (b) the station layout in regions R1–4 [45].	78
3.19	IsMuon efficiency and misidentification probabilities, as a function of momentum and in ranges of transverse-momentum: efficiency of IsMuon on muons, ϵ_{IM} , (a), IsMuon mis-id rate ($p \rightarrow \mu$), (b), IsMuon mis-id rate ($\pi \rightarrow \mu$), (c), and IsMuon mis-id rate ($K \rightarrow \mu$), (d) [46].	80
3.20	Efficiency of selecting muons against the efficiency for selecting pions for the $DLL_{\mu\pi}$ and muDLL variables, (a). Efficiency of selecting muons against the efficiency for selecting kaons for the $DLL_{\mu\pi}$ and muDLL variables, (b).	81
4.1	Different probe tracks used to measure the efficiencies of different subdetectors, showing the probe track used to measure the VELO efficiency (a), the probe track used to measure the T station efficiency (b), and the probe track used to measure the VELO and T station efficiencies simultaneously (c).	86
4.2	The efficiency for data and simulation as a function of track multiplicity, N_{track}	87
4.3	Ratio of absolute efficiencies between data and simulated events for long tracks. Ratios are for 2012 conditions [67].	87

4.4	The downstream tracking efficiency as a function of momentum, (a), (b). The vertex efficiency ratio of the K_s^0 for data over simulation (where simulation is indicated by the initials MC in the figure) for the cases where both pions are downstream track types (DD) and both pions are long track types (LL), (c). Figures are taken from Ref [66].	89
4.5	A sketch of the extrapolation in z of the number of expected K_s^0 candidates for a given momentum, p , and a mass, m . The line at $z = 140$ mm indicates the end of the z_{valid} region.	91
4.6	The number of decays as a function of z for long tracks with (black) and without (red) the cuts on the primary vertex of the K_s^0 applied, for data events with $10 < p < 20$ GeV, $2.0 < \eta < 3.2$ (a), $2.0 < \eta < 3.2$ (b).	95
4.7	The number of decays for downstream and long tracks as a function of z for both reconstructed simulated events (blue) and data (green), for events with $10 < p < 20$ GeV/c, $2.0 < \eta < 3.2$. The line at 90 mm indicates the lower edge of the lowest z_{ref} bin. The long tracks are those which peak around the red line at 90 mm and the downstream tracks are those which peak at higher z	96
4.8	The fit to the $m_{\pi\pi}$ mass distributions for downstream ((a), (b)) and long ((c), (d)) tracks for events with $1200 < z < 1210$ mm (a), $850 < z < 860$ mm (b), $90 < z < 100$ mm (c), $300 < z < 310$ mm (d).	98
4.9	The distributions in p, η for the bin $10 < p < 20$ GeV/c, $2.0 < \eta < 3.2$ for the quantities $n_{\text{reco},p,\eta}$, (a), $n_{\text{reco},p,\eta} \times \omega_{(\mu^+\mu^-)p_{\mu^\pm},\eta_{\mu^\pm}}$, (b), $N_{\text{gen},p,\eta}$, (c), and $\epsilon_{p,\eta}$, (d) from Equation 4.4 along with $n_{\text{data},p,\eta}$, (e), and $n_{\text{data},p,\eta}/\epsilon_{p,\eta}$, (f).	100
4.10	The (p, η) distribution for data, (a) and the (p, η) distribution for reconstructed simulated events after weights are applied, (b), for events with $90 < z < 100$ mm, $10 < p < 20$ GeV and $2.0 < \eta < 3.2$	102

4.11 The value of N_0 used in the calculation for the z_{ref} bin $90 < z < 100$ mm, (a), which shows the N_0 in the z_{ref} bin, $90 < z < 100$ mm, and the calculated value, match exactly. The calculated (solid distribution) and the distribution for generator-level simulated events in z for bin $10 < p < 20$ GeV, $2.0 < \eta < 3.2$, $90 < z < 100$ mm (b).	103
4.12 The difference between the final efficiencies with and without randomly varying the value taken as the central value in momentum p for simulation, top, and data, bottom. For events with $10 < p < 20$ GeV/c, $2.0 < \eta < 3.2$, left, $3.2 < \eta < 5.0$, right. The y axis shows the difference in efficiencies. The downstream tracks are shown with green errors and the long tracks are shown with blue errors.	105
4.13 The difference between the final efficiencies with and without randomly varying the value taken as the central value in z for simulation, top, and data, bottom. For events with $10 < p < 20$, $2.0 < \eta < 3.2$, left, $3.2 < \eta < 5$, right. The y axis shows the difference in efficiencies.	106
4.14 The signal and mass distributions for the same z bin but different p and η bins for downstream tracks. For events with $10 < p < 20$ GeV/c, $2.0 < \eta < 3.2$, (a), $10 < p < 20$ GeV/c, $3.2 < \eta < 5.0$, (b), $20 < p < 40$ GeV/c, $2.0 < \eta < 3.2$, (c), $20 < p < 40$ GeV/c, $3.2 < \eta < 5.0$, (d), and with a decay vertex between $1070 < z < 1080$ mm.	107
4.15 The efficiency distributions for data and reconstructed simulation. For events with $10 < p < 20$ GeV/c, $2.0 < \eta < 3.2$ (a) and $3.2 < \eta < 5.0$ (b), $20 < p < 40$ GeV/c, $2.0 < \eta < 3.2$ (c) and $3.2 < \eta < 5.0$ (d) and $40 < p < 100$ GeV/c, $3.2 < \eta < 5.0$ GeV/c, (e).	109
4.16 The ratio of efficiency distributions for data and reconstructed simulation. For events with $10 < p < 20$ GeV/c, $2.0 < \eta < 3.2$ (a) and $3.2 < \eta < 5.0$ (b), $20 < p < 40$ GeV/c, $2.0 < \eta < 3.2$ (c) and $3.2 < \eta < 5.0$ (d) and $40 < p < 100$ GeV/c, $3.2 < \eta < 5.0$ GeV/c, (e). The results of a zero-order fit to the downstream tracking efficiency as a function of z are shown.	110

5.1	Feynman diagram for the decay $\Lambda_b^0 \rightarrow J/\psi p\pi^-$	115
5.2	$\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation distributions, before and after reweighting compared to sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data for Λ_b^0 p_T (a), Λ_b^0 p (b), track multiplicity (c), $\Lambda_b^0 \tau$ (d).	124
5.3	$\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation distributions, before and after reweighting compared to sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data for Λ_b^0 p_T (a), Λ_b^0 p (b), track multiplicity (c), $\Lambda_b^0 \tau$ (d).	125
5.4	The distribution of the proton DLL _{pπ} variable for resampled $\Lambda_b^0 \rightarrow J/\psi pK^-$ simulation and $\Lambda_b^0 \rightarrow J/\psi pK^-$ sWeighted data (a), and the distribution of the pion DLL _{Kπ} variable for resampled $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ sWeighted data (b).	127
5.5	The proton DLL _{pπ} -DLL _{Kπ} variable for $\Lambda_b^0 \rightarrow J/\psi pK^-$ sWeighted data and $\Lambda_b^0 \rightarrow J/\psi pK^-$ resampled simulation.	128
5.6	Shape of all data candidates, prior to any PID selection and within 60 MeV/c ² of the Λ_b^0 mass, under different mass hypotheses (a), and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ signal simulation under these same hypotheses, (b). The shaded regions indicated the areas within which a tighter PID selection is applied. The lower shaded region in mass corresponds to events falling near the B^0 mass and the higher region corresponds to events falling near the B_s^0 mass. The widths of these shaded regions are detailed in Table 5.7.	131
5.7	The signal mass distribution with and without the mass-dependent PID cuts, as listed in Table 5.7(a), applied, (a). The difference between the two histograms in (a) (when normalised), (b).	132
5.8	$\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ simulation with the region in which tighter PID cuts are imposed indicated by the shading and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation under the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ mass hypothesis.	135
5.9	The mass distribution of $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation with and without the mass-dependent PID cut to reduce $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ contributions applied.	135
5.10	The Feynman diagram for the decay $\Lambda_b^0 \rightarrow p\pi^- \pi^+ \pi^-$	137

5.11 Fitted mass distributions for blinded $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data (a) and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data (b)	139
5.12 Combined mass of ($p\pi^-(\mu^+ \rightarrow \pi^+)$) for $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ lower mass side band data where the muon with the same sign as the proton has been given the mass of a pion (a), combined mass of $pK^-(\mu^+ \rightarrow \pi^+)$ for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ lower mass side band data where the muon with the same sign as the proton has been given that of a pion (b). The peaks at $\sim 2300 \text{ MeV}/c^2$ are due to the Λ_c^+ baryon appearing via $\Lambda_b^0 \rightarrow \Lambda_c^+(pK^-(\pi^-)\pi^+)X$ decays.	141
5.13 The mass of the combination ($pK(\mu^+ \rightarrow \pi^+)$) against the mass of the combination of ($pK\mu^-\mu^+$) for the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data set after the complete selection (a) and the mass of the combination ($p\pi^-(\mu^+ \rightarrow \pi^+)$) against the mass of the combination of ($p\pi^-\mu^-\mu^+$) for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data set after the complete selection (b). The red lines in the x and y direction indicate the nominal Λ_c^+ and Λ_b^0 masses respectively.	142
5.14 $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data with and without a cut on $\cos(\theta_{ll})$, where θ_{ll} is defined as the angle between the momentum of the lepton with the same sign as the proton and the sum of the lepton momenta, both in the frame of the Λ_b^0 mother, (a), the distribution of $\cos(\theta_{ll})$ for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data (b).	143
5.15 Comparison of the separately normalised Λ_b^0 mass distribution in the lower-mass side band for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data.	144
6.1 A schematic of a non-signal track, i , and a $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ candidate.	146
6.2 The isolation BDT response for $\Lambda_b^0 \rightarrow J/\psi pK^-$ data, representing the signal proxy, and the response for the upper mass side band of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data, representing the background.	147
6.3 The first six variables used in the combinatorial BDT for signal and background proxies. The red stripped histograms are background and the solid blue histograms are signal. All distributions shown are normalised. The variable $\Lambda_b^0 \text{ EV}_{\chi^2}$ refers to Λ_b^0 vertex χ^2/ndof in Table 6.2.	150

6.4	The second six variables used in the combinatorial BDT for signal and background proxies. The red stripped histograms are background and the solid blue histograms are signal. All distributions shown are normalised.	151
6.5	The final five variables used in the combinatorial BDT for signal and background proxies. The red stripped histograms are background and the solid blue histograms are signal. All distributions shown are normalised.	
	The variable Λ_b^0 DTF χ^2 refers to Λ_b DecayTreeFitter χ^2 in Table 6.2.	152
6.6	The BDT response for all 5 K-Folds and the signal efficiency against background rejection curve, integrated over all K-Folds. The red line on the signal efficiency against background rejection curve indicates the optimal working point.	153
6.7	The efficiency of the BDT as a function of BDT cut value using $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ sWeighted data. The dashed line shows the efficiency at the working point of 0.25.	155
6.8	The Punzi FOM for background values that are varied randomly within their error.	155
6.9	Fitted data for blinded $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ events (a) (same as Figure 5.11), and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ events (b).	156
7.1	The fit of a double CB to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation with a linear scale, (a), and a log scale, along with the number of standard deviations the data points lie from the fit function, (b).	158
7.2	The q^2 distribution for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data against $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ simulation, where the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ simulation has been reweighted in q^2 using the $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ differential branching fraction predictions.	159

7.3	The q^2 distribution of $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation after reweighting in q^2 , (a). The fit of a double CB function to $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation with q^2 -reweighting applied, (b). The red curve in (b) indicates the right CB function and the green curve the left CB function. The number of standard deviations the data points lie from the fit function are shown below the fit in (b).	160
7.4	The $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel in simulation, normalised to the expected yield of 1017 ± 41 events, along with the $B^0 \rightarrow J/\psi K^{*0}$ component under the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass hypothesis, normalised to the expected yield of 69 ± 9 events, for the cases where the mass-dependent PID cuts are applied and not applied, (a). A zoom-in of (a) is shown in (b). Here, the normalisation value of 69 ± 9 for the $B^0 \rightarrow J/\psi K^{*0}$ component is only correct for the red curve but both blue and red curves are normalised to the same value to allow a meaningful comparison of their shapes. A fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data with the yield of the $B^0 \rightarrow J/\psi K^{*0}$ component Gaussian constrained to 69 ± 9 events is shown on a linear scale in (c) and on a log scale in (d). The shape of $B^0 \rightarrow J/\psi K^{*0}$ simulation, with only initial PID cuts applied, under the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass hypothesis is shown in (e).	164
7.5	Fitted data for blinded $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ (a) and $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ (b).	165
7.6	Example pseudo experiments for signal and background hypotheses, (left) and background only fits, (right), with an exponential background with all fit parameters allowed to float (a), and the RooExpAndGauss fit parameters constrained from the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ background fit (b). The no. S and no. B in the legends refer to the number of signal events and background events respectively.	169
7.7	The significance distributions from fits to pseudo experiments where the fits in question use a background model which is either a free exponential or a RooExpAndGauss and exponential combined, where the RooExpAndGauss function is constrained.	170

7.8	Significance of pseudo experiments fits, using pseudo experiments generated from a constrained RooExpAndGauss and fitted back with the same model.	170
7.9	A comparison between the q^2 distribution taken from phase space simulation and from the $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ branching fraction predictions of Ref. [10].	172
7.10	Relative combined stripping and reconstruction efficiencies between $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation as a function of q^2	173
7.11	Absolute efficiencies for the preselection, trigger and PID selection, as a function of q^2 for $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation. The dashed lines indicate the efficiency of the ($9 < q^2 < 10 \text{ GeV}^2/c^4$) bin for the efficiency distribution of the same colour. There is no efficiency value for the last bin due to a lack of data in this bin.	174
7.12	The total relative efficiency assuming a flat BDT efficiency in q^2 . The shaded areas indicate the vetoed regions in q^2	175
7.13	The absolute efficiency for the preselection and PID selections as a function of $m_{p\pi}$, (a), and both $m_{p\pi}$ and q^2 , (b).	177
8.1	The fit to the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ mass distribution.	178
8.2	The fit to the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ mass distributions for data candidates with $2256 < m_{p\pi\mu\rightarrow\pi} < 2316 \text{ MeV}/c^2$ events vetoed and with the requirement $\text{DLL}_{\mu\pi} > 0$ placed on both muons.	180
8.3	The difference between data and simulation for the $\text{DLL}_{K\pi}$ variable, (a), and the difference between data and simulation for the $\text{DLL}_{p\pi}$ variable, (b).	184
8.4	The $\text{DLL}_{p\pi}$ distribution for sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ after the application of all selections.	185
8.5	The q^2 distribution for sWeighted $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ data fitted with a polynomial function.	186
8.6	The total efficiency as a function of q^2 multiplied by the $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ q^2 distribution and assuming either a flat or varying BDT efficiency distribution in q^2	188

8.7	The total relative efficiency calculated with weighted simulation, compared against the efficiency calculated using non-weighted simulation.	189
8.8	The $p\pi$ mass spectrum for sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data and phase space (phsp) $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation.	190
8.9	The $p\pi$ mass spectrum for $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data and $m_{p\pi}$ weighted $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation, with a cut placed at $m_{p\pi} < 2500 \text{ MeV}/c^2$	191
8.10	The efficiency as function of q^2 for $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation with $m_{p\pi}$ weights applied and not applied, with a cut placed at $m_{p\pi} < 2500 \text{ MeV}/c^2$	191
8.11	Examples of pseudo experiments showing background and signal fits, (left) and background only fits, fitted back with a RooExpAndGauss and exponential combined model with none of the fit components constrained, (right). The significance is 1.04 for the fits in (a) and 4.32 for the fits in (b). The no. S and no. B in the legends refer to the number of signal and background events respectively.	193
8.12	The background-only fit for the nominal background model over the full range of $5100 - 7000 \text{ MeV}/c^2$ and a background-only fit with a single exponential over the restricted range of $5500 - 7000 \text{ MeV}/c^2$	194
8.13	The fit to the distribution of the pseudo experiment's pulls for the signal channel fit.	195
8.14	The fit to the distribution of the pseudo experiment's pulls for the normalisation channel fit.	196
A.1	An example decision tree. The S and B stand for ‘Signal-like’ and ‘Background-like’. The β_i variables refer to the cut values chosen by the machine learning algorithm after the tree has been trained on signal and background samples. The blue ovals represent final nodes called leafs, which each leaf having an associated purity, i.e. the fraction of the weight of a leaf due to signal events.	211

List of Tables

2.1	Showing the masses for different possible N^* resonances	55
3.1	The HLT2 lines used in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis based on two identified muons [28].	64
4.1	Criteria applied to the vertexing procedure used to make a standard LHCb K_s^0 candidate from two downstream pions. The track DOCA is the distance of closest approach between the two daughter tracks and the mass windows are either side of the known K_s^0 mass of $497.6 \text{ MeV}/c^2$ [66].	88
4.2	Results from the study in Ref [66] for the ratio between data and simulation of the downstream tracking and vertexing efficiency as a function of the missing track momentum and the K_s^0 momentum, respectively.	90
4.3	Criteria to select K_s^0 candidates for the analysis.	93
4.4	The value of z where the cut is applied to the primary vertex of the K_s^0 . Results are shown in mm for each bin.	95
4.5	The ratio between data and simulation downstream tracking efficiency averaged over z , from $z > 1000 \text{ mm}$. The first error quoted is statistical and the second error is due to the mismodelling of hadronic material interactions in simulation.	111

5.1	Comparison of branching fractions between the $\mu^+\mu^-$ and J/ψ channels for various decays. The ratio between the resonant and non-resonant channels is estimated assuming that $\mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) = 0.0596 \pm 0.0003$. All values for the branching fractions are taken from Ref. [7], with the exception of the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ branching fraction measurements, which are taken from Refs [69] and [70] respectively.	116
5.2	The selection criteria used in the stripping line.	119
5.3	Trigger requirements.	120
5.4	Preselection applied.	121
5.5	Vetoos in q^2 for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel and the selection in q^2 for the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ normalisation channel.	122
5.6	Decay types used to produce the calibration samples used for resampling.	126
5.7	Vetoos used to reject reflections and their rejection rate, calculated using $B \rightarrow J/\psi X$ simulation, along side their efficiency on $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation, relative to the initial PID cuts already placed.	133
5.8	The veto requirements used to reduce the background from charmonium resonances. The notation $m_{X(Y \rightarrow Z)}$ implies the combined invariant mass of the X and Z particle, where the Y particle is given the mass of Z .	138
6.1	The variables used in the isolation BDT.	146
6.2	List of BDT input variables. The variables are listed in order of their separation power. The seperation values are also listed.	149
7.1	The widths of the CB functions from the fit to $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation, expressed relative to the width from the fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation. The terms left or right refer to the direction of the tail of the CB function.	161
7.2	The fit parameters for the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ fit showing how these parameters are handled in the fit.	162
7.3	The fit parameters for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ fit showing how these parameters were handled in the fit.	166

7.4	The fit parameters for a Gaussian function fitted to the distributions in Figure 7.7.	168
8.1	The trigger efficiency obtained using the TISTOS method in simulation and data.	182
8.2	Value of the fit constant in the zero-order polynomial fit to the difference between data and resampled simulation as a function of PID variable. . .	183
8.3	The different sources of systematic uncertainty and the total systematic uncertainty assigned.	198

List of abbreviations and definitions

DLL_{Kπ} The difference in an events total likelihood, given the distribution of hits in the RICH, when the hypothesis of the track in question is changed from being that of a pion to a kaon.

DLL_{μπ} The difference in an events total likelihood, given the distribution of hits in the RICH, when the hypothesis of the track in question is changed from being that of a pion to a muon.

DLL_{pπ} The difference in an events total likelihood, given the distribution of hits in the RICH, when the hypothesis of the track in question is changed from being that of a pion to a proton.

FDχ² The FD χ^2 is defined for two vertices (generally the PV and SV) as the change in χ^2 when the two vertices are combined into a single vertex fit.

IPχ² The IP χ^2 is the difference in the χ^2 of the fit to the PV, when the track whose IP χ^2 is being measured is added and then removed.

N_o The number of K_s^0 decays in a certain z_{ref} bin after efficiency corrections have been applied.

m_{corr} corrected mass, a function of the visible mass and the missing transverse-momentum for a decay which features unreconstructed track(s).

m_{pπ} The combined mass of the proton and the pion.

q^2 The four-momenta of the dimuon system.

$z_{K_s^0}$ The position along the z -axis past which no new K_s^0 particles are created.

z_{ref} A bin in z , with a 10mm width, lying in the region of z_{valid} .

z_{valid} The region along the z -axis within which the long track efficiency values are deemed to be valid.

ALICE A Large Ion Collider Experiment.

ATLAS A Toroidal LHC ApparatuS.

BBDT Bonsai BDT, a BDT used in the topological trigger lines which takes discrete input variables.

BDT Boosted Decision Tree, a BDT employs multivariate analysis techniques to combine a set of weakly discriminating variables into a single discriminating variable.

BSM Beyond the Standard Model.

CB Crystal Ball function.

CMS Compact Muon Solenoid.

DD Events where both daughter tracks are downstream track types.

DIRA The cosine angle between the momentum vector of a particle and the displacement vector between the particle's decay vertex and the **PV**.

DOCA The Distance of Closest Approach between two tracks.

downstream track Downstream tracks have no VELO track segment. They are reconstructed using hits in the T and TT stations.

DTF Decay Tree Fitter.

ECAL Electromagnetic calorimeter.

EM Electromagnetism.

FCNC Flavour Changing Neutral Currents. In the **SM** these are denoted $\Delta F = 2$, referring to the two internal W boson vertices required.

FD Flight Distance, how far a particle flies before decaying.

FOM Figure of Merit.

ghost track A ghost track has less than 70% of its hits originating from a single particle.

GP General Purpose.

HCAL hadronic calorimeter.

HLT High Level Trigger. The HLT is the software trigger which is applied after the **L0** trigger.

inMuon A binary selection of muon candidates based on whether or not the candidate fell inside the acceptance of the muon stations.

IP Impact Parameter. The IP is defined as the distance between a track and the **PV** at the track's closest point of approach.

isMuon A binary selection of muon candidates based on the penetration of the muon candidates through the calorimeters and iron filters.

IT Inner trackers, the inner section of the T stations.

L0 Level-0 trigger. The L0 is the first trigger to be applied and uses hardware to make decisions on events.

LHC Large Hadron Collider.

LHCb The Large Hadron Collider beauty experiment.

LL Events where both daughter tracks are long track types.

long track Long tracks traverse the entire detector. They are defined as tracks which have hits in the VELO and the T stations. Hits in the TT stations are optional.

M1-5 The muon stations.

MFV Minimal Flavour Violation.

NP New Physics.

OT Outer trackers, the outer section of the T stations.

part-reco Partially reconstructed background, backgrounds where a decay has been mis-identified as a signal candidate and not all the final states of the mis-identified decay are included in the final reconstruction.

PID Particle IDentification.

prompt decays prompt decays are the decays of particles that were produced at the **PV**.

PS pre-shower.

PV Primary Vertex, the pp interaction vertex.

QCD Quantum Chromodynamics.

QED Quantum Electrodynamics.

reflections Decays coming from a b -hadron with a mis-identified particle that can accumulate at a certain B or Λ_b^0 mass.

RICH Ring Imaging Cherenkov detectors, provide identification of hadrons by using Cherenkov radiation.

SM Standard Model.

SPD Scintillator Pad Detectors.

SU Special Unitary group.

SV Secondary Vertex.

TIS Events which are Triggered Independent of Signal.

TISTOS Events which require both the presence of signal and the rest of the event to fire the trigger.

TOS Events which are Triggered On Signal.

VELO VErtex LOcator. One of the LHCb detector subsystems, placed around the pp interaction point, used to realise the precise measurements of vertices and tracks.

vertex χ^2/ndof The χ^2 per degree of freedom for a vertex fit for the combination of a set of daughter tracks.

To Nain, on behalf of all four of us

Chapter 1

Introduction

The Standard Model (**SM**) is an effective theory which describes fundamental particles and their interactions to an impressive precision. An example of this precision can be seen in the prediction of the electron magnetic moment using Quantum Electrodynamics (**QED**). The electron magnetic moment is characterised by the dimensionless quantity, g , which, at tree level, is predicted to be $g = 2$. The deviation due to higher order effects is expressed as $a = \frac{g-2}{2}$. This has been predicted to 10th order [1] to be

$$a_{\text{theory}} = (1.15965218178 \pm 0.00000000077) \times 10^{-12} \quad (1.1)$$

and has been experimentally measured [2] to be

$$a_{\text{experiment}} = (1.159652180730 \pm 0.00000000028) \times 10^{-12} \quad (1.2)$$

where the numbers in brackets indicate the uncertainties on the last two digits. As can be seen there is agreement between the two values to 12 significant figures.

Despite the success of the SM, there are areas where it fails. For example, there is substantial astronomical evidence for Dark Matter [3] but the **SM** offers no Dark Matter candidate. There is also a problem with hierarchy, that is, why the Higgs boson should be so much lighter than the Planck mass. This could be explained by a cancellation between the quadratic radiative corrections and the bare mass but such fine-tuning is thought to be unnatural.

The SM has no way of incorporating gravity and it can be safely left-out due its to comparatively weak strength but this further highlights that the SM is by no means a complete theory. Moreover, although the SM allows for a small degree of asymmetry between matter and anti-matter there is still no explanation for the much larger amount of asymmetry between matter and anti-matter¹ observed in the universe.

These failings motivate theories introducing beyond the standard model (**BSM**) physics, the existence of which could be inferred from the observation of BSM particles. For example, there are many BSM theories which predict new weakly interacting massive particles or light supersymmetric particles as solutions to the Dark Matter problem [5].

Previous measurements in flavour physics suggest that unless it is assumed that the flavour structure of new physics (**NP**) is near-identical to that of the SM, the masses of these new particles are so high as to be out of the reach of direct searches at current accelerators [6]. Thus this motivates searches for NP particles via indirect means.

One way of searching for new particles indirectly is by looking for their appearance as virtual BSM particles in rare decays. Given that rare decays are more suppressed in the SM, they could be more sensitive to BSM effects. The existence of BSM particles could be inferred by measuring a difference between the observed value and that predicted in the SM for the branching fraction of a decay.

The analysis presented in this thesis uses data taken at the Large Hadron Collider (**LHC**) using the **LHCb** detector. The LHCb detector exploits the large forward $b\bar{b}$ production from pp collisions and the boost of b -hadrons along the direction of the beam pipe in order to make precision measurements of b -hadron decays. The LHCb detector has been used to study extensively the decay modes and properties of B -mesons but the decays of b -baryons are still relatively unexplored. This motivates measurements of heavily suppressed Λ_b^0 decays, which may be sensitive to the existence of new particles.

In addition, while particles with lifetimes up to $\sim 10^{-12}$ seconds, characteristic of B -decays, have been widely studied at LHCb, longer-lived particles are also of interest for new

¹the level of asymmetry between matter and anti-matter allowed for in the SM falls short of explaining the observed level by several orders of magnitude [4]

physics searches. The efficiency to reconstruct such long-lived particles is comparatively poorly known and this motivates the use of new methods to establish the reconstruction efficiency as a function of position in the detector.

Chapter 2

Theory

The form of the full **SM** Lagrangian is dictated by the internal symmetries $SU(3) \times SU(2) \times U(1)$, where **SU** denotes the special unitary group with determinant one. In the SM, Quantum Chromodynamics (**QCD**) is governed by an $SU(3)$ symmetry whereas the symmetry $SU(2) \times U(1)$ acts on the Higgs field and Electroweak sector.

This chapter will focus on the Electroweak $SU(2) \times U(1)$ sector. However two points of QCD are particularly relevant: confinement and asymptotic freedom.

Confinement refers to the strength of the colour force as a function of distance, R , which increases with larger separation between quarks. This means that a lone colour charge can never be observed as in the limit of $R \rightarrow \infty$ the total energy, E , would also go as $E(R) \rightarrow \infty$. Instead, when two quarks separate, the most energetically favoured state is to create two new quarks from the resulting colour field. This process is referred to as hadronisation.

Asymptotic freedom is a product of the energy-dependent coupling strength in QCD. More specifically, it refers to the fact that the coupling strength becomes asymptotically weaker with increasing energy. Thus, if two quarks have a high enough energy they will feel almost no colour potential and can be treated using perturbation methods. Unfortunately, many of the experimentally interesting or accessible QCD processes are not in this limit. As a result either effective theories or non-perturbative

methods are needed to produce QCD predictions. These methods will be discussed in more detail in [section 2.4](#).

The rest of this chapter will focus on the Electroweak force, which is responsible for the Flavour (F) Changing Neutral Currents (**FCNC**) in the decay $A_b^0 \rightarrow p\pi^-\mu^+\mu^-$. These FCNC's are denoted as $\Delta F = 2$, given that two changes in quark flavour are required within the decay. There will be a brief explanation of how fermions and the Electroweak propagators acquire mass, followed by a discussion on flavour and the so-called Minimal Flavour Violation (**MFV**) hypothesis.

2.1 Local gauge invariance

The Lagrangian density for a free fermion in natural units is given as

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (2.1)$$

which yields the equation of motion

$$i\partial_\mu\bar{\psi}\gamma^\mu - m\psi = 0. \quad (2.2)$$

This is the Dirac equation describing a particle of spin $\frac{1}{2}$ and mass, m , here represented by the field ψ .

If the spinor ψ is transformed by a global phase such that $\psi \rightarrow e^{i\theta}\psi$, [Equation 2.1](#) remains invariant, because the phase term commutes past the partial derivative. However if the phase θ is dependent on space-time points, x^μ , such that $\psi \rightarrow e^{i\theta(x)}\psi$, the resulting Lagrangian transform is $\mathcal{L} \rightarrow \mathcal{L} - \partial_\mu\theta(x)\bar{\psi}\gamma^\mu\psi$. Thus [Equation 2.1](#) breaks *local gauge invariance*.

Requiring that a complete Lagrangian be invariant under local gauge transformations requires the addition of new gauge fields to [Equation 2.1](#).

Local gauge invariance can be regained by adding the new field, A_μ ,

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - (q\bar{\psi}\gamma^\mu\psi)A_\mu \quad (2.3)$$

and requiring A^μ to transform as $A^\mu \rightarrow A^\mu + \partial_\mu \lambda(x)$, where $\lambda(x) = \frac{\theta(x)}{q}$.

This gauge transformation can be expressed more concisely by using the covariant derivative defined as

$$D_\mu = \partial_\mu - iq\lambda(x)A_\mu \quad (2.4)$$

which transforms as the field itself, giving

$$\begin{aligned} \phi' &\rightarrow e^{-iq\lambda(x)}\phi \\ D'_\mu \phi'(x) &\rightarrow e^{-iq\lambda(x)}D_\mu \phi(x). \end{aligned} \quad (2.5)$$

In the case of QED, A^μ is the electromagnetic potential and $q = e$. The current density can be read off from the interaction term in the Lagrangian, $(q\bar{\psi}\gamma^\mu\psi)A_\mu$, as

$$J^\mu = q\bar{\psi}\gamma^\mu\psi. \quad (2.6)$$

Introducing the A^μ field to [Equation 2.1](#) also requires an additional new term describing a free photon. The free photon term consists of a kinematic part, $F^{\mu\nu}F_{\mu\nu}$, and a mass term, $-m^2A^\mu A_\mu$. Whereas the kinematic term for the free photon is locally gauge invariant, the mass term is not and this therefore forbids a photon mass.

Thus the construction of a locally gauge invariant U(1) QED Lagrangian is fairly straightforward because of certain key experimental observations:

1. The photon mass is observed to be zero. Thus the mass term of the boson, $-m^2A^\mu A_\mu$, which is not locally gauge invariant, does not break the gauge symmetry because $m = 0$.
2. The photon has been observed to be invariant under a parity transformation and has no Electromagnetic ([EM](#)) charge. This allows the EM force to be modelled using a simple U(1) symmetry, under which the fermion mass term, $-m\psi\bar{\psi}$, in [Equation 2.1](#), remains locally gauge invariant.

2.2 Generation of fermion masses

Both points made previously in [section 2.1](#) do not apply to the case of the full $SU(2) \times U(1)$ Electroweak interaction, due to a number of experimentally observed differences. Firstly,

all three gauge bosons, W^\pm and Z^0 , are massive. Secondly the W^\pm bosons are observed to violate parity, only interacting with left-handed particles, whereas the Z^0 boson is observed to interact with both right and left-handed particles.

In order to describe this behaviour, a model is required which allows for parity violation whilst still preserving Lorentz invariance. The fermion field can be written in terms of its left- and right-handed components by using the the projection operators, defined as

$$P_R = \frac{1 + \gamma^5}{2} \quad (2.7)$$

and

$$P_L = \frac{1 - \gamma^5}{2}. \quad (2.8)$$

Thus the fermion spinor can be written as

$$\psi = \frac{1 - \gamma^5}{2}\psi + \frac{1 + \gamma^5}{2}\psi = \psi_L + \psi_R \quad (2.9)$$

where $\bar{\psi}_L\psi_L = \bar{\psi}_R\psi_R = 0$ (as $P_i^2 = P_i$ and $P_R + P_L = 1$). This gives the mass term as

$$-m\bar{\psi}\psi = -m[\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R], \quad (2.10)$$

where ψ_L is an isospin doublet and ψ_R is an isospin singlet.

A gauge field which allows for this parity violation is also required. To achieve this, three gauge fields, W^{1-3} , are introduced. These fields are triplet or adjoint representations of the SU(2) group and only interact with left-handed particles in the SM.

The generators of the SU(2) group are the Pauli matrices give by:

$$\text{SU}(2) : \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.11)$$

which gives the SU(2) covariant derivative as

$$D_\mu = \partial_\mu + i\frac{g}{2}W_i^\mu\sigma^i = \partial_\mu + \begin{pmatrix} W_3^\mu & W_1^\mu - iW_2^\mu \\ W_1^\mu + iW_2^\mu & -W_3^\mu \end{pmatrix} = \partial_\mu + i\frac{g}{2} \begin{pmatrix} W_\mu^3 & \sqrt{2}W_\mu^+ \\ \sqrt{2}W_\mu^- & -W_\mu^3 \end{pmatrix} \quad (2.12)$$

where $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)$ and the \pm superscript on the W_μ^\pm indicates the electric charge carried by the gauge boson. The effect of the W_μ^\pm charged field on the fermion fields is therefore to change between the lower and upper components of the SU(2) left-handed doublets. Writing the SU(2) left-handed doublets as

$$q_{L_{u,d}} = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad l_{L_{e,\nu_e}} = \begin{pmatrix} e_L \\ \nu_L \end{pmatrix} \quad (2.13)$$

allows the charged currents to be written as

$$\begin{aligned} J_{W^+}^\mu &= \frac{1}{\sqrt{2}}(\bar{\nu}_L \gamma^\mu e_L + \bar{u}_L \gamma^\mu d_L) \\ J_{W^-}^\mu &= \frac{1}{\sqrt{2}}(\bar{e}_L \gamma^\mu \nu_L + \bar{d}_L \gamma^\mu u_L). \end{aligned} \quad (2.14)$$

These W_μ^\pm bosons represent the charged W bosons of the weak force. The SU(2) neutral current does not however transform between upper and lower SU(2) doublet components as the W_3^μ generator only has diagonal matrix-elements.

Since ψ_L is a weak isospin doublet ($\mathbf{I} = \frac{1}{2}$) but ψ_R is an isospin singlet ($\mathbf{I} = 0$) they will behave differently under rotations and thus the combination of left- and right-handed terms, $-m\psi\bar{\psi} = -m[\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R]$, is not gauge invariant. To regain gauge invariance in the Lagrangian, a very specific potential that keeps the full Lagrangian invariant under $SU(2) \times U(1)$ but which breaks the symmetry of the vacuum, is introduced. This potential is given by

$$V(\phi) = -\mu^2\phi^*\phi + \lambda|\phi^*\phi|^2. \quad (2.15)$$

This is the Higgs potential, where μ and λ are constants and where λ is the Higgs self-coupling. The complete Higgs Lagrangian is given as

$$\mathcal{L}_{Higgs} = -(D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) + \mathcal{L}_Y, \quad (2.16)$$

where \mathcal{L}_Y are referred to as the Yukawa interactions, discussed in detail at the end of this section. An SU(2) doublet of complex scalar fields can be represented as

$$\begin{aligned} \phi &= \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_2 \end{pmatrix} \\ \tilde{\phi} &= \begin{pmatrix} \phi^{0*} \\ \phi^{+*} \end{pmatrix}, \end{aligned} \quad (2.17)$$

where the superscripts on each component indicate the electric charge. Stable minima are found when $\mu^2 < 0$ which yields an infinite number of degenerate minima states which satisfy

$$\phi^\dagger \phi = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{\nu}{\sqrt{2}}. \quad (2.18)$$

Here ν is a real constant measured to be 246 GeV [7]. By choosing $\mu^2 < 0$, the vacuum expectation value (i.e the value of $|\phi|$ at the minima) is now non-zero and the symmetry has been spontaneously broken.

Without any loss of generality, the expectation of the Higgs field can be written as

$$\begin{aligned} \langle \phi \rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix} \\ \langle \tilde{\phi} \rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} \nu \\ 0 \end{pmatrix}. \end{aligned} \quad (2.19)$$

The introduction of the Higgs potential can now be used to generate the masses of the fermions whilst still preserving the local gauge invariance of the Lagrangian. This is done by combining the Higgs potential, as given in [Equation 2.19](#), with the left- and right-handed fermion fields to give terms of the form $L\phi\bar{R}$, an SU(2) singlet.

The term, $L\phi\bar{R}$, is referred to as a Yukawa interaction, i.e. an interaction between a Dirac and scalar field. The Yukawa interaction terms for quarks are given as

$$\mathcal{L}_Y^q = a_{ij}\bar{q}_{Li}\tilde{\phi}u_{Rj} + b_{ij}\bar{q}_{Li}\phi d_{Rj} + h.c. = \bar{u}_{Li}\frac{\nu}{\sqrt{2}}a_{ij}u_{Rj} + \bar{d}_{Li}\frac{\nu}{\sqrt{2}}b_{ij}d_{Rj} + h.c. \quad (2.20)$$

where a_{ij} and b_{ij} are the Yukawa couplings strengths and from which the following matrices are defined as

$$Y_{ij}^u = \frac{a_{ij}\nu}{\sqrt{2}}$$

and

$$Y_{ij}^d = \frac{b_{ij}\nu}{\sqrt{2}}. \quad (2.21)$$

Making the change of notation $u_{L,R} \rightarrow u'_{L,R}$, $d_{L,R} \rightarrow d'_{L,R}$ and defining

$$u'_L \equiv \begin{bmatrix} u'_L \\ c'_L \\ t'_L \end{bmatrix}, u'_R \equiv \begin{bmatrix} u'_R \\ c'_R \\ t'_R \end{bmatrix}, d'_L \equiv \begin{bmatrix} d'_L \\ s'_L \\ b'_L \end{bmatrix}, d'_R \equiv \begin{bmatrix} d'_R \\ s'_R \\ b'_R \end{bmatrix} \quad (2.22)$$

allows [Equation 2.20](#) to be re-written as

$$\mathcal{L}_{Yukawa}^q = -\frac{\nu}{\sqrt{2}} \bar{d}'_L Y_d d'_R + \bar{u}'_L Y_u u'_R + h.c. \quad (2.23)$$

The weak interaction eigenstates however, indicated by primed vectors in [Equation 2.23](#), are not the same as the Higgs-interaction (mass) eigenstates. In order to diagonalise $Y_{u,d}$ in the mass basis, a unitary transformation is performed, given as

$$q'_A = U_{A,q} q_A, \quad q = u, d, \\ \text{with} \\ A = R, L, \quad (2.24)$$

where primed vectors again indicate weak eigenstates and non-primed vectors indicate mass eigenstates.

This transformation allows the Lagrangian in [Equation 2.23](#) to be written in terms of the diagonalised Yukawa matrices:

$$\mathcal{L}_{Yukawa}^q = -\frac{\nu}{\sqrt{2}} \bar{d}_L U_{L,d}^\dagger Y_d U_{R,d} + \bar{u}_L U_{L,u}^\dagger Y_u U_{R,u} + h.c., \quad (2.25)$$

Thus, [Equation 2.25](#) now features the diagonalised matrices, $\frac{\nu}{\sqrt{2}} U_{L,u,d}^\dagger Y_{u,d} U_{R,u,d}$, which are referred to as the mass matrices, $m_{u/d_{ij}}$:

$$m_{u_{ij}} \equiv \frac{\nu}{\sqrt{2}} \begin{bmatrix} m_u & & & \\ & m_c & & \\ & & m_t & \\ & & & \end{bmatrix} \\ m_{d_{ij}} \equiv \frac{\nu}{\sqrt{2}} \begin{bmatrix} m_d & & & \\ & m_s & & \\ & & m_b & \end{bmatrix} \quad (2.26)$$

Using the vector definitions in [Equation 2.22](#) and inserting [Equation 2.24](#) into the expression for weak charged currents shown in [Equation 2.14](#) (in this example just considering the

positive current $J_W^{\mu+}$) gives:

$$\begin{aligned}
J_W^{\mu+} &= \frac{1}{\sqrt{2}} \bar{u}_L' \gamma^\mu d_L' = \frac{1}{\sqrt{2}} \bar{u}_L U_{L,u}^\dagger \gamma^\mu U_{L,d} d_L \\
&= \frac{1}{\sqrt{2}} \bar{u}_L \gamma^\mu (U_{L,u}^\dagger U_{L,d}) d_L \\
&= \frac{1}{\sqrt{2}} \begin{pmatrix} \bar{u}_L & \bar{c}_L & \bar{t}_L \end{pmatrix} \gamma_\mu V_{CKM} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} \\
&= \frac{1}{\sqrt{2}} \begin{pmatrix} \bar{u}_L & \bar{c}_L & \bar{t}_L \end{pmatrix} \gamma_\mu \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}.
\end{aligned} \tag{2.27}$$

As stated previously, there are no FCNC's due to only having diagonal terms in the W_μ^3 generator. This remains the case after diagonalising the mass matrices. Because there are no neutral, or $\Delta F = 1$, FCNC's in the SM, no FCNC can occur at tree-level and instead must be mediated via loop diagrams containing at least two W_μ^\pm vertices.

The elements of the CKM matrix, V_{CKM} , introduced in [Equation 2.27](#), have a hierarchical structure where the most off-diagonal terms (i.e. V_{td} and V_{ub}) are the smallest ($\sim 10^{-3}$) and the diagonal terms are the largest (\sim unity). The origin of the hierarchical structure of the CKM matrix is not understood in the SM.

As the CKM matrix is a 3×3 complex matrix there are initially 18 free parameters. The requirement of unitarity reduces this to nine. Each quark field could also go under a phase transform, $q_i \rightarrow e^{i\Phi_j} q_i$, such that the product $\frac{1}{\sqrt{2}} \bar{u}_L \gamma^\mu V_{L,u}^\dagger V_{L,u} u_L$ remains unchanged in the Lagrangian. This phase transformation affects the CKM matrix as:

$$V_{CKM} = V_{\alpha j} \rightarrow \begin{pmatrix} e^{-i\Phi_u} & & \\ & e^{-i\Phi_c} & \\ & & e^{-i\Phi_t} \end{pmatrix} \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} e^{i\Phi_d} & & \\ & e^{i\Phi_s} & \\ & & e^{i\Phi_b} \end{pmatrix} \rightarrow V_{\alpha j} e^{i(\Phi_j - \Phi_\alpha)} \tag{2.28}$$

As demonstrated in [Equation 2.28](#), a global phase transform of each quark gives five relative phases between the six quarks in the CKM matrix, reducing the nine degrees of freedom to

four. These four parameters can be expressed in the Wolfenstein parameterisation, which to third order in the parameter λ gives [8]

$$\begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} \quad (2.29)$$

where

$$\begin{aligned} A &= 0.808_{-0.015}^{+0.022}, \\ \lambda &= 0.2253 \pm 0.0007, \\ \rho &= +0.135_{-0.016}^{+0.031}, \\ \eta &= 0.349 \pm 0.017. \end{aligned} \quad (2.30)$$

A key area of flavour physics is focused on measuring the parameters of the CKM matrix to a high level of precision.

2.2.1 Introducing weak hypercharge

In the arguments laid out thus far, the three $SU(2)$ gauge bosons W_μ^{1-3} have provided a satisfactory model for the left-handed nature of the W^\pm bosons. The Z^0 boson interacts with both left- and right-handed fermions however, as does the massless photon of the A_μ gauge field.

A new massive gauge field B_μ is introduced, which transforms under a $U(1)$ symmetry. The group generator of this $U(1)$ symmetry is denoted Y and is generally referred to as weak hypercharge. As will be shown in the following section, the currents generated by the massive W_μ^3 and B_μ bosons mix to give the massless photon and massive Z^0 boson. The addition of the B^μ gauge boson to the theory gives the full $SU(2) \times U(1)$ covariant derivative in [Equation 2.16](#) as

$$D_\mu = \partial_\mu + ig\frac{1}{2}\vec{\sigma} \cdot \vec{W}_\mu + ig'\frac{1}{2}YB_\mu \quad (2.31)$$

where g and g' are free parameters.

2.3 The generation of gauge boson masses

Parts of the following sections are based on [lecture notes](#) provided by the Dutch National Institute for Subatomic Physics.

2.3.1 Illustrating the unitary gauge using a U(1) symmetry

The choice of $\mu^2 < 0$ in the Higgs Lagrangian results in a non-zero vacuum expectation value which leads to spontaneous symmetry-breaking. A consequence of this symmetry breaking is the generation of so-called Goldstone bosons, which can be transformed away under an appropriate gauge choice.

As an illustrative example of this, taking a simple U(1) complex scalar field, given by $\phi_{u(1)} = \phi_1 + i\phi_2$, the field $\phi_{u(1)}$ can be written in terms of the fields (ξ, η) which are shifted to the vacuum minima, as sketched in [Figure 2.1](#). This gives

$$\phi_0 = \frac{1}{\sqrt{2}}[(v + \eta) + i\xi], \quad (2.32)$$

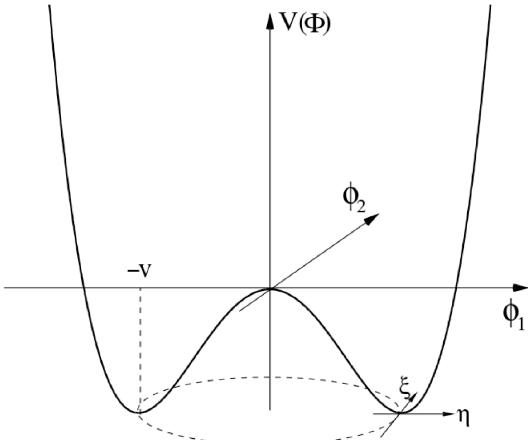
meaning that perturbations around the minima are no longer symmetric in η , as $V(\eta) \neq V(-\eta)$. A gauge can now be chosen such that the ξ terms vanish. This requires $\phi_{u(1)}$ to be rotated by $-\frac{\xi}{\nu}$. Assuming such a rotation is infinitesimal such that terms $\mathcal{O}(\xi^2 \eta^2 \xi)$ can be dropped leaves

$$\begin{aligned} \phi'_{u(1)} &\rightarrow e^{-i\xi/\nu} \phi_{u(1)} = e^{-i\xi/\nu} \frac{1}{\sqrt{2}}[(v + \eta) + i\xi] = \\ &e^{-i\xi/\nu} \frac{1}{\sqrt{2}}[v + \eta] e^{+i\xi/\nu} = \frac{1}{\sqrt{2}}[v + h] \end{aligned} \quad (2.33)$$

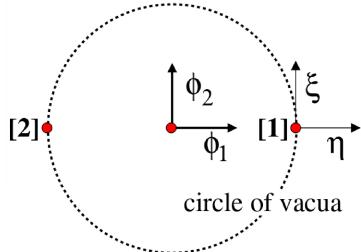
where h is a real field. This choice of gauge where the ξ terms disappear is referred to as the unitary gauge and the particle ξ (which is massless) is referred to as a Goldstone boson¹. The real field h corresponds to a massive scalar particle.

Goldstone's theorem states that for each broken generator of an original symmetry group a massless spin-zero particle will appear. In the Higg's mechanism, the gauge is

¹In the Higgs mechanism, under the full $SU(2) \times U(1)$ symmetry, there is more than one Goldstone boson



(a)



(b)

Figure 2.1: The resulting $V(\phi)$ potential for the case of $\mu^2 < 0$, (a) and the shift of fields to the minima, (b), [9].

transformed such that these spin-zero particles are removed and their degrees of freedom appear as the longitudinal components of other bosons. Therefore, if the symmetries associated with a gauge boson are broken then said gauge boson will acquire a mass via the Higgs mechanism².

²Although not demonstrated here, the broken symmetry in this U(1) example would give rise to a massive photon

2.3.2 Giving mass to the W^\pm and Z^0 bosons

In the case of $SU(2) \times U(1)$ symmetry there are 4 generators. These are

$$\left\{ \frac{\sigma_1}{2}, \frac{\sigma_2}{2}, \frac{\sigma_3}{2} - Y, \frac{\sigma_3}{2} + Y \right\}. \quad (2.34)$$

The last two generators in the list in [Equation 2.34](#) correspond to a mixing between the $SU(2)$ W_μ^3 and $U(1)$ B_μ^Y fields.

For the $SU(2) \times U(1)$ case, ϕ_0 is written in the unitary gauge as

$$\phi_0 = \begin{pmatrix} 0 \\ \nu + h \end{pmatrix}. \quad (2.35)$$

Invariance of ϕ_0 under a symmetry with a generator Z implies that $e^{i\alpha Z}\phi_0 = \phi_0$. Again dropping higher orders implies that for ϕ_0 to remain invariant under such a transformation requires $Z\phi_0 = 0$.

Thus it can be seen that for 3 of these generators $\left\{ \frac{\sigma_1}{2}, \frac{\sigma_2}{2}, \frac{\sigma_3}{2} - Y \right\}$ their associated gauge bosons will acquire mass:

$$\begin{aligned} \sigma_1\phi &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix} = +\frac{1}{\sqrt{2}} \begin{pmatrix} v+h \\ 0 \end{pmatrix} \neq 0 \rightarrow \text{broken} \\ \sigma_2\phi &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix} = -\frac{i}{\sqrt{2}} \begin{pmatrix} v+h \\ 0 \end{pmatrix} \neq 0 \rightarrow \text{broken} \\ \left[\frac{\sigma_3}{2} - Y \right] \phi_0 &= \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix} \neq 0 \rightarrow \text{broken}. \end{aligned} \quad (2.36)$$

The generator $\frac{\sigma_3}{2} + Y$ does not have an associated mass though as

$$\left[\frac{\sigma_3}{2} + Y \right] \phi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix} = 0 \rightarrow \text{unbroken}. \quad (2.37)$$

Therefore, by mixing the third component of $SU(2)_L$ with $U(1)_Y$, there is one broken generator and one unbroken generator. These correspond to the photon and the Z^0 boson.

Writing the Z^0 and γ in terms of B_μ^Y and W_μ^3 gives

$$Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}} (gW_\mu^3 - g'B_\mu^Y) \quad (2.38)$$

and

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}} (gW_\mu^3 + g'B_\mu^Y) \quad (2.39)$$

where g and g' are free parameters as in [Equation 2.31](#). Substituting

$$W_\mu^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}} \quad (2.40)$$

and expanding [Equation 2.31](#) gives :

$$(D^\mu\phi)^\dagger(D_\mu\phi) = \frac{1}{8}\nu^2[g^2(W_\mu^+)^2 + g^2(W^{-\mu})^2 + (g^2 + g'^2)Z_\mu^2 + 0.A_\mu^2], \quad (2.41)$$

from which the masses of the gauge bosons can be read off as $\frac{1}{2}M^2V_\mu^2$ giving

$$\begin{aligned} M_\gamma &= 0 \\ M_{W_-} &= M_{W_+} = \frac{1}{2}\nu g \\ M_Z &= \frac{1}{2}\nu\sqrt{g^2 + g'^2}. \end{aligned} \quad (2.42)$$

This leads to the expression

$$\frac{M_W^2}{M_Z^2 \cos^2(\theta_W)} = 1,$$

where

$$\cos(\theta_W) = \frac{M_W}{M_Z}. \quad (2.43)$$

Thus in summary, the degrees of freedom given by the Goldstone bosons that were generated when the symmetry was broken provide the longitudinal components of the W^\pm and Z^0 bosons, allowing them to acquire mass, whilst still preserving local gauge invariance. The massive scalar field h is the Higgs boson and has a mass given as

$$m_h = \sqrt{2\lambda\nu^2}. \quad (2.44)$$

The SM does not predict the values of g and g' . Experimental measurements of the masses give:

$$\begin{aligned} M_W &= 80.385 \pm 0.015 \text{ GeV}/c^2 \\ M_Z &= 91.1875 \pm 0.0021 \text{ GeV}/c^2 \end{aligned} \quad (2.45)$$

and

$$M_{\text{Higgs}} = 125.09 \pm 0.24 \text{ GeV}/c^2, \quad (2.46)$$

where all mass values are taken from [7]. The fact that the W^\pm and Z^0 propagators are massive explains why the weak force is comparatively weak compared to the Electromagnetic force.

2.4 Form factors for hadronic transitions

Electroweak decays such as $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ also have contributions from non-perturbative QCD contributions. These QCD contributions are expressed as form factors, which describe the QCD effects as a generic functional form and are dependent on the final and initial hadron states. As these are non-perturbative they are difficult to calculate and there are various models and approximations used to do so. The most relevant for $b \rightarrow dll$ decays, where l refers to any lepton, are Heavy Quark Effective Theory (HQET) and lattice QCD. Heavy Quark Effective Theory is sufficient when the transitioning quarks in both the initial and final state hadron are much heavier than the spectator quarks. In this scenario, instead of the light quark interacting directly with the heavy quark, the light quark can be treated as interacting with a colour potential whose source is the (effectively stationary) heavy quark. It is assumed that this potential is unchanged when the quark transition occurs. This works very well for $b \rightarrow c$ transitions but less well for $b \rightarrow s$ and $b \rightarrow d$ transitions. The HQET can be combined with lattice QCD for light quark transitions however, as used in the calculation of the differential branching fraction of $\Lambda_b^0 \rightarrow \Lambda^0\mu^-\mu^+$ [10].

Lattice QCD can be used in the case when there is no appropriate effective theory or perturbative alternative. The idea of lattice QCD is to express the matrix-elements of interest as correlation functions, and then use numerical integration to solve for the path integral required to evaluate these correlation functions. This numerical integration is carried out on a grid or lattice of points in space time. At each lattice site a field, representing a quark, is defined and the link between each site represents the gluon. Monte Carlo methods are then used to evaluate the path integrals, or gauge links, for different lattice configurations. The fact that there is a finite spacing, given by the lattice spacing, means that discrete, as opposed to continuous, symmetries are involved and the theory remains renormalizable. Lattice QCD has proved to a powerful technique, for example, it has allowed the proton mass to be calculated to a value within 2% of its measured value [11].

2.5 The flavour problem

Loop-level FCNC's, like the decay $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$, contain both QCD and weak contributions. Due to the large separation in distance and time scales of these two forces, the total Lagrangian can be written as an effective theory whereby the physics can be separated at a certain energy, μ , such that the effective Hamiltonian is factorised into the short distance (energy $> \mu$) and long distance (energy $< \mu$) parts. The loop contributions within the FCNC decay can then be integrated out and replaced with effective couplings, parameterised by the so-called Wilson coefficients. The Wilson coefficients can be calculated precisely in the SM and NP models.

Some of the tightest limits on NP models with generic flavour structure come from meson oscillations, such as $K^0 - \bar{K}^0$ and $B_{(d/s)}^0 - \bar{B}_{(d/s)}^0$ oscillations. The diagrams for B oscillations are shown in Figure 2.2. In the case of kaon mixing the b quark is replaced

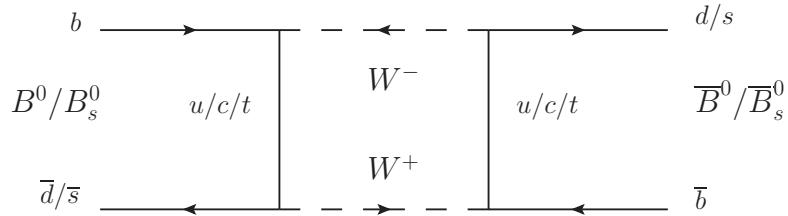


Figure 2.2: The lowest order Feynman diagram for B^0 and B_s^0 mixing.

with an s quark in Figure 2.2. Such mixing occurs via FCNC processes and hence cannot occur at tree level. Given that experimental measurements of such meson mixing are in agreement with the SM predictions [12], it must be the case that $|\mathcal{A}_{NP}^{\Delta F=2}| < |\mathcal{A}_{SM}^{\Delta F=2}|$,

where \mathcal{A} indicates the amplitude. This sets the mass scale of NP, Λ , to be [6]

$$\Lambda \gtrsim \begin{cases} 1.3 \times 10^4 \text{ TeV} \times |c_{sd}|^{1/2} \\ 5.1 \times 10^2 \text{ TeV} \times |c_{bd}|^{1/2} \\ 1.1 \times 10^1 \text{ TeV} \times |c_{bs}|^{1/2} \end{cases}, \quad (2.47)$$

where c_{ij} refers to the coupling strength of the NP physics contribution between the i^{th} and j^{th} quark flavour. Assuming a generic structure where c_{ij} is unity (where i, j indicate a u -, d -type quark) sets a value for Λ beyond the energy scales accessible to current accelerators. Alternatively, if $\Lambda \sim 1 \text{ TeV}$ then $c_{ij} \leq 10^{-5}$. This could indicate that NP exists at $\Lambda \sim 1 \text{ TeV}$ but that the coupling constants for NP contributions to $\Delta F = 2$ operators have a strong hierarchy (and thus these particles have evaded detection thus far) or that the coupling constants are generic (i.e. no hierarchy) but NP lies at a much higher energy scale, or some combination of both these factors.

There is a scheme however which would allow for a $\sim \text{TeV}$ NP energy scale and accommodate the experimental constraints on $\Delta F = 2$ processes. This is to assume that the unique source of flavour symmetry breaking beyond the SM is also from the Yukawa couplings. This is an attractive solution as it naturally gives small effects to $\Delta F = 2$ processes. This hypothesis is referred to as the Minimal Flavour Violation (MFV) hypothesis. MFV does not however offer an explanation for the observed pattern of masses and mixing angles of quarks.

The assumption that NP is minimally flavour violating can be tested by verifying that the values of the CKM matrix-elements measured in decays mediated via loops are compatible with those measured in tree-level decays. This comparison is interesting because it could be possible for NP particles to appear virtually in higher-order decays, but not at tree level, which could cause discrepancies between the two.

2.6 Using $b \rightarrow d l l$ decays to search for new physics

As discussed in section 2.5, precision measurements of CKM elements can be used as a probe for potential new physics effects and to test the MFV hypothesis. Assuming the effects of any new physics are small, the more highly suppressed decays in the SM may be more sensitive to new physics. This means rarer decays, mediated via smaller CKM elements, could provide a good handle on NP. The CKM element V_{td} is the second smallest element in the CKM matrix, being off-diagonal in both matrix indices. It is also useful to compare V_{td} decays against their V_{ts} equivalent because by calculating the ratio, $|V_{td}/V_{ts}|$, some experimental and theoretical errors cancel. The most precise measurement of $V_{tq=d,s}$ comes from $B_q^0 - \bar{B}_q^0$ mixing, the Feynman diagrams for which are shown in Figure 2.2. The diagrams for the internal u -type quarks appearing in Figure 2.2 will destructively interfere giving amplitudes that go as $m_t^2 - m_u^2$. This means that diagrams featuring an internal top quark dominate due to the large mass difference between the top quark and other quarks. This gives

$$\begin{aligned}\Delta M_{B_d} &\sim m_t^2 |V_{tb} V_{td}|^2 \sim m_t^2 \mathcal{O}(\lambda^6), \\ \Delta M_{B_s} &\sim m_t^2 |V_{tb} V_{ts}|^2 \sim m_t^2 \mathcal{O}(\lambda^4),\end{aligned}\tag{2.48}$$

where λ is the Wolfenstein parameterisation, see section 2.2.

Similarly, semileptonic $b \rightarrow d(s)ll$ FCNC decays will also be sensitive to the V_{tq} CKM elements, via both box and loop diagrams. The Feynman diagrams for the semileptonic $b \rightarrow d(s)ll$ decays, $B^- \rightarrow \pi^-(K^-)\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow p\pi(K)\mu^+\mu^-$, are shown in Figure 2.3.

2.6.1 Measuring V_{ts} and V_{td} using tree-level process

It is difficult to measure precisely the CKM elements, $V_{tq=d,s}$, using tree-level decays, as the rates for $t \rightarrow d, s$ processes are small and final states involving d and s quarks are difficult to detect in hadron colliders.

The values of V_{tq} for tree-level processes are inferred by using the unitary constraints on the CKM triangle [13]. This allows a comparison between values for V_{tq} obtained via

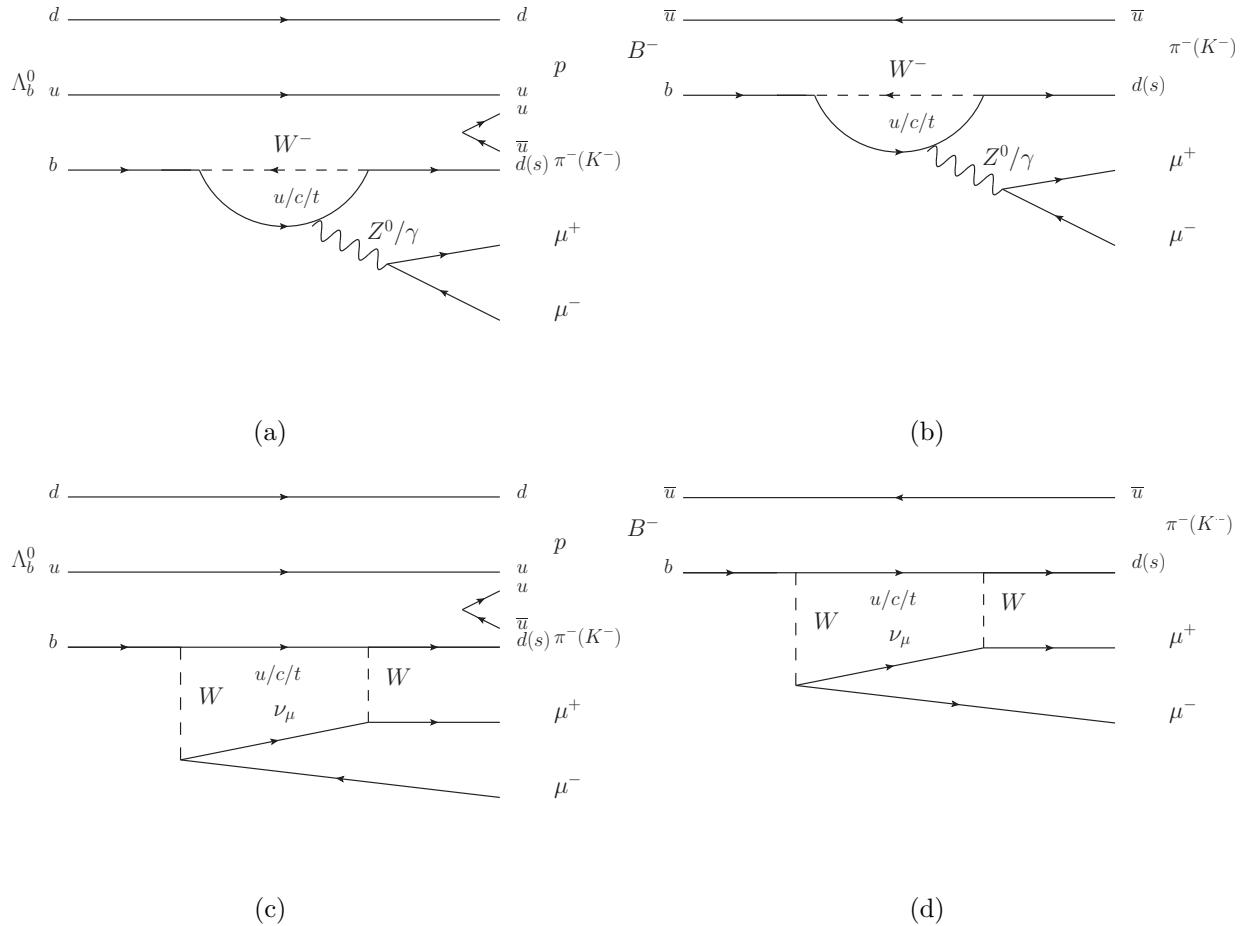


Figure 2.3: Feynman diagrams for $\Lambda_b^0 \rightarrow p \pi^-(K^-) \mu^+ \mu^-$ via a loop, (a), $B^- \rightarrow \pi^-(K^-) \mu^+ \mu^-$ via a loop, (b), $\Lambda_b^0 \rightarrow p \pi^-(K^-) \mu^+ \mu^-$ via a box diagram, (c), $B^- \rightarrow \pi^-(K^-) \mu^+ \mu^-$ via a box diagram, (d).

tree-level and higher order decays.

2.6.2 Discrepancies between tree-level and higher-order decays

The observed values for the mass differences, measured using B_q^0 mixing, are given below in terms of the corresponding $B_q^0 - \overline{B_q^0}$ oscillation frequency [7] [14]

$$\begin{aligned} \Delta M_{B_d} &= (0.5055 \pm 0.0020) \text{ ps}^{-1} \\ \Delta M_{B_s} &= (17.757 \pm 0.021) \text{ ps}^{-1}. \end{aligned} \tag{2.49}$$

The largest source of error on the values of V_{tq} , extracted from the measured values of ΔM_{B_q} shown in [Equation 2.49](#), are due to theoretical uncertainties on the hadronic B_q^0 -mixing matrix-elements.

Recent work [15] has used improved techniques within the context of lattice QCD to increase the precision on these hadronic matrix-element calculations. Using the experimental values shown in [Equation 2.49](#), along with the increased precision on matrix element calculations, the values of $|V_{td}|$ and $|V_{ts}|$ are calculated in Ref. [15] to give

$$\begin{aligned} |V_{td}| &= (7.94 \pm 0.31 \pm 0.08) \times 10^{-3}, \\ |V_{ts}| &= (38.8 \pm 1.1 \pm 0.4) \times 10^{-3}, \\ |V_{td}/V_{ts}| &= 0.2047 \pm 0.0029 \pm 0.0010, \end{aligned} \quad (2.50)$$

where the errors are from the lattice mixing matrix-elements and other theory assumptions and the measured value of ΔM_q . The error on $|V_{td}/V_{ts}|$ is smaller than that of its products in the case of B_q^0 -mixing, due to the cancellation of some dominating errors on the B_d^0 - and B_s^0 -mixing hadronic matrix-elements.

The values in [Figure 2.4](#) show $|V_{td}/V_{ts}|$ and $|V_{tq=d,s}|$, calculated using ΔM_q , with the improved mixing matrix-elements implemented, along with the average given by the Particle Data Group (PDG) [7] before this improvement. The values for $|V_{td}/V_{ts}|$ and $|V_{tq}|$ as determined from $B \rightarrow \pi^+(K^+)\mu^-\mu^+$ decays are also shown, along with the total combined values for $|V_{td}/V_{ts}|$ and $|V_{tq}|$, calculated using CKM unitary conditions for tree-level processes (labelled tree) and all processes (labelled full). A full discussion of using semileptonic rare decays to measure $|V_{td}/V_{ts}|$ follows in [section 2.7](#) and [section 2.8](#).

Using these improved values for $|V_{td}|$, $|V_{ts}|$, the analysis in Ref. [15] finds that the values for $|V_{td}|$ and $|V_{ts}|$ from ΔM_q measurements lie 2.4σ and 1.3σ , respectively, away from those deduced from tree-level processes, with a difference of 3σ in the case of the ratio. For the case of the semileptonic B decays, the values of $|V_{td}|$ and $|V_{ts}|$ lie 2.0σ and 2.9σ respectively below those deduced from tree values. There is therefore mild tension between values for $|V_{td}|$ and $|V_{ts}|$ taken from tree-level processes and those taken from decays mediated by FCNC's. Despite the improvements in Ref. [15], the dominant error

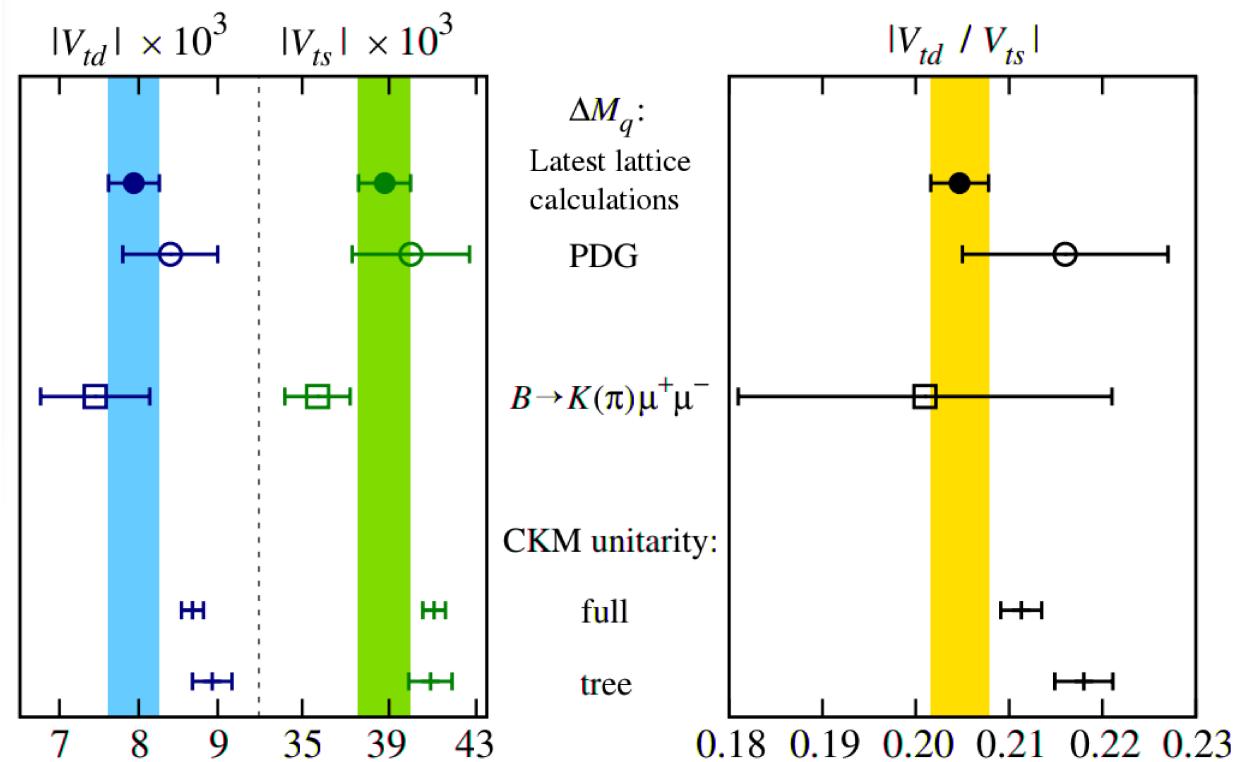


Figure 2.4: Comparing measurements of V_{td} , V_{ts} , $|V_{td}/V_{ts}|$ using (top to bottom) current experimental measurements of $\Delta M_{q=d/s}$ with the latest theory calculations, as given in the PDG [7], using semileptonic decays, and the combined value for tree and loop level processes, using unitary constraints (CKM fitter [13]). Plot taken from Ref. [15].

on the values of $|V_{td}|$, $|V_{ts}|$, calculated using ΔM_q , is still due to the hadronic B_q^0 -mixing matrix-elements.

An alternative method to improve the understanding of this discrepancy between loop- and tree-level $|V_{td}|$, $|V_{ts}|$ values is to introduce more decays mediated via FCNC's which are sensitive to $|V_{td}|$ or $|V_{ts}|$, such as the semileptonic decay $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ for $|V_{td}|$ and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ for $|V_{ts}|$.

2.7 Measuring V_{td} using semileptonic mesonic decays

In Ref. [16], the value of $\mathcal{B}(B^+ \rightarrow \pi^+\mu^+\mu^-)$ is measured to be $(1.83 \pm 0.25) \times 10^{-8}$. This channel is the equivalent process of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ but with one fewer spectator quark, as demonstrated in Figure 2.3. In Ref. [16], the combination of $\mathcal{B}(B^+ \rightarrow \pi^+\mu^+\mu^-)$ with $\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-)$ [17] is used to calculate:

$$\frac{|V_{ts}|}{|V_{td}|} = \frac{\mathcal{B}(B \rightarrow K^+\mu^-\mu^+)}{\mathcal{B}(B \rightarrow \pi^+\mu^-\mu^+)} \times \frac{\int F_K dq^2}{\int F_\pi dq^2} = 0.24^{+0.05}_{-0.04}, \quad (2.51)$$

where $F_{K/\pi} dq^2$ is a combination of Wilson coefficients, phase space factors and form factors. In order to extract $|V_{td}/V_{ts}|$ it is necessary to calculate $F_{K/\pi} dq^2$, which requires knowledge of the relevant form factors. There has however been much theoretical work around this decay channel [12, 18–22]. The study in Ref. [16] takes the form factors for $B \rightarrow \pi$ from Ref. [12] and Ref. [22] and the form factors for $B \rightarrow K$ from Ref. [23]. This provides the most accurate determination of $|V_{td}/V_{ts}|$ from a decay that is mediated via both penguin and box diagrams.

2.8 Measuring V_{td} using semileptonic baryonic decays

More measurements of $|V_{td}/V_{ts}|$ are needed to resolve the discrepancies between the value of $|V_{td}/V_{ts}|$ measured via either tree-level or loop-level processes. There is currently ongoing work to measure $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and the work in this thesis outlines the search for, and branching fraction measurement of, $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$. Prior to this work, there has not been an observation of a $b \rightarrow dll$ process in the baryonic sector. One challenge with this analysis is that there are no theory predictions for the underlying form factors. This means that any simulation used will not correctly model the distribution of the invariant mass squared of the dimuon system, q^2 , nor that of the dihadron mass, $m_{p\pi}$.

The branching fraction as a function of q^2 has been calculated however for the decay $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ [10]. The $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ q^2 distributions for theory and experiment are

shown in [Figure 2.5](#).

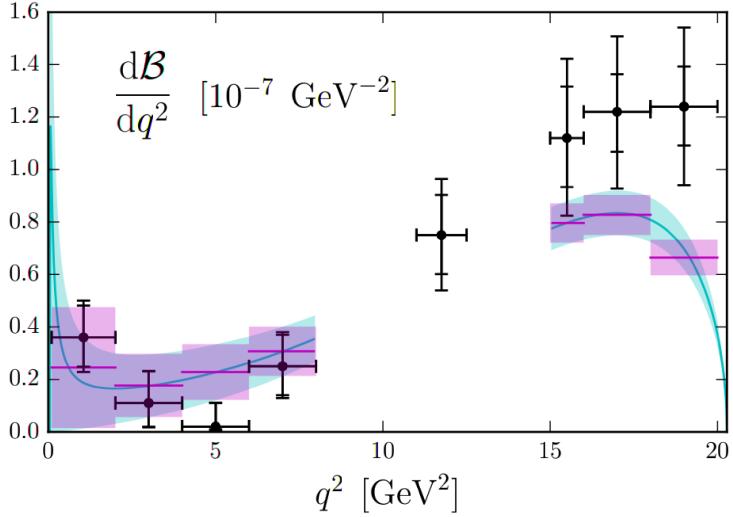


Figure 2.5: Branching fraction as a function of q^2 for $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ decays [\[24\]](#).

In the absence of a reliable model for the q^2 distribution of $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ decays, the q^2 distribution of the decay $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ is used as a proxy for $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$. For both $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ decays the maximum value of q^2 , denoted Q^2 , is similar:

$$Q_{\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-}^2 = [M_{\Lambda_b^0} - \sqrt{(M_p^2 + M_\pi^2 + 2M_p M_\pi)}]^2 \sim 20.75 \text{ GeV}/c^2 \quad (2.52)$$

$$Q_{\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-}^2 = [M_{\Lambda_b^0} - M_{\Lambda^0}]^2 \sim 20.30 \text{ GeV}/c^2.$$

A key difference however is the presence of resonances in the case of $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$. There is the possibility of the $p\pi$ coming from an N^* resonance giving $\Lambda_b^0 \rightarrow N^*(\rightarrow p\pi) \mu^+ \mu^-$. All N^* resonances have isospin $I = 1/2$. The equivalent process via the Δ baryons ($I = 3/2$), i.e. $\Lambda_b^0 \rightarrow \Delta^0(\rightarrow p\pi) \mu^+ \mu^-$, is forbidden due to the need of the spectator quarks (which only interact via the strong interaction) to conserve isospin. The properties of different N^* states are shown in [Table 2.1](#). The lightest N^* has a mass of $1440 \text{ MeV}/c^2$.

In this analysis the q^2 distribution is modelled using the theoretical predictions for $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ decays and the q^2 distribution taken from $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ data, although

J^p	Mass/ MeV/ c^2	J^p	Mass/ MeV/ c^2
$\frac{1}{2}^+$	1440	$\frac{1}{2}^+$	1710
$\frac{3}{2}^-$	1520	$\frac{3}{2}^+$	1720
$\frac{1}{2}^-$	1535	$\frac{3}{2}^-$	1875
$\frac{1}{2}^-$	1650	$\frac{3}{2}^+$	1900
$\frac{5}{2}^-$	1675	$\frac{7}{2}^-$	2190
$\frac{5}{2}^+$	1680	$\frac{9}{2}^+$	2220
$\frac{3}{2}^-$	1700	$\frac{9}{2}^-$	2250

Table 2.1: Showing the masses for different possible N^* resonances

the low statistics are problematic in the latter case. The effect of the different q^2 proxy choices on the final result is taken as a systematic uncertainty (see [chapter 8](#)).

Chapter 3

The LHCb detector

The Large Hadron Collider (LHC) is a particle accelerator which accelerates protons along a $\sim 27\text{km}$ long tunnel which runs under the Franco-Swiss border. There are four detectors located at four different points on the [LHC](#). These are [ATLAS](#), [CMS](#), [ALICE](#) and [LHCb](#). At each one of these four points along the LHC, the protons beams are collided at a pp collision or interaction point.

In this section, there will be a general overview of the LHCb detector, highlighting the aspects of the detector's design which allow for good b -physics performance. This is followed by a more detailed break down of the subdetectors relevant for this thesis, as well as an overview of the trigger system.

The [LHCb](#) detector is designed primarily to study b -physics. A profile of the detector is shown in [Figure 3.1](#). At the LHC, the majority of $b\bar{b}$ quark pairs are produced at low values of θ , where θ is defined as being equal to 0 along the z axis in [Figure 3.1](#). The LHCb detector is a single-arm spectrometer in the sense that there is no instrumentation about $\theta = \pi$.

By placing the detector instrumentation along the beamline, the LHCb detector exploits the fact that the $b\bar{b}$ quark pairs are produced at low values of θ , or high values of pseudorapidity, η , where

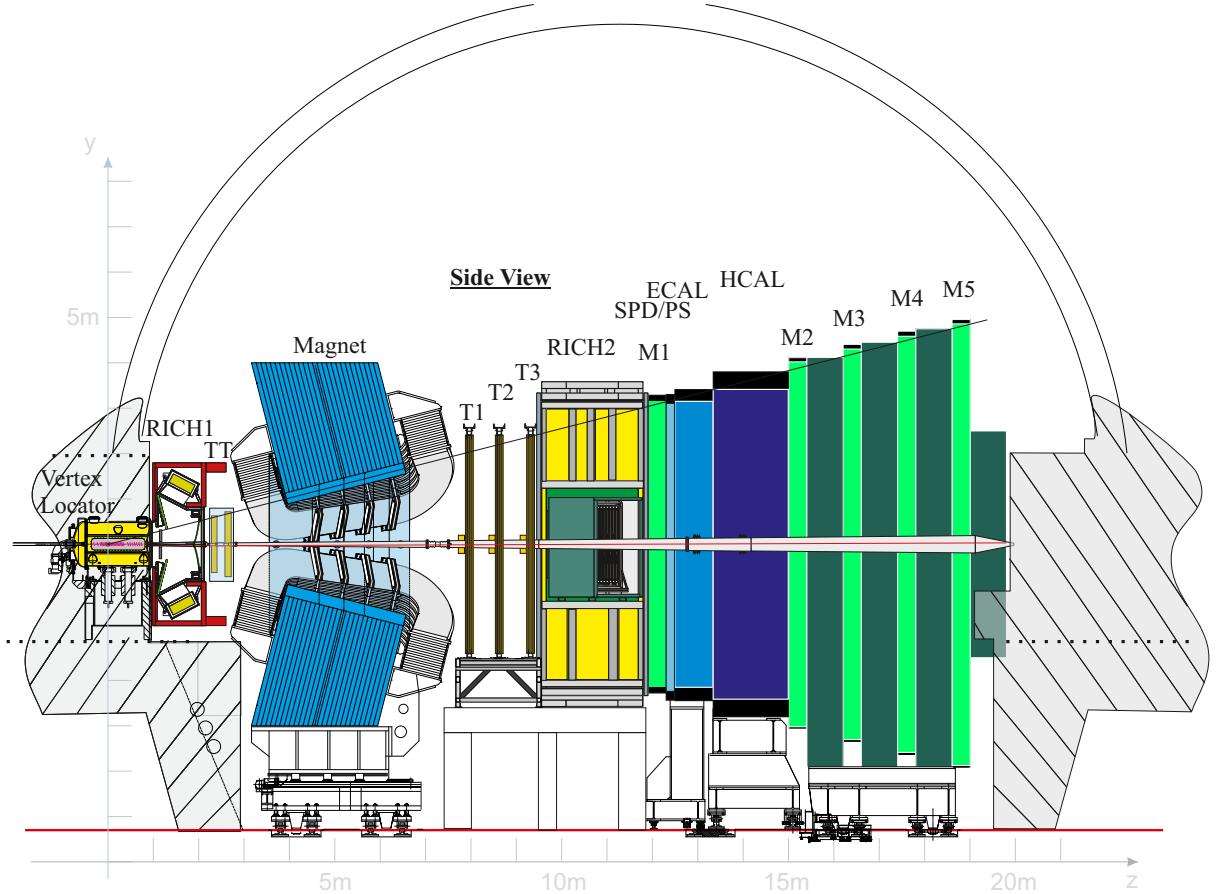


Figure 3.1: A profile of the LHCb detector [25]. Any reference to x, y, z directions in this thesis refers to this diagram. The interaction point is located inside the Vertex Locator.

$$\eta = -\ln[\tan \frac{\theta}{2}]. \quad (3.1)$$

The production of $b\bar{b}$ pairs as a function of η can be seen in Figure 3.2. The LHCb coverage in η ($2 < \eta < 5$) is also shown in Figure 3.2, along with a comparison of the η coverage of the general purpose (GP) detectors, ATLAS and CMS. Despite LHCb only covering 1.8% of the total solid angle, 27% of $b\bar{b}$ pairs fall within the detector acceptance [26].

A key method for distinguishing b -hadrons from other particles is to measure the displacement of particles in the detector, as b -hadrons tend to be highly boosted. A high

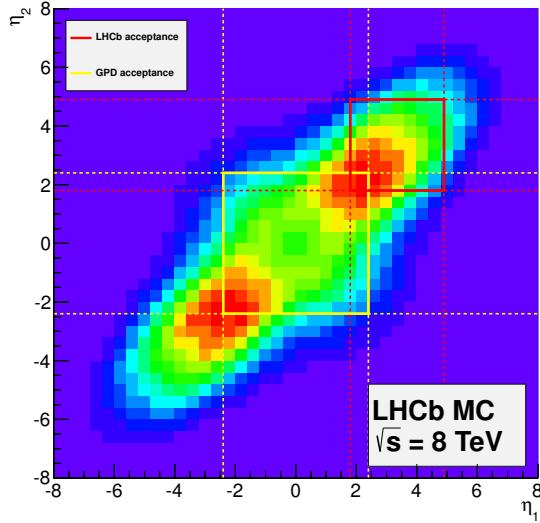


Figure 3.2: The production of $b\bar{b}$ quarks as a function of η . The acceptance for LHCb and General Purpose (**GP**) detectors are overlaid on top. Figure taken from the LHCb public page.

number of tracks at any time makes obtaining a good resolution on the displacement of these vertices difficult. As a consequence, the luminosity collected by the LHCb detector is lower than other LHC experiments. At the LHCb experiment an integrated luminosity of 1 fb^{-1} of data was collected during 2011, with $\sqrt{s} = 7 \text{ TeV}$. During 2012, 2 fb^{-1} of data was collected, with $\sqrt{s} = 8 \text{ TeV}$. The data collected during the years 2011 and 2012 at the LHC is referred to as Run-1 data. Data collected during the years 2015 and 2016 is referred to as Run-2 data.

The detector subsystems, all marked in [Figure 3.1](#), can be outlined as follows:

VELO The Vertex Locator system (**VELO**) is a high precision tracking system which provides good resolution on the pp interaction vertex, or primary vertex (**PV**), and flight distance (**FD**). The resolution of the flight distance (i.e. how far a particle flies before decaying) for particles produced at the PV (so-called **prompt** decays) is a function of the PV and the Secondary Vertex (SV). The VELO is discussed in more detail in [section 3.3](#).

TT and T stations The tracking system is completed by the Tracker Turicensis (TT) station and Tracking stations (T stations), which lie upstream and downstream of the magnet respectively. The tracking system is discussed in more detail in [section 3.4](#)

RICH 1,2 The two Ring Imaging Cherenkov detectors (**RICH**) provide identification of charged hadrons by using Cherenkov radiation. The RICH information is important in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis, to help distinguish pions and protons from kaons. The RICH is discussed in more detail in [section 3.5](#).

The ECAL, HCAL and SPD/PS The calorimeter system provides identification of photons, electrons (**ECAL**) and hadrons (**HCAL**) as well as giving information on their position and energy. The pre-shower (**PS**) detectors and Scintillator Pad Detectors (**SPD**) are used to help electron identification. The information provided by the calorimetry system is less important to the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis than other subdetectors and is not discussed further.

Muon Chambers The muon chambers identify muons by employing large amounts of iron, which stop most particles, but not muons. The process of muon identification is discussed in more detail in [section 3.6](#).

In addition to these subdetectors there is also the magnet, as marked in [Figure 3.1](#), which provides a strong magnetic field, bending charged particles in the $x - z$ plane.

Before discussing the detectors subsystems in detail the key features of the trigger are outlined. The purpose of the trigger is to reduce the amount of data read out from the detector to a manageable level, whilst still preserving the most interesting physics events.

3.1 Trigger

The LHCb trigger consists of two levels, the hardware trigger , referred to as the Level-0 (**L0**) trigger and the software trigger referred to as the High Level Trigger (**HLT**). The L0 trigger makes decisions based on information from the muon systems and the calorimeter

and reduces the rate of events to below 1 MHz, at which point the whole detector can be read out. The HLT reduces the rate down to 5 kHz, at which point events can be stored. An overview of the HLT algorithms and the rates involved in the entire trigger process in 2012 can be seen in [Figure 3.3](#).

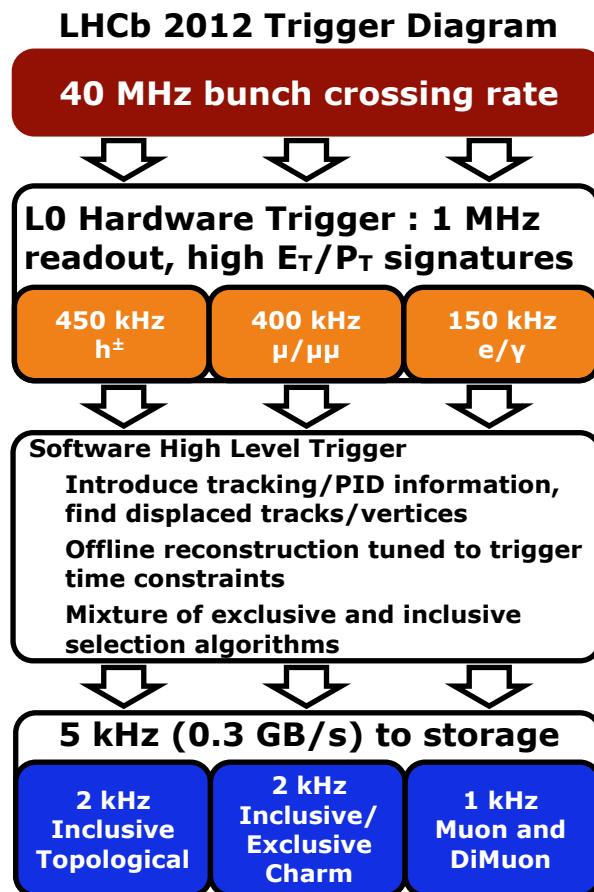


Figure 3.3: Overview of the LHCb trigger rates. Figure taken from the LHCb public page.

3.1.1 L0 trigger

The L0 trigger is divided into the L0-Calorimeter trigger, the L0-Muon trigger and L0-PileUp trigger, the latter being used only for the determination of the luminosity. The L0 trigger is fully synchronised with the 40 MHz bunch crossing rate of the LHC.

The events that pass the LHCb trigger can be split into three categories. Triggered On

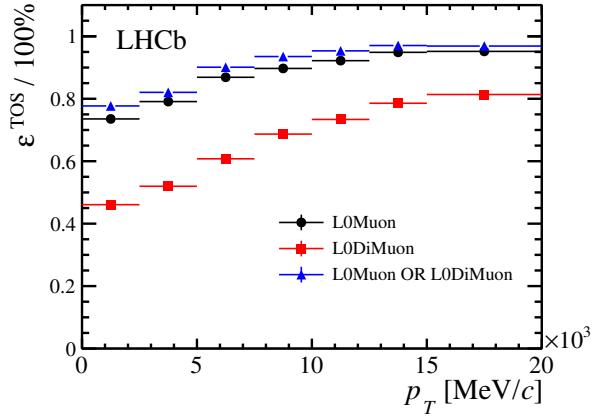


Figure 3.4: The 2011 L0Muon and L0DiMuon lines efficiency, ϵ_{TOS} , on $B^+ \rightarrow J/\psi (\mu^+\mu^-)K^+$ as a function of P_T (J/ψ) [29].

Signal (**TOS**) events are those for which the presence of the signal candidate is sufficient to fire the trigger. Triggered Independent of Signal (**TIS**) are those events whereby the trigger is fired without using any of the hits associated with the signal particles. Finally, there are events that are neither TIS nor TOS, i.e. neither the presence of the signal alone nor the rest of the event alone are sufficient to fire the trigger, but rather both are necessary [27]. These are referred to as **TISTOS** events.

In the $A_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis a TOS is required on the L0 line L0Muon, which requires the P_T of single muon in an event to be above 1.48 GeV/ c in 2011 data [28] and 1.76 GeV/ c in 2012 data.

The TOS efficiency of the 2011 L0Muon line, calculated using $B^+ \rightarrow J/\psi (\mu^+\mu^-)K^+$ data via the so-called TISTOS technique, outlined in subsection 3.2.1, is shown in Figure 3.4, as a function of the P_T of the J/ψ .

3.2 The HLT trigger

The HLT trigger consists of two levels, Level-1 and Level-2, referred to as HLT1 and HLT2 respectively. The HLT1 performs a partial event reconstruction and an inclusive selection of signal candidates, reducing the rate to 80 kHz. The HLT2 performs a full event

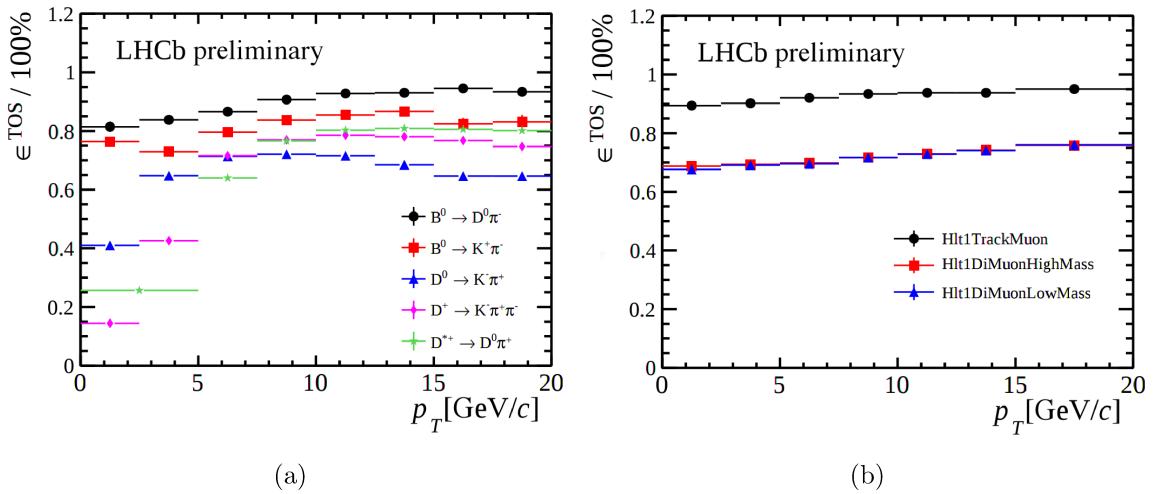


Figure 3.5: The `Hlt1TrackAllL0` TOS efficiency (ϵ_{TOS}) of the decays indicated in the legend as a function of the P_T of the relevant B or D meson, (a), the `Hlt1TrackMuon` (and `Hlt1DiMuonHighMass`, `Hlt1DiMuonLowMass`) ϵ_{TOS} as a function of the P_T of the J/ψ (b). Both plots show 2012 data [30].

reconstruction which is close to that used for the off-line reconstruction and reduces the rate to 5 kHz.

The HLT1 trigger decisions most relevant to the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis are the `Hlt1TrackAllL0` and `Hlt1TrackMuon` lines. The `Hlt1TrackAllL0` line selects VELO track candidates based on their P_T and displacement from the primary vertex. If the selected VELO track also matches with hits in the muon chamber and has a $P_T > 1 \text{ GeV}/c$ then it will be selected by the `Hlt1TrackMuon` line [30]. The efficiency of the `Hlt1TrackAllL0` and `Hlt1TrackMuon` trigger lines can be seen in Figure 3.5.

The HLT2 trigger consists of so-called topological lines, named `Hlt2Topo(N)Body`, which are designed to trigger on partially reconstructed b -hadron decays. The selection of partially reconstructed decays allows heavy flavour decays to be selected even in the cases where not all the final state particles are reconstructed. To allow for an inclusive cut on partially reconstructed events, a corrected mass $\textcolor{violet}{m}_{\text{corr}}$ is used that takes into account the amount of missing transverse-momentum with respect to the flight of the mother ($P_{T'miss}$).

This is defined as

$$m_{corr} = \sqrt{m^2 + P_{T'miss}^2} + P_{T'miss}. \quad (3.2)$$

The topological lines use a Boosted Decision Tree (**BDT**) [31] to discriminate between signal and background. A BDT employs multivariate analysis techniques to combine a set of discriminating variables into a single discriminating variable. A more detailed overview of BDTs is given in Appendix A. Due to the detector resolution and the need to account for any variation in, for example, the detector efficiency, a BDT which receives input variables that have been discretised is preferred. These BDTs which take discrete input variables are referred to as a Bonsai BDT (**BBDT**) [32]. The BBDT lines used in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis are `Hlt2Topo2BodyBBDT` and `Hlt2TopoMu2(3)BodyBBDT`, which select two or three body candidates respectively. The variables used in the BBDT are $\sum |P_T|$, P_T^{min} , *mass*, m_{corr} , **DOCA**, **$IP\chi^2$** and **$FD\chi^2$** [33] [34]. The $FD\chi^2$ is defined for two vertices (generally the PV and SV) as the change in χ^2 when the two vertices are combined into a single vertex fit. The DOCA is the distance of closest approach between two tracks. The impact parameter (**IP**) is defined as the distance between a track and the PV at the track's closest point of approach, as sketched in Figure 3.6. The $IP\chi^2$ is the difference in the χ^2 of the vertex fit to the PV, when the track whose $IP\chi^2$ is being measured is added and removed from the fit.

Similar to the HLT1 algorithm, the HLT2 algorithm has lines that select events with one or two identified muons in the final state. The criteria of the HLT2 muon lines relevant for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis are shown in Table 3.1. The criteria in Table 3.1 are split into **Detached** and **DetachedHeavy** lines, with the latter having a higher threshold on the dimuon mass.

3.2.1 The TISTOS method for efficiency estimations

There are two ways of estimating the trigger efficiency, using simulation to reproduce fully the physics and detector, or using a data driven approach. In the data driven case, as the detector records only events passing the trigger, the number of events that the

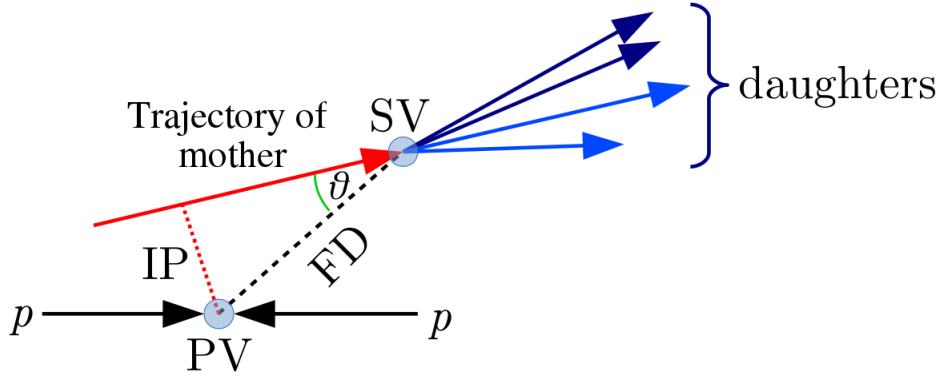


Figure 3.6: The Flight Distance (FD), Impact Parameter (IP) and Primary and Secondary Vertices (PV, SV) for an event. The black lines labelled p represent the colliding proton beams. Image altered from Ref. [35].

Hlt2DiMuon	Detached	DetachedHeavy
Track χ^2/ndof	< 5	< 5
Track IP_{χ^2}	> 9	-
Mass [GeV/c^2]	> 1	> 2.95
FD_{χ^2}	> 49	> 25
vertex $_{\chi^2}$	< 25	< 25
$P_T^{\mu\mu} [\text{GeV}/c]$	> 1.5	-

Table 3.1: The HLT2 lines used in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis based on two identified muons [28].

trigger processes is not directly observable. A solution to this is to look at and compare samples which fire the trigger in different ways. As previously discussed, the accepted trigger events can be split into three categories: TIS, TOS and TISTOS.

The TIS and TOS events can be used to predict the efficiency of the TOS sample, ϵ_{TOS} , if it is assumed that the TIS and TOS events are independent. Under this assumption then the trigger efficiency under TOS given that the event is TIS, $\epsilon_{TOS|TIS}$, is simply

$\epsilon_{TOS|TIS} = \epsilon_{TOS}$ giving

$$\epsilon_{TOS} = \frac{N_{TISTOS}}{N_{TIS}}. \quad (3.3)$$

Thus the efficiency for TOS events can be deduced by looking at the number of TOS events which are triggered under TIS.

3.3 The VELO

This section is based on sections 5.3 and 5.4 of Ref. [26].

The VELO provides R and ϕ coordinates by using silicon modules placed along the beam direction, as shown in the upper half of Figure 3.7. The modules act as sensors in the R and ϕ directions. Once beam conditions are stable, the VELO sensor is placed 7 mm from the collision point. This is narrower than the aperture required for injection by the LHC. Thus during injection the VELO is opened, again as indicated in Figure 3.7. The VELO is designed such that tracks within the nominal LHCb acceptance cross at least three VELO stations.

3.3.1 Measuring the impact parameter

The impact parameter is an important variable in the $A_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis, as it allows prompt background to be distinguished from signal candidates. The resolution depends on, amongst other things, the amount of multiple scattering of particles by the detector material and the resolution of the position hits in the detector from which the tracks are reconstructed. The resolution as a function of $1/p_T$ can be seen in Figure 3.8. Tracks with $p_T > 1 \text{ GeV}/c$ have a resolution of $< 35 \mu\text{m}$.

3.3.2 Measuring the PV position

A good measurement of the PV position is important for gaining good precision on the FD of a particle. The PV resolution is strongly correlated with the number of tracks, N ,

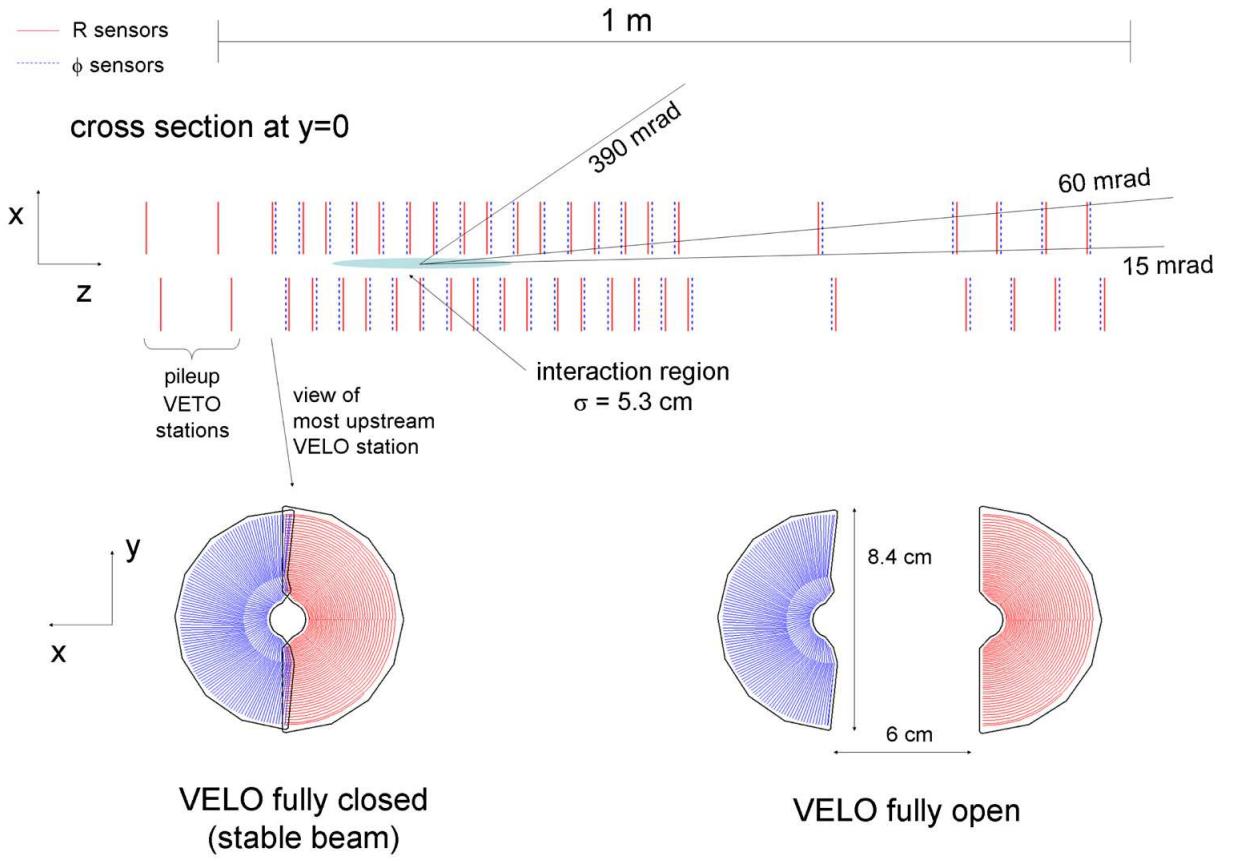


Figure 3.7: Above: the VELO stations as placed along the z direction, below: as shown in the $x - y$ plane in both the open (for injection) and closed (for stable beams) configuration [25].

in an event. At least five tracks are required to form a PV.

The resolution of the PV in the x, y and z directions is fitted using the function

$$\sigma_{PV} = \frac{A}{N^B} + C, \quad (3.4)$$

where A, B and C are constants. The resolution of the PV in x, y, z as a function of N can be seen in [Figure 3.9](#).

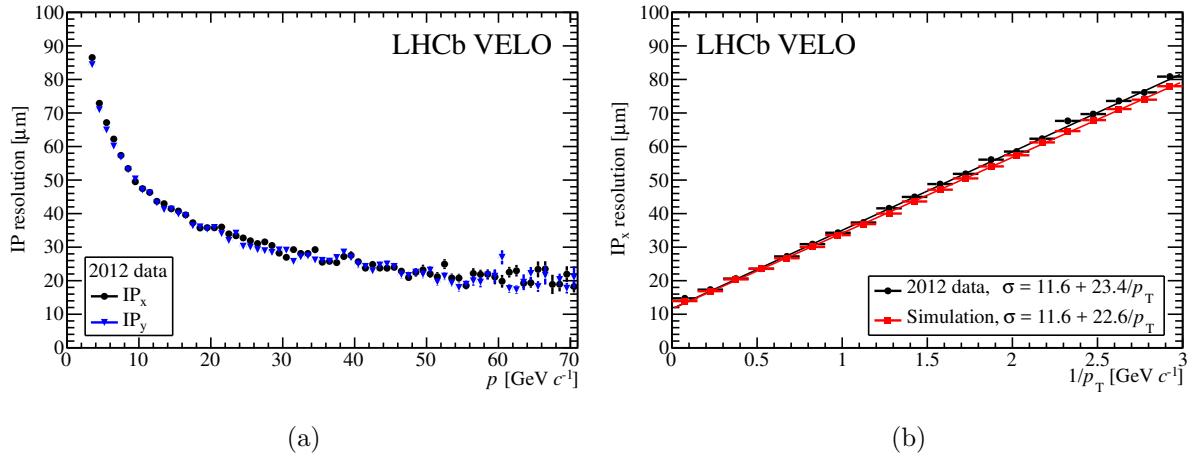


Figure 3.8: The resolution of IP_x and IP_y as a function of the momentum of the track, (a), the resolution of IP_x and IP_y as a function of $1/P_T$, (b) [26].

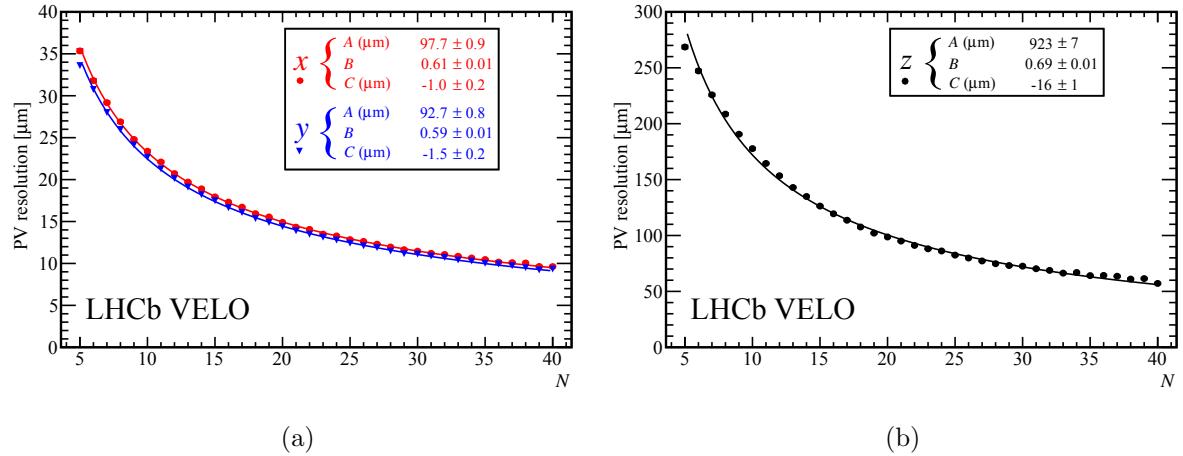


Figure 3.9: The PV resolution in the x, y directions as a function of track multiplicity, N , (a). The PV resolution in z direction as a function of N , (b). The values in the legends refer to the constants in Equation 3.4 [26].

3.3.3 Reconstructing tracks in the VELO

The track-finding algorithms in LHCb start in the VELO. Tracks reconstructed in the VELO are then matched with tracks reconstructed from hits in the tracking units further

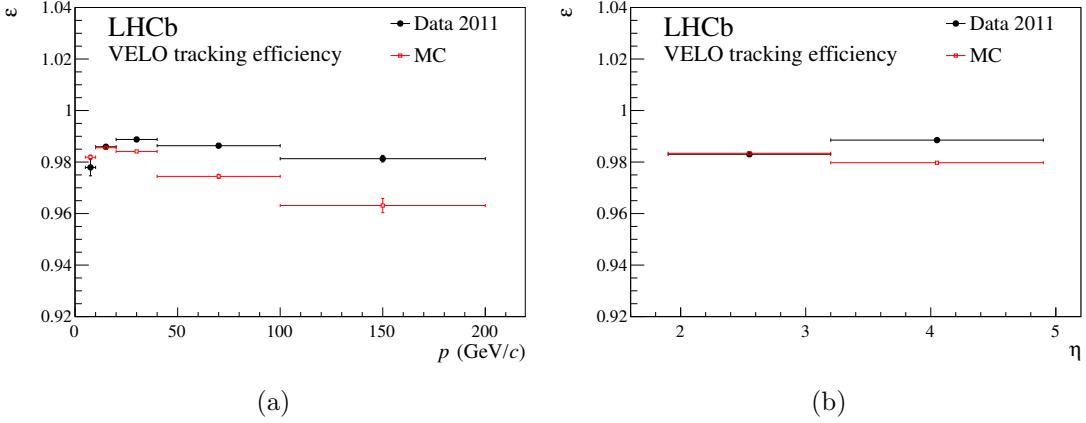


Figure 3.10: The tracking efficiency of the VELO for data and simulation as a function of momentum (left) and η (right).

downstream. The track reconstruction efficiency is defined as the probability that a charged particle passing through the detector (in this case, the VELO) is reconstructed.

The VELO algorithm searches for tracks by looking for clusters in the R sensors consistent with a straight line and then looks for a set of compatible Φ sensor clusters to reconstruct a trajectory. The minimum number of clusters required to find a track is six (three in the R sensors and three in the Φ sensors) and the average number of clusters is 11.

The efficiency of the track reconstruction in the VELO is deduced using a tag-and-probe method. Here samples of $J/\psi \rightarrow \mu^+ \mu^-$ decays are used, where one muon is tagged (fully-reconstructed) and the other probe track is partially reconstructed from hits in other parts of the detector. The momenta of the tag and the probe muon are combined to check that it falls within the J/ψ mass window, which is used to confirm the track selection. The VELO efficiency is then obtained by matching the partially reconstructed track to a fully reconstructed track that includes a VELO segment. The absolute efficiency for data for VELO tracks is $\sim 98\%$. The VELO tracking efficiency observed in the data and in the simulation is shown in Figure 3.10.

3.4 The TT station, T station and track reconstruction

The TT station sits upstream of the dipole magnet and the T stations sit downstream of the magnet. The TT station is roughly 150cm wide and 130cm high and uses silicon micro-strips to detect passing particles. The Inner Tracker (**IT**) of the T stations is also made of silicon micro-strips and the outer part (the Outer Tracker (**OT**)) of the T stations is made up of straw tubes. The TT, IT and the OT are shown in Figure 3.11. Each outer

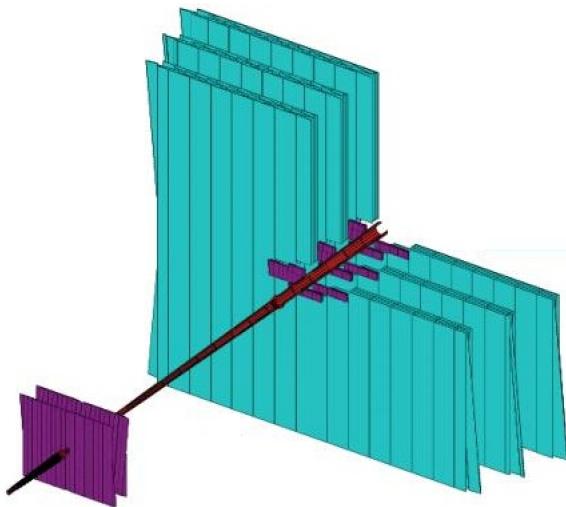


Figure 3.11: The Outer Tracker in the T stations (light blue) made up of straw-modules and the Inner Tracker and TT station (purple) made up of silicon micro-strips [36].

tracker for each T station consists of four layers rotated with respect to the vertical by $(0^\circ, -5^\circ, 5^\circ, 0^\circ)$.

Charged particles must have a momentum of at least $1.5 \text{ GeV}/c$ in order to traverse the magnetic field and reach the T stations. The trajectories of charged particles passing through the detector can be reconstructed from hits in the VELO, T and TT (IT and OT) stations. An example event showing such tracks reconstructed from assigned hits in these stations can be seen in Figure 3.12.

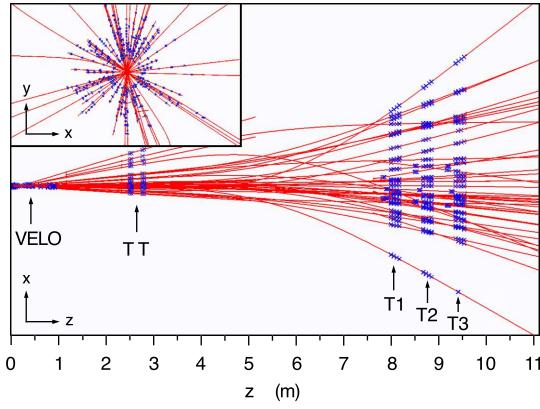


Figure 3.12: A display of the reconstructed tracks for an event along with the assigned hits in blue. The insert shows a zoom of the VELO region in the x - y plane [25].

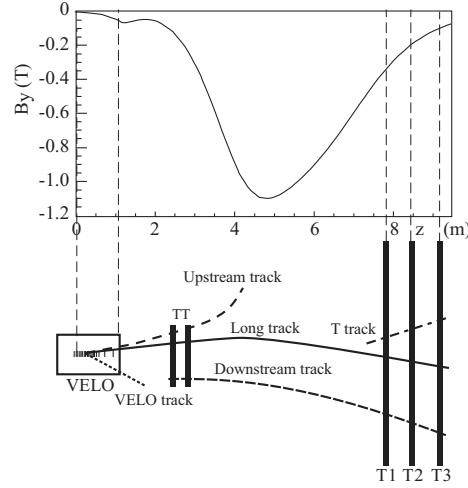


Figure 3.13: Different types of track in the LHCb detector along with the corresponding magnetic field strength in the y direction [25]

The tracks in LHCb can be characterised into different types, as outlined in Figure 3.13. These different track types are:

Long tracks which traverse the entire detector. They are defined as tracks which have hits in the VELO and the T stations. Hits in the TT stations are optional. As they traverse the full magnetic field, long tracks offer the most precise momentum measurement and are therefore of most importance to physics analyses. They are

also the best understood tracks in terms of efficiency measurements. There are on average 33 long-tracks per event [37].

Upstream tracks leave hits only in the VELO and the TT station. This is generally due to the tracks having too low a momentum to reach the T stations. These tracks are not used directly in physics analyses.

Downstream tracks like long tracks, are directly used in physics analyses for particles which are long-lived enough to decay outside the VELO, such as K_s^0 and Λ^0 . The reconstruction efficiency of these tracks is poorly understood and part of this thesis focuses on a method to improve this efficiency measurement. There are on average 14 downstream tracks per event.

VELO tracks pass only through the VELO and are used for PV reconstruction. They are not used directly in physics analyses.

T tracks pass only through the T stations and tend to be produced in secondary interactions. They are not used directly in physics analyses.

The reconstruction of VELO tracks has already been described. There are two complementary algorithms for adding tracks from downstream detectors to these VELO tracks to make a long track. In the first, so-called forward tracking, the momentum of a particle and its trajectory are entirely determined from a VELO track and a single hit in one of the T stations. Further hits in the T stations which are compatible with this trajectory are then added [38]. The second algorithm used is referred to as track-matching, where long tracks are made by combining VELO tracks with T tracks, which are found by a stand alone track-finding algorithm [39].

For both algorithms the candidate tracks are combined to form a final set of long tracks. Once combined, hits from the TT station consistent with the trajectory are also added to improve the momentum measurement [40].

The downstream tracks are reconstructed by taking a stand alone T track and then extrapolating this track through the magnetic field. Once having computed the point at

the centre of the magnetic field using the T track, a search for a list of TT measurements making a straight-line with this point is carried out. The main challenges for this method are *ghost* tracks (i.e. hits within the detector which are falsely reconstructed as a track) and the wide window across the TT station within which matching TT measurements must be searched for [41].

The work described later in this thesis will outline a method to measure the downstream track-finding efficiency using a data-driven method, as a function of p and η , and will provide a relationship between the downstream tracking efficiency observed in data and the downstream tracking efficiency according to simulation.

3.5 The RICH detectors

The main role of the **RICH** is to provide identification for the charged hadrons π , K and p . It also provides information on the charged leptons e and μ which complement the information provided by the calorimeter and muon systems.

The momentum of a particle is highly correlated to the polar angle θ , with high momentum particles having a smaller angle. The RICH1 detector system is placed just after the VELO and provides identification for particles with a low momentum range of 2–40 GeV/ c . The RICH2 detector system is placed between the T stations and the first *muon station*, M1, and covers a momentum range of 15–100 GeV/ c . The RICH2 detector system is placed further downstream, as the higher momentum particles are less affected by the magnet.

The RICH functions on the basis that if a particle travels faster than the speed of light in a medium it will emit Cherenkov radiation. The angle of this radiation is related to the speed of the particle and the refractive index of the medium. By measuring the Cherenkov angle θ_c , the particle mass can be deduced through the relation

$$\cos \theta_c = 1/\beta n = E/pn = \frac{\sqrt{m^2 + p^2}}{pn}. \quad (3.5)$$

The radiators used in RICH1 are silica aerogel and C_4F_{10} . The aerogel has a very low

density and high refractive index which makes it ideal for very low momentum particles with momenta $\sim \text{GeV}/c$. The aerogel is placed just after the VELO and the rest of the RICH1 is filled with C_4F_{10} . The optical system in both RICH detectors is similar. A primary spherical tilted focusing mirror and a plane secondary mirror focus and direct the Cherenkov light on to a plane of hybrid photon detectors (HPDs), as shown in [Figure 3.14](#).

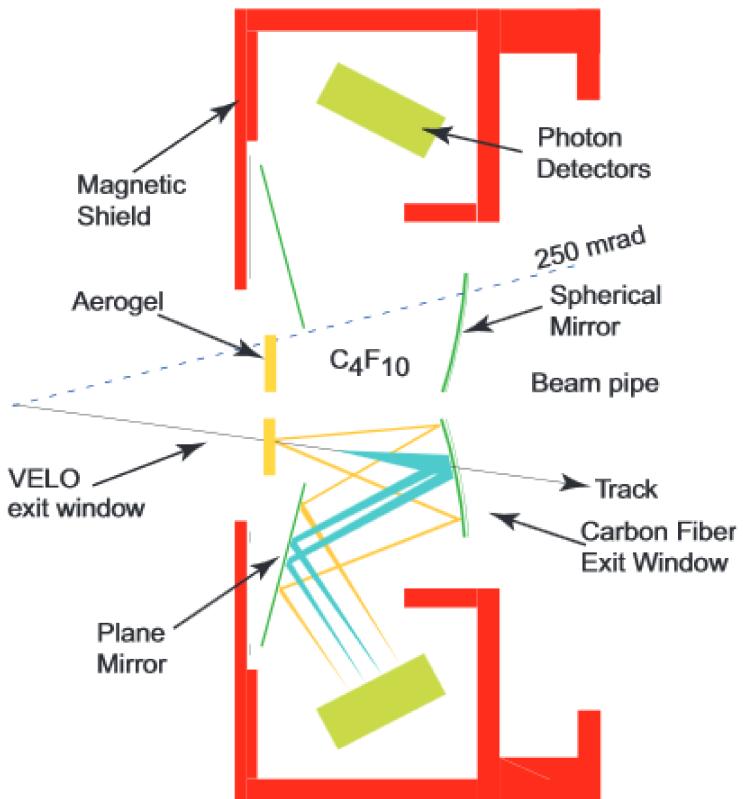


Figure 3.14: The optical set-up for the RICH1 detector [42].

3.5.1 The particle identification algorithm

A good understanding of the performance of the Particle IDentification (**PID**) of the RICH is important for the analyses described in this thesis. A precise understanding of the performance allows the efficiency and mis-identification (mis-id) rate of tracks to be calculated after the RICH information has been applied to the data candidates. To

identify effectively the particle type of a track in the RICH, an overall event log-likelihood algorithm is calculated, taking into account all possible hypotheses for all tracks in the event.

The likelihood minimisation procedure starts by assuming all particles are pions, as pions are the most abundant particle type. The overall event likelihood is calculated from the distribution of the photon hits (which will form rings referred to as Cherenkov rings) along with the event's associated tracks and their errors, assuming the pion mass hypothesis for every track. The mass hypothesis of each track is then changed, one track at a time, whilst leaving the hypothesis for every other track in the event unchanged. The mass hypotheses for e , μ , K and proton are applied and the change in mass hypothesis which gives the largest increase in the event likelihood is identified. This method is repeated until no improvement in the event likelihood can be found and thus each track has been set to its preferred hypothesis. Some modifications are made to the logic described above in order to reduce CPU time. The final output of this algorithm is a variable for each track in an event which shows the difference in the total event log-likelihood when the mass hypothesis of the track in question is changed from a pion to e , μ , K and proton. These variables are referred to as e.g. $DLL_{K\pi} = \Delta\text{Log}(K - \pi)$, $DLL_{p\pi} = \Delta\text{Log}(p - \pi)$ and $DLL_{\mu\pi} = \Delta\text{Log}(\mu - \pi)$.

3.5.2 The PID performance

The performance of the PID is assessed using data candidates which have been selected without using PID information. Thus decays are needed whereby the individual daughter states can be reconstructed with very little background, using kinematic information alone. The decays chosen include the channels $\Lambda^0 \rightarrow p\pi^-$ and $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$. These decays have a high purity as can be seen [Figure 3.15](#). The remaining background is removed using the *sPlot* technique [44], which assigns a so-called sWeight (which can be positive or negative) to each event, depending on how signal-like the event is deemed to be. The *sPlot* technique is summarised in [Appendix B](#).

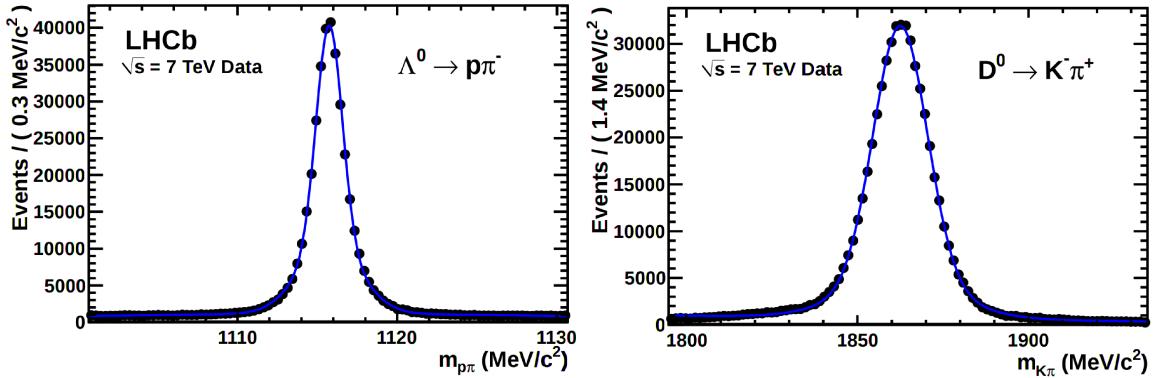


Figure 3.15: Fitted invariant mass distributions for $\Lambda^0 \rightarrow p\pi^-$, (a), and $D^0 \rightarrow K^-\pi^+$, (b) [43].

Using these control samples, the $\text{DLL}_{X\pi}$ distributions are considered for each track type. It is then possible to study the discriminating power of particular PID requirements. The discriminating power of the selection cuts $\text{DLL}_{K\pi} > 0$ and $\text{DLL}_{K\pi} > 5$ placed on tracks with a kaon mass hypotheses can be seen in Figure 3.16(a). Here the efficiency of these cuts on kaons is shown along with the misidentification of pions as kaons. The discriminating power of the selection cuts $\text{DLL}_{p\pi} > 0$ and $\text{DLL}_{p\pi} > 5$ on tracks with a proton mass hypotheses can be seen in Figure 3.16(b), along with the mis-id rate of pions to protons and finally the efficiency is shown for $\text{DLL}_{pK} > 0$, $\text{DLL}_{pK} > 5$ along with the mis-id rate of kaons to protons in Figure 3.16(c). In Figure 3.16(c) the variable DLL_{pK} is formed using the relation

$$\text{DLL}_{pK} = \text{DLL}_{p\pi} - \text{DLL}_{K\pi}. \quad (3.6)$$

The Cherenkov angle as a function of momentum for tracks passing through the C_4F_{10} radiator is shown in Figure 3.17

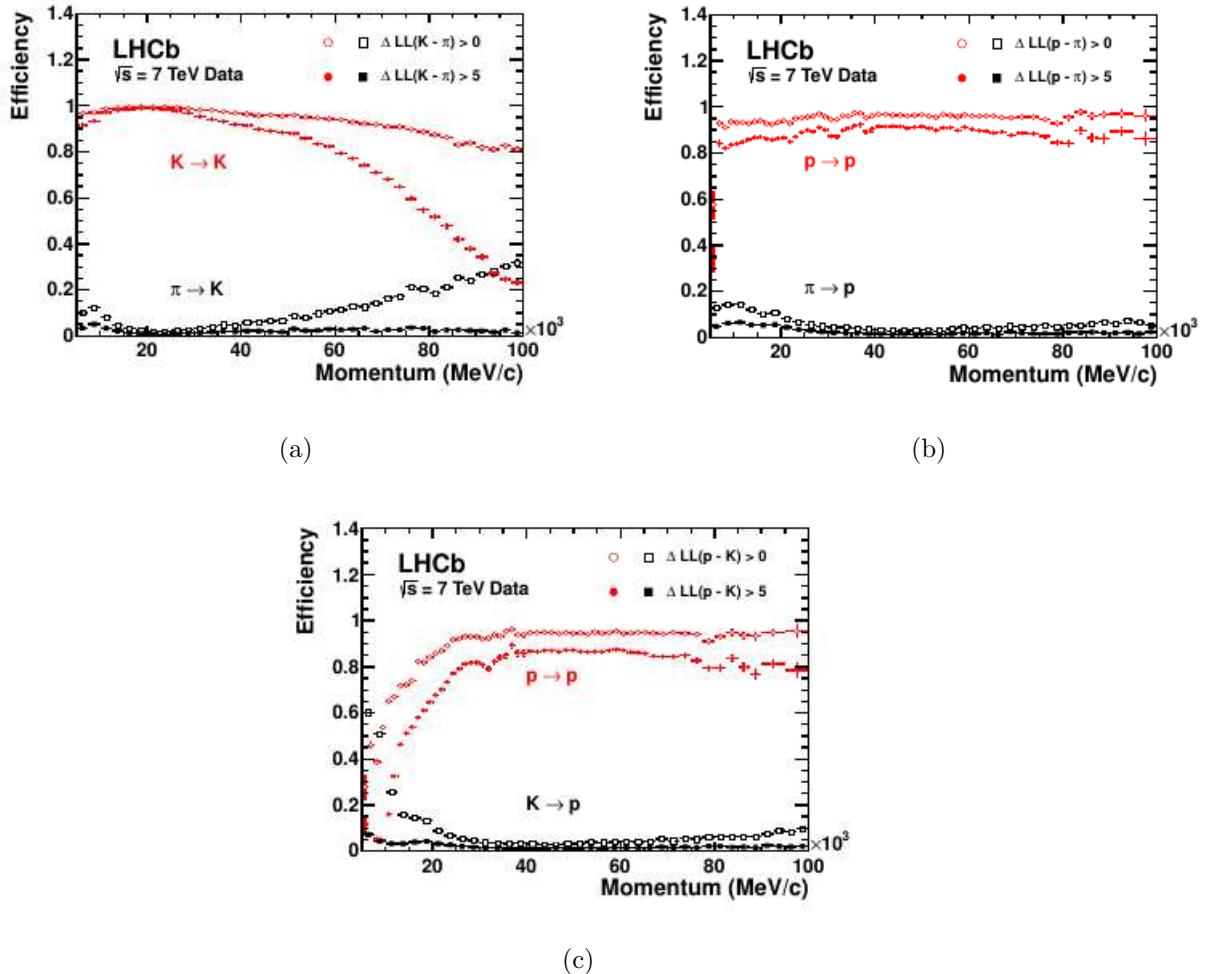


Figure 3.16: The kaon efficiency and the misidentification rate for pions as kaons for different PID cuts (filled and hollow markers) (a), the proton efficiency and the misidentification rate for pions as protons for different PID cuts (filled and hollow markers), (b) and the proton efficiency and the misidentification rate for kaons as protons for different PID cuts (filled and hollow markers), (c) [43].

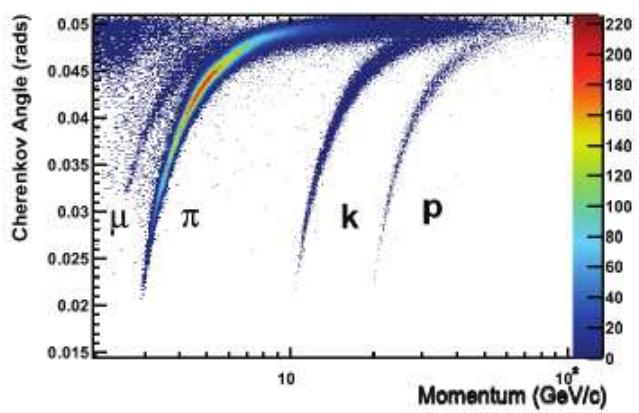


Figure 3.17: Reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator [43].

3.6 Muon identification

The selection of muons is performed using information from the five muon detector stations M1–5. Each station is divided up into four regions R1–4, all of which cover roughly the same acceptance, as shown in [Figure 3.18](#). The granularity of the station in each region R1–4 reflects the particle density in that region. There is also increased granularity in the x direction, as this is the plane in which particles are bent by the magnetic field.

The M1 station is placed before the [ECAL](#) and stations M2–5 are placed downstream of the ECAL. Stations M2–5 are interleaved with 80cm thick iron absorbers. These iron absorbers along with the muon stations M2–5 allow the identification of penetrating muons.

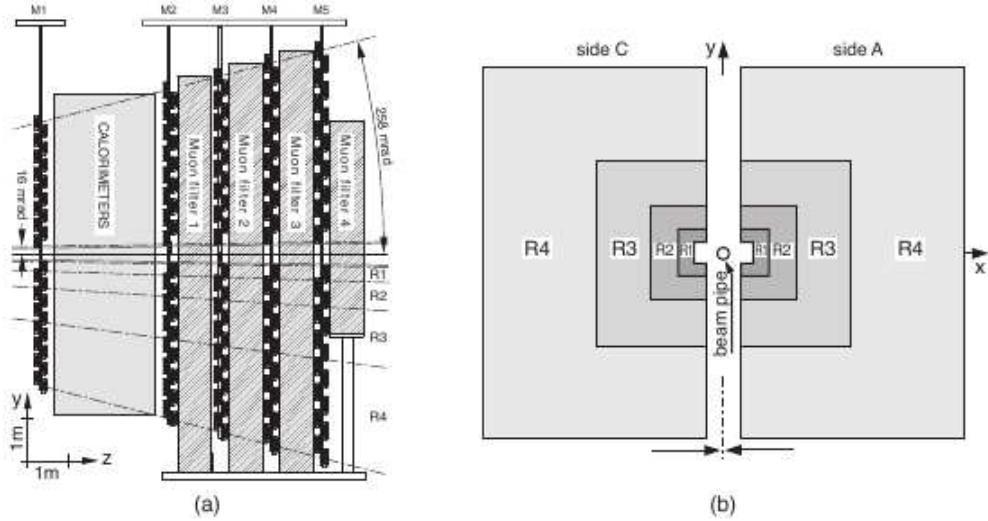


Figure 3.18: The five muon stations (a) (the muon filters refer to iron absorbers - see text) and (b) the station layout in regions R1–4 [45].

3.6.1 The efficiency of the muon stations

As the L0 algorithm requires hits in all five muon stations and an overall efficiency of $> 95\%$ is desirable, the efficiency of each individual station must be at least 99%. This is achieved by having a good time resolution and a high detector redundancy. This redundancy is

realised by having two independent readouts in many of the stations/regions and requiring a hit in at least one of them and the time resolution to be better than 4ns. All regions of all muon stations fulfil the $> 99\%$ requirement.

3.6.2 The muon identification procedure

The off-line muon identification strategy can be split into three steps [46]

IsMuon: A binary selection of muon candidates based on the penetration of the muon candidates through the calorimeters and iron filters is parameterised by the boolean IsMuon variable. There are also momentum requirements, depending on which muon stations have been traversed. This provides high efficiency, while reducing the misidentification probability of hadrons to the percent level.

muDLL: Based on the pattern of hits around the extrapolation to the different muon stations of the trajectories of charged particles reconstructed in the tracking system, the logarithm of the ratio between the muon and non-muon hypotheses is constructed and used as a discriminating variable, referred to as muDLL.

DLL _{$\mu\pi$} : The RICH information, along with information from the calorimeter, is combined with muDLL to give the discriminating variable DLL _{$\mu\pi$} .

3.6.3 Efficiency of the muon ID procedure

The efficiency of the muon selection process is extracted from data, using $J/\psi \rightarrow \mu^+ \mu^-$ decays combined with tag-and-probe techniques. The calculation of the mis-id rates of kaons, pions and protons to muons is performed using the decays outlined in subsection 3.5.2. The efficiency for IsMuon, ϵ_{IM} , is shown in Figure 3.19, along with the mis-id rates as a function of momentum and in ranges of transverse-momentum.

There is a weak efficiency dependence on transverse-momentum. The reason for a more dramatic loss of efficiency for low P_T muons is due to the scattering of tracks out of the detector. The mis-id rate also falls with increased transverse-momentum. This is because

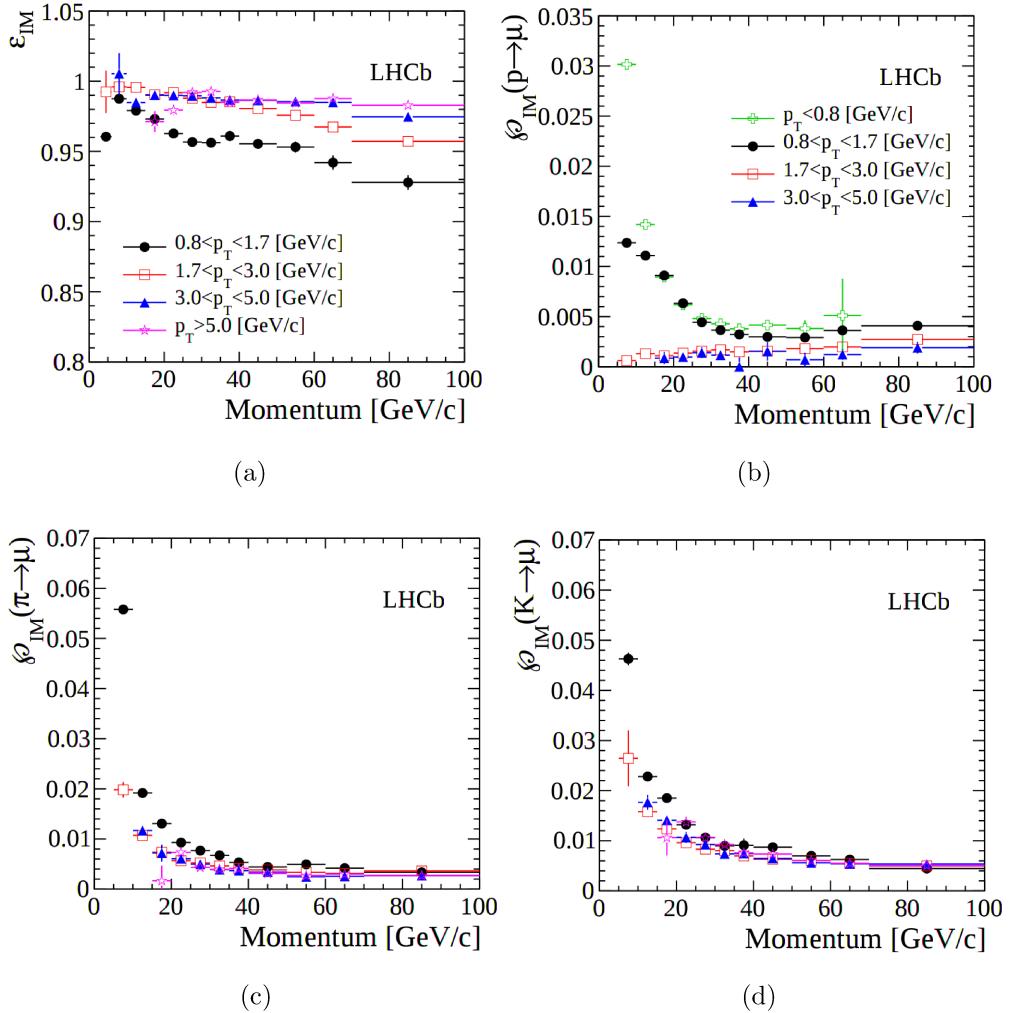


Figure 3.19: IsMuon efficiency and misidentification probabilities, as a function of momentum and in ranges of transverse-momentum: efficiency of IsMuon on muons, ϵ_{IM} , (a), IsMuon mis-id rate ($p \rightarrow \mu$), (b), IsMuon mis-id rate ($\pi \rightarrow \mu$), (c), and IsMuon mis-id rate ($K \rightarrow \mu$), (d) [46].

a higher P_T for a given momentum corresponds to a larger polar angle and the track then traverses lower occupancy detector regions [46].

The efficiency of the $DLL_{\mu\pi}$ variable on muons, compared to the mis-id rate for kaons and pions, along with the equivalent distributions for the mu DLL variable alone, are shown in Figure 3.20. The variables most used in the work presented in this thesis are the IsMuon

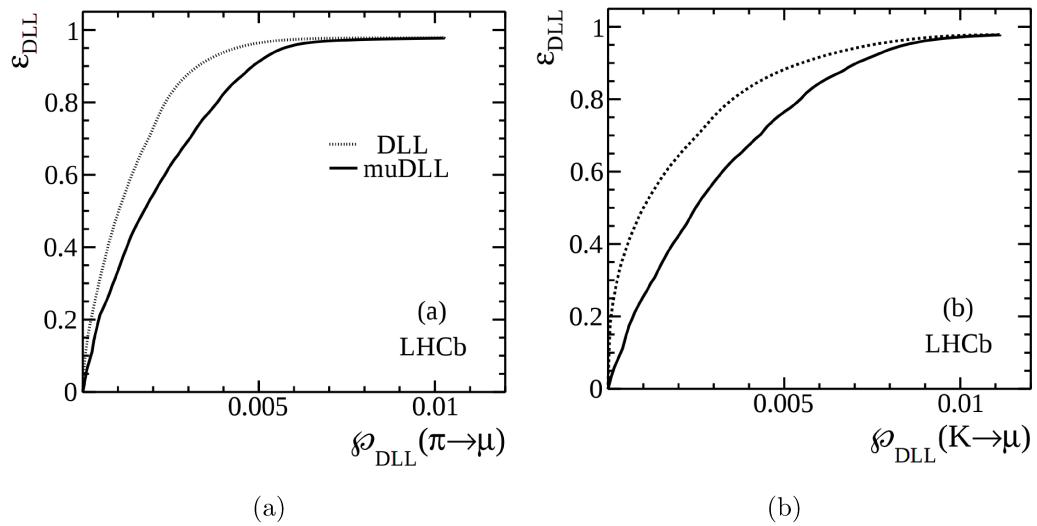


Figure 3.20: Efficiency of selecting muons against the efficiency for selecting pions for the $\text{DLL}_{\mu\pi}$ and muDLL variables, (a). Efficiency of selecting muons against the efficiency for selecting kaons for the $\text{DLL}_{\mu\pi}$ and muDLL variables, (b).

and $\text{DLL}_{\mu\pi}$ variables in combination.

3.7 Simulation

The analyses described in this thesis rely on an accurate simulation of the physics within the detector, and of the detector itself, in order to correctly model the detector output. The simulation of the detector output can be split into two stages; simulation of the particle decays; and simulation of the detector and the particles interactions with it. In the remainder of this thesis, the simulation of the particles themselves with no detector present is referred to as the generator-level simulation.

The first stage of simulation is to generate pp collisions. These are generated using Pythia 6.4 [47] and Pythia 8.1 [48], which produce sets of outgoing particles produced in the interactions between two incoming particles [47]. The simulated $b\bar{b}$ quark pair is hadronised repeatedly until the relevant b -hadron is produced, which for much of the simulation used in this thesis would be a Λ_b^0 baryon [35]. Once the relevant b -hadron has been simulated, its decay is described by the EvtGen package [49] which allows for the implementation of form factors and other different physics models. The final state radiation is generated using Photos [50]. All of these programs, Pythia, EvtGen and Photos are combined to create generator-level simulation. The LHCb detector and the material interactions of particles with the detector are simulated using the Geant4 [51] tool-kit. In the remainder of this thesis, simulation which includes both the detector and the physics simulation is referred to as reconstructed simulation or just simulation.

Chapter 4

Measuring the efficiency of long-lived particles

In this chapter, the reconstruction efficiency of the long-lived particle, K_s^0 , decaying via the mode, $K_s^0 \rightarrow \pi^+\pi^-$, is studied as a function of the position of the K_s^0 decays along the z -axis. As previously outlined in [chapter 3](#), the z -axis is defined as pointing along the beam pipe away from the interaction point. This study is of interest because the reconstruction efficiency of long-lived particles at LHCb is still poorly measured when compared to the equivalent measurement for shorter-lived particles.

The reconstruction efficiency is defined as the likelihood of the decay, $K_s^0 \rightarrow \pi^+\pi^-$, being successfully reconstructed, given that both pion tracks are in the detector acceptance. The measurement of the reconstruction efficiency for any decay in the LHCb detector is less precise for [downstream tracks](#), and for [long tracks](#) with secondary vertexes occurring past $z \sim 140$ mm. There are numerous examples of published LHCb papers which include long-lived particles [52–65] and as such, it would be desirable to have a better understanding of the downstream track reconstruction efficiency in both simulation and data. This is the first study to look in detail at the reconstruction efficiency as a function of the decay position in z for both downstream and long tracks, allowing for the calculation of the long track reconstruction efficiency for tracks originating along the entire length of the [VELO](#).

The aim of the work presented in this chapter is to use the well-understood measurement of the tracking efficiency of long tracks from the decay $J/\psi \rightarrow \mu^+\mu^-$, to deduce the efficiency of downstream tracks in the decay $K_s^0 \rightarrow \pi^+\pi^-$. This provides a complementary measurement to that given in Ref [66] of the downstream reconstruction efficiency.

In [section 4.1](#), the method with which the efficiency for the reconstruction of long tracks is calculated is outlined. This is followed by an overview of a previous study on the K_s^0 reconstruction efficiency for downstream tracks, referred to as the D^{*+} study. There is then an overview of the method used to calculate the downstream tracking efficiency in the analysis presented in this chapter. In [section 4.2](#), there is an overview of the data samples used in the analysis. There is further detail on the analysis method and a discussion on the reduction of backgrounds in [sections 4.3](#) and [4.4](#). The agreement between simulation and data is discussed in [section 4.5](#) and the verification of the calculations used is outlined in [section 4.6](#). There is a presentation of the systematic uncertainties and results in [section 4.7](#) and [section 4.8](#). Finally, there is a comparison between the results obtained with the study presented in this chapter and those from the D^{*+} study, outlined in [section 4.9](#).

4.1 Current reconstruction efficiency studies carried out at LHCb

This section will outline two previous LHCb efficiency studies and the study presented in this chapter. The first study involves the measurement of long track efficiencies and the results of this study are used in the analysis presented in this chapter. The second study outlined in this chapter is the so-called D^{*+} study, the results from which will be compared to the results obtained from the analysis in this chapter as a cross check.

4.1.1 Long track efficiency measurements

The current measurement of the long track reconstruction efficiency is carried out in Ref [67] using $J/\psi \rightarrow \mu^+\mu^-$ decays, where the J/ψ comes from a b -hadron. In the study in

Ref [67], a tag-and-probe technique is used to deduce the tracking efficiency of long tracks. The long track efficiency is initially calculated in bins of p , η and track multiplicity

In order to be classed as a long track, hits are required in the T stations and in the VELO. Hits in the TT stations are optional. The efficiencies for long tracks are deduced via two methods in Ref [67], the results of which are then averaged. The first method is referred to as the combined method. This method combines the efficiency of the VELO stations to reconstruct tracks with the efficiency of the T stations to reconstruct tracks. In the VELO case, downstream tracks are used as probes, as shown in Figure 4.1(a). In the T station case, VELO and muon tracks are used as probes, as shown in Figure 4.1(b). The product of the efficiency of the VELO stations and T stations then gives the total long track reconstruction efficiency. The second method, referred to as the long method, uses tracks that have hits in the TT stations and muon stations, as shown in Figure 4.1(c), to deduce directly the efficiency of the T stations and the VELO stations simultaneously.

There are small differences between the combined and long methods used to calculate the long track efficiency¹. As the effects causing the discrepancies between the two efficiency methods are common to both simulation and data, the efficiency is quoted as the ratio between simulation and data, thereby cancelling these discrepancies.

As the simulation and data tracking efficiencies as a function of track multiplicity are within 2σ of one another (see Figure 4.2), the ratio of the long track efficiencies between simulation and data is quoted as a function of just p and η . The worsening efficiency with increasing track multiplicity in Figure 4.2 is due to the increasing likelihood of *ghost* tracks with a higher number of tracks. The ratios between simulation and data for different p and η bins are shown in Figure 4.3. Given that the J/ψ used in the long track reconstruction efficiency measurements comes from a b -hadron decay, the values of z over which these efficiencies from Ref [67] are deemed to be valid is taken to be roughly equal to the primary vertex distribution of the b -hadrons, giving a region of validity of roughly $-140 < z < 140$ mm. The region is referred to as the long track efficiency validity

¹For example, in the long method, the efficiency is calculated for tracks passing through the TT stations. In the combined method, this is only true for the calculation of the VELO efficiencies

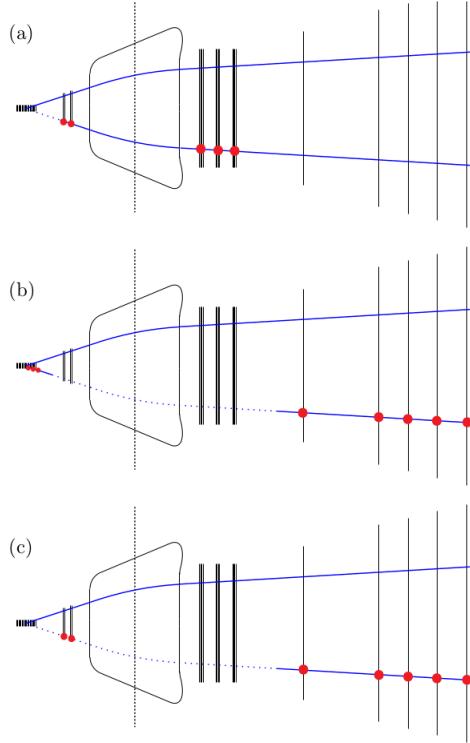


Figure 4.1: Different probe tracks used to measure the efficiencies of different subdetectors, showing the probe track used to measure the VELO efficiency (a), the probe track used to measure the T station efficiency (b), and the probe track used to measure the VELO and T station efficiencies simultaneously (c).

region, or z_{valid} , hereafter.

4.1.2 Downstream tracking efficiency measurements

In order for a track to be classed as a downstream track, hits are required in the T and TT stations and there must be no VELO track segment which matches the downstream track segment. Thus, given that there can be no hits in the VELO and that there are no suitable downstream muonic decays, it is not possible to form a downstream probe track. Therefore, the tag-and-probe technique used in Ref [67] cannot be used to measure the reconstruction efficiency of downstream tracks.

As a consequence, different approaches must be used to measure the downstream

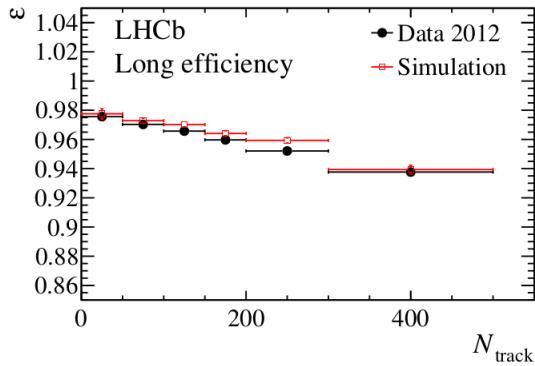


Figure 4.2: The efficiency for data and simulation as a function of track multiplicity, N_{track} .

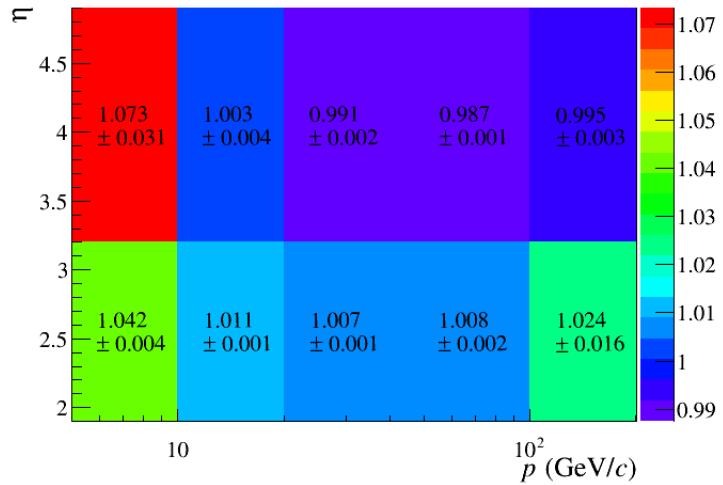


Figure 4.3: Ratio of absolute efficiencies between data and simulated events for long tracks. Ratios are for 2012 conditions [67].

track reconstruction efficiency. The difficulty of calibrating the downstream tracking efficiency means that the efficiency is less precisely known than for long tracks. This lack of knowledge means that analyses involving long-lived particles, such as K_s^0 or Λ^0 particles, tend to have a larger associated systematic uncertainty than those analyses involving only shorter-lived particles.

The study in Ref [66] measures the reconstruction efficiency of downstream K_s^0 mesons using $D^{*+} \rightarrow (D^0 \rightarrow \phi K_s^0(\rightarrow \pi^+ \pi^-))\pi^+$ decays. The tracking efficiency in this study is deduced by comparing the yields of partially and fully reconstructed decays. The partially

Quantity	Selection
vertex fit convergence	True
mass window (pre-fit)	$\pm 80 \text{ MeV}/c^2$
mass window (post-fit)	$\pm 64 \text{ MeV}/c^2$
vertex χ^2/ndf	<25
track DOCA χ^2	<25

Table 4.1: Criteria applied to the vertexing procedure used to make a standard LHCb K_s^0 candidate from two downstream pions. The track **DOCA** is the distance of closest approach between the two daughter tracks and the mass windows are either side of the known K_s^0 mass of $497.6 \text{ MeV}/c^2$ [66].

reconstructed decay does not reconstruct one of the pions from the K_s^0 . Instead, the missing four momentum is calculated using the reconstructed decay products and the flight direction of the D^0 .

The vertex efficiency, defined as the efficiency to vertex the two downstream pion tracks and pass the requirements on the vertex fit of the two downstream pions, is also calculated in Ref [66]. The vertex requirements are shown in Table 4.1, where post-fit and pre-fit refer to the value of the mass window before and after the vertex fit has been applied.

The results of the study, showing the ratio between data and simulation for the tracking efficiency and vertex efficiency as a function of missing track momentum and the K_s^0 momentum respectively, are shown in Figure 4.4. In Figure 4.4, the term unfolding refers to the unfolding of the efficiency in bins of momentum, correcting for cases where the momentum inferred from the partially reconstructed case lies in bin i , but it is known from fully reconstructed data that the correct momentum actually lies in bin j . The nominal result is when both the simulation and data have been unfolded. The reconstruction efficiency for the decay $K_s^0 \rightarrow \pi^+ \pi^-$ can be written as

$$\epsilon_{\text{reconstruction}} = \epsilon_{K_s^0 \text{ vertexing}} \times \epsilon_{\pi \text{ tracking}} \times \epsilon_{\pi \text{ tracking}}, \quad (4.1)$$

where ϵ denotes the relative efficiency between data and simulation.

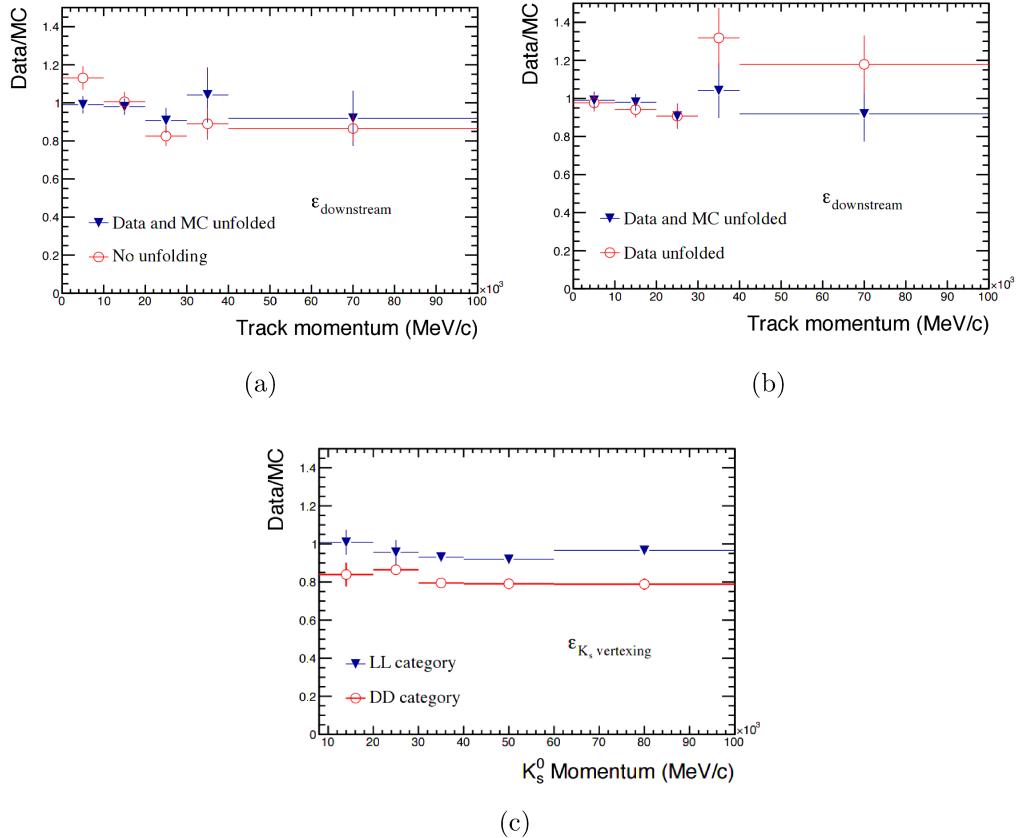


Figure 4.4: The downstream tracking efficiency as a function of momentum, (a), (b). The vertexing efficiency ratio of the K_s^0 for data over simulation (where simulation is indicated by the initials MC in the figure) for the cases where both pions are downstream track types (DD) and both pions are long track types (LL), (c). Figures are taken from Ref [66].

The results from Ref [66], averaged using a similar momentum binning scheme as used in Ref [67], are shown in Table 4.2. Averaging the vertexing efficiency over all K_s^0 momentum gives an average of 0.82 ± 0.03 . The average relative tracking efficiency between data and simulation, disregarding the higher momentum bin where the statistics are limiting, is 0.95 ± 0.04 . Using Equation 4.1 gives an average relative reconstruction efficiency between data and simulation of 0.74 ± 0.05 .

The dependence of the tracking efficiency on the momentum of the missing track is not overly strong, although the large errors on the highest momentum bin make it difficult

K_s^0 momentum bin/ GeV/ c	Vertex	missing track momentum bin/ GeV/ c	Tracking
$10 < p < 20$	0.84 ± 0.05	$0 < p < 20$	0.97 ± 0.03
$20 < p < 40$	0.83 ± 0.04	$20 < p < 40$	0.92 ± 0.07
$40 < p < 100$	0.78 ± 0.03	$40 < p < 100$	0.83 ± 0.15

Table 4.2: Results from the study in Ref [66] for the ratio between data and simulation of the downstream tracking and vertexing efficiency as a function of the missing track momentum and the K_s^0 momentum, respectively.

to reach any definite conclusion. For example, assuming the case where both the pions and the K_s^0 have a momentum greater than 40 GeV/ c gives a total efficiency of 0.55 ± 0.14 , which differs from the nominal value of 0.74 ± 0.05 by 25% but still agrees with the nominal value at 1.4σ . In addition, one weakness of the study in Ref [66] is that the value of the missing track momentum, calculated from other kinematic parameters in the decay, has a very poor resolution which means that the momentum dependence is difficult to quantify.

4.1.3 Measuring the reconstruction efficiency of $K_s^0 \rightarrow \pi^+\pi^-$ as a function of z

In the analysis in this chapter, the well-understood ratio between the long track reconstruction efficiency for data and simulated events at low z is used to correct the absolute data reconstruction efficiency in low z bins. This gives the efficiency-corrected number of K_s^0 particles in this bin, denoted N_0 . The z bins within which the efficiencies are applied to calculate N_0 are referred to as the z_{ref} bins. The width of z_{ref} is chosen to be 10 mm. This width is deemed to be narrow enough such that the change in $N(z)$ across the 10mm bin width is of the same order as the statistical error on that bin. The number of K_s^0 decays within the z_{ref} bin, combined with the momentum, lifetime and mass of the K_s^0 , can be used to calculate the predicted number of K_s^0 particles in a z bin further downstream through the relation

$$N(z, p) = N_0 e^{\frac{-z \cdot m}{p \cdot \tau}}, \quad (4.2)$$

where m is the mass of the K_s^0 , p is the momentum of the K_s^0 , τ is the true lifetime of the K_s^0 and N_0 is as previously defined. This calculation of $N(z, p)$ is further illustrated in Figure 4.5.

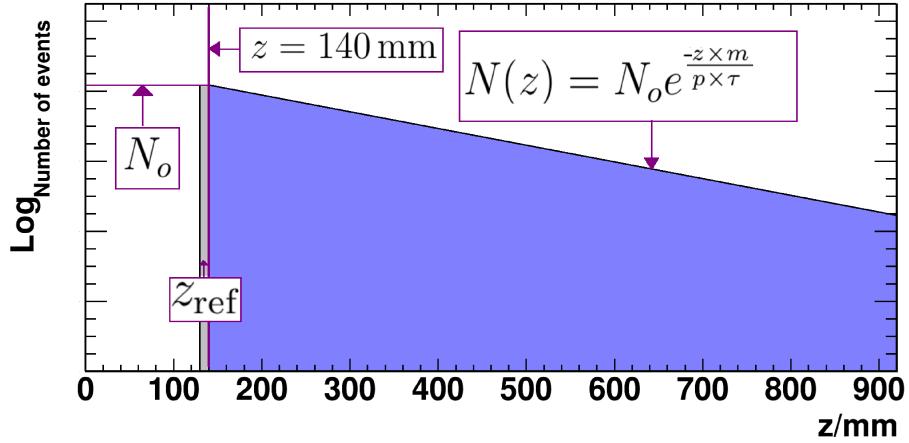


Figure 4.5: A sketch of the extrapolation in z of the number of expected K_s^0 candidates for a given momentum, p , and a mass, m . The line at $z = 140 \text{ mm}$ indicates the end of the z_{valid} region.

The reconstruction efficiency is calculated for data by comparing the number of expected K_s^0 decays, using the calculation in Equation 4.2, with the number of observed K_s^0 decays, for different bins in z . This method of calculating the reconstructed efficiency is referred to as the extrapolation method. The final downstream efficiency values obtained using this extrapolation method are quoted both as a function of z and averaged over z . The results are given in bins of the p and η of the K_s^0 , with momentum bins of $10 < p < 20 \text{ GeV}/c$, $20 < p < 40 \text{ GeV}/c$ and $40 < p < 100 \text{ GeV}/c$ and η bins of $2.0 < \eta < 3.2$ and $3.2 < \eta < 5.0$.

4.2 Data samples

This section outlines both the simulation and data samples used for this analysis.

4.2.1 Selecting K_s^0 particles from data

After the data output by the HLT2 trigger has been stored, an initial selection is placed on the data, referred to as the stripping selection. This selection is applied in order to reduce the amount of data written to disk where it is readily accessible for further analysis.

Different initial selection criteria are used to write out a set of so-called stripping lines. The stripping lines used through out this thesis are all predefined selections already implemented within the LHCb software. Normally, the stripping selection is designed to enhance the signal fraction in the dataset. However, given that $K_s^0 \rightarrow \pi^+\pi^-$ decays are abundant, even when the K_s^0 candidate has not been used in the trigger line or the stripping selection decision processes, there is roughly one $K_s^0 \rightarrow \pi^+\pi^-$ candidate per event.

As such, a stripping line which is not specifically designed to select K_s^0 candidates is used in this analysis. Selection criteria are then applied to K_s^0 candidates in the stripped dataset, as summarised in [Table 4.3](#). Here, **DIRA** refers to the cosine of the angle between the reconstructed K_s^0 momentum and the displacement vector between the primary vertex and the K_s^0 decay vertex. All other variables in [Table 4.3](#) have already been defined previously in [chapter 3](#).

The analysis in Ref [67] required the momenta of the daughter muons to be greater than $5000\,\text{MeV}/c$. In order to use the efficiency from this analysis correctly, the same cut is applied in the present analysis.

4.2.2 Simulation

Multiple samples of simulated events are used, with each featuring a different b -hadron decay. The K_s^0 that is studied is required not to come from the b -hadron but from elsewhere

Quantity	DD selection	LL selection
K_s^0 mass window (pre-fit)	$\pm 80 \text{ MeV}/c^2$	$\pm 50 \text{ MeV}/c^2$
K_s^0 mass window (post-fit)	$\pm 64 \text{ MeV}/c^2$	$\pm 35 \text{ MeV}/c^2$
K_s^0 vertex χ^2/ndf	< 9	< 9
K_s^0 track DOCA χ^2	< 25	< 25
K_s^0 DIRA	> 0.999995	> 0.999995
pion momenta	> 5000 MeV/c	> 5000 MeV/c
pion IP χ^2	> 9	> 9
pion p_T	> 0	> 250 MeV/c

Table 4.3: Criteria to select K_s^0 candidates for the analysis.

in the event. In total, around 85M events are used from which $\sim 3\text{M}$ K_s^0 candidates are reconstructed and selected.

4.3 Calculating the genuine number of K_s^0 decays as a function of z

In order to calculate the reconstruction efficiency using the extrapolation method, the long track efficiencies must first be unfolded in the so-called reference z bin, z_{ref} . This z_{ref} bin is taken from the z_{valid} region of the detector, defined as $-140 < z < 140$ mm, as previously detailed in [subsection 4.1.1](#).

The value of z for a given K_s^0 candidate is defined as the position of the vertex in z of the dipion system (referred to as the end vertex of the K_s^0 candidate), where the value of z shown in [Equation 4.2](#) is equal to the difference in z between the end vertex of the K_s^0 candidate and the centre of the z_{ref} bin.

In order to use $N(z_{\text{ref}}, p)$ to calculate $N(z, p)$ for all z , it is necessary that no new K_s^0 particles are created at a position beyond z_{ref} , otherwise it would not be a simple decaying exponential. It is also necessary to select a z_{ref} bin within the z_{valid} region, i.e. below $z \sim 140$ mm. Given that the point in z past which no new K_s^0 particles are created, referred to as $z_{K_s^0}$, is generally greater than 140 mm, steps must be taken to reduce the value of $z_{K_s^0}$. This reduction is achieved by cutting on the primary vertex (PV) distribution of the K_s^0 . This PV cut forces the maximum PV position in z of the K_s^0 to be lower, thus also shifting $z_{K_s^0}$ downwards in z . This is illustrated in [Figure 4.6](#) which shows the effect that a cut on the PV position has on the number of K_s^0 decays as a function of z . As the PV of a K_s^0 meson is dependent on both its p and η value, the cut applied depends on the p and η bin, as outlined in [Table 4.4](#).

[Figure 4.7](#) shows the z distributions for both data and reconstructed simulated events with the PV cuts applied. The red line indicates the lower edge of the lowest reference bin used. The agreement and comparative disagreement in [Figure 4.7](#) between data and reconstructed simulated events for long tracks and downstream tracks respectively has a direct effect on the final tracking efficiency values obtained.

momentum range	$10 < p < 20 \text{ GeV}$	$20 < p < 40 \text{ GeV}$	$40 < p < 100 \text{ GeV}$
η range			
$2.0 < \eta < 3.2$	55	20	20
$3.2 < \eta < 5.0$	20	-10	-30

Table 4.4: The value of z where the cut is applied to the primary vertex of the K_s^0 . Results are shown in mm for each bin.

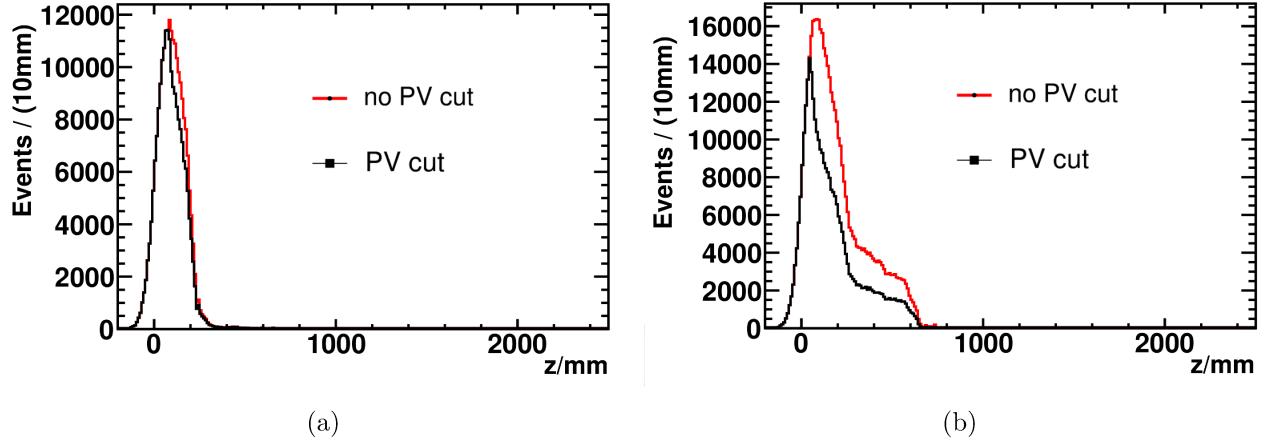


Figure 4.6: The number of decays as a function of z for long tracks with (black) and without (red) the cuts on the primary vertex of the K_s^0 applied, for data events with $10 < p < 20 \text{ GeV}$, $2.0 < \eta < 3.2$ (a), $2.0 < \eta < 3.2$ (b).

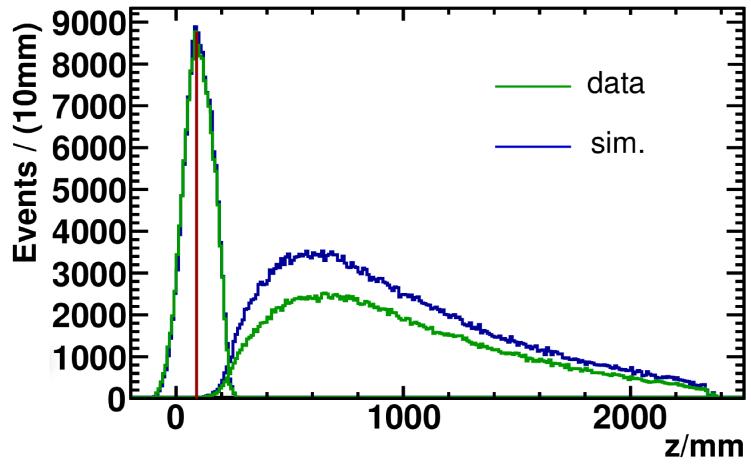


Figure 4.7: The number of decays for downstream and long tracks as a function of z for both reconstructed simulated events (blue) and data (green), for events with $10 < p < 20$ GeV/ c , $2.0 < \eta < 3.2$. The line at 90 mm indicates the lower edge of the lowest z_{ref} bin. The long tracks are those which peak around the red line at 90 mm and the downstream tracks are those which peak at higher z .

4.4 The removal of background

Background in the data sample is dealt with by fitting to the K_s^0 mass distribution in 10mm wide bins in z from $z = 0$ to $z = 2000$ mm. These fits are used to compute the number of signal events in the fit signal region, where the signal region is defined as being within 4σ of the mean of a double Crystal Ball (CB) function [68], fitted to the data. The CB function is a Gaussian function with a power-law tail below a certain threshold. It is defined as

$$f(x; \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}, & \text{if } \frac{(x-\bar{x})}{2\sigma} > -\alpha \\ A \cdot (B - \frac{(x-\bar{x})}{2\sigma})^{-n}, & \text{otherwise} \end{cases} \quad (4.3)$$

where A, B and N are all constants that depend on $\alpha, n, \bar{x}, \sigma$. Thus, if α is positive, the tail, $A \cdot (B - \frac{(x-\bar{x})}{2\sigma})^{-n}$, will start below the mean (giving a so-called left tail) and vice versa for the case where α is negative. The background is fitted using a first-order polynomial.

In order to form a background-subtracted distribution of some quantity X , e.g. momentum, the background distribution for this quantity is taken from the mass side bands, where the mass side bands are defined as being from $440 \text{ MeV}/c^2$ to $540 \text{ MeV}/c^2$ and excluding the signal region. The mass side band distribution is then scaled by the ratio of the proportion of background in the signal window over the proportion of background in the mass side band. As shown in [Figure 4.8](#), the widths of the signal mass distributions for long tracks are narrower than those of downstream tracks ($\sim 25\text{-}35 \text{ MeV}/c^2$ for long tracks, $\sim 55\text{-}65 \text{ MeV}/c^2$ for downstream tracks, depending on the z bin) indicating the better long track reconstruction resolution. The background distribution under the signal window tends to be fairly flat.

The levels of background present vary with z . Mass fits are performed only in bins of z as there are insufficient statistics to also bin in p and η . However the background shape is fairly similar for all momentum and η bins so it is assumed that the proportion of background under the signal window is independent of p and η and thus it suffices to bin only in z . There is further discussion of the validity of this assumption in [section 4.7](#).

In summary, during the background-subtraction process the proportion of background

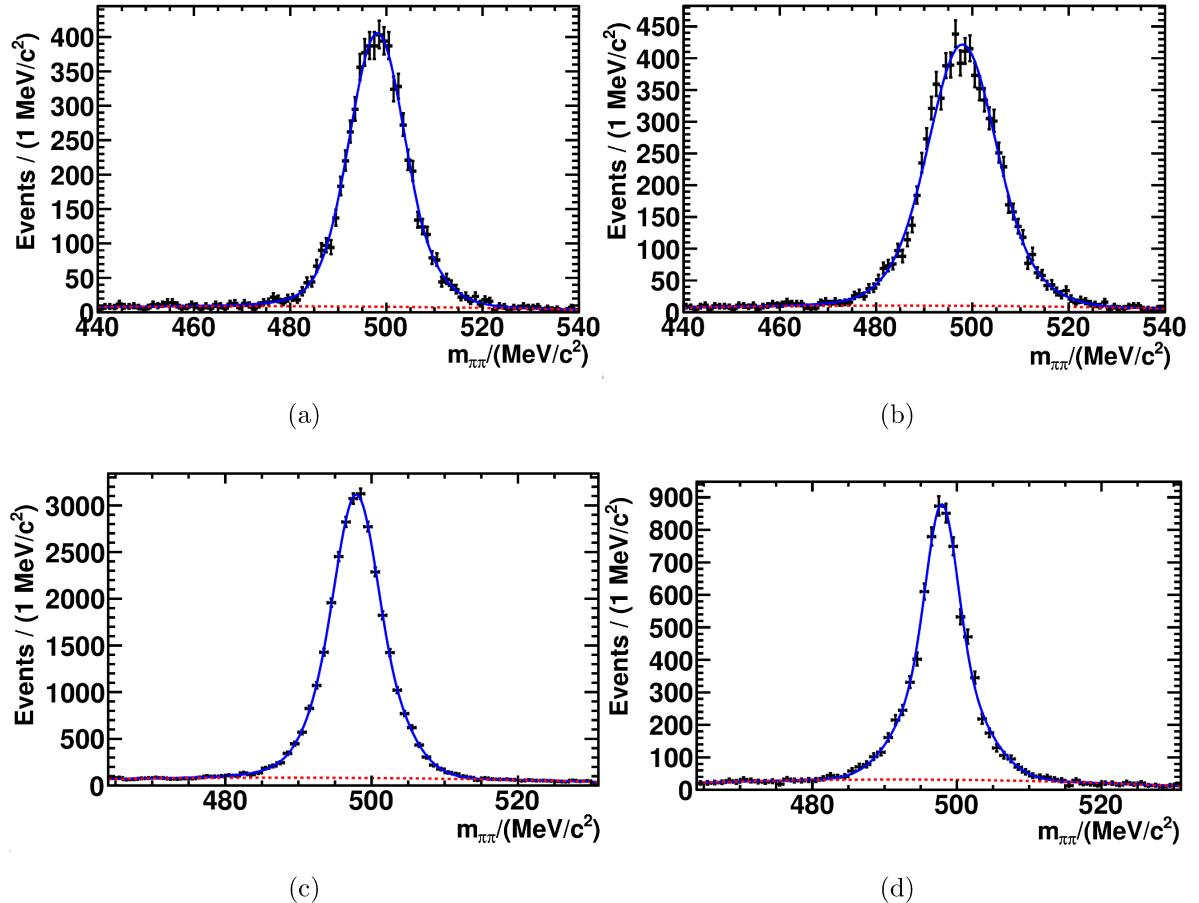


Figure 4.8: The fit to the $m_{\pi\pi}$ mass distributions for downstream ((a), (b)) and long ((c), (d)) tracks for events with $1200 < z < 1210$ mm (a), $850 < z < 860$ mm (b), $90 < z < 100$ mm (c), $300 < z < 310$ mm (d).

events present in the signal window is calculated integrated over all p and η and then this same proportion is used to multiply the mass side bands of distribution X for each individual p and η bin, for a given z bin.

4.4.1 Unfolding the reconstruction efficiency in the z_{ref} bins

For each z_{ref} bin, the long track efficiency from Ref [67], which is cited as the ratio between data and simulation, is used to extract the absolute efficiency from the data. The results

of the analysis carried out in this chapter are quoted in the same p and η bins as used in Ref [67], as discussed previously in subsection 4.1.3. In addition, in order to compare the results from this study to those from Ref [66], no reweighting of the track multiplicity in simulation is applied.

As previously discussed in chapter 3, there are two categories of LHCb simulated events. Reconstructed simulated events are those whose simulation includes the presence of the fully simulated LHCb detector, whereas so-called generator-level simulated events are those whose simulation does not include the LHCb detector. To extract the absolute reconstruction efficiency in data from the ratio of reconstruction efficiencies between data and simulation, the number of reconstructed simulated events, n_{reco} , for a given region in p and η , is compared to N_{gen} , the number of generator-level simulated events present in the same region of p and η . By combining these numbers and applying the corrections given in Ref [67] (and shown in Figure 4.3) the efficiency for data, $\epsilon_{p,\eta}$, for a given p , η (of the K_s^0) is computed using

$$\epsilon_{p,\eta} = \frac{n_{reco,p,\eta} \times \omega_{(\pi^+\pi^-)p_{\pi^\pm},\eta_{\pi^\pm}}}{N_{gen,p,\eta}}, \quad (4.4)$$

where $\omega_{(\pi^+\pi^-)p_{\pi^\pm},\eta_{\pi^\pm}}$ refers to the correction weights applied to both pions taken from the J/ψ tag-and-probe method, valid for long tracks, which are dependent on the p and η value of the pion, not the K_s^0 .

By multiplying the number of data events, $n_{data,p,\eta}$, by $1/\epsilon_{p,\eta}$, within a chosen z_{ref} bin (and subtracting the background as discussed in section 4.4), the value for N_0 , as defined in Equation 4.2, can be calculated for a given value of p and η . The reconstruction efficiency in the z_{ref} bin is computed in bins of width 1 GeV/c in p and bins of 0.1 in η . These finer bins are combined into coarser ones for the results.

The quantities $\epsilon_{p,\eta}$, $n_{reco,p,\eta}$, $n_{reco,p,\eta} \times \omega_{(\pi^+\pi^-)p_{\pi^\pm},\eta_{\pi^\pm}}$ and $N_{gen,p,\eta}$ from Equation 4.4 along with $n_{data,p,\eta}$ and $n_{data,p,\eta}/\epsilon_{p,\eta}$, are shown in Figure 4.9.

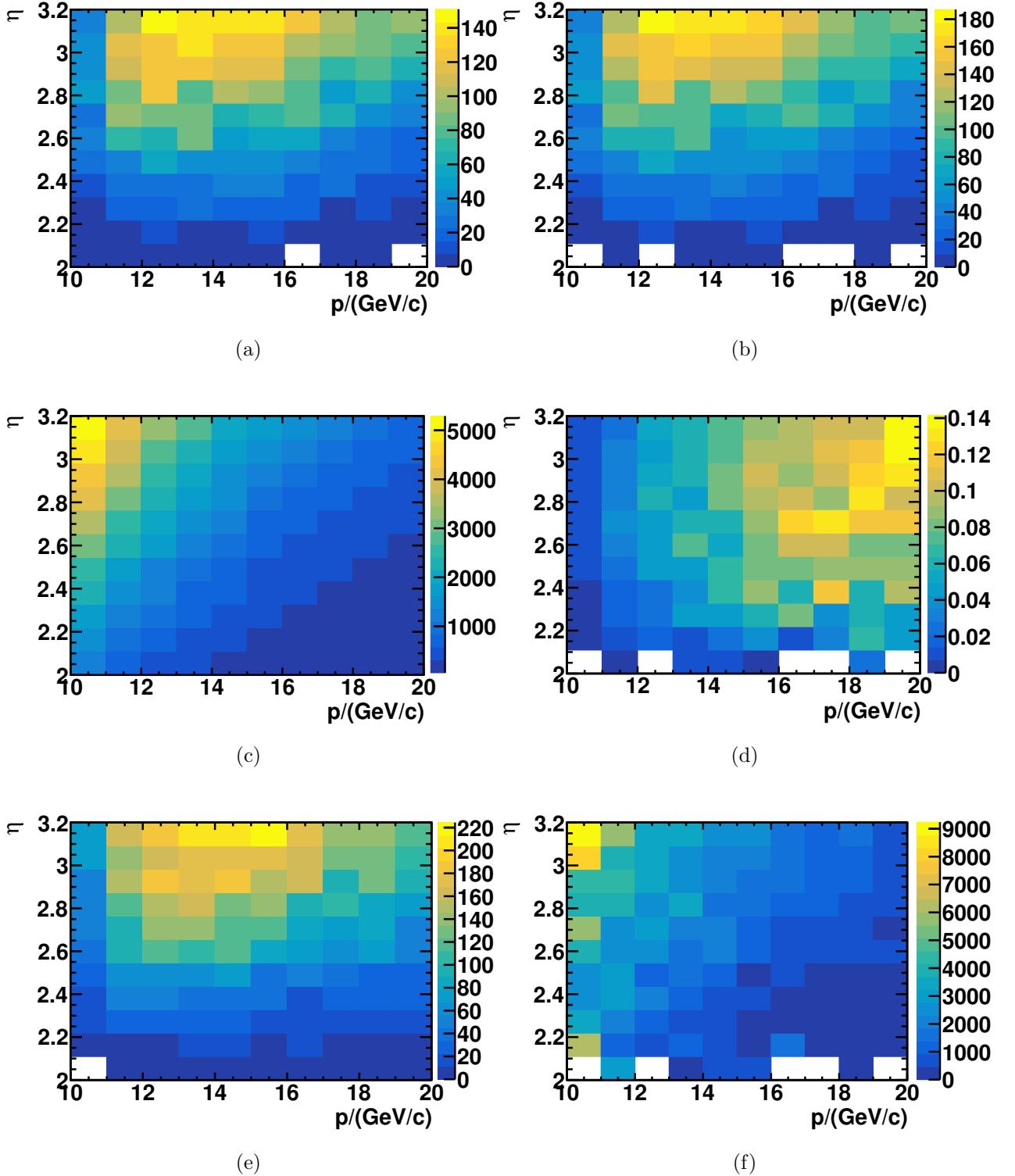


Figure 4.9: The distributions in p, η for the bin $10 < p < 20 \text{ GeV}/c$, $2.0 < \eta < 3.2$ for the quantities $n_{reco,p,\eta}$, (a), $n_{reco,p,\eta} \times \omega_{(\mu^+\mu^-)p_{\mu^\pm},\eta_{\mu^\pm}}$, (b), $N_{gen,p,\eta}$, (c), and $\epsilon_{p,\eta}$, (d) from Equation 4.4 along with $n_{data,p,\eta}$, (e), and $n_{data,p,\eta}/\epsilon_{p,\eta}$, (f).

4.5 Comparison of simulated events and data distributions

The extrapolation method is also applied to reconstructed simulated events in order to validate the extrapolation method. The distribution of the number of reconstructed events in p and η is not exactly matched between simulated events and data even after the corrections for differences between long track efficiencies from Ref [67] are applied, although they are highly similar, as shown in [Figure 4.9\(b\)](#) and [Figure 4.9\(e\)](#). This slight difference is maybe due to the physics modelling of K_s^0 mesons in Pythia.

As the reconstructed simulated events do not have exactly the same distribution of $N(z_{\text{ref}}, p, \eta)$ as in the data, the distribution of $N(p, \eta)$ for reconstructed simulated events is weighted over all z , for all sub-bins of p and η , using weights calculated from the $N(z_{\text{ref}}, p, \eta)$ distributions in background-subtracted data. These are referred to as simulation correction weights and they are given as

$$\omega_{p,\eta,z_{\text{ref}}} = \frac{n_{\text{reco},p,\eta,z_{\text{ref}}}}{n_{\text{data},p,\eta,z_{\text{ref}}}}. \quad (4.5)$$

The effect of weighting the reconstructed simulated events is shown in [Figure 4.10](#). In [Figure 4.10](#), a scale factor has also been applied to simulation in order to normalise the number of simulation events in each z_{ref} bin to the number of data events. This scale factor is applied to all simulation events.

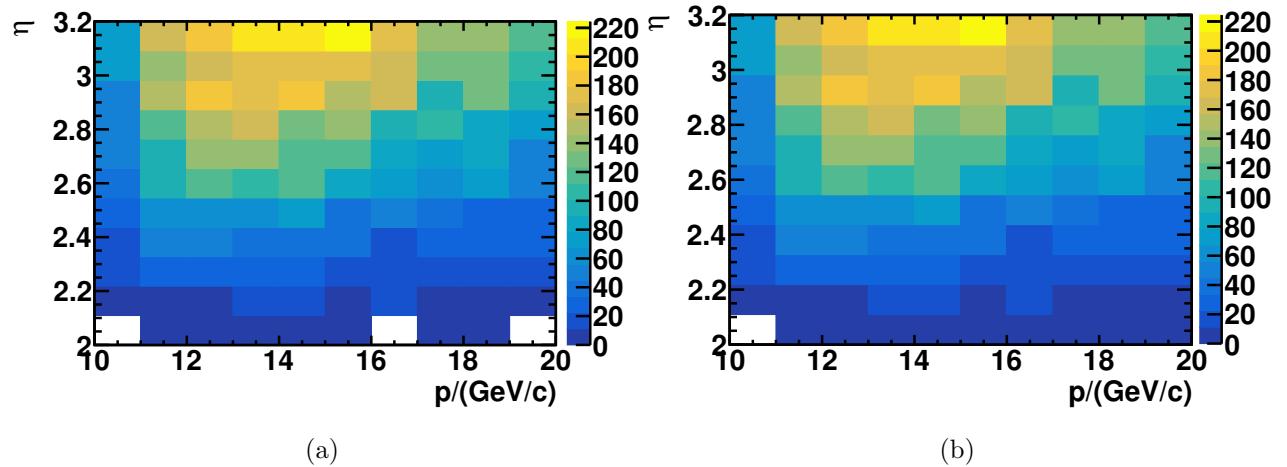


Figure 4.10: The (p, η) distribution for data, (a) and the (p, η) distribution for reconstructed simulated events after weights are applied, (b), for events with $90 < z < 100$ mm, $10 < p < 20$ GeV and $2.0 < \eta < 3.2$.

4.6 Calculation of $N(z, p)$

The extrapolation method works on the basis that the p distribution within the z reference bin, along with the number of K_s^0 decays in the z_{ref} bin, N_0 , can be used to calculate the distribution of the number of decays as a function of z and p further down the detector. This is done using Equation 4.2. The extrapolation is repeated for every z reference bin.

When calculating $N(z, p)$, the finer the momentum bin used the better. However, bins smaller than $1 \text{ GeV}/c$ result in too low statistics. The systematic uncertainty arising from the binning choice of p is investigated in section 4.7.

4.6.1 Verifying the calculation using generator-level simulated events

The calculation in Equation 4.2 is verified by taking $N(z_{\text{ref}}, p)$ in generator-level reconstructed events and using this to calculate $N(z, p)$. One can then check that the calculated distribution matches the distribution of the generator-level simulated events exactly in the

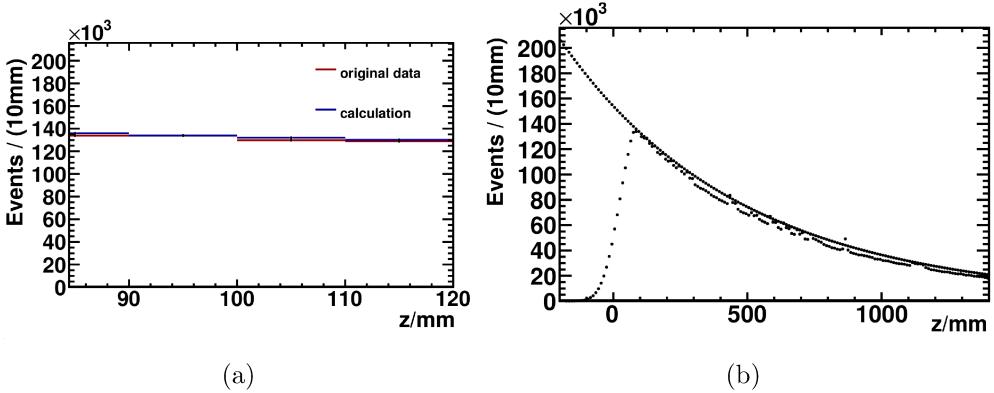


Figure 4.11: The value of N_0 used in the calculation for the z_{ref} bin $90 < z < 100$ mm, (a), which shows the N_0 in the z_{ref} bin, $90 < z < 100$ mm, and the calculated value, match exactly. The calculated (solid distribution) and the distribution for generator-level simulated events in z for bin $10 < p < 20$ GeV, $2.0 < \eta < 3.2$, $90 < z < 100$ mm (b).

reference bin used, shown in Figure 4.11(a). The comparison of the calculated distribution with generator-level simulation across all z is shown in Figure 4.11(b).

Despite the good agreement in slope between the calculated distribution and the simulated generator-level events, there are slight differences between them due to the occasional upward fluctuation in the simulation. These fluctuations are attributed to material interactions simulated in the generator-level simulated distribution which will not be modelled in the calculation ². The material interactions are further discussed in subsection 4.7.3.

²If a true generator-level simulation were used then there would be no material interactions modelled. However, here the generator-level simulation has been obtained by truth-matching the simulation, as in LHCb software Geant4 is required to decay long-lived particles

4.7 Systematic checks

The main sources of potential systematic uncertainties are from certain presumptions made in the method applied and mismodelling in the simulation, as well as from the finite binning used in z and p . These sources of potential systematic uncertainty are outlined below.

4.7.1 Finite binning of p and z

In order to quantify the effect of having finite bin widths, the central value of the bin used in both p and z is varied randomly, the result recalculated, and then compared with the original results. As shown in [Figure 4.12](#) and [4.13](#) the difference is consistent with zero. The variation in the size of the error bars is due to the changing statistics as a function of z .

4.7.2 Assuming that the background shape is the same across all p and η

As outlined in [section 4.4](#), in order for the mass fits of the K_s^0 mass distributions, which are integrated over all p and η , to be used to subtract the right amount of background in each p and η bin, it is necessary for the background distribution to have a similar shape over all p and η bins. Only the ratio between the amount of background in the mass side bands and in the signal window is important and the level of background can vary across these bins. To check the effect of this assumption, mass fits are performed in bins of p and η within each z bin and the background shapes within each p and η sub-bin are compared, as shown in [Figure 4.14](#). Across all p, η bins the background is flat. The same process was checked with a range of z bins and yielded similar results. Therefore no systematic uncertainty is attributed to this assumption.

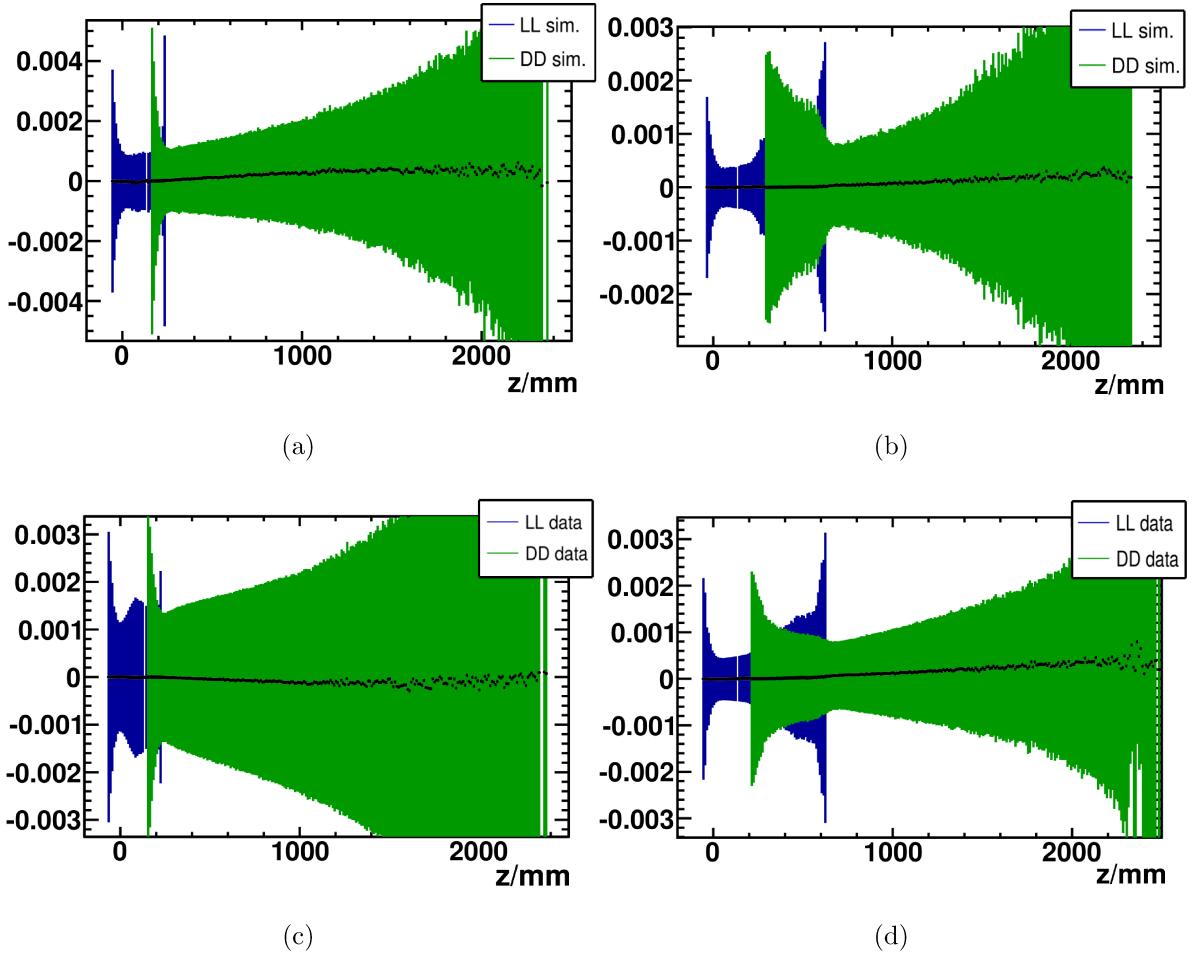


Figure 4.12: The difference between the final efficiencies with and without randomly varying the value taken as the central value in momentum p for simulation, top, and data, bottom. For events with $10 < p < 20 \text{ GeV}/c$, $2.0 < \eta < 3.2$, left, $3.2 < \eta < 5.0$, right. The y axis shows the difference in efficiencies. The downstream tracks are shown with green errors and the long tracks are shown with blue errors.

4.7.3 The modelling of material interactions in simulation

As the long track reconstruction efficiencies are calculated using leptonic final states, the simulation will better model $J/\psi \rightarrow \mu^+\mu^-$ decays than $K_s^0 \rightarrow \pi^+\pi^-$ decays, due to material interactions in the hadronic case. The analysis of Ref [67] recommends adding

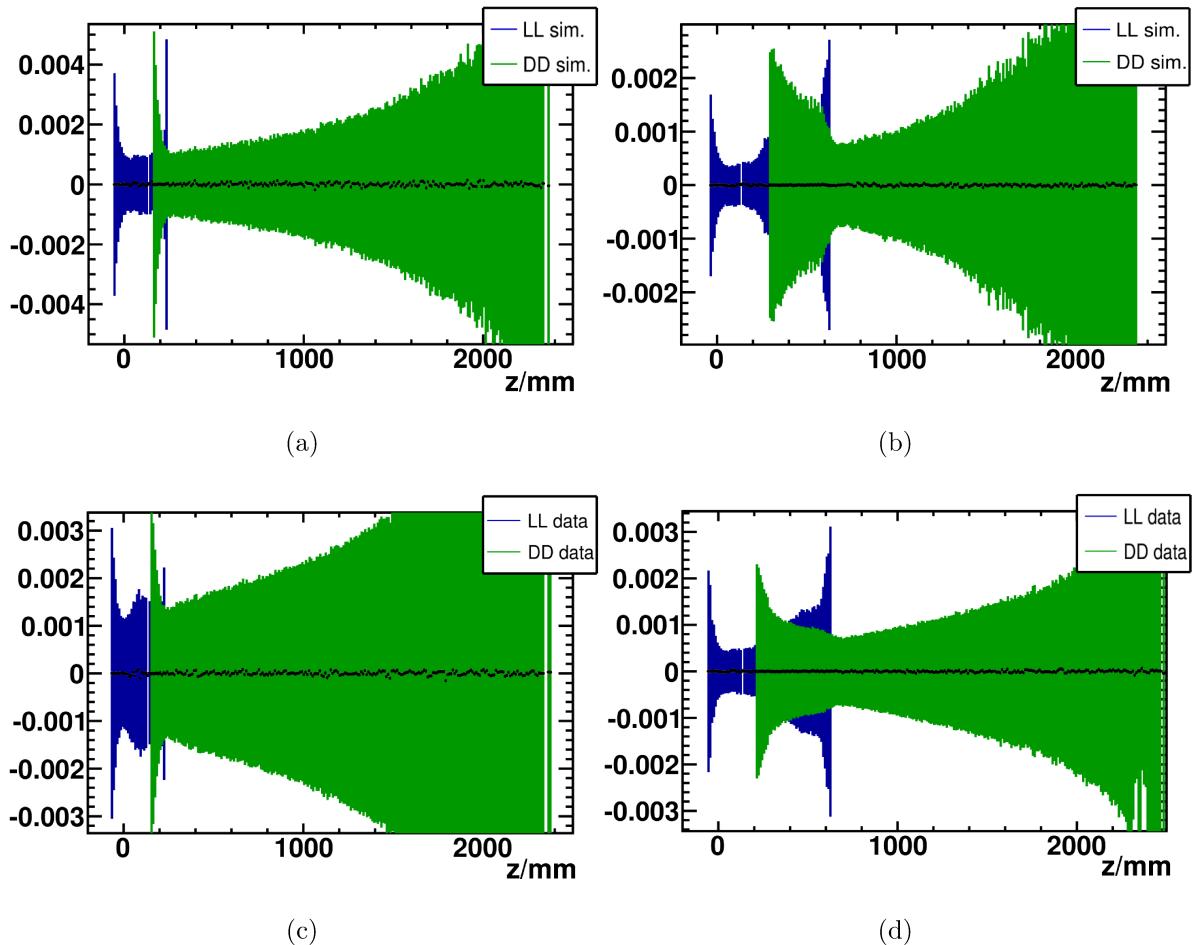


Figure 4.13: The difference between the final efficiencies with and without randomly varying the value taken as the central value in z for simulation, top, and data, bottom. For events with $10 < p < 20$, $2.0 < \eta < 3.2$, left, $3.2 < \eta < 5$, right. The y axis shows the difference in efficiencies.

an additional 1.4% error to account for the poor modelling of material interactions for hadronic final states.

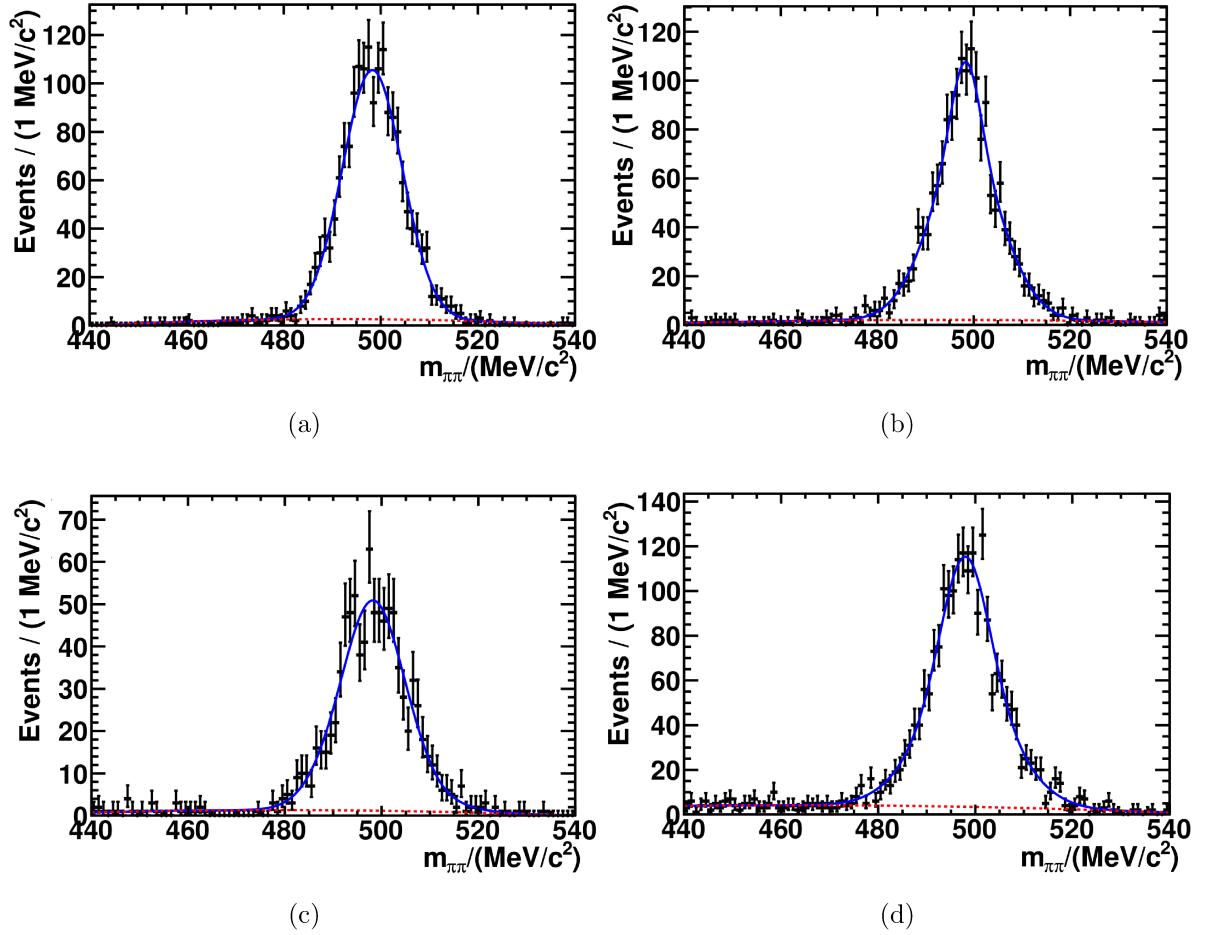


Figure 4.14: The signal and mass distributions for the same z bin but different p and η bins for downstream tracks. For events with $10 < p < 20 \text{ GeV}/c$, $2.0 < \eta < 3.2$, (a), $10 < p < 20 \text{ GeV}/c$, $3.2 < \eta < 5.0$, (b), $20 < p < 40 \text{ GeV}/c$, $2.0 < \eta < 3.2$, (c), $20 < p < 40 \text{ GeV}/c$, $3.2 < \eta < 5.0$, (d), and with a decay vertex between $1070 < z < 1080 \text{ mm}$.

4.8 Results

The tracking efficiencies are first calculated for each $1 \text{ GeV}/c$ sub-momentum bin. The efficiency is then taken as a weighted average, according to the error on the efficiency in each sub-momentum and z_{ref} bin. The final result is shown in Figure 4.15. There are no results for the bin $40 < p < 100 \text{ GeV}/c$, $2.0 < p < 3.2$ due to the limited statistics. Finally,

the ratio between simulation and data is taken, as shown in [Figure 4.16](#).

Due to the nature of the calculation of $N(z, p)$ the uncertainties are correlated, which is why the fluctuations in the points in [Figure 4.15](#) do not reflect the size of the error bars. Much of the shape in z of the efficiency distributions in [Figure 4.15](#) can be explained by the position of the sensors in the VELO, as shown in [Figure 3.7](#). The drop off at low η of the long track efficiency at $\sim 200\text{mm}$ is due to the sensors in the VELO being more spaced apart after this point. At lower η values, and thus larger track angle with respect to the beam pipe, no tracks make the VELO-track requirement of passing through at least three VELO sensors past $\sim 200\text{mm}$. At higher η , this wider spacing at $\sim 200\text{mm}$ again causes the efficiency to drop off, but some tracks still pass through three sensors. The difference in shape between the distribution in the $3.2 < \eta < 5.0$ bin within the highest momentum bin and the $3.2 < \eta < 5.0$ bin within the lower momentum bins is not due directly to the change in momentum but because the η bins are very broad, and very high momentum tracks will tend to have higher η values. The higher the η value the less sensitive the tracking efficiency will be to the spacing between the VELO sensors.

The efficiency of the downstream tracks is fairly flat in z . The gradual increase in tracking efficiency at higher z is most likely due to two factors. Firstly, it is more difficult to vertex tracks which originate at a distance from the TT stations, particularly for tracks with lower η values. Lower η tracks at lower z are also more likely to pass through the Outer Tracker, as opposed to the Inner Tracker in the T stations, where the Outer Tracker provides a worse resolution. A worse resolution means that K_s^0 mesons are less likely to pass the vertex requirements. Secondly, higher z values imply a higher momentum within a given momentum bin, which is associated with a higher efficiency. The difference in shape at low z in the downstream tracking between lower and higher η bins is just a reflection of the ability to upgrade a track to a long track, which is dependent on η due to the reasons already outlined.

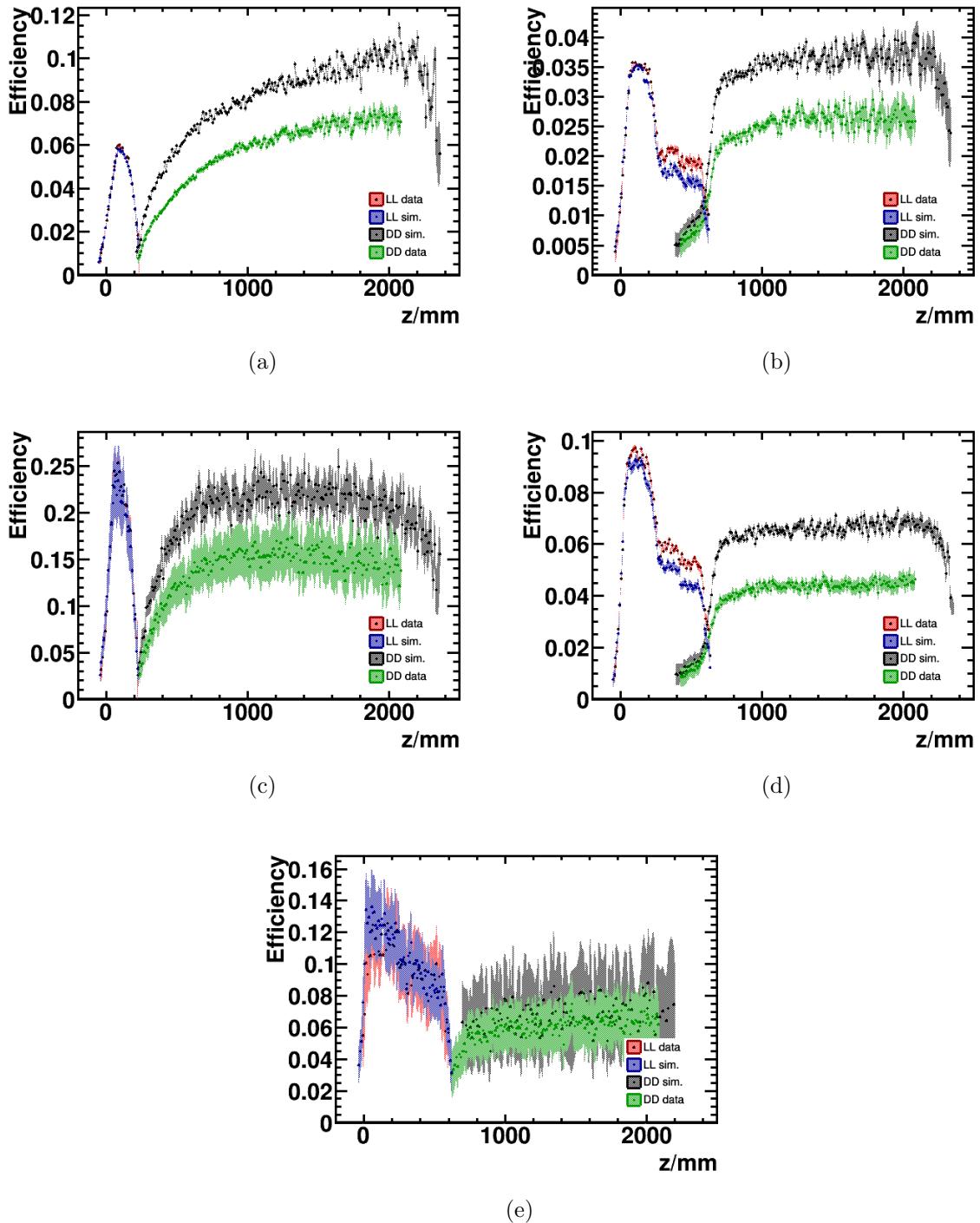


Figure 4.15: The efficiency distributions for data and reconstructed simulation. For events with $10 < p < 20 \text{ GeV}/c$, $2.0 < \eta < 3.2$ (a) and $3.2 < \eta < 5.0$ (b), $20 < p < 40 \text{ GeV}/c$, $2.0 < \eta < 3.2$ (c) and $3.2 < \eta < 5.0$ (d) and $40 < p < 100 \text{ GeV}/c$, $3.2 < \eta < 5.0 \text{ GeV}/c$, (e).

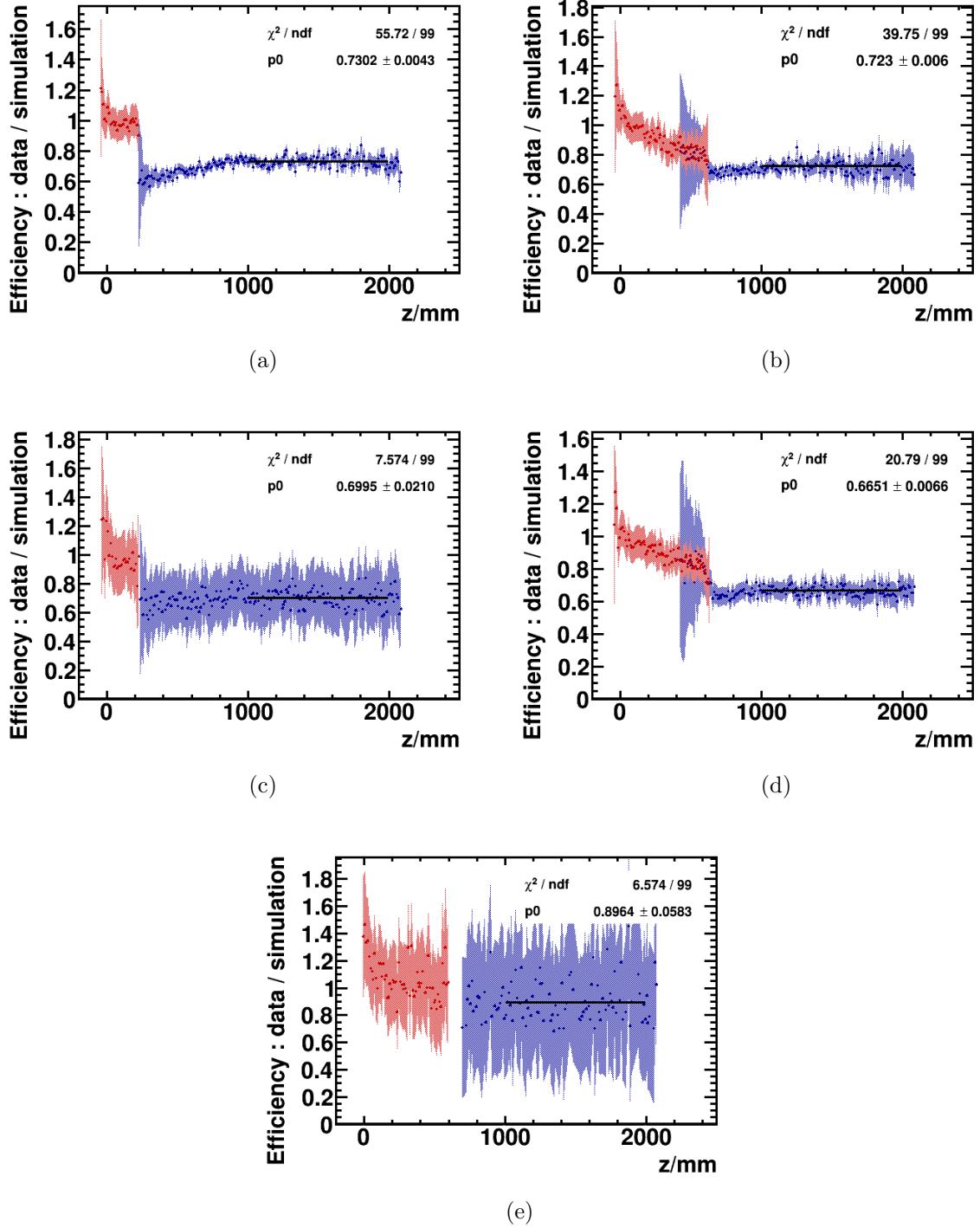


Figure 4.16: The ratio of efficiency distributions for data and reconstructed simulation. For events with $10 < p < 20 \text{ GeV}/c$, $2.0 < \eta < 3.2$ (a) and $3.2 < \eta < 5.0$ (b), $20 < p < 40 \text{ GeV}/c$, $2.0 < \eta < 3.2$ (c) and $3.2 < \eta < 5.0$ (d) and $40 < p < 100 \text{ GeV}/c$, $3.2 < \eta < 5.0 \text{ GeV}/c$, (e). The results of a zero-order fit to the downstream tracking efficiency as a function of z are shown.

4.9 Comparison between the results from this analysis and previous studies

The ratio between both the long and downstream absolute efficiencies is as expected, i.e. the absolute efficiency of long tracks is higher than for downstream tracks. The ratio between simulation and data is equal to unity over approximately the region expected, i.e. at low z below $\sim 140\text{mm}$, where the long track efficiencies are well modelled in simulation. The same ratio between data and simulation for downstream tracks is always less than unity, suggesting that the efficiency of reconstructing downstream tracks is higher in simulation than data.

To compare this result to that of Ref [66], the downstream efficiency as a function of z is fitted with a zero-order polynomial, to give an average over z for the region $z > 1000\text{mm}$, as shown in Figure 4.16. These averages are shown in Table 4.5.

All bins agree with the results from Ref [66] within 2σ , with the majority of bins agreeing within 1σ . The results also follow a similar trend with momentum, with higher momentum bins having generally lower efficiency ratios. The exception to this is the highest momentum bin where the statistics are limited.

The results from this study have already been used in a number of LHCb analyses and there is a need for the study to be repeated for Run-2 data. In addition, the technique developed in this analysis could be applied to other common long-lived particles such as Λ^0

	$2.0 < \eta < 3.2$	$3.2 < \eta < 5.0$
$10 < p < 20\text{ GeV}/c$	$0.73 \pm 0.03 \pm 0.01$	$0.72 \pm 0.03 \pm 0.01$
$20 < p < 40\text{ GeV}/c$	$0.70 \pm 0.05 \pm 0.01$	$0.66 \pm 0.02 \pm 0.01$
$40 < p < 100\text{ GeV}/c$	—	$0.90 \pm 0.09 \pm 0.01$

Table 4.5: The ratio between data and simulation downstream tracking efficiency averaged over z , from $z > 1000\text{ mm}$. The first error quoted is statistical and the second error is due to the mismodelling of hadronic material interactions in simulation.

baryons. The average efficiency values of the value from Table 4.5, excluding the highest momentum bin, is 0.70 ± 0.02 .

Chapter 5

Selection and backgrounds in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis

The Λ_b^0 is a heavy baryon containing a bottom quark. The decay mode $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ is a $b \rightarrow d$ FCNC transition and can therefore only occur at loop level in the SM. In new physics models, new heavy particles can contribute and can significantly change the branching fraction of this process.

In this chapter, an outline of the selections used to increase the fraction of signal events with respect to the background is presented, along with the types of backgrounds considered. In the next chapter, the multivariate analysis techniques used to further reduce background are presented. In [chapter 7](#) there will be a discussion of the fit models used and the efficiencies of the selections applied and finally the results of this analysis are presented in [chapter 8](#).

In this chapter in [section 5.1](#), a brief reiteration of the motivation for searching for $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ is presented, along with an outline of the analysis method. This is followed by a summary of the data and simulation samples used in this analysis in [section 5.2](#), and then the details of the initial selection applied to increase the signal fraction of events with respect to the background in [section 5.3](#). The agreement between data and simulation samples is discussed in [section 5.4](#). The selections used to remove decays from

B reflections, i.e. decays with a mis-identified particle that can accumulate at a certain B mass, and background from the Cabibbo-favoured channel $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$, are discussed in section 5.5. Finally, there is a discussion of partially reconstructed backgrounds in section 5.6.

This analysis was performed blind, meaning that events within the signal region of the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass spectrum were removed from the dataset during all selection processes outlined in this chapter and during the optimisation process outlined in chapter 6. Therefore, all plots of the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass spectrum in chapters 5–7 do not show the signal region.

5.1 Overview

The aim of the analysis presented in the next four chapters is to measure the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction. The $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ decay mode has never been observed and neither has any other baryon decay mediated by a $b \rightarrow d$ transition. The $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction is measured relative to the resonant decay mode $\Lambda_b^0 \rightarrow p\pi^-J/\psi$ ($\rightarrow\mu^+\mu^-$), which has already been observed [69]. It is desirable to measure the branching fraction relative to the branching fraction of another decay with the same final state, as many sources of systematic uncertainty cancel. For example the absolute number of Λ_b^0 baryons produced within LHCb does not affect the ratio of $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)/\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-J/\psi (\rightarrow\mu^+\mu^-))$. The $\Lambda_b^0 \rightarrow p\pi^-J/\psi (\rightarrow\mu^+\mu^-)$ channel is hereafter denoted simply $\Lambda_b^0 \rightarrow J/\psi p\pi^-$.

The Feynman diagram for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ decay is shown in Figure 2.3(a) and the diagram for the hadronic decay mode $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ is shown in Figure 5.1. For both the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ decay modes, the proton and pion can come from strongly decaying N^* resonances.

Table 5.1 shows the comparison of the branching fractions of various decays of the form $\frac{H_b \rightarrow X_q \mu\mu}{H_b \rightarrow X_q J/\psi (\rightarrow\mu\mu)}$, where H_b is a hadron containing a b -quark and X_q is any hadron. The majority of the ratios, $\frac{H_b \rightarrow X_q J/\psi (\rightarrow\mu\mu)}{H_b \rightarrow X_q \mu\mu}$, in Table 5.1, are of order 100. Taking the ratio between the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fractions to also be of order 100,

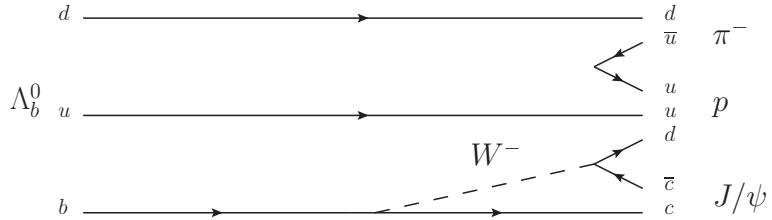


Figure 5.1: Feynman diagram for the decay $\Lambda_b^0 \rightarrow J/\psi p \pi^-$.

the 2000 $\Lambda_b^0 \rightarrow J/\psi p \pi^-$ decays observed in 3 fb^{-1} of LHCb data in Ref. [69] would then lead to an expectation of $\sim 10\text{--}20$ $\Lambda_b^0 \rightarrow p \pi^- \mu^+ \mu^-$ signal events, depending on the relative efficiencies between the $\Lambda_b^0 \rightarrow p \pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p \pi^-$ channels. If the ratio of the Λ_b^0 decay modes, $\frac{\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-}{\Lambda_b^0 \rightarrow \Lambda J/\psi}$ is used however, an event yield of 30–60 is expected. In either case, given the small number of signal events expected, the primary challenge for this analysis is the reduction of backgrounds. The branching fractions for the decays $\Lambda_b^0 \rightarrow J/\psi p \pi^-$ and $\Lambda_b^0 \rightarrow J/\psi p K^-$ are also shown in Table 5.1. There is currently no branching fraction measurement for the decay $\Lambda_b^0 \rightarrow p K^- \mu^+ \mu^-$, although it is an ongoing LHCb analysis at the time of writing. Once the $\Lambda_b^0 \rightarrow p K^- \mu^+ \mu^-$ analysis at LHCb is completed, the $\Lambda_b^0 \rightarrow p K^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow p \pi^- \mu^+ \mu^-$ branching fractions will be combined to deduce the ratio, \mathcal{R} , between them. Note that there is little to gain by deducing the $\Lambda_b^0 \rightarrow p K^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow p \pi^- \mu^+ \mu^-$ branching fractions in a combined analysis (which would allow for the cancellation of some systematic uncertainties when calculating \mathcal{R}) as the error on \mathcal{R} will be statistically dominated due to the limited statistics available, particularly in the $\Lambda_b^0 \rightarrow p \pi^- \mu^+ \mu^-$ channel. As such, the work outlined in this thesis only measures the branching fraction measurement for $\Lambda_b^0 \rightarrow p \pi^- \mu^+ \mu^-$ and does not attempt to calculate \mathcal{R} .

Decay	Branching Fraction	Ratio of branching fractions $J/\psi \rightarrow \mu^+\mu^-/\mu\mu$
$B^0 \rightarrow K^{*0}\mu^+\mu^-$	$(1.05 \pm 0.10) \times 10^{-6}$	74.9 ± 7.9
$B^0 \rightarrow K^{*0}J/\psi$	$(1.32 \pm 0.06) \times 10^{-3}$	
$B^+ \rightarrow K^+\mu^+\mu^-$	$(4.43 \pm 0.24) \times 10^{-7}$	137.2 ± 8.6
$B^+ \rightarrow K^+J/\psi$	$(1.02 \pm 0.03) \times 10^{-3}$	
$B^0 \rightarrow \pi^+\mu^+\mu^-$	$(2.3 \pm 0.6) \times 10^{-8}$	106 ± 30
$B^0 \rightarrow \pi^+J/\psi$	$(4.1 \pm 0.4) \times 10^{-5}$	
$\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$	$(1.1 \pm 0.3) \times 10^{-6}$	34 ± 12
$\Lambda_b^0 \rightarrow \Lambda^0J/\psi$	$(6.2 \pm 1.4) \times 10^{-4}$	
$\Lambda_b^0 \rightarrow J/\psi p\pi^-$	$(2.6 \pm 0.5) \times 10^{-5}$	—
$\Lambda_b^0 \rightarrow J/\psi pK^-$	$(3.0 \pm 0.6) \times 10^{-4}$	—

Table 5.1: Comparison of branching fractions between the $\mu^+\mu^-$ and J/ψ channels for various decays. The ratio between the resonant and non-resonant channels is estimated assuming that $\mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) = 0.0596 \pm 0.0003$. All values for the branching fractions are taken from Ref. [7], with the exception of the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ branching fraction measurements, which are taken from Refs [69] and [70] respectively.

5.1.1 Analysis strategy

The overall idea of this analysis is to fit the Λ_b^0 mass distribution for signal and background components after applying background reduction criteria in the form of specific vetoes and a BDT. The remaining backgrounds, after these specific vetoes and the BDT have been applied, are combinatorial background and partially reconstructed (part-reco) background. Combinatorial background arises from cases when tracks coming from different mother particles are combined to form a candidate. Part-reco background refers to cases when a decay has been mis-identified as a signal candidate and not all the final state particles of the mis-identified decay are included in the final reconstruction. In this analysis, the part-reco background predominantly comes from semi-leptonic cascades of the form $\alpha \rightarrow \beta(\rightarrow X)\mu\nu$, where the decay products X contain a hadron that has been mis-identified as a muon. The $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow \psi(2S)(\rightarrow \mu^+\mu^-)p\pi^-$ channels are vetoed during the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ selection by applying cuts on the q^2 spectrum.

Using $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ as the normalisation channel, the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction can be determined from the expression

$$\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-) = \frac{N_{\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-}}{N_{\Lambda_b^0 \rightarrow J/\psi p\pi^-}} \times \frac{\epsilon_{\Lambda_b^0 \rightarrow J/\psi p\pi^-}}{\epsilon_{\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-}} \times \mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-) \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-), \quad (5.1)$$

where ϵ_X refers to the total selection efficiency of channel X . The terms N_X in [Equation 5.1](#) are extracted by fitting the Λ_b^0 mass distributions in data and ϵ_X are calculated using simulation, with the exception of the BDT efficiency. The order of selection applied in the analysis is outlined as follows:

- Application of an initial loose selection and removal of the $\Lambda_b^0 \rightarrow \psi(2S)p\pi^-$ channel by placing vetoes in q^2 .
- Application of a selection dedicated to the removal of [reflections](#), i.e. decays with a mis-identified particle that can accumulate at a certain B mass.
- Application of a BDT to further reduce combinatorial and reflection backgrounds.

- Selection of either the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ or the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel by applying vetoes or selections respectively in q^2 .

5.2 Data samples and simulation

The data-set analysed during this study represents an integrated luminosity of 3 fb^{-1} of pp -collisions recorded over the 2011 and 2012 data-taking periods.

5.2.1 Data and simulation samples

Signal and background simulated events were generated with PYTHIA 6.4 and 8.1 [47, 48], parameterised as specified in Ref. [71]. Unstable particles were decayed with EVTGEN [49] and the detector response was simulated with GEANT4 [72, 73]. All candidates are required to be in the detector acceptance. All simulation samples used in this analysis are generated using phase space only models. In phase space simulation, the decay rate for the simulated process is calculated disregarding the form factors arising from the presence of hadronic matrix elements in the complete expression for the decay rate, which have not yet been predicted. The distributions generated in phase space simulation are dependent only on the kinematics of the decay, meaning that resonances and other QCD effects will not be modelled.

5.3 Preselection

The term preselection refers to the initial selection placed on data candidates to increase the fraction of signal events with respect to the background. This section discusses all the preselection criteria applied.

5.3.1 Stripping cuts

As previously discussed in [chapter 4](#), after the data output by the LHCb HLT2 trigger has been stored, an initial selection is placed on the data, referred to as the stripping selection. The stripping line used for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis is the B2XMuMu line. This line features the selections shown in [Table 5.2](#). In [Table 5.2](#), the variables m_B , $m_{\mu^+\mu^-}$ refer to the invariant mass of $K^+\pi^-\mu^+\mu^-$ and of the dimuon system, respectively. The B vertex χ^2/ndof refers to the vertex fit for the combination of the daughter tracks of the B , where B refers to the mother particles, which in the stripping line is assumed to be a B meson. All other variables in [Table 5.2](#) have already been defined in [section 3.2](#).

Selection	Criteria
m_B	$4800 < m_{K^+\pi^-\mu^+\mu^-} < 7100 \text{ MeV}/c^2$
$m_{\mu^+\mu^-}$	$< 7100 \text{ MeV}/c^2$
μ $IP\chi^2$	> 9
h $IP\chi^2$	> 6
B $IP\chi^2$	< 16
B vertex χ^2/ndof	< 8
B $FD\chi^2$	> 121
μ $DLL_{\mu\pi}$	> -3
μ isMuon	True

Table 5.2: The selection criteria used in the stripping line.

The stripping line does not apply hadron PID requirements and events are selected using the stripping line criteria in [Table 5.2](#) with the mass hypothesis of the daughters being taken as $K^+\pi^-\mu^+\mu^-$. Once selected, the mass hypothesis of the kaon is substituted with that of a proton. This means that the cut at $4.8 \text{ GeV}/c^2$ on the $K^+\pi^-\mu^+\mu^-$ mass in the stripping line is not a clean cut off under the $p\pi^-\mu^+\mu^-$ mass hypothesis but rather gives a shoulder rising up from $4.8 \text{ GeV}/c^2$. Due to this rise, when fitting the Λ_b^0 mass distribution, only events above $5100 \text{ MeV}/c^2$ are considered.

5.3.2 Trigger requirements

The trigger requirements placed on all data and simulation samples are listed in [Table 5.3](#). At least one line from each of the L0, HLT1 and HLT2 triggers must be passed. All trigger lines in [Table 5.3](#) have been previously discussed in sections [3.1](#) and [3.2](#) in [chapter 3](#).

Trigger Decisions
LambdabL0MuonDecisionTOS
LambdabHlt1TrackAllL0DecisionTOS
LambdabHlt1TrackMuonDecisionTOS
LambdabHlt2TopoMu2BodyBBDTDecisionTOS
LambdabHlt2TopoMu3BodyBBDTDecisionTOS
LambdabHlt2Topo2BodyBBDTDecisionTOS
LambdabHlt2DiMuonDetachedDecisionTOS
LambdabHlt2DiMuonDetachedHeavyDecisionTOS

Table 5.3: Trigger requirements.

5.3.3 Preselection criteria

Further preselection cuts have been applied as outlined in [Table 5.4](#). These cuts are placed to reduce combinatorial background, and to eliminate some backgrounds arising from the mis-identification of one or more daughter particles, i.e. reflection backgrounds. These reflection backgrounds are discussed in full in [section 5.5](#), along with the additional selections used to further reduce them. Similarly, the hard cut made on the proton momentum is imposed because of the increased mis-identification rate of kaons as protons at low momentum when applying the PID DLL_{pK} requirement, where $DLL_{pK} = DLL_{p\pi} - DLL_{K\pi}$. Similarly, the hard cut made on DLL_{pK} is because of the large background from $B^0 \rightarrow K^{*0} (\rightarrow K^+ \pi^-) \mu^+ \mu^-$ decays. The lower bound cut on the dihadron mass, $m_{p\pi}$, is to ensure that there is no contribution from $\Lambda_b^0 \rightarrow \Lambda^0 (\rightarrow p\pi^-) \mu^+ \mu^-$ decays. This cut is

99.99% efficient on the signal channel, according to simulation. There is no upper bound placed on $m_{p\pi}$, where values go up to just under $5000 \text{ MeV}/c^2$ according to simulation. The branching fraction measurement performed in this analysis is therefore defined as being valid within the region of $1120 \text{ MeV}/c^2 < m_{p\pi} < 5000 \text{ MeV}/c^2$. The q^2 range used in the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ channel is from 0 to $20 \text{ GeV}^2/c^4$, where the upper limit represents approximately the phase space limit. Additional vetoes are placed on the q^2 spectrum to suppress background, as discussed in [subsection 5.3.4](#).

The efficiency of the total preselection on signal is $(67.7 \pm 0.5)\%$, with the largest efficiency loss being due to the PID selection.

Quantity	Selection
$p p_T$	$> 400 \text{ MeV}/c$
πp_T	$> 400 \text{ MeV}/c$
p momenta	$> 7500 \text{ MeV}/c$
π momenta	$> 2000 \text{ MeV}/c$
Λ_b^0 vertex χ^2	< 4
Λ_b^0 DIRA	> 0.9999
$m_{p\pi}$	$1120 < m_{p\pi} < 5000 \text{ MeV}/c^2$
<hr/>	
p DLL $_{p\pi}$	> 0
p (DLL $_{p\pi}$ -DLL $_{K\pi}$)	> 8
π^- DLL $_{K\pi}$	< -5
π^- isMuon	False
p isMuon	False

Table 5.4: Preselection applied.

5.3.4 Selections in q^2

The majority of the selection process during this analysis is performed on the combined $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ data sets, with the only selection on q^2 initially being

the removal of the $\psi(2S)$ region. In the final stages of the analysis additional cuts are placed on the q^2 distributions to select either the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ or $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ candidates, with the criteria shown in [Table 5.5](#).

Cut	q^2 veto / (GeV $^2/c^4$)
J/ψ veto	8.0–11.0
$\psi(2S)$ veto	12.5–15.0
Cut used to select the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel	9.0–10.0

Table 5.5: Vetoing in q^2 for the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ channel and the selection in q^2 for the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ normalisation channel.

5.4 Data and simulation agreement

Simulation is used to estimate the selection efficiency, however some variables, such as the PID likelihoods, are known to be poorly described in the simulation. This section describes the techniques used to improve the agreement between data and simulation.

There are two techniques used: the reweighting of the simulation, discussed in [subsection 5.4.1](#), and the resampling of the simulation, discussed in [subsection 5.4.2](#).

5.4.1 Simulation reweighting

The simulation is reweighted in order to better reproduce the distributions of variables used in the selection. These variables are the track multiplicity (n_t), momentum (p), transverse momentum (p_T) and lifetime (t) of the Λ_b^0 . The $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ decay is used as a signal proxy. The weights are calculated by comparing $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation. The $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data candidates have all selection criteria applied. These weights are then applied to both $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation.

The distributions of Λ_b^0 p , p_T , t and n_t for $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data along with $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation before and after reweighting are shown in [Figure 5.2](#). The equivalent distribu-

tions for $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation, again shown against $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data, are shown in [Figure 5.3](#).

Both the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation agree with $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data within errors after reweighting. The effect of the simulation reweighting on the final efficiency value is discussed in more detail as a source of systematic uncertainty in [subsection 8.2.5](#).

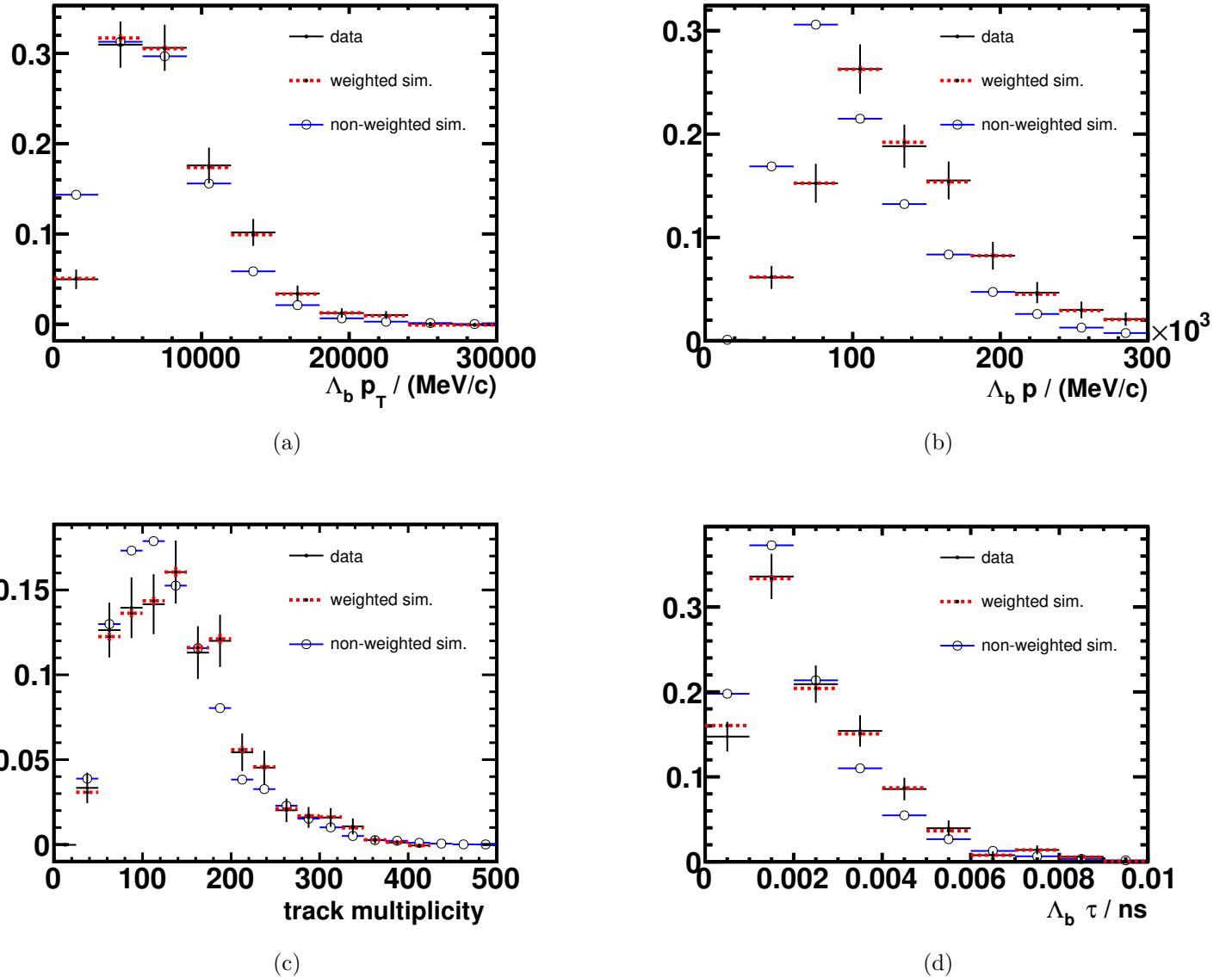
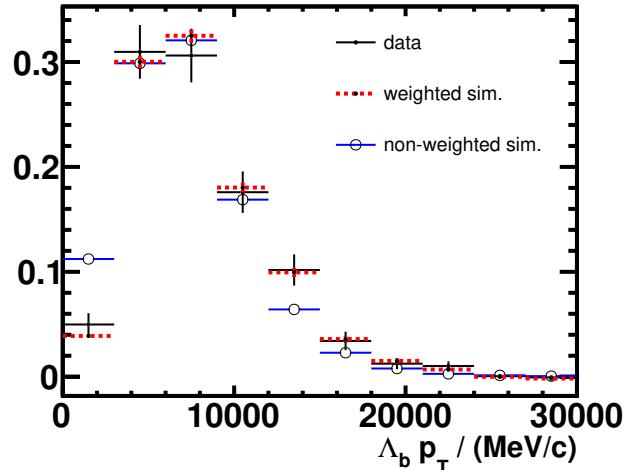
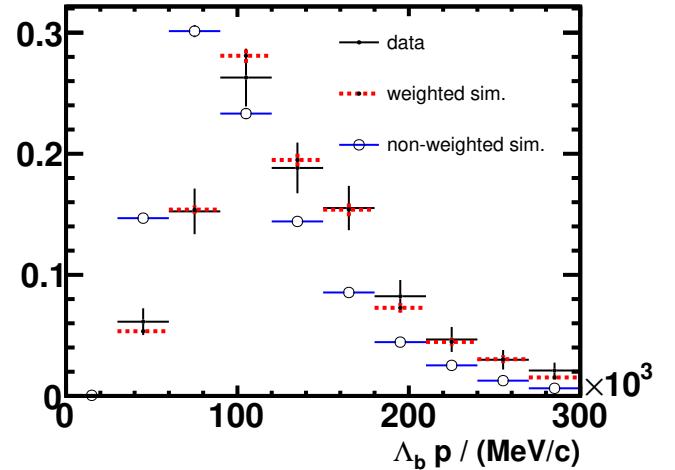


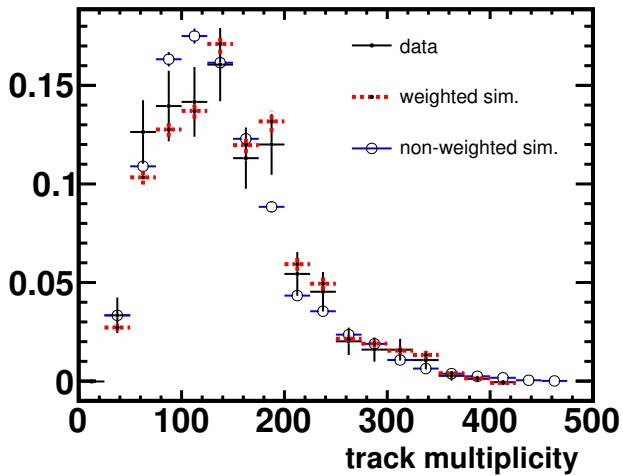
Figure 5.2: $\Lambda_b^0 \rightarrow J/\psi p \pi^-$ simulation distributions, before and after reweighting compared to sWeighted $\Lambda_b^0 \rightarrow J/\psi p \pi^-$ data for $\Lambda_b^0 p_T$ (a), $\Lambda_b^0 p$ (b), track multiplicity (c), $\Lambda_b^0 \tau$ (d).



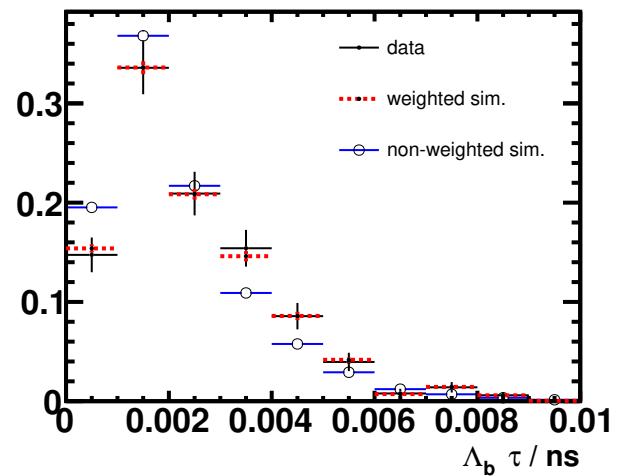
(a)



(b)



(c)



(d)

Figure 5.3: $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation distributions, before and after reweighting compared to sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data for $\Lambda_b^0 p_T$ (a), $\Lambda_b^0 p$ (b), track multiplicity (c), $\Lambda_b^0 \tau$ (d).

5.4.2 Resampling the simulation

The PID variables, $\text{DLL}_{X\pi}$, express the difference in an event's total likelihood when the hypothesis of the track in question is changed from being that of a pion to particle X. The distributions of these variables in simulation are resampled, meaning that the PID value for an event is replaced with a value picked at random from the PID distribution for a given track multiplicity, pseudo rapidity (η) and momentum taken from the data¹. The data samples used to produce these distributions are obtained from large calibration samples where different species of particles have been selected without using any PID information. The PID variables of interest in the present analysis are $\text{DLL}_{p\pi}$, $\text{DLL}_{K\pi}$ and $\text{DLL}_{pK} = \text{DLL}_{p\pi} - \text{DLL}_{K\pi}$. The calibration samples used to produce the relevant PID distributions in data are shown in [Table 5.6](#). The data calibration samples are provided by a dedicated LHCb software package.

Particle	Decays used	PID variables
K	$D^{+*} \rightarrow D^0 (\rightarrow K^- \pi^+) \pi^+$	$\text{DLL}_{K\pi}$
π	$D^{+*} \rightarrow D^0 (\rightarrow K^- \pi^+) \pi^+$	$\text{DLL}_{K\pi}$
p	$\Lambda^0 \rightarrow p \pi^-$	$\text{DLL}_{p\pi}$

Table 5.6: Decay types used to produce the calibration samples used for resampling.

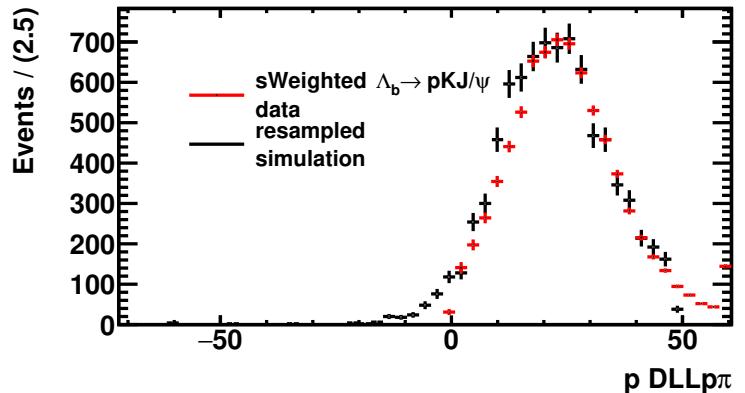
Verifying the resampled simulation against data

Resampled simulation samples are checked against $\Lambda_b^0 \rightarrow J/\psi pK^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ weighted data. Comparisons for the pion $\text{DLL}_{K\pi}$ and the proton $\text{DLL}_{p\pi}$ PID variables between simulation and data are shown in [Figure 5.4](#). The simulation has been reweighted in these plots as outlined in [subsection 5.4.1](#). The agreement between data and simulation is good for both PID variables. The discrepancy at high values for the proton $\text{DLL}_{p\pi}$ is due to an upper cut of 50 in the data calibration samples. The effect that this cut has on

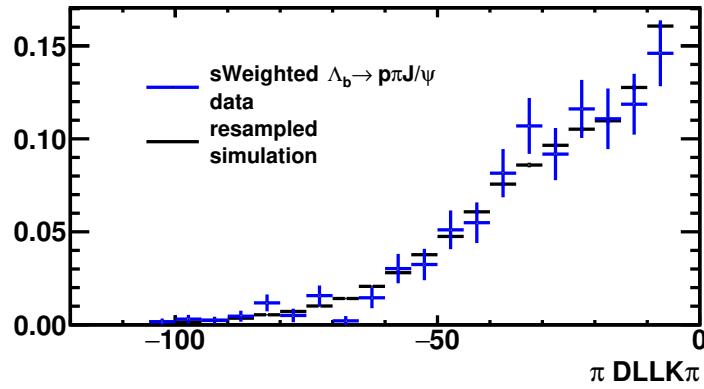
¹The advantage of using this technique is that the PID distributions are not correlated after resampling, meaning that they can still be used to train BDTs (although this is not the case in this analysis)

the final relative efficiency between the signal and normalisation channel is discussed in section 8.2.2.

More generally, any discrepancies between data and simulation are not considered problematic, as the same PID variables are featured in the signal and normalisation channels and the simulation is only used to estimate the relative efficiency between the two.



(a)



(b)

Figure 5.4: The distribution of the proton $DLL_{p\pi}$ variable for resampled $\Lambda_b^0 \rightarrow J/\psi pK^-$ simulation and $\Lambda_b^0 \rightarrow J/\psi pK^-$ sWeighted data (a), and the distribution of the pion $DLL_{K\pi}$ variable for resampled $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ sWeighted data (b).

The proton DLL_{pK} distribution, as seen in Figure 5.5, is poorly replicated in simulation.

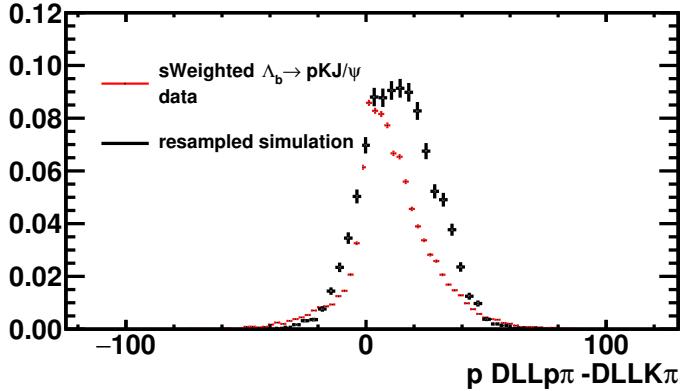


Figure 5.5: The proton $DLL_{p\pi}$ - $DLL_{K\pi}$ variable for $\Lambda_b^0 \rightarrow J/\psi p K^-$ sWeighted data and $\Lambda_b^0 \rightarrow J/\psi p K^-$ resampled simulation.

This is due to the fact that binning the PID variables in the quantities n_t , η and p preserves only the correlation in these binned quantities after resampling. This presents a problem with the variable DLL_{pK} as in the data samples provided by LHCb software it is expressed as the difference between $DLL_{K\pi}$ and $DLL_{p\pi}$, meaning it is not possible to resample the variable DLL_{pK} directly. These two variables, $DLL_{K\pi}$ and $DLL_{p\pi}$, are correlated: if a particle, with a given value of n_t , η and p , creates a ring in the RICH detectors which appears more pion-like when considering $DLL_{p\pi}$, it is also more likely to be classified as more pion-like when considering $DLL_{K\pi}$.

Due to this poor replication, the DLL_{pK} variable cannot be modelled with resampled simulation. Where the variable is used in data, the selection efficiency is computed directly using the data calibration samples shown in [Table 5.6](#).

5.5 Reflection backgrounds

A reflection, or peaking, background is a decay with a mis-identified daughter. Despite the daughter mis-identification, such decays can accumulate in mass, causing a peaking structure. However, owing to baryon number conservation, for B meson decays to form peaking backgrounds under the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass hypothesis, a kaon or pion must be

mis-identified as a proton. The difference in mass between the true and false hypotheses is so large as to wash-out much of the peaking structure.

In this section, the removal of B reflections, that is, decays of type $B \rightarrow hh\mu\mu$ or $B \rightarrow hhJ/\psi$, where h represents hadrons, is discussed in [subsection 5.5.1](#) and the number of reflections remaining after these vetoes are applied is estimated. This is followed by an examination of the removal of $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ reflections in [subsection 5.5.2](#). Finally, there is a discussion of the removal of double mis-identified peaking backgrounds in [subsection 5.5.3](#) and double mis-identified charmonium resonances in [subsection 5.5.4](#).

To reduce the amount of background as much as possible and to make the background shape easier to model, all of the major reflection channels are vetoed. All cuts are applied to both the signal channel $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and the normalisation channel $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ identically.

5.5.1 B reflection vetoes

The B reflections considered are:

- $B^0 \rightarrow (K^+ \rightarrow p)\pi^-\mu^+\mu^-$
- $B^0 \rightarrow (\pi^+ \rightarrow p)(K^- \rightarrow \pi^-)\mu^+\mu^-$
- $B^0 \rightarrow (\pi^+ \rightarrow p)\pi^-\mu^+\mu^-$
- $B_s^0 \rightarrow (\pi^+ \rightarrow p)\pi^-\mu^+\mu^-$
- $B_s^0 \rightarrow (K^+ \rightarrow p)(K^- \rightarrow \pi^-)\mu^+\mu^-$

where, in each case, the dimuon pair are either non-resonant or from a J/ψ and the notation ($X \rightarrow Y$) denotes X as being the true ID of the particle and Y as the PID hypothesis used to reconstruct the event.

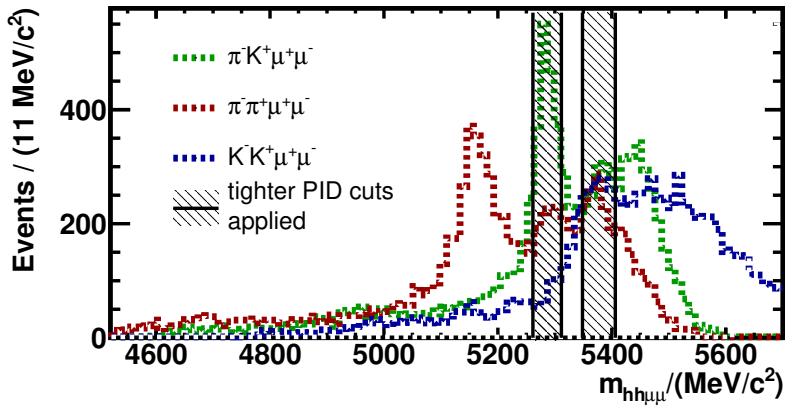
All reflections are vetoed by changing the track's mass hypothesis and re-evaluating the relevant invariant mass. If an event is close to the relevant mother particle mass a harder PID cut is then applied to the event. For example, the mass of the proton is changed to

the mass of a kaon for $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)\mu^+\mu^-$ events, and events with mass, $m_{K\pi\mu\mu}$, close to that of the B^0 mass, have a harder PID cut applied. These cuts to the daughter hadrons are referred to as mass-dependent PID cuts. The strength of the cut applied is dependent on the size of the contribution of the reflection background in question. For the case of the decay $B_s^0 \rightarrow \phi(\rightarrow K^+K^-)\mu^+\mu^-$, a veto is placed around the ϕ mass. The regions in B mass where harder PID cuts are applied are shown in [Figure 5.6](#), as indicated by the shaded regions. These regions are detailed in [Table 5.7](#). [Figure 5.6\(a\)](#) shows data within $\pm 60 \text{ MeV}/c^2$ of the Λ_b^0 mass, under different mass hypotheses with only preselection cuts applied, but no PID criteria. [Figure 5.6\(b\)](#) shows $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ signal simulation under these same mass hypotheses. The dominance of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ channel² is apparent in [Figure 5.6\(a\)](#). The exact vetoes used for each reflection background are shown in [Table 5.7](#), along with the fraction of the relevant background that is removed by these cuts.

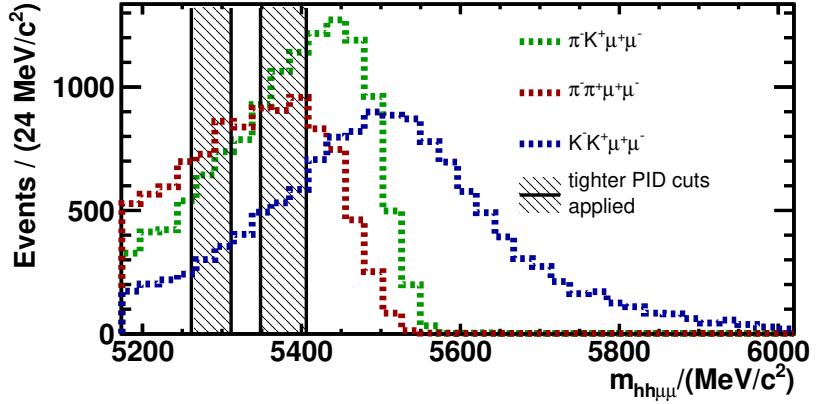
The effect of these mass-dependent PID cuts on the signal mass distribution when the daughters are given the correct mass hypothesis of $p\pi^-\mu^+\mu^-$ is shown in [Figure 5.7](#). In [Figure 5.7\(a\)](#) the signal mass distribution is shown with and without the mass-dependent cuts, as listed in [Table 5.7](#), applied. [Figure 5.7\(b\)](#) shows the difference between the two histograms in [Figure 5.7\(a\)](#), where each histogram has been normalised before taking the difference between them. The PID cuts in [Table 5.7](#) are dependent on the mass of the daughters when one or more of the daughters are given an alternative mass hypothesis. Due to the large difference in mass between the proton and the mass of a pion or kaon, the shaded regions indicated in [Figure 5.6\(b\)](#), when plotted under the signal mass hypothesis, are much wider. As a result, there is no strong dependence on the signal mass of the PID cuts in [Table 5.7](#).

The width of the windows within which harder PID cuts are applied were chosen by rerunning the whole analysis with different window widths and optimising using the Figure of Merit (**FOM**) $\frac{S}{\sqrt{S+B}}$, where S is the expected number of signal events and B is the

²Where for the case of [Figure 5.6\(a\)](#) the dimuon pair can either be decaying via a J/ψ resonance or non-resonantly



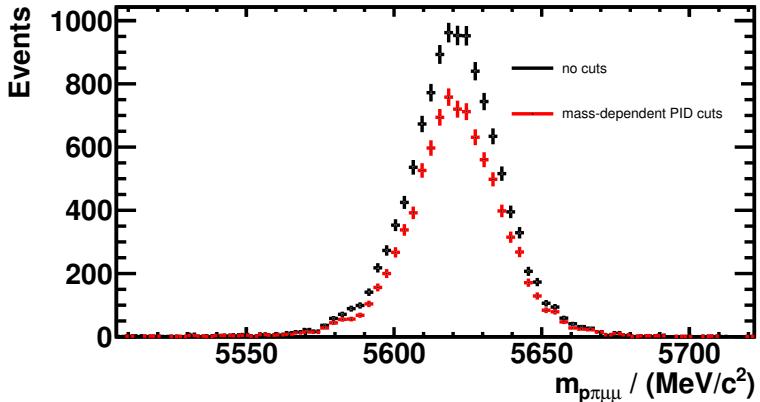
(a)



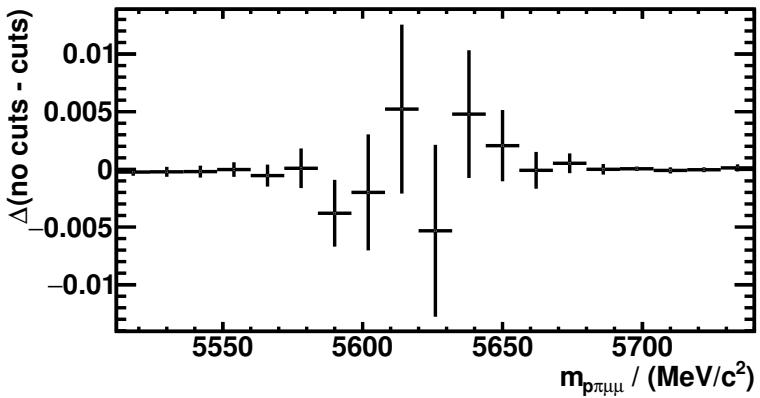
(b)

Figure 5.6: Shape of all data candidates, prior to any PID selection and within $60 \text{ MeV}/c^2$ of the Λ_b^0 mass, under different mass hypotheses (a), and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ signal simulation under these same hypotheses, (b). The shaded regions indicated the areas within which a tighter PID selection is applied. The lower shaded region in mass corresponds to events falling near the B^0 mass and the higher region corresponds to events falling near the B_s^0 mass. The widths of these shaded regions are detailed in Table 5.7.

expected number of background events. Once the cuts in Table 5.7 have been applied, the number of remaining events in each reflection channel, after all selection has been applied, is estimated. This selection includes the mass-dependant PID cuts, the preselection and the application of the BDT to reduce background, as discussed in chapter 6. The BDT



(a)



(b)

Figure 5.7: The signal mass distribution with and without the mass-dependent PID cuts, as listed in Table 5.7(a), applied, (a). The difference between the two histograms in (a) (when normalised), (b).

cut used during the optimisation process for the reflection background selection criteria is, however, not the same as the nominal BDT cut used throughout the rest of this analysis. This is due to the BDT cut being optimised again after the decisions on the reflection selection criteria had been made. The BDT cut used during the optimisation of the reflection background selection criteria is 0.2 and, as such, is not dissimilar for the nominal BDT cut value.

Daughters	Window/ MeV/ c^2	Cut	Rejection	Efficiency for $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ sim.
$K^+ \pi^-\mu^+\mu^-$	5246 – 5330	$p \text{ DLL}_{pK} > 17$	0.97	0.90
$\pi^+ \pi^-\mu^+\mu^-$	5247 – 5329 5348 – 5406	$p \text{ DLL}_{p\pi} > 5$	0.60	0.98
$K^+ K^-\mu^+\mu^-$	5348 – 5406	$\pi \text{ DLL}_{K\pi} < -10$	0.99	0.96
$K^+ K^-$	< 1025	removed		

Table 5.7: Vetoos used to reject reflections and their rejection rate, calculated using $B \rightarrow J/\psi X$ simulation, along side their efficiency on $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation, relative to the initial PID cuts already placed.

To estimate the number of remaining events in each channel, a data-driven approach is used. PID cuts on the data are reversed, so as to favour a particular reflection background channel. All other cuts are left the same. For example, to deduce the yield of $B^0 \rightarrow K^+\pi^-\mu^+\mu^-$ the PID cut on the proton would be reversed to make it more kaon like. These cuts are referred to as reversed-PID cuts.

The reversed-PID cuts allow the background peaks to be fitted with the correct mass hypothesis. The initial signal yield is then taken from these fits and resampled simulation is used to calculate the efficiency that results from unfolding the effect of the reversed-PID cuts and applying the standard PID selection.

In chapter 6, there is a discussion of the application of a BDT to data, the purpose of which is to reduce both combinatorial and reflection backgrounds. To estimate the number of remaining reflection backgrounds after the BDT has been applied, the BDT efficiency is deduced by applying the BDT to the relevant reflection background with the reverse-PID cut still applied. By comparing the yield from the fit to the relevant reflection background, before and after the BDT cut has been applied, the value of the BDT efficiency on the given reflection background can be estimated. Combining all this information allows an estimation of the remaining number of events for each reflection background,

after all selection has been applied. This estimation process is carried out for the J/ψ reflection channels, where a large sample is available in the data. The ratio of branching fractions between the J/ψ mode and the equivalent $\mu^+\mu^-$ mode is used to estimate the remaining number of events in the equivalent $\mu^+\mu^-$ channel. The resulting yields for the J/ψ reflection channels are insignificant for all background channels considered with the exception of the $B^0 \rightarrow K^+\pi^-J/\psi$ channel, the yield for which is 69 ± 9 events.

Assuming that the ratio of PID and selection efficiencies for the $B^0 \rightarrow K^+\pi^-J/\psi$ and $B^0 \rightarrow K^+\pi^-\mu^+\mu^-$ channels are the same as those between the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channels and taking the ratio between the relevant branching fractions (as shown in [Table 5.1](#)) gives an expected yield for the $B^0 \rightarrow K^+\pi^-\mu^+\mu^-$ background in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel of ~ 0.2 events, which is considered negligibly small.

In conclusion, all reflection channels are negligible after all selection has been applied, with the exception of $B^0 \rightarrow K^+\pi^-J/\psi(\rightarrow \mu^+\mu^-)$. The $B^0 \rightarrow K^+\pi^-J/\psi(\rightarrow \mu^+\mu^-)$ reflection background is initially added as a fit component to the final fit, with its yield constrained to the expected yield of 69 ± 9 , as discussed in [section 7.1](#).

5.5.2 Background from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ decays

The Cabibbo-favoured modes $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ will form backgrounds for the signal and normalisation channels respectively. As for the B reflection channels, the background from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ is reduced by the application of mass-dependent PID cuts. The mass region in which a tighter PID cut of $DLL_{K\pi} < -15$ is applied (in addition to the nominal PID cut of $DLL_{K\pi} < -5$ applied to all data) is shown in [Figure 5.8](#), along with the signal simulation under the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ hypothesis.

In [Figure 5.9](#) the effect of the mass-dependent cuts applied to reduce the Cabibbo-favoured mode $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ on the signal channel is shown via the comparison of the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass distribution in simulation with and without the relevant mass-dependent PID cut applied.

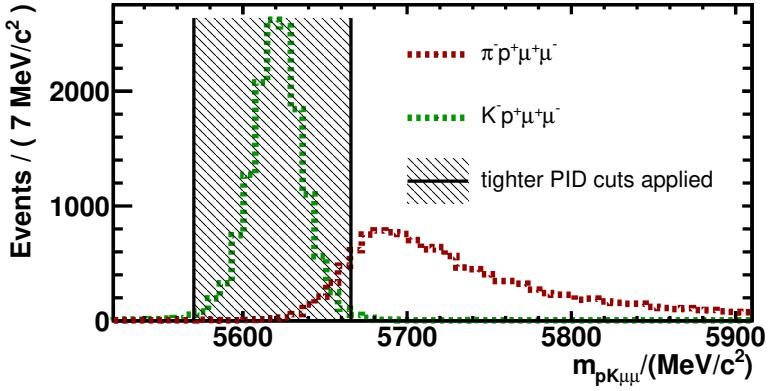


Figure 5.8: $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ simulation with the region in which tighter PID cuts are imposed indicated by the shading and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation under the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ mass hypothesis.

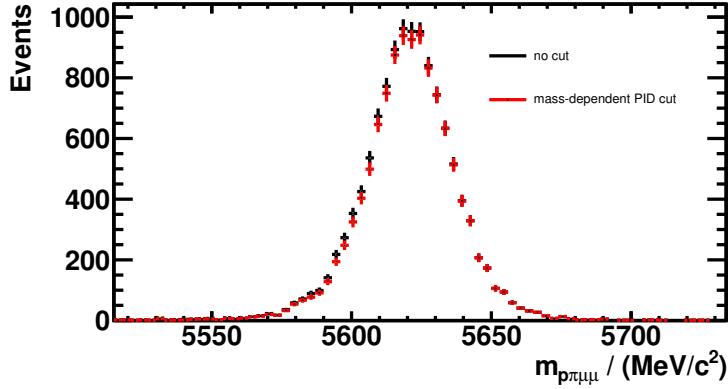


Figure 5.9: The mass distribution of $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation with and without the mass-dependent PID cut to reduce $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ contributions applied.

The number of remaining $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ events, after the mass-dependent vetoes and all other selections have been applied, is again estimated using a data-driven method. The number of $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ events present with a reverse-PID cut on the meson of $DLL_{K\pi} > 5$, which favours the kaon hypothesis, is of order 300 events. The calibrated simulation is then used to deduce the number of $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ events remaining once this cut is removed and the mass-dependent PID cut has been re-applied. This gives

1 ± 1 residual $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ events and retains $\sim 75\%$ of signal events. Given the low expected $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ yield, the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ component is not added to the final signal fit, although the effect of not adding it is evaluated as systematic uncertainty in [chapter 8](#).

The estimation of the number of remaining $\Lambda_b^0 \rightarrow J/\psi pK^-$ events is also approached using a data-driven method. To estimate the number of $\Lambda_b^0 \rightarrow J/\psi pK^-$ events, the known $\Lambda_b^0 \rightarrow J/\psi pK^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ branching fractions are used, as shown in [Table 5.1](#), along with the efficiencies of the $DLL_{K\pi}$ selections on the $\Lambda_b^0 \rightarrow J/\psi pK^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels (taken from calibrated simulation) and the number of observed $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ events in [Figure 6.9](#). This gives the expected number of $\Lambda_b^0 \rightarrow J/\psi pK^-$ events as being 84 ± 10 . A $\Lambda_b^0 \rightarrow J/\psi pK^-$ component is therefore added to the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ fit, as discussed in [section 7.1](#).

5.5.3 Double mis-identification of muons as pions

A further source of background, with a double mis-id of the pions as muons, is from events of the form $\Lambda_b^0 \rightarrow p\pi^-\pi^+\pi^-$. The Feynman diagram for this process is shown in [Figure 5.10](#). Due to the small mass difference between muons and pions, this background would peak around the Λ_b^0 mass under the signal mass hypothesis.

In order to deduce the number of $\Lambda_b^0 \rightarrow p\pi^-\pi^+\pi^-$ present after selection, a data-driven method is used to estimate the efficiency of the muon selection when placed on a true pion. This selection is namely `isMuon == True` and $DLL_{\mu\pi} > -3$. The combined efficiency to mis-identify both pions as muons is found to be $\sim 0.01\%$. There is no branching fraction measurement for $\Lambda_b^0 \rightarrow p\pi^-\pi^+\pi^-$. Instead, the $\Lambda_b^0 \rightarrow p\pi$ branching fraction is used as a proxy. This branching fraction is a suitable proxy for $\Lambda_b^0 \rightarrow p\pi^-\pi^+\pi^-$ as $\Lambda_b^0 \rightarrow p\pi$ has a related Feynman diagram. The $\Lambda_b^0 \rightarrow p\pi$ branching fraction is of order 10^{-6} . Applying the muon selection efficiency gives an effective branching fraction of order 10^{-10} .

Given that the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ branching fraction is of order 10^{-6} , the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction, assuming that it is 100 times smaller than the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ branching

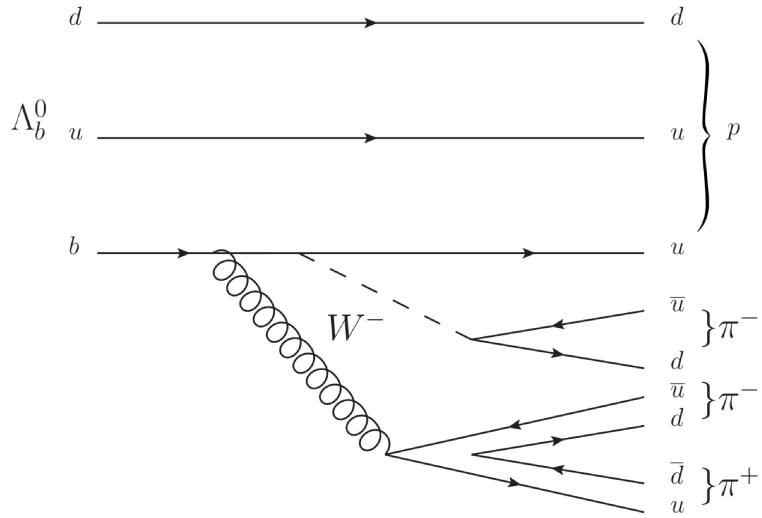


Figure 5.10: The Feynman diagram for the decay $\Lambda_b^0 \rightarrow p\pi^-\pi^+\pi^-$.

fraction, will be $\sim 10^{-8}$. This is 100 times larger than the expected contribution from $\Lambda_b^0 \rightarrow p\pi^-\pi^+\pi^-$ after the muon selection criteria have been applied. Consequently, no fit component is required for this $\Lambda_b^0 \rightarrow p\pi^-\pi^+\pi^-$ contribution.

More recently, the number of $\Lambda_b^0 \rightarrow p\pi^-\pi^+\pi^-$ decays in 3 fb^{-1} of LHCb data has been measured in Ref. [74] and the observed number is compatible with the estimate of the branching fraction outlined above.

As an additional post-unblinding check, the double mis-id contamination is further investigated by checking the effect of tightening the selection on the $\text{DLL}_{\mu\pi}$ variable on the signal sample. No significant effect is observed.

5.5.4 Mis-identification from swaps between muons and hadrons

Charmonium resonances can have double mis-identification such that the muon with the same sign as the pion is misidentified as the pion, and the pion is misidentified as the muon. The same process can occur between the proton and the muon with the same sign as the proton. The possibility of these backgrounds is dealt with by tightening the selection on events where the invariant mass of the pion (proton) candidate and the muon

Selection / (GeV^2/c^4)	Criteria
$8.0 < m_{\mu^+(\pi^- \rightarrow \mu^-)}^2 < 11.0$	π inMuon: True
$12.5 < m_{\mu^+(\pi^- \rightarrow \mu^-)}^2 < 15.0$	π inMuon: True
$8.0 < m_{(p \rightarrow \mu^+) \mu^-}^2 < 11.0$	p inMuon: True
$12.5 < m_{(p \rightarrow \mu^+) \mu^-}^2 < 15.0$	p inMuon: True

Table 5.8: The veto requirements used to reduce the background from charmonium resonances. The notation $m_{X(Y \rightarrow Z)}$ implies the combined invariant mass of the X and Z particle, where the Y particle is given the mass of Z .

of opposite sign to the pion (proton), form a mass close to the nominal J/ψ or $\psi(2S)$ masses. The additional requirements placed on $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ candidates to reduce swap backgrounds are shown in [Table 5.8](#). The extra `inMuon` requirement reduces the rate of mis-identification, in order for the muon to have been mis-identified as a pion (proton) it has to fulfil the `isMuon==False` requirement, which it would do if it fell outside of the muon acceptance. All pion and proton tracks are already required to have `isMuon == False`. The residual background from all swaps after these selections have been applied is found to be negligible at ~ 0.1 events.

5.6 Partially reconstructed backgrounds in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel

The background in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel is largely combinatorial and as such the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass distribution can be well-modelled by a single exponential. However, this is not the case for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel, which has an additional contribution from `part-reco` backgrounds, as shown in [Figure 5.11](#), alongside a fit to the mass distribution from the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ channel. All fits in this analysis are unbinned extended maximum-likelihood fits. The mass shapes are discussed in detail in [chapter 7](#).

The background in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel must originate from muons coming from

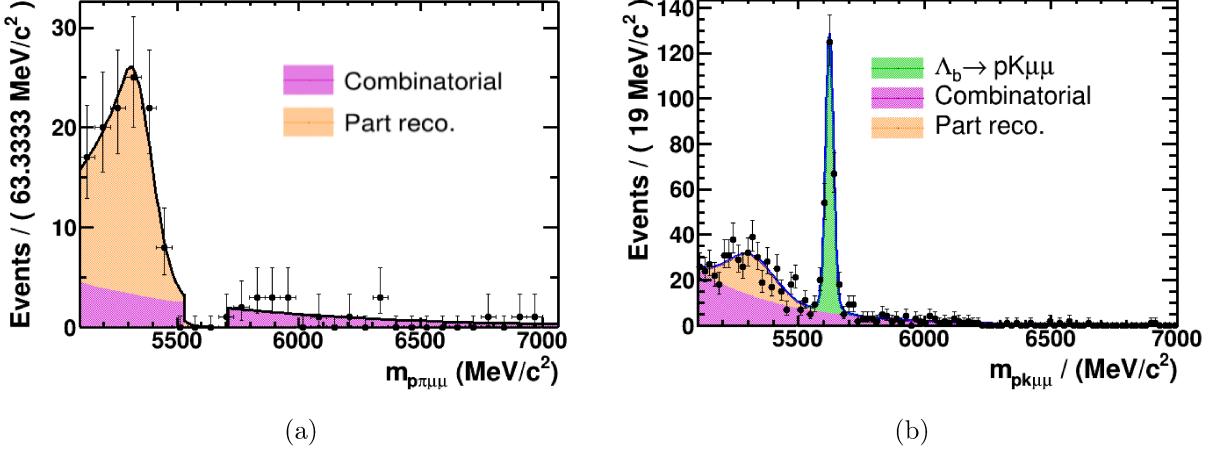


Figure 5.11: Fitted mass distributions for blinded $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data (a) and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data (b).

different vertices, otherwise it would also be present in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ case. In addition, the distribution of the momentum of the positive muon with respect to the direction of the dimuon system is highly asymmetric in the lower mass side band of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data, suggesting that the part-reco background component is dominated by semi-leptonic processes.

In this section the composition of the backgrounds in both the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data and the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data are considered. The $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ background is studied because the backgrounds for both the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ datasets will be similar but there are more statistics in the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ channel. The $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ candidates are selected in the same way as the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ candidates but with the $DLL_{K\pi}$ requirement on the kaon candidate changed from $DLL_{K\pi} < -5$ to $DLL_{K\pi} > 5$ and the mass-dependent PID veto used to remove $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ candidates not applied.

To check for potential background arising from mis-identified particles, alternative mass hypotheses are tried in combinations of pairs and triplets for the muon, proton and pion (kaon) candidates in the region $m_{p\pi(K)\mu\mu} < 5500$ MeV/c² in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ ($\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$) datasets. The only narrow mass peak that is formed by trying all of

the available combinations corresponds to cascade decays involving a Λ_c^+ . These decays are most clearly visible with the BDT selection relaxed, as shown in [Figure 5.12](#). The Λ_c^+ mass peaks observed in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ datasets come from the decays $\Lambda_b^0 \rightarrow \Lambda_c^+(\rightarrow p\pi^+X)Y$ where $X = \pi^-$ or K^- , $Y = \mu^-\nu$ or π^- and the pion with the same sign as the proton is mis-identified as a muon.

The background from decays with $Y=\mu^-\nu$ rather than $Y=\pi^-$ are expected to be dominant, as only one mis-id is required. This is verified by looking at the mass, $m_{p\pi(\mu \rightarrow \pi)}$, of the Λ_c^+ candidates (now with the nominal BDT cut in place) against the mass, $m_{p\pi\mu\mu}$, of the Λ_b^0 candidates. These distributions are shown in [Figure 5.13](#) for both $m_{p\pi(\mu \rightarrow \pi)}$ and $m_{pK(\mu \rightarrow \pi)}$. All the Λ_c^+ candidates have $m_{p\pi\pi}$ significantly less than $m_{\Lambda_b^0}$, suggestive of the large mass difference associated with the $Y=\mu^-\nu$ hypothesis rather than the $Y=\pi^-$ hypothesis. In the latter case, the small mass difference between pions and muons would give a value of $m_{p\pi\pi\pi}$ much closer to the Λ_b^0 mass. The full range in $m_{p\pi\mu\mu}$ is shown in the x axis of [Figure 5.13\(b\)](#). Note that this does not unblind the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ dataset, as the range in $m_{p\pi\pi}$ is restricted to less than $2850 \text{ MeV}/c^2$. As a post-unblinding check against $\Lambda_b^0 \rightarrow \Lambda_c^+(\rightarrow p\pi^+\pi^-)\pi^-$ decays, which would peak around the Λ_b^0 mass under the $m_{p\pi\mu\mu}$ hypothesis, the events with a value of $m_{p\pi(\mu \rightarrow \pi)}$ which fall around the nominal Λ_c^+ mass are removed and the effect on the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ signal peak evaluated. There is no significant effect observed.

The assumption that the part-reco is semi-leptonic in nature is also supported by the fact that the $\cos(\theta_{ll})$ distribution, where θ_{ll} is defined as the angle between the momentum of the lepton with the same sign as the proton and the sum of the lepton momenta, both in the frame of the Λ_b^0 mother, is asymmetric. This is a signature for semileptonic cascade decays where one lepton is much harder than the other. This is demonstrated for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ candidates in [Figure 5.14](#), which shows the $\cos(\theta_{ll})$ distribution, along with a comparison of the mass distribution with and without a cut on $\cos(\theta_{ll})$. Again, the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data is used to demonstrate these features given the larger sample size.

In order to model the partially reconstructed backgrounds, the similarity between the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ channels is exploited and the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$

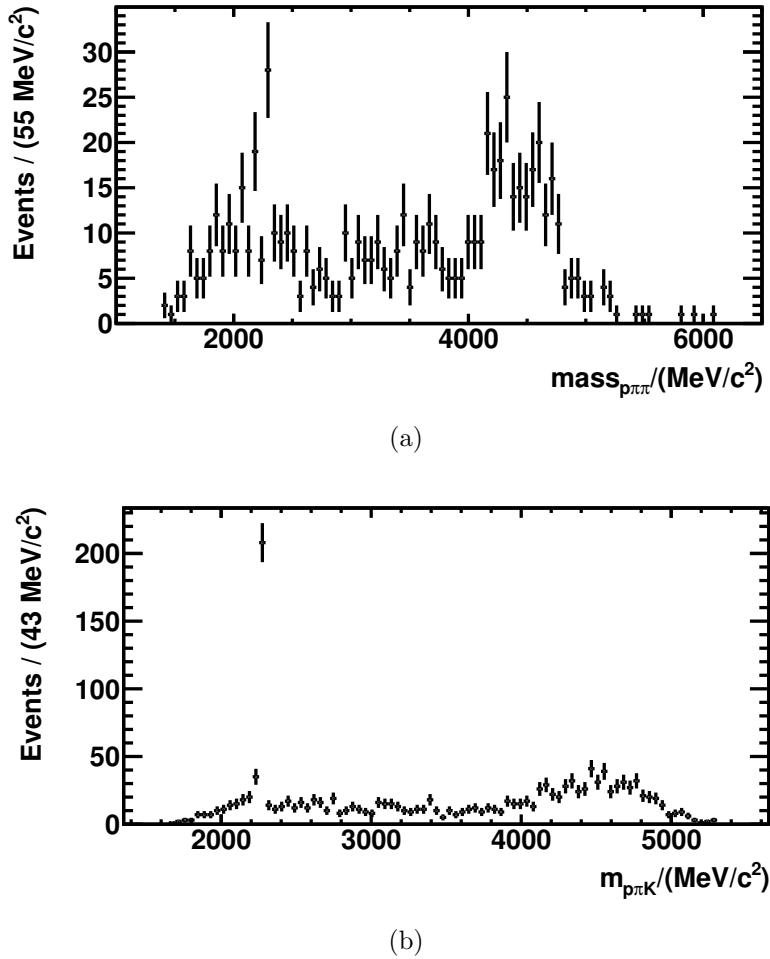


Figure 5.12: Combined mass of ($p\pi^-(\mu^+ \rightarrow \pi^+)$) for $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ lower mass side band data where the muon with the same sign as the proton has been given the mass of a pion (a), combined mass of $pK^-(\mu^+ \rightarrow \pi^+)$ for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ lower mass side band data where the muon with the same sign as the proton has been given that of a pion (b). The peaks at ~ 2300 MeV/ c^2 are due to the Λ_c^+ baryon appearing via $\Lambda_b^0 \rightarrow \Lambda_c^+(pK^-(\pi^-)\pi^+)X$ decays.

candidates are used to define the part-reco background shape for the fit to $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data. A comparison between the lower mass side band distributions for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ datasets is shown in Figure 5.15.

In summary, there is a background component in the lower mass side band of both

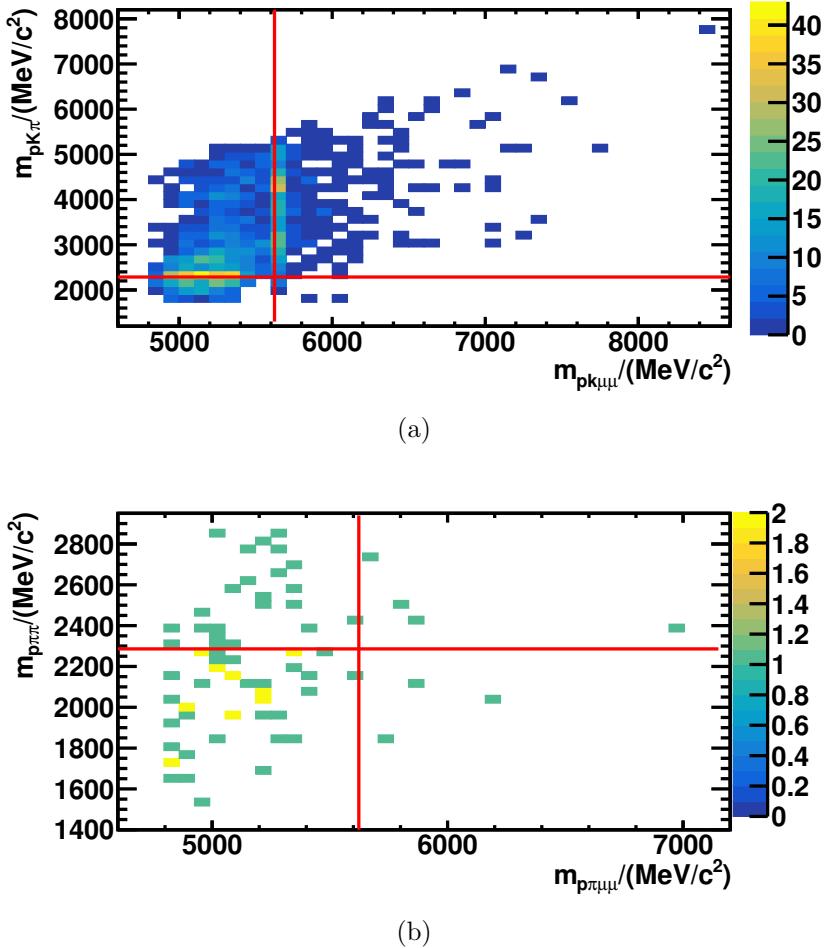
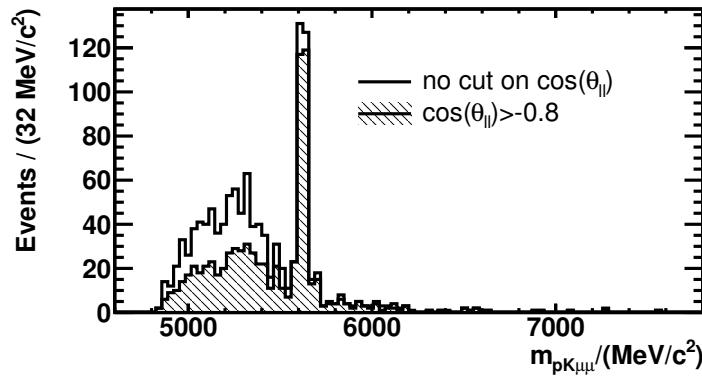
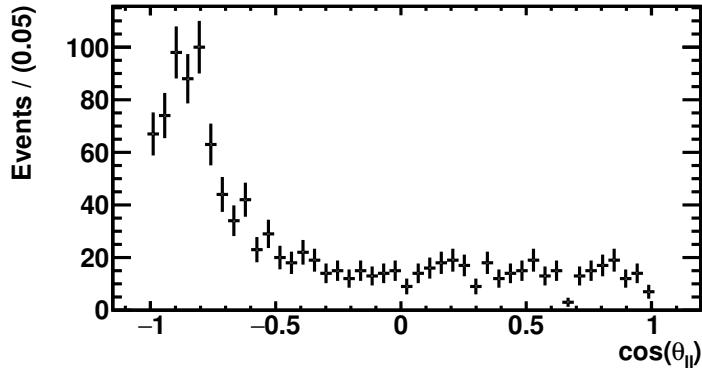


Figure 5.13: The mass of the combination ($pK(\mu^+ \rightarrow \pi^+)$) against the mass of the combination of ($pK\mu^-\mu^+$) for the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data set after the complete selection (a) and the mass of the combination ($p\pi^-(\mu^+ \rightarrow \pi^+)$) against the mass of the combination of ($p\pi^-\mu^-\mu^+$) for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data set after the complete selection (b). The red lines in the x and y direction indicate the nominal Λ_c^+ and Λ_b^0 masses respectively.

$\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data, that is not present in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data. The exact composition of this background is not deduced, although there are signs both that it is semi-leptonic and that it contains Λ_c^+ cascade decays. Instead, the background is fitted in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel by exploiting the similarity between the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ mass distributions. The effect of this assumption on the observed



(a)



(b)

Figure 5.14: $\Lambda_b^0 \rightarrow p K^- \mu^+ \mu^-$ data with and without a cut on $\cos(\theta_{ll})$, where θ_{ll} is defined as the angle between the momentum of the lepton with the same sign as the proton and the sum of the lepton momenta, both in the frame of the Λ_b^0 mother, (a), the distribution of $\cos(\theta_{ll})$ for $\Lambda_b^0 \rightarrow p K^- \mu^+ \mu^-$ data (b).

signal yield taken from the fit is considered as a systematic uncertainty, as discussed in chapter 8.

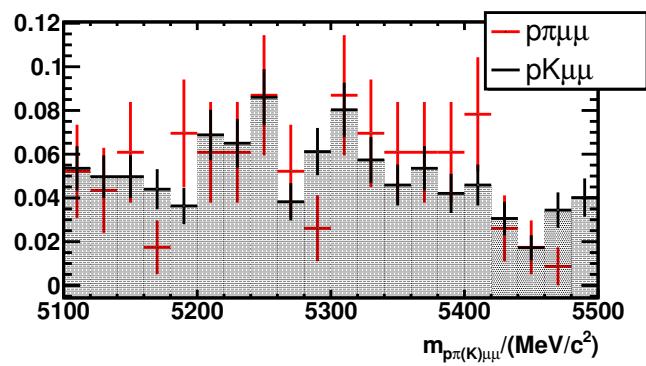


Figure 5.15: Comparison of the separately normalised Λ_b^0 mass distribution in the lower-mass side band for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data.

Chapter 6

Multivariate methods used in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis

This chapter will outline the multivariate analysis (MVA) methods used in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis to reduce background. The MVA methods used in this analysis are in the form of BDTs, which are trained and implemented using a dedicated MVA software package [75]. A brief overview of MVA methods using BDTs can be found in Appendix A. There are two BDTs used in this analysis. The first BDT, referred to as the isolation BDT, is designed to determine how isolated the final state tracks of the signal channel are, and is discussed in detail in section 6.1. The isolation BDT output is then taken as an input for the BDT used to reduce combinatorial background, discussed in section 6.2. The optimisation process for the cut placed on the combinatorial BDT output is discussed in section 6.3.

6.1 Isolation of the final state tracks

The isolation BDT used in this analysis was originally trained for the analysis in Ref [76]. The algorithm used iterates over every track in an event which is not a candidate signal track, as illustrated in Figure 6.1, where the non-signal candidate track is denoted i . The BDT algorithm then computes how signal-like each track is, depending on the variables

given in [Table 6.1](#). In [Table 6.1](#), the ghost probability (ghost prob.) is an estimate of the probability that a track is a so-called [ghost](#) track. Such a track has less than 70% of its hits originating from a single particle, where the number of hits is evaluated only from simulation.

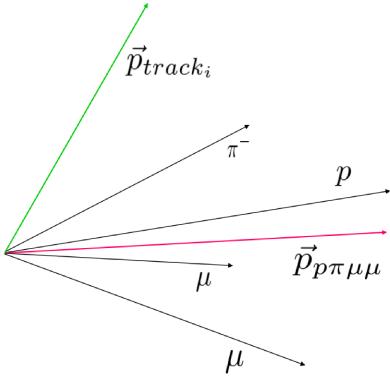


Figure 6.1: A schematic of a non-signal track, i , and a $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ candidate.

Variables used in the isolation BDT
$p_{T\text{track}}$
$\vec{p}_{\text{track}} \cdot \vec{p}_{p\pi\mu\mu}$
ghost prob. _{track}
χ^2_{track}
$IP\chi^2_{\text{track}} \text{ (} p\pi\mu\mu \text{ vertex)}$

Table 6.1: The variables used in the isolation BDT.

In Ref [76], the isolation BDT was trained using $\Lambda_b^0 \rightarrow p\mu^-\nu$ simulation as the signal proxy and $\Lambda_b^0 \rightarrow (\Lambda_c^+ \rightarrow pX\mu^-\nu)\mu^-\nu$ as the background proxy. The resulting BDT weights, which were not obtained by myself but simply taken the analysis in [76], were then applied to this analysis. Although the training samples differ in many aspects from $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$, the variables used in the training, as shown in [Table 6.1](#), are generic enough such that the

resulting BDT can be applied to a range of channels. The BDT weights for the isolation BDT are computed prior to any selection.

6.1.1 Isolation BDT response

The isolation BDT output of the non-candidate track which is most signal-like is used as the nominal isolation BDT output value. The output of the isolation BDT applied to sWeighted $\Lambda_b^0 \rightarrow J/\psi pK^-$ data (signal) and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ candidates with a mass above 6000 MeV/ c^2 (background), is shown in Figure 6.2. The peak in the signal output at -2

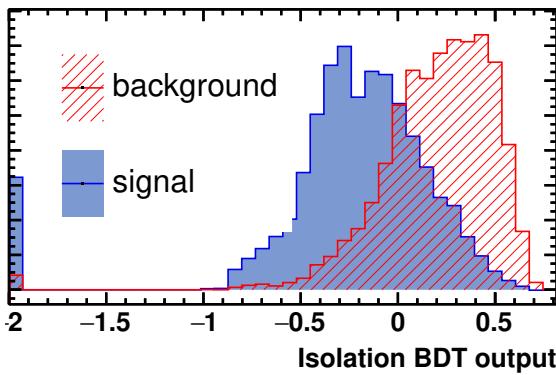


Figure 6.2: The isolation BDT response for $\Lambda_b^0 \rightarrow J/\psi pK^-$ data, representing the signal proxy, and the response for the upper mass side band of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data, representing the background.

corresponds to cases where no non-signal track in the event passes the preselection applied, so the signal candidates are completely isolated.

6.2 The combinatorial Boosted Decision Tree

In order to reduce combinatorial and reflection background, a further BDT is trained using $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data above 6000 MeV/ c^2 as a proxy for the combinatorial background and sWeighted $\Lambda_b^0 \rightarrow J/\psi pK^-$ data as a proxy for the signal events. To choose which variables were used as inputs to the BDT, a large range of variables were initially selected and those

that were deemed least discriminating, based on how often the variables were used to split a tree and the size of the separation resulting from such a split, were disregarded. The final selection of variables, ordered by separation power, is shown in [Table 6.2](#). The N^* in [Table 6.2](#) refers to the combination of the $p\pi$ as if they came from a N^* decay, $\Lambda_b^0 \rightarrow N^*(\rightarrow p\pi^-)\mu^+\mu^-$. Variables that assume that the $p\pi$ come from an N^* are used because much of the signal will genuinely decay via this resonance, whereas this is not the case in combinatorial background. The μ^+ $DLL_{\mu\pi}$ refers to whichever muon has the same sign as the proton, so in the conjugate case this would be μ^- . The difference in separation power amongst the weakest eight BDT variables is minimal. The μ^+ $DLL_{\mu\pi}$ is slightly more discriminating than the equivalent μ^- $DLL_{\mu\pi}$ variable, and therefore only the former was included as an input variable. This is thought to be due to the mis-identification of pions as muons in the combinatorial background coming from Λ_c^+ decays. The DecayTreeFitter ([DTF](#)) χ^2 refers to the χ^2 per degree of freedom of the simultaneous fit to the complete decay chain. The PID variables are used in the BDT in order to exploit the varying PID performance with kinematics and thus help further reduce contributions from reflection backgrounds.

The distribution of the input variables for the signal and background training samples are shown in Figures [6.3](#), [6.4](#) and [6.5](#).

Rank	Variable	Seperation / 0.1	Rank	Variable	Seperation / 0.1
1	A_b^0 vertex $\text{vertex } \chi^2/\text{ndof}$	9.0	10	$\pi^- \text{IP} \chi^2$	0.7
2	A_b DIRA	3.4	11	$\mu^+ \text{ DLL}_{\mu\pi}$	0.6
3	Isolation BDT output	2.9	12	$p \text{ IP} \chi^2$	0.6
4	$A_b \tau$	2.8	13	$p \text{ IP}$	0.5
5	$A_b p_T$	2.4	14	$p \text{ DLL}_{p\pi}$	0.4
6	A_b DecayTreeFitter χ^2	1.8	15	$A_b \text{ IP} \chi^2$	0.3
7	$N^* \text{ IP} \chi^2$	1.8	16	$p \text{ DLL}_{pK}$	0.1
8	$N^* p_T$	1.6	17	$p \vec{p} $	0.1
9	$A_b \vec{p} $	0.9			

Table 6.2: List of BDT input variables. The variables are listed in order of their separation power. The seperation values are also listed.

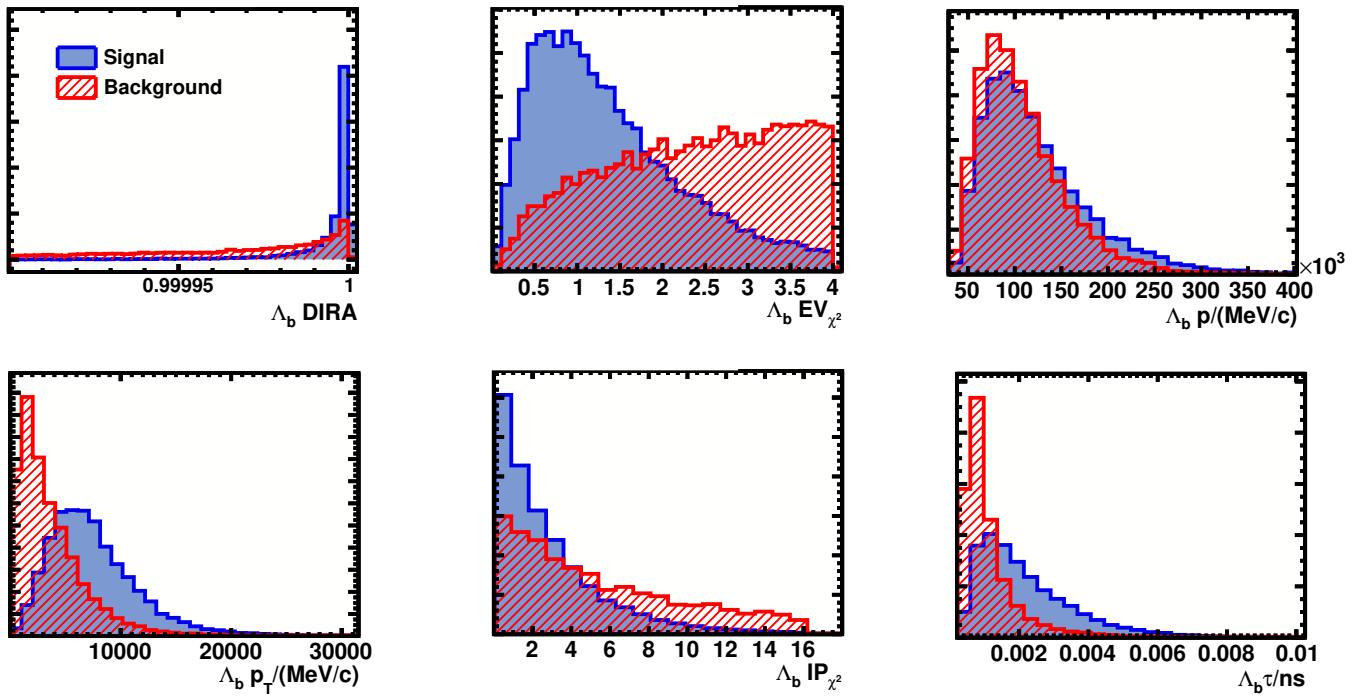


Figure 6.3: The first six variables used in the combinatorial BDT for signal and background proxies. The red stripped histograms are background and the solid blue histograms are signal. All distributions shown are normalised. The variable Λ_b^0 EV χ^2 refers to Λ_b^0 vertex χ^2/ndof in Table 6.2.

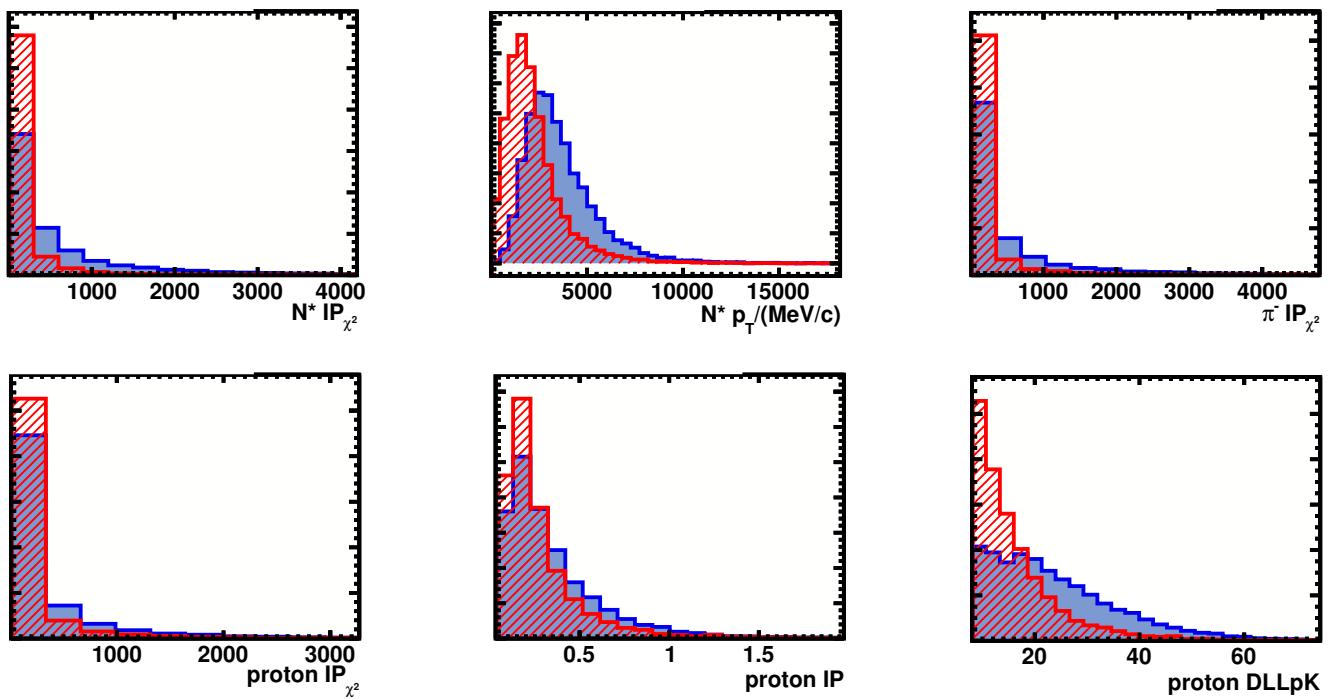


Figure 6.4: The second six variables used in the combinatorial BDT for signal and background proxies. The red stripped histograms are background and the solid blue histograms are signal. All distributions shown are normalised.

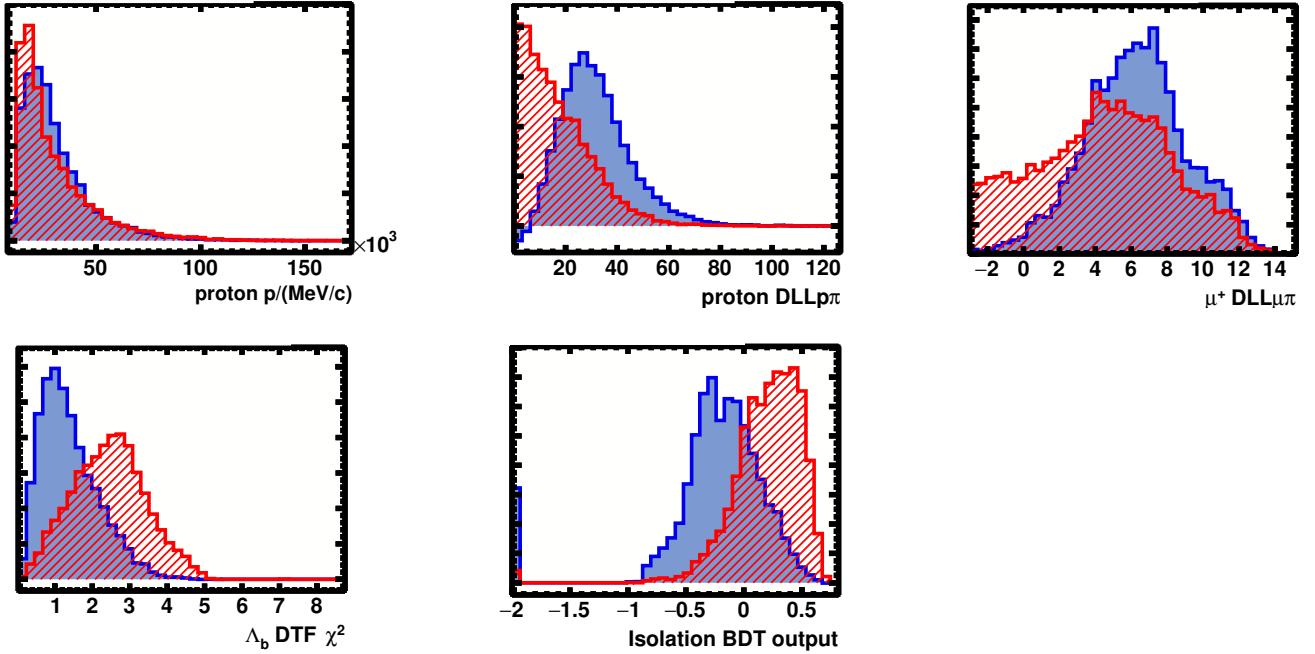


Figure 6.5: The final five variables used in the combinatorial BDT for signal and background proxies. The red stripped histograms are background and the solid blue histograms are signal. All distributions shown are normalised. The variable Λ_b^0 DTF χ^2 refers to Λ_b DecayTreeFitter χ^2 in [Table 6.2](#).

The BDT is trained using the K-Folding technique [77]: each event is randomly assigned an integer, i , between 0 and the number of K-Folds, n , (in this case $n = 5$) and n BDTs are trained with the i^{th} BDT being tested on the i^{th} data set and trained on the remaining data. Employing this technique allows $(n-1)/n$ of the data set to be used to train the BDT.

6.3 The BDT performance and optimisation

The BDT response for each K-Fold, as well as the signal efficiency against background rejection, integrated over all K-Fold's, is shown in [Figure 6.6](#). The red line in [Figure 6.6\(f\)](#) indicates the BDTs optimal working point. The choice of working point is discussed later

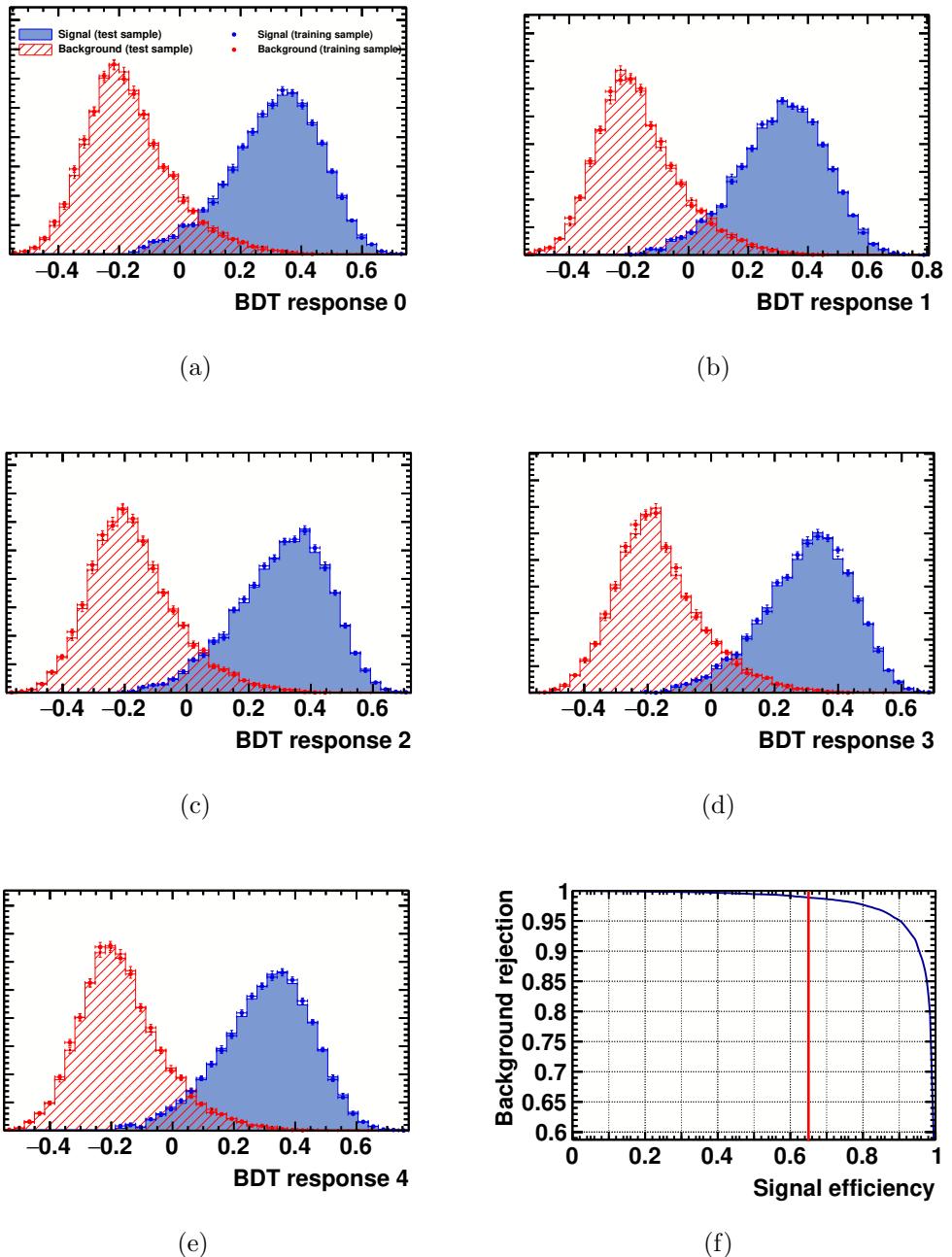


Figure 6.6: The BDT response for all 5 K-Folds and the signal efficiency against background rejection curve, integrated over all K-Folds. The red line on the signal efficiency against background rejection curve indicates the optimal working point.

in this section.

As there is no branching fraction prediction, the BDT is optimised using the Punzi FOM [78] defined as

$$\text{FOM}_{\text{PUNZI}} = \frac{\epsilon_{\text{selection}}}{\sigma/2 + \sqrt{B}}, \quad (6.1)$$

where $\epsilon_{\text{selection}}$ refers to the efficiency of the selection and B refers to the number of background events. The Punzi FOM is also a function of the target significance, σ . The target significance is 3σ . There is no change in the chosen working point if this is varied by $\pm 0.5\sigma$.

In order to deduce B , a fit is performed to the blinded mass spectrum of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data. The blinded region in mass is taken as $5530 < m_{p\pi\mu\mu} < 5710$ MeV/ c^2 , which removes all signal events according to simulation. The data is fitted either side of the blinded region and the number of background events is taken by extrapolating the background probability density function (PDF) across the blinded region. The details of the fit model used are discussed in [chapter 7](#).

As the number of background events across the signal window is very low (~ 5 or less for BDT cuts greater than ~ 0.3), the FOM is calculated many times, each time randomly choosing a yield of background events within one standard deviation of the true value. This is done with the aim of reducing the sensitivity to the large uncertainty on the shape of the combinatorial background.

To obtain the value of $\epsilon_{\text{selection}}$ in [Equation 6.1](#), the BDT efficiency is calculated using $\Lambda_b^0 \rightarrow J/\psi pK^-$ data, and is also compared to the efficiency derived from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data. This BDT efficiency is then multiplied by the rest of the selection efficiency (see [chapter 7](#)). The BDT efficiency calculated with both $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ data as a function of BDT output value is shown in [Figure 6.7](#). To deduce the BDT efficiency, the yield extracted from the fit to $\Lambda_b^0 \rightarrow J/\psi pK^-$ data is compared to that extracted from a subsequent fit to $\Lambda_b^0 \rightarrow J/\psi pK^-$ data with the relevant BDT cut applied. In the case of $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data, the same method is used but a BDT cut of 0 is initially placed on the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data, as without this cut the background to signal ratio is too large and the fit becomes difficult to perform.

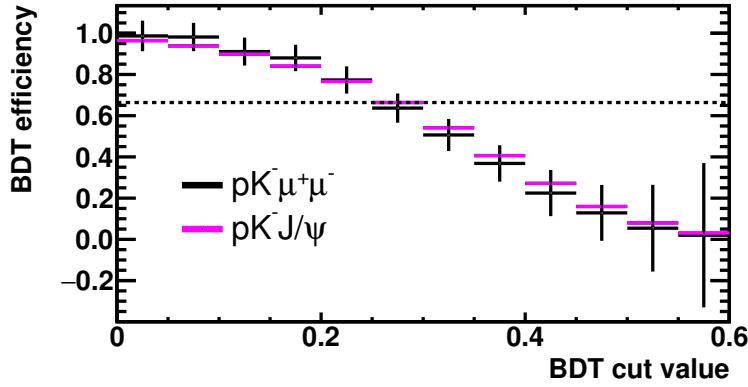


Figure 6.7: The efficiency of the BDT as a function of BDT cut value using $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ sWeighted data. The dashed line shows the efficiency at the working point of 0.25.

The values of the Punzi FOM as a function of the BDT cut value are shown in Figure 6.8. Each line represents a different assumption for the background yield. The

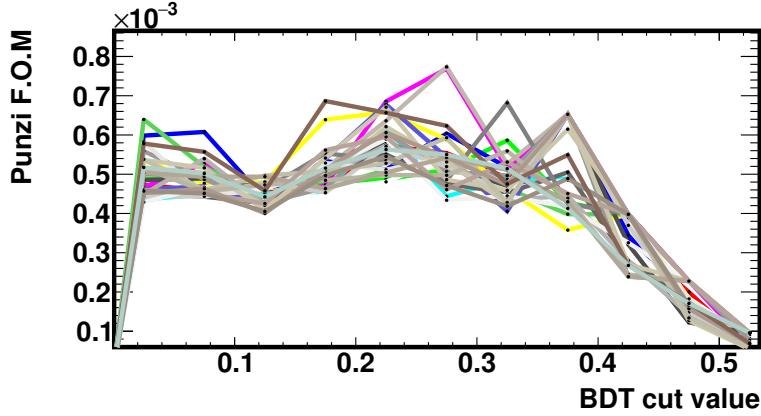


Figure 6.8: The Punzi FOM for background values that are varied randomly within their error.

optimal working point is chosen to be at a BDT cut of 0.25. This point is chosen because it has the consistently highest FOM. The blinded $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data at this optimal cut point, along side $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ at the same cut point, are shown in Figure 6.9. The points above the fit line around $5700 \text{ MeV}/c^2$ in Figure 6.9(b) could be due to a statistical

fluctuation, most likely caused by a downward fluctuation just below the mass range around $5700 \text{ MeV}/c$. Alternatively the discrepancy between the fit and data could be caused by a mis-modelling of the CB tails in simulation.

There are 1017 ± 41 $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ signal events observed. Again, the fit models are discussed in [chapter 7](#).

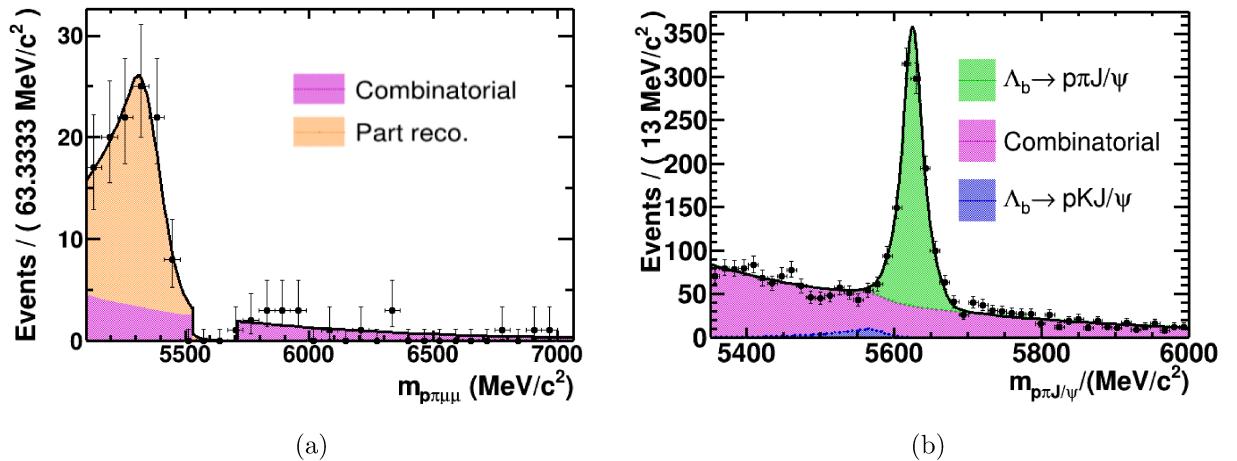


Figure 6.9: Fitted data for blinded $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ events (a) (same as [Figure 5.11](#)), and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ events (b).

Chapter 7

Mass fits and efficiency calculations for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis

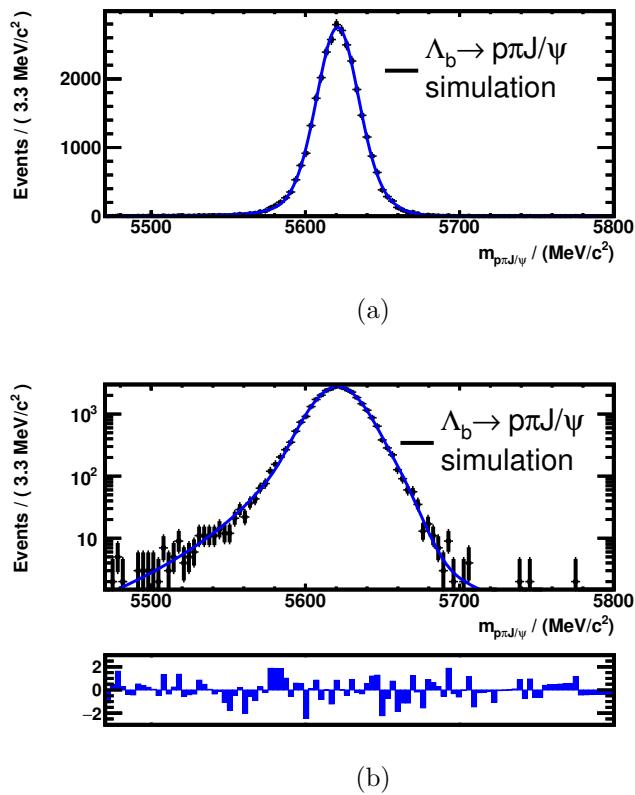
In this chapter there will be a discussion of the fit templates used to model the mass distributions for the signal channel $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$, and the normalisation channel, $\Lambda_b^0 \rightarrow J/\psi p\pi^-$, in [section 7.1](#). This is followed by a summary of the relative efficiency of the selections applied to the signal channel and the normalisation channel in [section 7.2](#).

7.1 Mass fits

In this section, the signal fit models used for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels are discussed in [subsection 7.1.1](#). In [subsection 7.1.2](#) and [subsection 7.1.3](#) there is an examination of the background models used in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channels respectively, as well as a study of the effect that the choice of fit model used has on the expected signal significance. Finally, in [subsection 7.1.4](#) the expected signal significance is detailed.

7.1.1 Signal fits

The $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ signal shape is modelled by a double Crystal Ball (CB) function [68], as previously defined in [Equation 4.3](#). The mean and Gaussian widths of the CB functions (denoted as \bar{x} and σ in [Equation 4.3](#) respectively) are allowed to float and the tail parameters (α, n in [Equation 4.3](#)) are fixed from a fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation, as shown in [Figure 7.1](#).



[Figure 7.1](#): The fit of a double CB to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation with a linear scale, [\(a\)](#), and a log scale, along with the number of standard deviations the data points lie from the fit function, [\(b\)](#).

In the case of the signal fit to $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ data, the parameters are all fixed from the fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data. However, the width of the CB functions in the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ fit are adjusted, prior to the fit, based on the variation of the widths with q^2 seen in the simulation.

As the CB widths will depend on q^2 , the width is taken from $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation, where, in the absence of any form factor predictions, the q^2 distribution of the simulation has been reweighted using the $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ differential branching fraction predictions. To check the validity of using $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ differential branching fraction predictions as a proxy for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ q^2 distribution, the same reweighting in q^2 applied to $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation is applied to $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ simulation. The reweighted $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ simulation is then compared to $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data, where any background contamination has been removed using the sPlot technique. This comparison is shown in [Figure 7.2](#). The differences between the q^2 distributions in $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data and the $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ differential branching fraction predictions are covered by a systematic uncertainty as discussed in [section 8.2](#).

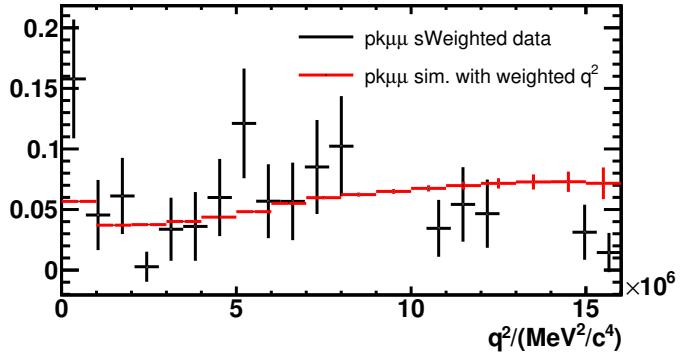


Figure 7.2: The q^2 distribution for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data against $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ simulation, where the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ simulation has been reweighted in q^2 using the $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ differential branching fraction predictions.

Due to the lack of statistics at high q^2 in $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ phase space simulation, events with a q^2 value higher than $15 \text{ GeV}^2/\text{c}^4$ are sensitive to single event fluctuations, as shown in [Figure 7.3\(a\)](#). Given that these events will receive the largest weight, the lack of statistics in these high q^2 bins has a sizeable effect on the resulting mass distribution. As such, only events with a q^2 value less than $15 \text{ GeV}^2/\text{c}^4$ are used in the fit to the q^2 -reweighted $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass distribution. The corresponding fit is shown in [Figure 7.3\(b\)](#). When the double CB function is fitted to $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data, using the widths deduced with

simulation and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data, the full q^2 range is fitted.

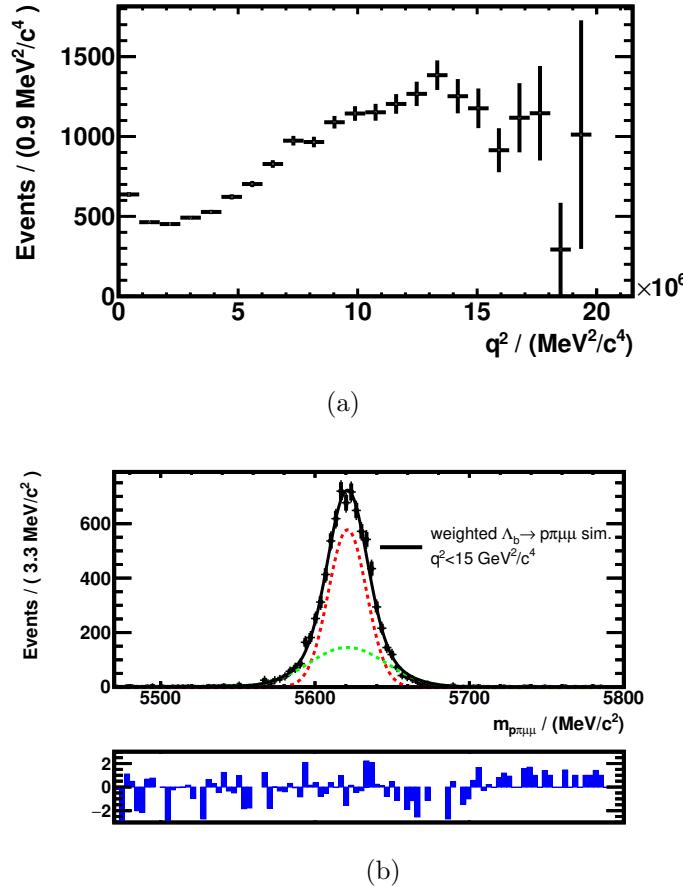


Figure 7.3: The q^2 distribution of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation after reweighting in q^2 , (a). The fit of a double CB function to $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation with q^2 -reweighting applied, (b). The red curve in (b) indicates the right CB function and the green curve the left CB function. The number of standard deviations the data points lie from the fit function are shown below the fit in (b).

The widths for both left and right CB functions for the reweighted $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulated events, relative to those of the equivalent CB functions from the fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation, are shown in [Table 7.1](#). The absolute widths for the CB functions taken from $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation are also shown.

Function	$\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ relative width	$\Lambda_b^0 \rightarrow J/\psi p\pi^-$ width/ MeV/ c^2
CB right	0.97	12.7
CB left	1.12	22.2

Table 7.1: The widths of the CB functions from the fit to $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation, expressed relative to the width from the fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation. The terms left or right refer to the direction of the tail of the CB function.

7.1.2 Complete fit to the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel

The background in the fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data is modelled using an exponential function for the combinatorial background and an additional component for the Cabibbo-favoured mode, $\Lambda_b^0 \rightarrow J/\psi pK^-$, which can be misidentified as $\Lambda_b^0 \rightarrow J/\psi p\pi^-$.

The fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data was shown previously in [Figure 6.9](#) in [chapter 6](#). The shape for the $\Lambda_b^0 \rightarrow J/\psi pK^-$ component is taken from $\Lambda_b^0 \rightarrow J/\psi pK^-$ simulation, plotted under the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass hypothesis. The $\Lambda_b^0 \rightarrow J/\psi pK^-$ simulation is fitted with a Gaussian which becomes an exponential at a certain point in mass, denoted α . The complete expression for the fit function is given in Equation 7.1:

$$gXexp(x) = \begin{cases} \mathcal{C}e^{-\beta x}, & \text{if } x \leq \alpha \\ e^{-\frac{1}{2}(\frac{x-\bar{x}}{k})^2}, & \text{otherwise} \end{cases}$$

where

$$\mathcal{C} = e^{-\frac{1}{2}(\frac{\alpha-\bar{x}}{k})^2} e^{-\beta\alpha}; \beta = \frac{\bar{x}-\alpha}{k^2}, \quad (7.1)$$

and is referred to as a RooExpAndGauss hereafter. The $\Lambda_b^0 \rightarrow J/\psi pK^-$ yield is Gaussian constrained to the expected yield of 84 ± 10 , derived in [section 5.5](#) in [chapter 5](#). A summary of all the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ fit parameters is shown in [Table 7.2](#).

Reflection components from $B^0 \rightarrow J/\psi K^{*0}$

As discussed in [chapter 5](#), the only B reflection component to remain after the selection is $B^0 \rightarrow J/\psi K^+\pi^-$. The yield in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel is expected to be 69 ± 9 events.

Fit Parameters	Constrained, free or fixed
Number of signal events	Free
Number of background events	Free
Exponential para. for combinatorial background	Free
α (Equation 7.1)	Fixed from $\Lambda_b^0 \rightarrow J/\psi pK^-$ simulation
\bar{x} (Equation 7.1)	Fixed from $\Lambda_b^0 \rightarrow J/\psi pK^-$ simulation
k (Equation 7.1)	Fixed from $\Lambda_b^0 \rightarrow J/\psi pK^-$ simulation
Number of $\Lambda_b^0 \rightarrow J/\psi pK^-$ events	Gaussian-constrained to expected yield
Signal peak: mean	Free
Signal peak: CB tails ($\alpha_{CB}, \alpha'_{CB}, n_{CB}, n'_{CB}$)	Fixed from $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation
Signal peak: fraction between CB's	Free
Signal peak: width of CB's ($\sigma_{CB}, \sigma'_{CB}$)	Free
Total no. of parameters in fit	15
Total no. of free parameters in fit	8

Table 7.2: The fit parameters for the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ fit showing how these parameters are handled in the fit.

To investigate the effect that the remaining $B^0 \rightarrow J/\psi K^+\pi^-$ events could have on the total normalisation yield derived from the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass fit, a $B^0 \rightarrow J/\psi K^+\pi^-$ component was added to the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass fit, with the $B^0 \rightarrow J/\psi K^+\pi^-$ yield Gaussian constrained to the expected yield of 69 ± 9 .

This $B^0 \rightarrow J/\psi K^+\pi^-$ component shape was modelled using $B^0 \rightarrow J/\psi K^{*0} (\rightarrow K^+\pi^-)$ simulation. This assumes that the shape of non-resonant decays, that is $B^0 \rightarrow J/\psi K^+\pi^-$ decays that do not decay via K^{*0} , is similar to that of the resonant decays $B^0 \rightarrow J/\psi K^{*0} (\rightarrow K^+\pi^-)$. This is a reasonable assumption as, under the $J/\psi p\pi^-$ mass hypothesis, the large difference between the mass of the proton and the kaon causes much of the original peaking structure of the decay to be lost. Thus, any sensitivity to the difference in the widths between the resonant and non-resonant case under the $J/\psi K^+\pi^-$ mass difference

will be greatly reduced under the $J/\psi p\pi^-$ hypothesis.

The mass-dependent PID cuts complicate the fit to the $B^0 \rightarrow J/\psi K^{*0}$ mass component considerably. Therefore, in order to test the effect that adding the $B^0 \rightarrow J/\psi K^{*0}$ component to the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass fit has on the normalisation yield, simulation was used without the mass-dependent PID cuts applied. The mass shape can then be modelled by a much simpler function, as shown in [Figure 7.4\(e\)](#).

The presence of the mass-dependent PID cuts however causes the $B^0 \rightarrow J/\psi K^{*0}$ to flatten out and dip to the left of the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ signal peak. This is demonstrated in [Figures 7.4\(a\) and \(b\)](#), which show the $B^0 \rightarrow J/\psi K^{*0}$ component under the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass hypothesis, with and without the mass-dependent PID cuts applied. Both the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $B^0 \rightarrow J/\psi K^{*0}$ components in [Figures 7.4\(a\) and \(b\)](#) are taken from simulation and both components have been normalised to their expected yields.

[Figures 7.4\(c\) and 7.4\(d\)](#) show a fit to the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass with the $B^0 \rightarrow J/\psi K^{*0}$ component included, where the component shape used is that shown in [Figure 7.4\(e\)](#). The effect of adding this $B^0 \rightarrow J/\psi K^{*0}$ component on the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ signal yield derived from the fit is 0.2%. Given the negligible size of this effect¹, no further study is carried out and the $B^0 \rightarrow J/\psi K^{*0}$ component is not included in the final fit to the normalisation channel. No $B^0 \rightarrow K^{*0}\mu^+\mu^-$ component is included in the signal fit, as the yield of expected $B^0 \rightarrow K^{*0}\mu^+\mu^-$ events in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel, calculated using the expected $B^0 \rightarrow J/\psi K^{*0}$ yield and the relevant branching fractions, is consistent with zero.

¹The expected number of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ signal events is of order 10 and thus a statistical error of $\sim 30\%$ on the branching fraction is expected

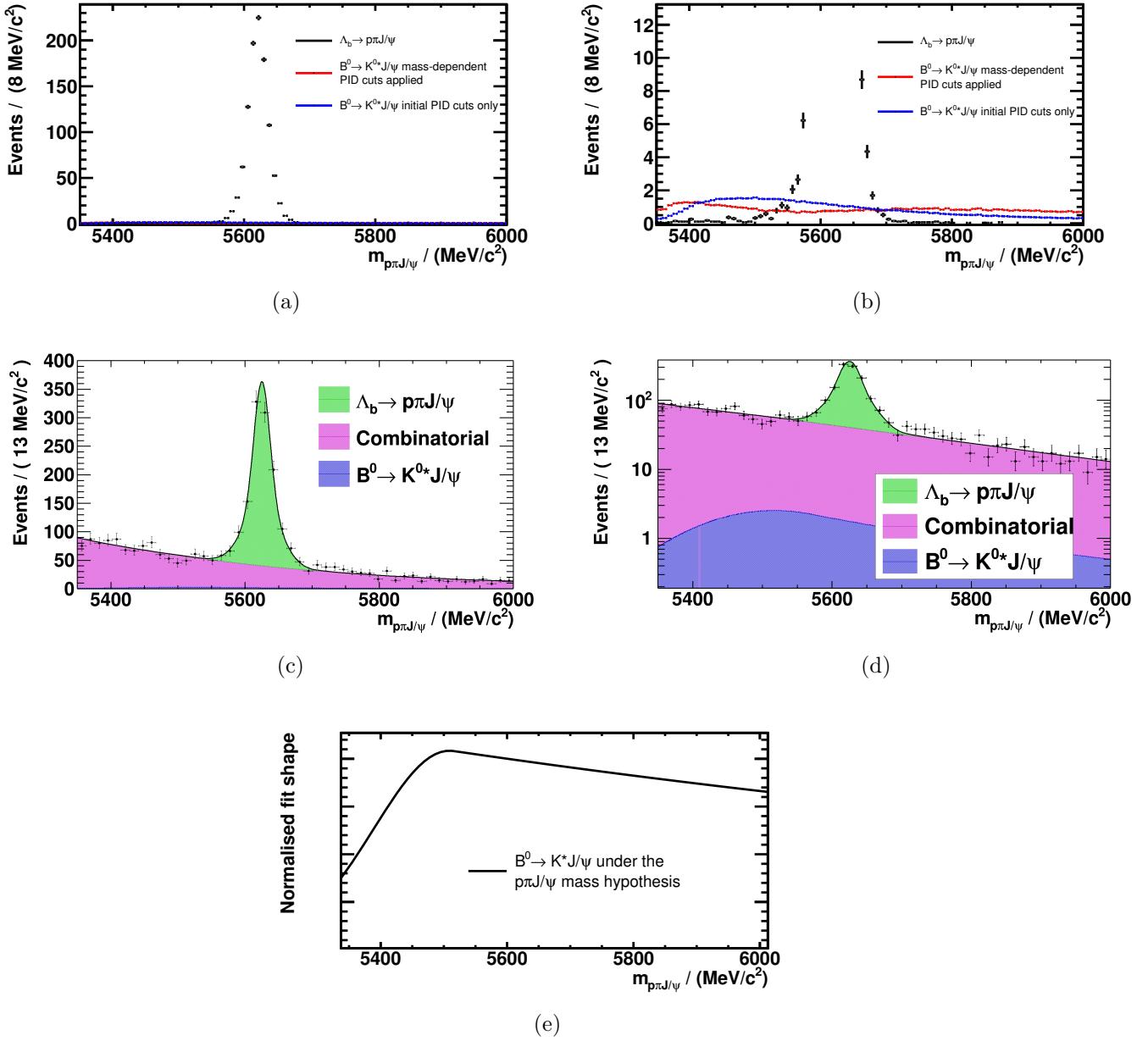


Figure 7.4: The $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel in simulation, normalised to the expected yield of 1017 ± 41 events, along with the $B^0 \rightarrow J/\psi K^{*0}$ component under the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass hypothesis, normalised to the expected yield of 69 ± 9 events, for the cases where the mass-dependent PID cuts are applied and not applied, (a). A zoom-in of (a) is shown in (b). Here, the normalisation value of 69 ± 9 for the $B^0 \rightarrow J/\psi K^{*0}$ component is only correct for the red curve but both blue and red curves are normalised to the same value to allow a meaningful comparison of their shapes. A fit to $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data with the yield of the $B^0 \rightarrow J/\psi K^{*0}$ component Gaussian constrained to 69 ± 9 events is shown on a linear scale in (c) and on a log scale in (d). The shape of $B^0 \rightarrow J/\psi K^{*0}$ simulation, with only initial PID cuts applied, under the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ mass hypothesis is shown in (e).

7.1.3 Complete fit to the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel

The background for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel is made up of a combinatorial component and a part-reco component.

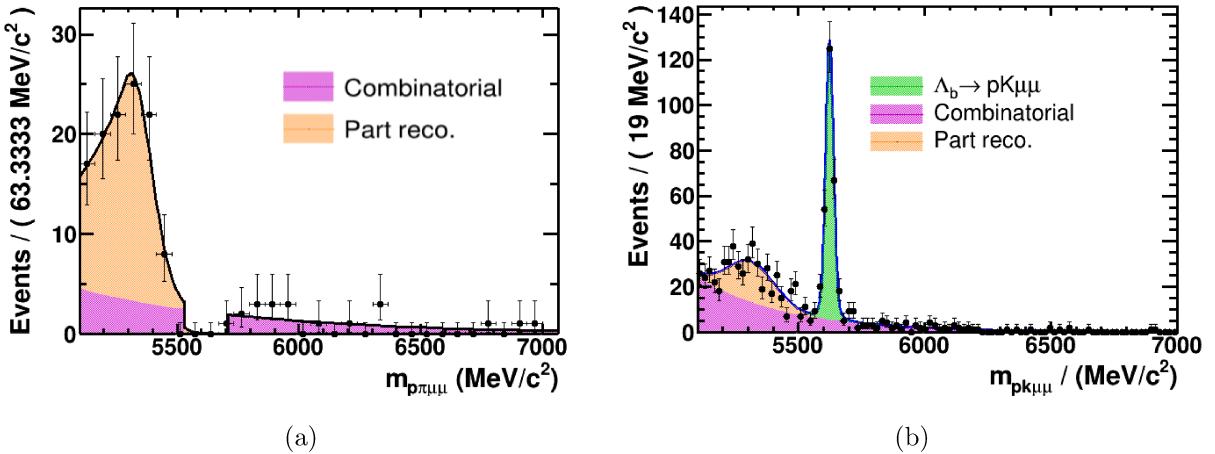


Figure 7.5: Fitted data for blinded $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ (a) and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ (b).

The shape for the part-reco background for $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$, shown in Figure 7.5(a), is a RooExpAndGauss, as defined in Equation 7.1. The value of the parameters for the RooExpAndGauss are taken from the fit to the part-reco component of $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data, shown in Figure 7.5(b), with the errors on the parameters of the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ fit used as Gaussian-constraints on the RooExpAndGauss used to model the part-reco component in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ fit.

The combinatorial background is modelled by a single exponential. The exponential parameter for the combinatorial component is allowed to float, as is the relative size between the RooExpAndGauss and exponential components. A summary of which parameters in the complete fit to $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ candidates are either fixed, free, or constrained is shown in Table 7.3.

The expected $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ yield in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel is 1 ± 1 event. No $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ component is therefore added to the nominal fit but the inclusion of a $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ component is considered as a source of systematic uncertainty in

Fit Parameters	Constrained, free or fixed
Number of signal events	Free
Number of background events	Free
Frac. between background components	Free
Exponential parameter for combinatorial background	Free
α (Equation 7.1)	Constrained from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$
\bar{x} (Equation 7.1)	Constrained from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$
k (Equation 7.1)	Constrained from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$
Signal peak: mean	Fixed from $\Lambda_b^0 \rightarrow J/\psi p\pi^-$
Signal peak: CB tails ($\alpha_{CB}, \alpha'_{CB}, n_{CB}, n'_{CB}$)	Fixed from $\Lambda_b^0 \rightarrow J/\psi p\pi^-$
Signal peak: fraction between CB's	Fixed from $\Lambda_b^0 \rightarrow J/\psi p\pi^-$
Signal peak: width of CB's ($\sigma_{CB}, \sigma'_{CB}$)	Fixed from $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and adjusted using $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ sim.
Total no. of parameters in fit	15
Total no. of free parameters in fit	7
No. of degrees of freedom in signal plus background fit	7
No. of degrees of freedom in background-only fit	6

Table 7.3: The fit parameters for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ fit showing how these parameters were handled in the fit.

chapter 8.

Effect of the choice of background proxy on the signal significance

The RooExpAndGauss fit shape is the nominal model used for the part-reco background but due to the low background statistics, regardless of the shape used to describe the distribution of the background present, there is little effect on the final fit significance.

To study the effect of the choice of background proxy on the signal significance, pseudo experiments are generated from a different background shape taken from a fit to the

blinded $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ dataset, and from a signal shape deduced using simulation, as detailed in [subsection 7.1.1](#).

For the purpose of this study, a single generic exponential component is used to fit the entire background component in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ dataset, instead of the nominal RooExpAndGauss shape plus exponential component, and pseudo experiments are then generated from this fit. This is so as not to bias the shape of the distribution of data points in the pseudo experiments generated from this fit to be too like that of the nominal fit model used. Note that a generic polynomial PDF fitted to the small $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ dataset gives unphysical behaviour in the mass distribution for high mass values.

The number of expected signal events is deduced by taking the number of signal events from the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ signal fit, adjusting for the relative efficiencies between $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$, taking the value for $|V_{ts}/V_{td}|^2$ from Ref. [15], and assuming that the ratio of form factors (and phase space differences) between the two channels cancel. The relative efficiency adjustment is straightforward, as both channels have the same selection, with the exception of the different cuts on the $DLL_{K\pi}$ variable. The expected number of signal events in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel is estimated as ~ 9 . This differs from the estimated values given in [chapter 2](#) as in this case the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ channel is used to estimate the number of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ instead of the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel. The $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ channel was used here because at the time of doing this study the relevant efficiencies between $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ had not yet been calculated.

The number of background events is deduced by integrating over the entire background fit performed on blinded $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data in [Figure 7.5\(a\)](#).

The pseudo experiments were fitted using a background shape that was either a RooExpAndGauss and an exponential combined, or a single exponential. The RooExpAndGauss parameters are Gaussian constrained from the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ background shape and there is no constraint on the exponential parameter or on the relative size between the RooExpAndGauss and exponential components, as summarised in [Table 7.3](#).

The significance for each fit to the pseudo data is calculated using Wilk's theorem,

applied across the entire fit range. Wilk's theorem states that the likelihood ratio between two hypotheses, Θ and Θ_0 , when Θ_0 is a special case of Θ , will be distributed as a χ^2 distribution with degrees of freedom equal to the difference in dimensionality of Θ and Θ_0 . In this case, the difference in the number of degrees of freedom between the signal and background and background-only hypotheses is one.

Example pseudo experiment fits can be seen in [Figure 7.6](#). The significance for the fits to these example pseudo experiments are 4.0σ for the single exponential fit, [Figure 7.6\(a\)](#), and 4.1σ for the constrained RooExpAndGauss fit, [Figure 7.6\(b\)](#).

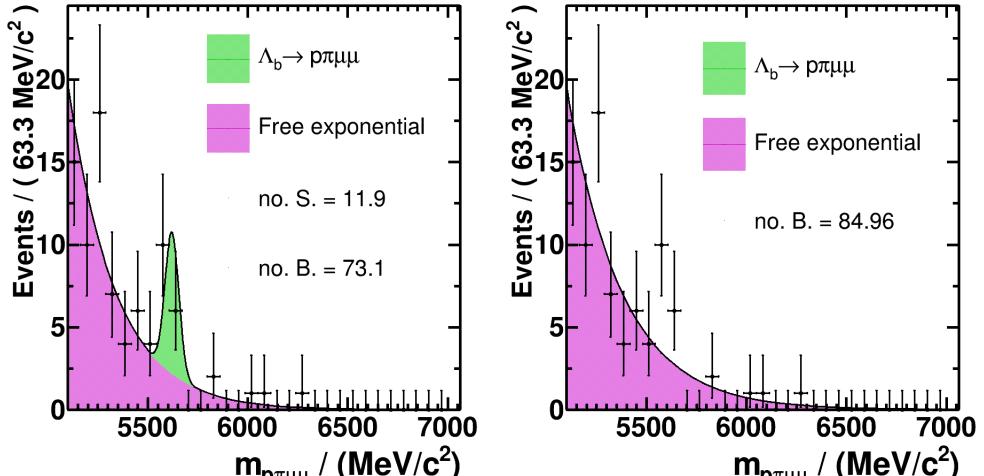
The outputs of the significance for fits to 1000 pseudo experiments, for each fit case, are shown in [Figure 7.7](#). The two significance distributions are fitted with a Gaussian and these Gaussian fit parameters are shown in [Table 7.4](#). The results in [Table 7.4](#) demonstrate that, due to the low statistics of the data set, the effect of the choice of the background model on the final significance is small. The effect of the background shape on the signal yield is discussed as a source of systematic uncertainty in [chapter 8](#).

Fit	Mean	Standard Dev.
RooExpAndGauss and exponential	2.8	1.2
exponential only	2.5	1.2

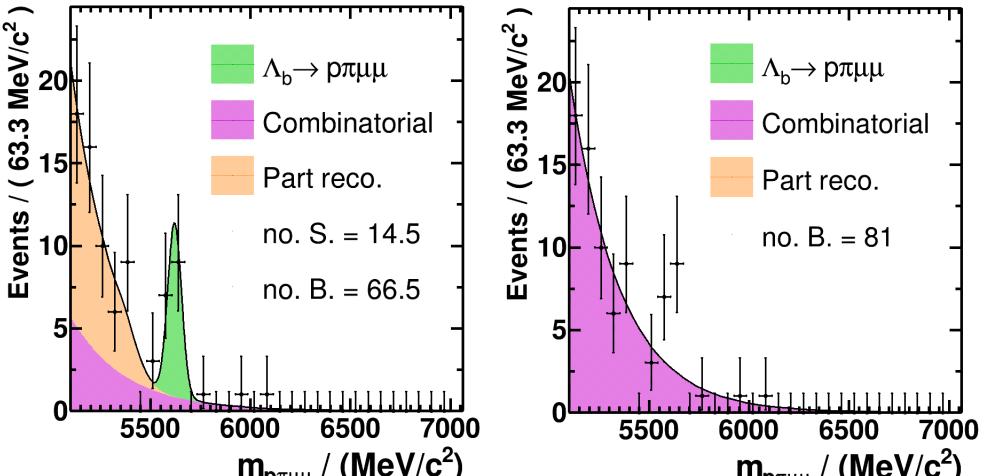
Table 7.4: The fit parameters for a Gaussian function fitted to the distributions in [Figure 7.7](#).

7.1.4 Expected significance for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel

The expected significance is calculated using the same pseudo experiment-based method as outlined in [section 7.1.3](#), again with the expected signal yield taken as 9, but this time the pseudo experiments are initially generated by fitting the background with the nominal constrained RooExpAndGauss model and then fitting back with this same model. This gives a very similar average significance as in the case where the pseudo experiments were generated from an exponential-only background model. The average significance for



(a)



(b)

Figure 7.6: Example pseudo experiments for signal and background hypotheses, (left) and background only fits, (right), with an exponential background with all fit parameters allowed to float (a), and the RooExpAndGauss fit parameters constrained from the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ background fit (b). The no. S and no. B in the legends refer to the number of signal events and background events respectively.

pseudo experiments generated using the nominal RooExpAndGauss model can be seen in Figure 7.8, which shows the Gaussian fit to the output of 1000 pseudo experiments, along

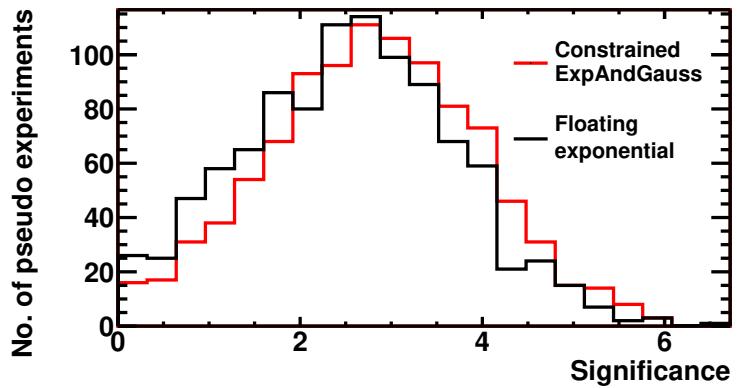


Figure 7.7: The significance distributions from fits to pseudo experiments where the fits in question use a background model which is either a free exponential or a RooExpAndGauss and exponential combined, where the RooExpAndGauss function is constrained.

with the parameters of the fitted Gaussian. Based on these pseudo experiment studies, a signal significance of at least 3σ will be seen 44% of the time.

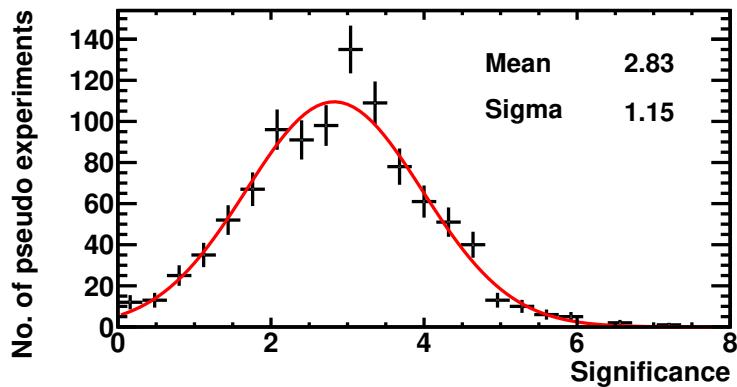


Figure 7.8: Significance of pseudo experiments fits, using pseudo experiments generated from a constrained RooExpAndGauss and fitted back with the same model.

7.2 Selection efficiency

The relative selection and reconstruction efficiency between the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels is necessary to deduce the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction. The branching fraction is calculated from the expression

$$\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-) = \frac{N_{\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-}}{N_{\Lambda_b^0 \rightarrow J/\psi p\pi^-}} \times \frac{\epsilon_{\Lambda_b^0 \rightarrow J/\psi p\pi^-}}{\epsilon_{\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-}} \times \mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-) \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-). \quad (7.2)$$

In this section the methods used to quantify $\epsilon_{\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-}/\epsilon_{\Lambda_b^0 \rightarrow J/\psi p\pi^-}$ will be outlined. This quantity is referred to hereafter as the relative efficiency. All efficiencies are calculated as a function of q^2 and by definition the relative efficiencies between the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels should be equal to unity in the dimuon mass bin ($9 < q^2 < 10$ GeV $^2/c^4$), which corresponds to the square of the mass of the J/ψ .

In this section, the relative efficiency is calculated for the detector acceptance and track reconstruction and the entire selection, including the trigger requirements, and the BDT. The total efficiency integrated over q^2 will depend on the final q^2 distribution that is assumed.

The simulation and data have different q^2 distributions due to the simulation being generated with a phase space model. In particular, no form factors are taken into account and, given the distribution seen in both $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ data and theory predictions, these are thought to cause a rise in the q^2 distribution at higher q^2 . This is illustrated in [Figure 7.9](#)². The efficiency, therefore, is calculated in bins of q^2 and the distribution of q^2 events is taken from $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ differential branching fraction predictions, taken from Ref. [10]. A comparison between the q^2 distributions taken from phase space simulation and those from the $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ branching fraction predictions is shown in [Figure 7.9](#).

The efficiency is not weighted in terms of the $p\pi$ mass or the angle of the leptons.

²The $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ theory predictions used in the analysis are taken from Ref. [10]. The theory predictions have since been updated in February 2016 in Ref. [24] as shown in [Figure 2.5](#). Although the previous predictions are still used, any effect that a variation in the q^2 distribution will have on the relative efficiency is already well covered by an assigned systematic uncertainty, as discussed in [section 8.2](#).

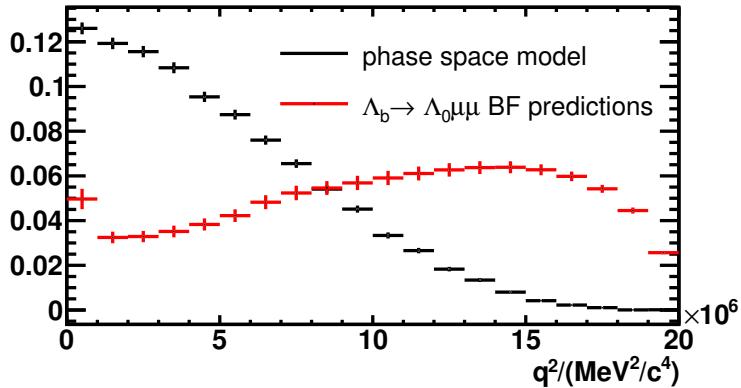


Figure 7.9: A comparison between the q^2 distribution taken from phase space simulation and from the $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ branching fraction predictions of Ref. [10].

However, the effect of not reweighting in the $p\pi$ mass on the final relative efficiency is considered as a systematic uncertainty in chapter 8. It is not necessary to obtain a highly accurate model of the selection and reconstruction efficiencies, as, due to the low statistics in the signal channel, the systematic uncertainty applied to account for the limited knowledge of the efficiency variation with $p\pi$, adds little to the total uncertainty.

7.2.1 Detector acceptance cuts

The detector acceptance criteria requires that the daughters lie within the range $10^\circ < \theta < 400^\circ$ mrad of the beam axis and only events within this range are simulated. The effect of the detector level acceptance cuts are calculated by taking the q^2 distribution of generator-level simulation with and without the detector acceptance criteria applied to the decay daughters.

7.2.2 Stripping and reconstruction efficiency

The total efficiency of the stripping selection and reconstruction is calculated using $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation, by comparing the distribution of simulated events as a function of q^2 before and after the application of both the stripping line and reconstruction.

[Figure 7.10](#) shows the resulting relative efficiency between the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels as a function of q^2 for the stripping selection and the reconstruction.

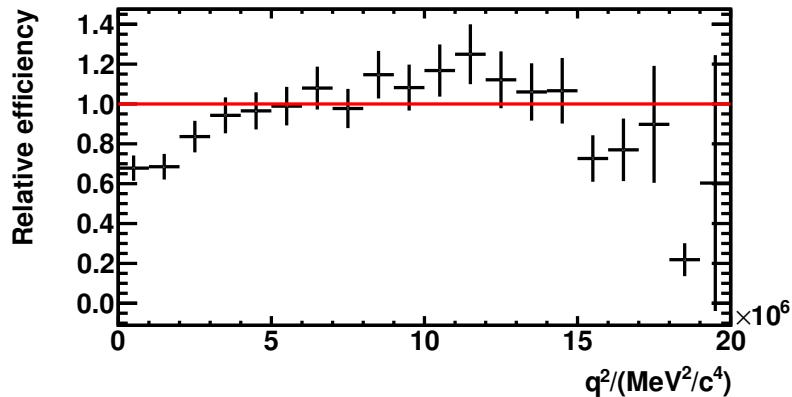


Figure 7.10: Relative combined stripping and reconstruction efficiencies between $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation as a function of q^2 .

7.2.3 Trigger efficiency, preselection and PID cuts

The trigger efficiency is calculated using simulation, by applying the relevant TOS requirements, as discussed in [section 5.2](#). The preselection efficiency is calculated using resampled simulation, as is the effect of the mass-dependent PID cuts. However, any selection involving the DLL_{pK} variable is not calculated using resampled simulation, due to the simulation's poor replication of the DLL_{pK} variable (see [subsection 5.4.2](#)). Instead, the efficiency values for a given PID cut are calculated directly from the calibration samples shown in [Table 5.6](#).

The absolute efficiency distribution for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel for the trigger, preselection and PID selection can be seen in [Figure 7.11](#). The dashed lines indicate the efficiency of the ($9 < q^2 < 10 \text{ GeV}^2/c^4$) bin of the relevant distribution. The trigger varies with q^2 as expected, with the trigger being more efficient for harder muons. The slight drop off at high q^2 for the PID reflection cuts is due to this region of q^2 corresponding to softer hadrons, which will have poorer separation in the RICH.

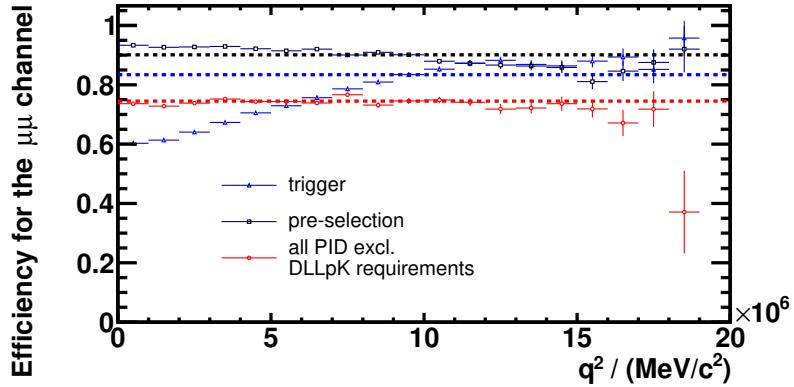


Figure 7.11: Absolute efficiencies for the preselection, trigger and PID selection, as a function of q^2 for $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation. The dashed lines indicate the efficiency of the ($9 < q^2 < 10 \text{ GeV}^2/\text{c}^4$) bin for the efficiency distribution of the same colour. There is no efficiency value for the last bin due to a lack of data in this bin.

7.2.4 BDT efficiency

The value of the BDT efficiency is assumed to be flat in q^2 , and therefore to have no effect on the relative efficiency. The difference to the total relative efficiency between the case where the efficiency is assumed to be flat and the case where the BDT efficiency is taken as a function of q^2 from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data is taken as a systematic uncertainty (see chapter 8).

7.2.5 Total relative efficiency

The total integrated relative efficiency, assuming a given q^2 distribution, is given by the sum $\sum_i (\epsilon_i \times N_{q_i^2}/N)$, where the i refers to the i^{th} bin in q^2 , ϵ_i refers to the relative efficiency between the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels for the i^{th} bin, and the distribution of $N_{q_i^2}/N$ is taken from $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ branching fraction predictions. The total error is taken by combining the errors on the q^2 efficiency distribution, which is dependent on the simulation statistics, and the errors on the $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ branching fraction theory predictions. Due to there being only a handful of events in the last q^2 bin

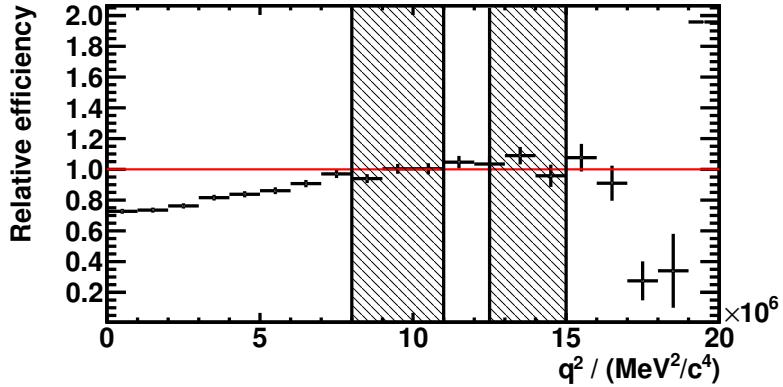


Figure 7.12: The total relative efficiency assuming a flat BDT efficiency in q^2 . The shaded areas indicate the vetoed regions in q^2 .

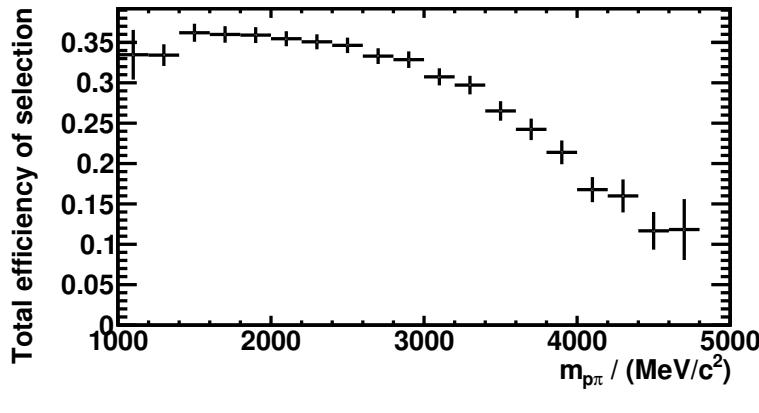
($19 < q^2 < 20$) GeV^2/c^4 in phase space simulation before any selections have been applied, and the large weight this bin gets from the $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ branching fraction predictions, this bin can not be included in the integration over q^2 . According to $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ theory predictions for the differential branching fraction, this last bin in q^2 contains 2.5% of all events. Instead, the integration is done only up to $19 \text{ GeV}^2/c^4$, with a scaling factor applied to renormalise the $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ branching fraction predictions, and the difference between the efficiency achieved by integrating either up to $19 \text{ GeV}^2/c^4$ or up to $20 \text{ GeV}^2/c^4$ is taken as a systematic, as discussed in [chapter 8](#).

Integrating up to $19 \text{ GeV}^2/c^4$, not including the areas of q^2 that are vetoed, gives a total integrated relative efficiency of 0.487 ± 0.022 . The effect of simulation reweighting on the total relative efficiency, as well as the choice of the q^2 distribution used to give $N_{q_i^2}/N$, are both evaluated as systematic uncertainties, as outlined in [chapter 8](#).

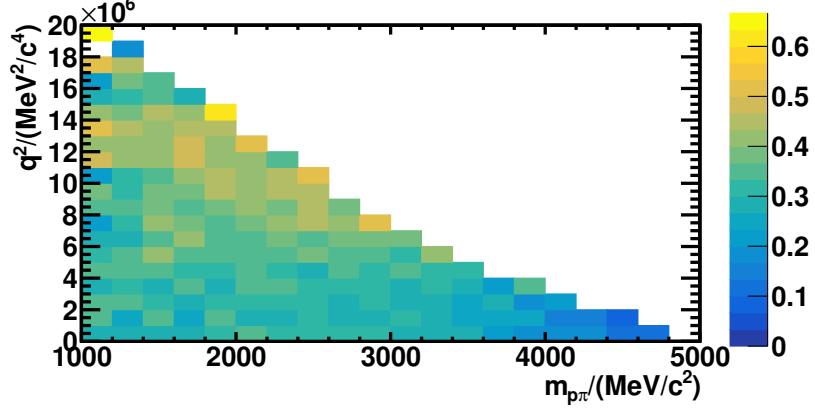
7.2.6 Efficiency as a function of $m_{p\pi}$

It is also of interest to study the efficiency as a function of the dihadron mass spectrum, $m_{p\pi}$. The efficiency as a function of $m_{p\pi}$ will be highly correlated with the efficiency as a function of q^2 , as events with a high value for $m_{p\pi}$ must correspond exclusively to low q^2 events, and vice versa. As such, given that there is a rising efficiency with low q^2 events, a

falling efficiency is expected for high $m_{p\pi}$ events. This expected behaviour is shown in [Figure 7.13\(a\)](#), which shows the efficiency for the preselection and PID selections as a function of $m_{p\pi}$. The same preselection and PID efficiency can be seen as a function of both q^2 and $m_{p\pi}$ in [Figure 7.13\(b\)](#). As shown in [Figure 7.13\(b\)](#), the efficiency in $m_{p\pi}$, for a given value of q^2 , is fairly flat. This is relevant, as the $m_{p\pi}$ spectrum is poorly replicated in phase space simulation, due to the presence of N^* resonances, which are ignored in the phase space model. The effect of the use of the phase space model for the dihadron mass spectrum on the total efficiency is considered as a source of systematic uncertainty in [chapter 8](#).



(a)



(b)

Figure 7.13: The absolute efficiency for the preselection and PID selections as a function of $m_{p\pi}$, (a), and both $m_{p\pi}$ and q^2 , (b).

Chapter 8

Results

In this chapter the result of the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction measurement will be presented in [section 8.1](#). This will be followed by an overview of the systematic uncertainties on the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction measurement in [section 8.2](#).

8.1 The $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction measurement

The mass fit to $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data is shown in [Figure 8.1](#).

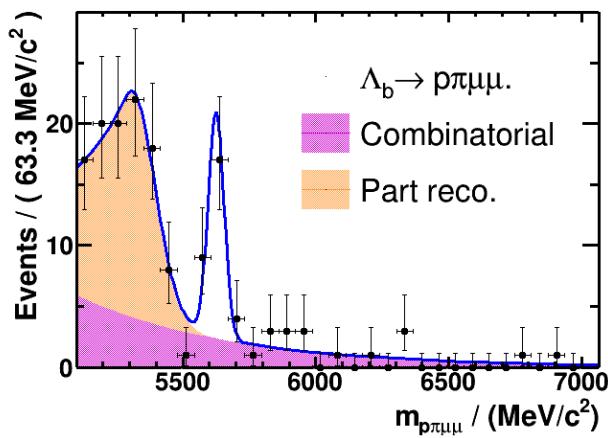


Figure 8.1: The fit to the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass distribution.

The $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ signal peak is observed with a 5.5σ significance, where the significance is calculated using Wilk's theorem, as previously discussed in [section 7.1.3](#).

The $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ event yield taken from the fit to the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass distribution is 22 ± 6 events. Given that 1017 ± 41 events are observed in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ fit and $\epsilon_{\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^- / \Lambda_b^0 \rightarrow J/\psi p\pi^-} = 0.487 \pm 0.022$, the value of $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)$ is given as

$$\begin{aligned}\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-) &= 22 \times (1/0.487) \times (1/1017) \times \mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-) \times \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) \\ &= (6.9 \pm 1.9 \pm 1.1^{+1.3}_{-1.0}) \times 10^{-8}\end{aligned}\tag{8.1}$$

where the first error is the statistical uncertainty, the second is the systematic uncertainty and the third is the uncertainty from $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-)$. The systematic uncertainties are discussed in detail in [section 8.2](#).

There are no theoretical predictions for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction. However, the yield of 22 ± 6 is consistent with the estimated yield of 9 events at the 2σ level. The order of magnitude of the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction is two orders of magnitude less than the $\Lambda_b^0 \rightarrow p\pi^- J/\psi (\rightarrow \mu^+\mu^-)$ branching fraction and the ratio between $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)$ and $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-)$ is consistent with that of $\mathcal{B}(\Lambda_b^0 \rightarrow \Lambda^0 \mu^+\mu^-)$ and $\mathcal{B}(\Lambda_b^0 \rightarrow \Lambda^0 J/\psi)$.

A series of separate post-unblinding checks were performed to check for contamination from pions misidentified as muons. The PID requirements on both muons were tightened from $DLL_{\mu\pi} > -3$ to $DLL_{\mu\pi} > 0$ and events with $2256 < m_{p\pi\mu\rightarrow\pi} < 2316 \text{ MeV}/c^2$ are vetoed in order to remove Λ_c^+ contributions. The $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass distribution was then refitted, as shown in [Figure 8.2](#). After these additional $DLL_{\mu\pi}$ selections the signal yield reduces by less than an event and the additional Λ_c^+ veto also reduces the signal by \sim an event. The significance after both these selections reduces to 5.3σ and the branching fraction remains consistent with that calculated without these additional selections applied.

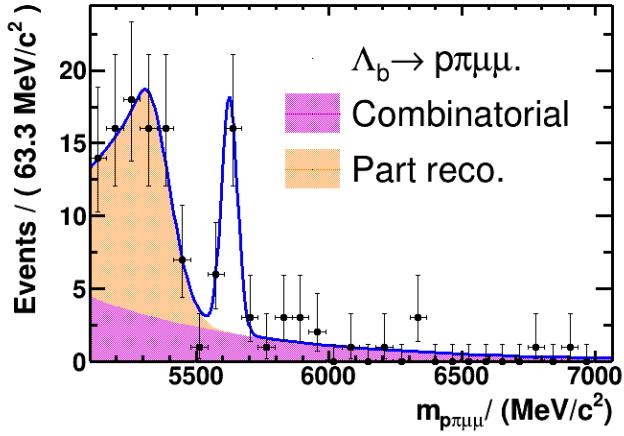


Figure 8.2: The fit to the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ mass distributions for data candidates with $2256 < m_{p\pi\mu\mu} < 2316 \text{ MeV}/c^2$ events vetoed and with the requirement $\text{DLL}_{\mu\pi} > 0$ placed on both muons.

8.2 Systematic uncertainties on $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)$

For the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis, a range of sources of systematic uncertainty is studied. All systematic uncertainties are quoted as a percentage of the value of the quantity whose uncertainty is being deduced, namely either the relative efficiency between the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels, or the uncertainty on the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ ($\Lambda_b^0 \rightarrow J/\psi p\pi^-$) yields extracted from the maximum-likelihood fits. In some cases, the systematic uncertainty associated with a certain factor is deduced by comparing the nominal value of this factor against a value deduced by changing some aspects used in the calculation of this factor. In these cases, the error is quoted as either the difference between the nominal and changed value or as the error on the changed value, whichever is largest. The main systematic uncertainties considered for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis are:

Trigger The TISTOS method, as discussed in subsection 3.2.1, is applied to data and simulation to give an estimate of how accurate the simulation trigger efficiency estimation is.

PID The PID variables in the $\Lambda_b^0 \rightarrow J/\psi pK^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data channels are

compared against the equivalent in simulation in order to quantify how well the PID distributions are modelled in simulation.

The assumed q^2 distribution The q^2 distributions from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ sWeighted data is compared against that of $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ theoretical predictions in order to quantify the effect that changing the assumed q^2 distribution has on the total relative efficiency.

The assumed $m_{p\pi}$ distribution The $m_{p\pi}$ distribution from $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ sWeighted data is compared to that from phase space $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation in order to quantify the effect that changing the assumed $m_{p\pi}$ distribution has on the total relative efficiency.

The effect of the choice of fit model on the signal yield To quantify the effect of the choice of fit model on the signal yield, constraints on the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ fit model were lifted and the fit redone. The model background was also altered and the fit repeated.

Simulation reweighting and BDT efficiency The effect that reweighting the simulation has on the final relative efficiency value is quantified by comparing the relative efficiency with and without the simulation weights applied. The total relative efficiency is also recalculated using the BDT efficiency taken from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data.

8.2.1 The TISTOS method

The TISTOS method, as discussed in [subsection 3.2.1](#), refers to the use of data to calculate the trigger efficiency by assuming that TIS (Triggered Independent of Signal) and TOS (Triggered On Signal) events are independent. Thus the efficiency for the selection of TOS candidates can be expressed as

$$\epsilon_{TOS} = \frac{N_{TISTOS}}{N_{TIS}}. \quad (8.2)$$

The trigger efficiency values for data and simulation are compared by applying the TISTOS method to background-subtracted $\Lambda_b^0 \rightarrow J/\psi pK^-$ data and $\Lambda_b^0 \rightarrow J/\psi pK^-$ simulation. The $\Lambda_b^0 \rightarrow J/\psi pK^-$ channel is used due to the increased statistics in the $\Lambda_b^0 \rightarrow J/\psi pK^-$ data channel compared to the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channels. It is assumed that the difference between the $\Lambda_b^0 \rightarrow J/\psi pK^-$ data and simulation will be similar to the difference between $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ data and simulation. The trigger efficiency values in [Table 8.1](#) are all given relative to the previous trigger level. The L0 and HLT1 trigger TISTOS efficiencies agree between data and simulation within 1σ and the HLT2 trigger TISTOS efficiencies agree within 2σ .

Trigger Level	simulation	data
L0	0.867 ± 0.002	0.869 ± 0.020
HLT1	0.942 ± 0.001	0.952 ± 0.022
HLT2	0.965 ± 0.003	0.922 ± 0.027
Total	0.788 ± 0.004	0.762 ± 0.017

Table 8.1: The trigger efficiency obtained using the TISTOS method in simulation and data.

The total difference in trigger efficiency between data and simulation is 3.4% and this is assigned as a systematic uncertainty.

8.2.2 PID

Resampled simulation is used to produce the distributions of the $p\text{DLL}_{p\pi}$ and $\pi\text{DLL}_{K\pi}$ variables, as discussed in [subsection 5.4.2](#). To quantify how well the PID is modelled in resampled simulation, the distributions of the PID variables, $\text{DLL}_{p\pi}$ and $\text{DLL}_{K\pi}$, are compared between data and resampled simulation using $\Lambda_b^0 \rightarrow J/\psi pK^-$ data which has better statistics than $\Lambda_b^0 \rightarrow J/\psi p\pi^-$. The data are shown in [Figure 5.4](#).

The difference is then taken between data and resampled simulation for each bin in PID value, and the difference is fitted with a zero-order polynomial to extract the average

difference between data and simulation. The difference between data and simulation is looked at between the range of 0 to 50 for the $\text{DLL}_{p\pi}$ variable, which disregards the cut-off at high $\text{DLL}_{p\pi}$ values discussed in [subsection 5.4.2](#). The effect of this cut-off on the final efficiency is discussed below. The difference between the data and simulation distributions for the $\text{DLL}_{p\pi}$ and $\text{DLL}_{K\pi}$ PID variables are shown in [Figure 8.3](#). The values of the fitted constants in the zero-order polynomial fits shown in [Figure 8.3\(a\)](#) and [Figure 8.3\(b\)](#) are shown in [Table 8.2](#). The results of both fits give a value for the fit constant that is consistent with zero¹ and no systematic uncertainty is assigned.

	$\text{DLL}_{p\pi}$	$\text{DLL}_{K\pi}$
Fit constant	$(3.0 \pm 18.0) \times 10^{-4}$	$(1.0 \pm 16.0) \times 10^{-4}$

Table 8.2: Value of the fit constant in the zero-order polynomial fit to the difference between data and resampled simulation as a function of PID variable.

The effect of the upper cut-off in the data calibration samples on the final relative efficiency

As discussed in [subsection 5.4.2](#), there is an upper cut-off at 50 on all the PID calibration samples from data used to resample the simulation. This translates to an effective cut-off in the PID distributions in resampled simulation.

This does not affect the majority of the PID distributions used in this analysis, with the exception of the $p\text{DLL}_{p\pi}$ distribution, where the cut at 50 removes 4.5% of the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ signal events. This value of 4.5% is calculated using the $\text{DLL}_{p\pi}$ distribution given by sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data, as shown in [Figure 8.4](#).

How much of an effect this cut-off at 50 in resampled simulation has on the total

¹in the case of [Figure 8.3\(b\)](#) the distribution is not particularly flat, although the average is consistent with zero. However, even if there is some small deviation from a flat hypothesis for this PID variable, given that the same PID variables are used in the signal and normalisation channels, the effect on the relative efficiency of any mis-modelling of the PID distributions in the simulation will still be small

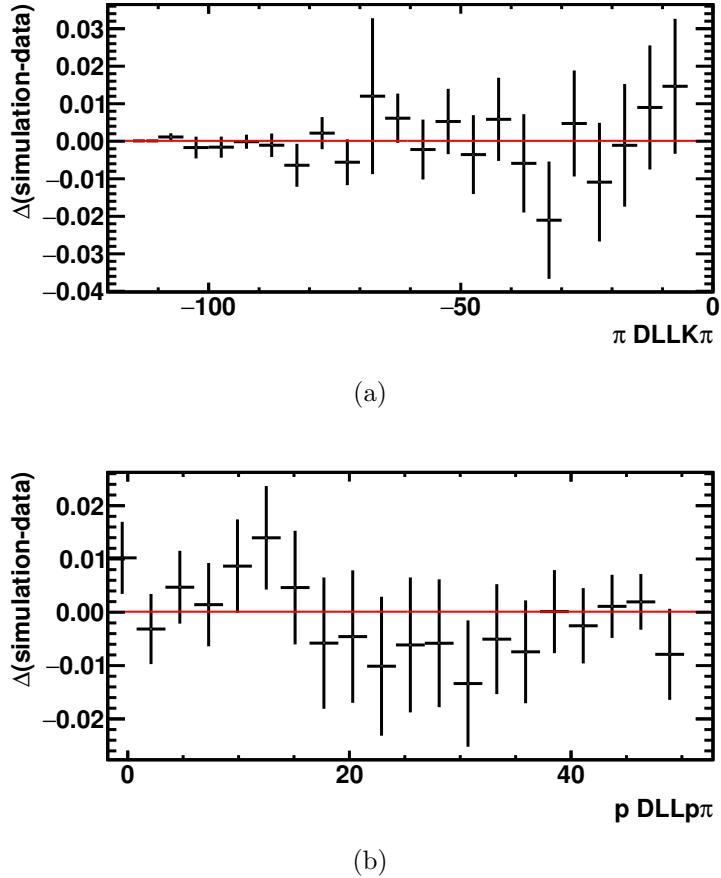


Figure 8.3: The difference between data and simulation for the $\text{DLL}_{K\pi}$ variable, (a), and the difference between data and simulation for the $\text{DLL}_{p\pi}$ variable, (b).

relative efficiency depends largely on two factors. Firstly, if the efficiency distribution for these events in the $\text{DLL}_{p\pi} > 50$ window is flat or varied with q^2 , relative to the J/ψ bin in q^2 , and secondly, whether or not the proportion of events in this $\text{DLL}_{p\pi} > 50$ window is similar for both the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels.

Given that protons with a higher momentum will be easier to identify in the RICH, high values of $\text{DLL}_{p\pi}$ are more likely to be associated with low values of q^2 , where it is found that the PID efficiency is fairly flat. However, if the extreme case were assumed, i.e. that the efficiency across this region of $\text{DLL}_{p\pi} > 50$ in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel differs from the efficiency in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channel by the maximal amount observed in

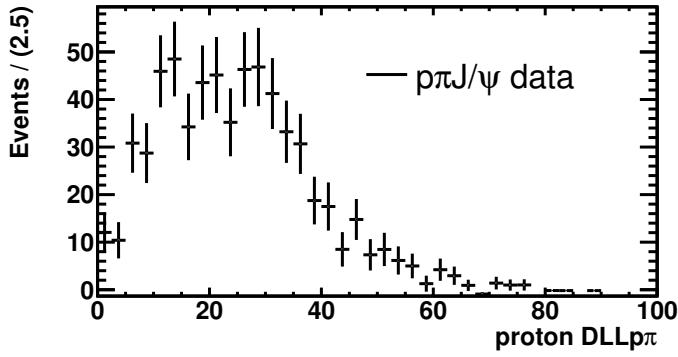


Figure 8.4: The $\text{DLL}_{p\pi}$ distribution for sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ after the application of all selections.

Figure 7.11, this would yield an efficiency difference of $\sim 20\%$ between the two channels. Thus, if it is assumed that the proportion of events in the $\text{DLL}_{p\pi} > 50$ window is the same in the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels, the maximum effect on the total integrated relative efficiency that this cut-off at 50 could have would be 20% of 4.5%, giving 0.9%.

To deduce whether or not the proportion of events in the $\text{DLL}_{p\pi} > 50$ window is similar between the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ channel, the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ channels in data are used as a proxy. The proportion of events in this $\text{DLL}_{p\pi} > 50$ window is compared between the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ channel and it is assumed that these proportions will be similar in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ ($\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$) case. For $\Lambda_b^0 \rightarrow J/\psi pK^-$ data, $94 \pm 2\%$ of events fall outside the $\text{DLL}_{p\pi} > 50$ window, and in the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ case it is $93 \pm 15\%$ of events. Taking the 15% statistical error as the difference in proportions of $\text{DLL}_{p\pi} > 50$ events between the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi pK^-$ channels, and combining this with the previous value of 0.9%, gives a maximum difference in the total integrated efficiency of 1.0%. This is, of course, an approximate estimate, but it indicates that it is reasonable to assume that the effect of this upper cut-off is small. This 1.0% is assigned as a systematic uncertainty.

8.2.3 The effect of the choice of the q^2 distribution on the total relative efficiency

The total efficiency will depend on which q^2 distribution is used as a proxy for the signal q^2 distribution. The nominal proxy used is taken from the $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ differential branching theory predictions, which were shown in [Figure 7.9](#).

As an additional study, the q^2 distribution from $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ data is used as an alternative to calculate the total integrated efficiency, although the statistics are limited. The $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ kinematics should be similar to those from $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$, although the range in q^2 for $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ will be smaller, due to the heavier kaon. The total integrated efficiency calculated using the q^2 distribution from $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ data and $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ branching fraction predictions are compared. The $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ data is sWeighted and, in order to deduce the proportion of $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ events that fall into the q^2 vetoes, the q^2 distribution is fitted with a generic polynomial, as shown in [Figure 8.5](#). The fit is applied across the whole range in q^2 , including the vetoed regions.

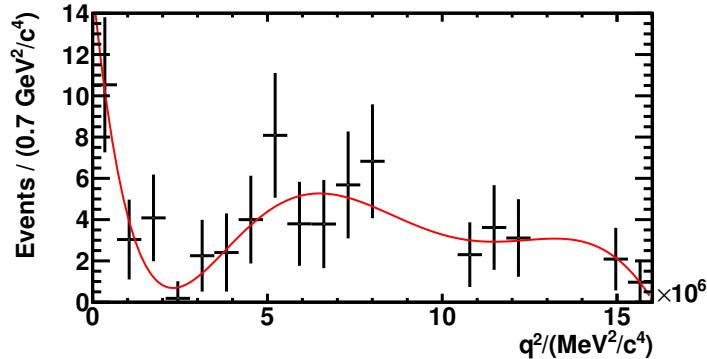


Figure 8.5: The q^2 distribution for sWeighted $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ data fitted with a polynomial function.

It is estimated from the fit in [Figure 8.5](#) that 67.8% of $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ events remain following the q^2 vetoes, giving a total efficiency integrated over q^2 of 0.528 ± 0.042 . The error quoted here considers only the statistical error on the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ data and not the statistical error on the simulation, which is considered as a separate source of

systematic uncertainty in [subsection 8.2.5](#). This efficiency value is consistent with the nominal value of 0.487 ± 0.022 computed in [section 7.2](#). However the statistical error on the relative efficiency calculated using the q^2 distribution from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data is taken as an estimate of the systematic uncertainty on the relative efficiency associated with the modelling of the q^2 distribution. This gives an uncertainty of 7.9%.

Effect of limiting the range in q^2 on the total integrated efficiency

As discussed in [subsection 7.2.5](#), the final q^2 bin of $(19 < q^2 < 20) \text{ GeV}^2/c^4$ has extremely limited statistics in phase space simulation after selection, meaning it cannot be included in the total efficiency integration over q^2 . However, all bins above $15 \text{ GeV}^2/c^4$ have a relative efficiency of less than unity. Assuming therefore that the maximum relative efficiency this q^2 bin could have is unity, and the minimum value of efficiency this bin could have is zero, the maximum possible change to the relative efficiency is 2.5%. This value is derived from $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ theory predictions which have 2.5% of events in the $(19 < q^2 < 20) \text{ GeV}^2/c^4$ q^2 bin. This 2.5% is taken as a systematic uncertainty on the relative efficiency value.

8.2.4 The effect of the choice of the BDT efficiency proxy on the total relative efficiency

The nominal BDT efficiency was computed from $\Lambda_b^0 \rightarrow J/\psi pK^-$ events and assumed to be flat in q^2 . To quantify the effect that this assumption has on the relative efficiency, the BDT efficiency is recalculated using $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data. The distribution $\epsilon_i \times N_{q_i^2}/N$, defined in [subsection 7.2.5](#), for the BDT efficiency taken from either $\Lambda_b^0 \rightarrow J/\psi pK^-$ or $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data, combined with the q^2 shape taken from $\Lambda_b^0 \rightarrow \Lambda^0\mu^+\mu^-$ branching fraction predictions, is shown in [Figure 8.6](#). For events above $15 \text{ GeV}^2/c^4$ in $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data, the nominal value from $\Lambda_b^0 \rightarrow J/\psi pK^-$ data is used, as there are less than five events in $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data in this bin, making an efficiency calculation using $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ data impractical.

The total integrated efficiency assuming a BDT efficiency taken from $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$

data is 0.514 ± 0.048 , which is consistent with the 0.487 ± 0.022 taken using $\Lambda_b^0 \rightarrow J/\psi p K^-$ data as a proxy for the BDT efficiency. The difference between these two values is quoted as a fraction of the nominal efficiency of 0.487 ± 0.022 , giving a systematic uncertainty of 5.6%.

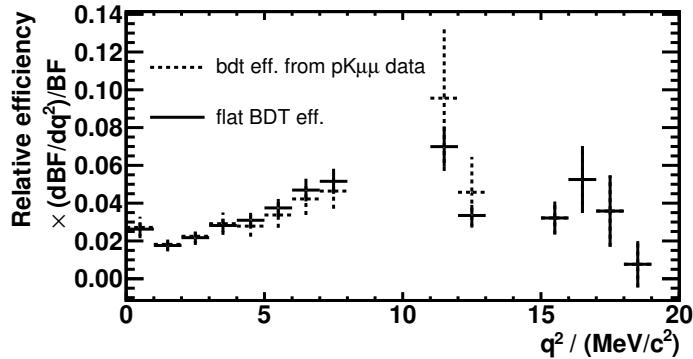


Figure 8.6: The total efficiency as a function of q^2 multiplied by the $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ q^2 distribution and assuming either a flat or varying BDT efficiency distribution in q^2 .

8.2.5 Effect of weighting the simulation and the simulation statistics on the total integrated efficiency value

The final relative efficiency is 0.487 ± 0.022 which gives a 4.4% uncertainty on the relative efficiency value. This error comes from the combination of the uncertainty on the efficiency values on each q^2 bin, due to the simulation statistics, and also the error on the q^2 distribution itself, taken from the theoretical error on the $\Lambda_b^0 \rightarrow \Lambda^0 \mu^+ \mu^-$ branching fraction predictions.

The effect that reweighting the simulation has on the total integrated efficiency value is studied by comparing the total integrated relative efficiency computed where the simulation has been weighted and not weighted, as outlined in subsection 5.4.1. The comparison of these two relative efficiencies as a function of q^2 is shown in Figure 8.7. The integrated efficiency for the weighted case is 0.487 ± 0.022 and for the unweighted case it is 0.492 ± 0.023 . This gives a difference of 1.1% between the two values, which is assigned as a systematic

uncertainty.

In addition, the efficiency calculated by weighting simulation using $\Lambda_b^0 \rightarrow J/\psi pK^-$ data, instead of $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data, is calculated and compared with the nominal relative efficiency value. This is found to be 0.484 ± 0.023 which gives a difference of 0.6% with the nominal value which is also assigned as a systematic uncertainty. This gives a total systematic uncertainty, due to the reweighting on the simulation, of 1.3%.

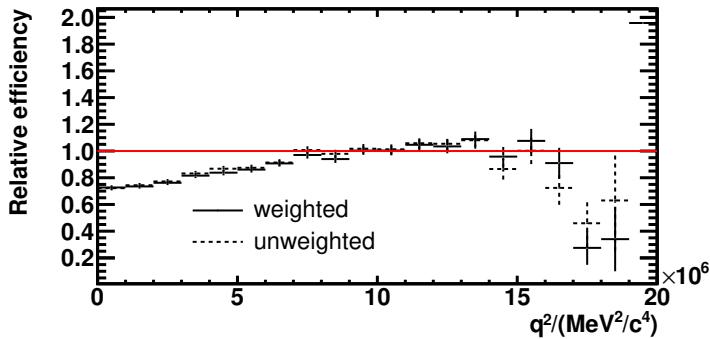


Figure 8.7: The total relative efficiency calculated with weighted simulation, compared against the efficiency calculated using non-weighted simulation.

8.2.6 Effect of reweighting the $p\pi$ mass spectrum in simulation on the total integrated efficiency

The $p\pi$ mass spectrum, $m_{p\pi}$, in phase space simulation is known to be incorrectly modelled for both $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$. This is shown in Figure 8.8, which shows sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data and phase space (phsp) simulation for both the $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ channels.

To quantify the effect of this incorrect modelling, $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ sWeighted data is used to reweight the $p\pi$ mass spectrum in $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation. As the range in the dihadron mass spectrum for $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data is smaller than for $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ data, due to the narrower dimuon spectrum in the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ case, the reweighting is done with a cut placed at $m_{p\pi} < 2500 \text{ MeV}/c^2$ on both $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data

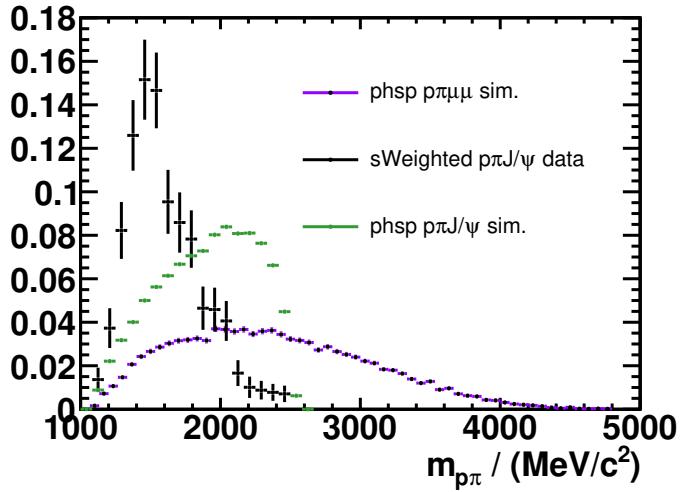


Figure 8.8: The $p\pi$ mass spectrum for sWeighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data and phase space (phsp) $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation.

and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ simulation. The comparison between $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation after reweighting in $m_{p\pi}$ is shown in Figure 8.9 and the relative efficiency as a function of q^2 for the $m_{p\pi}$ weighted and unweighted case is shown in Figure 8.10.

The $m_{p\pi}$ weighted $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ and $\Lambda_b^0 \rightarrow p\pi^- \mu^+ \mu^-$ simulation is used to calculate the efficiency, relative to the J/ψ q^2 bin. The comparison between the relative efficiency with and without the dihadron weights applied is taken as a percentage of the unweighted value, giving a difference of 7.7%. This is assigned as the systematic uncertainty arising from the mismodelling of the $m_{p\pi}$ spectrum in simulation.

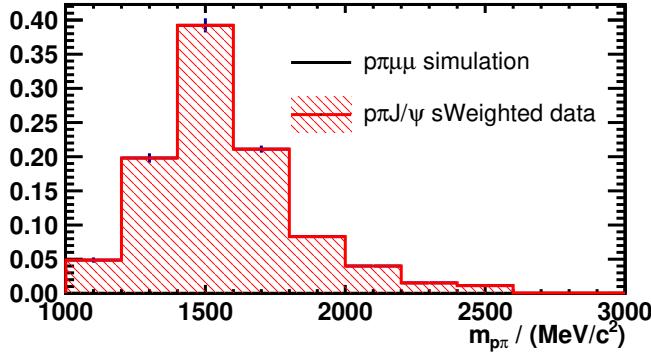


Figure 8.9: The $p\pi$ mass spectrum for $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ data and $m_{p\pi}$ weighted $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation, with a cut placed at $m_{p\pi} < 2500 \text{ MeV}/c^2$.

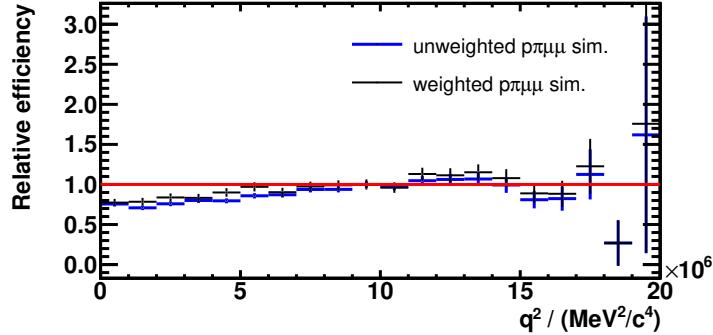


Figure 8.10: The efficiency as function of q^2 for $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ simulation with $m_{p\pi}$ weights applied and not applied, with a cut placed at $m_{p\pi} < 2500 \text{ MeV}/c^2$.

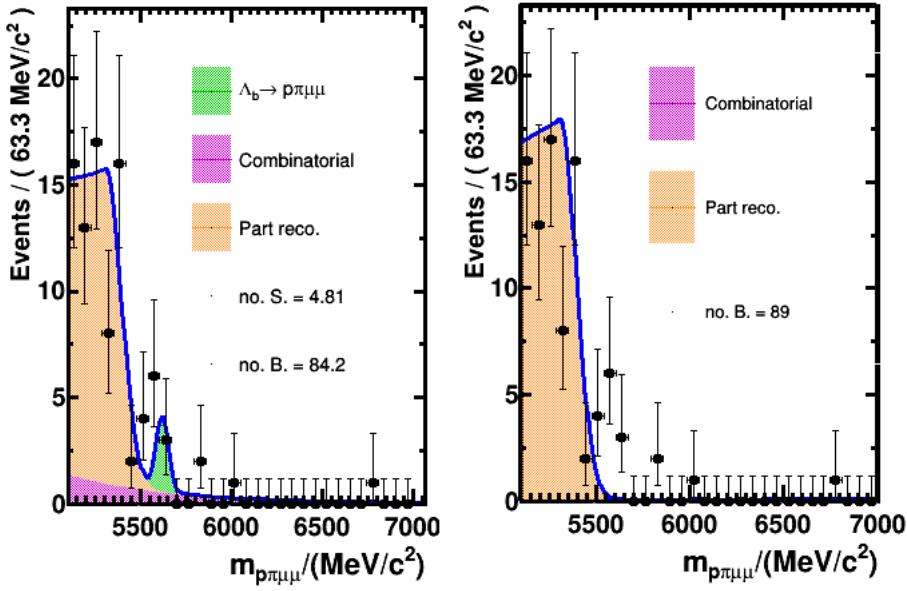
8.2.7 Effect of the fit shape on the fitted signal yield

The effect of the chosen fit shape model on the fitted signal yield is evaluated via various approaches. The first approach outlined in this section involves the comparison between the number of signal events given by the fit when the fit is constrained using Gaussian constraints on the `RooExpAndGauss` part-reco component and when these constraints are lifted. The second approach considers the effect of adding the $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ mass component to the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ signal fit. The third approach involves a complete change of background model from a `RooExpAndGauss` and exponential to a single exponential and a change in the fit range.

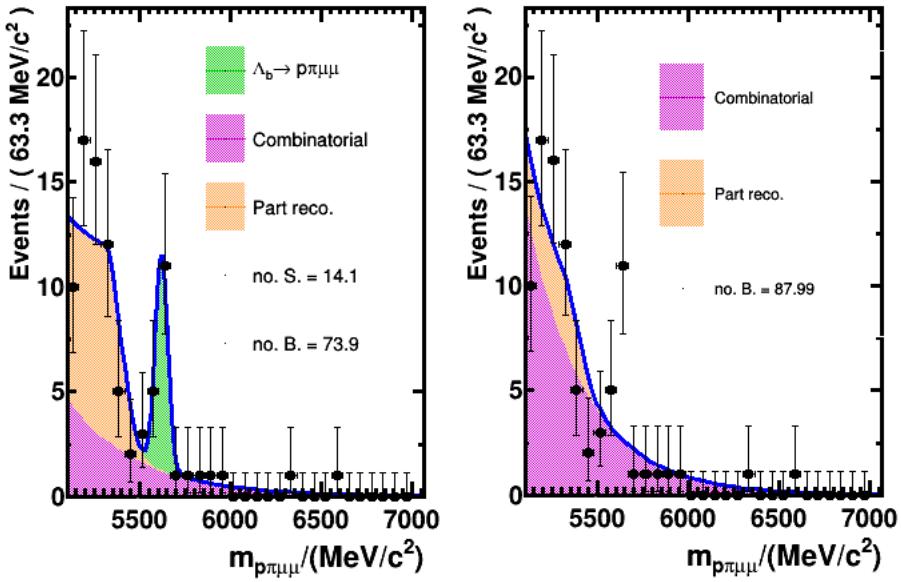
Lifting the constraints on the signal fit

To compare the number of signal events given by the fit when the fit is constrained using Gaussian constraints on the RooExpAndGauss part-reco component and when these constraints are lifted, the number of signal events for either case are taken from fits to 1000 pseudo experiments. The pseudo experiments are generated using the same procedure as outlined in [subsection 7.1.1](#), using the expected signal yield, n_{sig} , of $n_{sig} = 9$. Some examples of pseudo experiments fitted with no constraints are shown in [Figure 8.11](#).

The average number of signal events for the case where the fit is not constrained is 8.37 ± 0.17 , and for the case where the fit is constrained it is 8.98 ± 0.14 . The difference in the average number of signal events between a free and constrained fit is therefore 0.62, which, when quoted as a fraction of the average number of signal events from the nominal constrained fit, corresponds to an error of 6.9%. This is taken as a systematic uncertainty on the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ signal yield.



(a)



(b)

Figure 8.11: Examples of pseudo experiments showing background and signal fits, (left) and background only fits, fitted back with a `RooExpAndGauss` and exponential combined model with none of the fit components constrained, (right). The significance is 1.04 for the fits in (a) and 4.32 for the fits in (b). The no. S and no. B in the legends refer to the number of signal and background events respectively.

Addition of a $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ component to the fit

The expected number of $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ events in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ channel after selection is 1 ± 1 event and only $\sim 20\%$ of $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ events are expected to fall in the signal region. Given that only 0.2 events are expected, no fit component is added. The effect on the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ yield of adding or removing the $\Lambda_b^0 \rightarrow J/\psi pK^-$ component to the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ fit is 1.6%. This 1.6% is added as a systematic uncertainty on the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ signal yield.

Changing the background model

A fit to background-only data, where the blinded region is not included in the fit, is repeated twice. The fit is performed once with the nominal RooExpAndGauss plus exponential model, across the full fit range of $5100 - 7000 \text{ MeV}/c^2$, as indicated by the blue curve in Figure 8.12, and once with an exponential-only model with a restricted range of $5500 - 7000 \text{ MeV}/c^2$, indicated with the red curve in Figure 8.12. This restricted range is designed to avoid the inclusion of the part-reco shoulder at low mass.

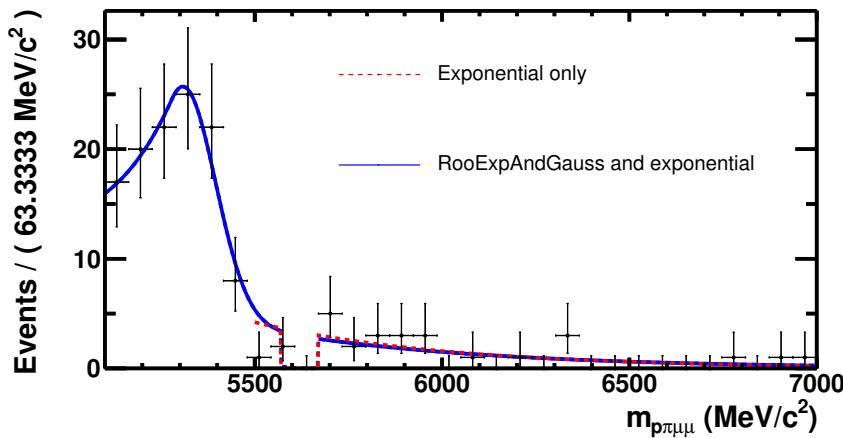


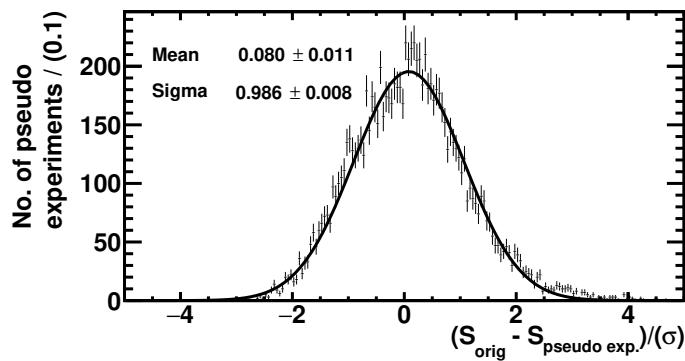
Figure 8.12: The background-only fit for the nominal background model over the full range of $5100 - 7000 \text{ MeV}/c^2$ and a background-only fit with a single exponential over the restricted range of $5500 - 7000 \text{ MeV}/c^2$.

The number of fitted events, combined with the relevant fit parameters, is used to

calculate the number of events in the blinded region. In the case of the nominal model, 1.89 events are expected in the signal region. For the exponential-only model it is 1.93 events. This correspond to a change of 2.1% on the nominal expected background level and a change of 0.4% on the number of signal events, assuming $n_{sig} = 9$. Given the negligible size of this error, no systematic uncertainty is assigned.

8.2.8 Verifying the fit coverage and bias for the signal channel

To verify the fit coverage and bias, the pull is calculated for the nominal fit model fitted using pseudo data, which are again generated with $n_{sig} = 9$. The pull is defined by the difference in the number of signal events given by the pseudo experiment fits and the original number of signal events the pseudo experiments were generated with, divided by the error on the number of signal events given by the fit. Due to the low statistics in the signal channel, the assumption that the error on the signal yield given by the fit is symmetric is not valid. Instead, the asymmetric error is used to calculate the pull, where the uncertainty taken from above the signal yield is used for pseudo experiments with a signal yield lower than the generated number, and vice versa for pseudo experiments with a signal yield lower than the generated number. The pull is calculated for 10000 pseudo data sets, and the fit to the distribution of the pulls is shown in [Figure 8.13](#).



[Figure 8.13](#): The fit to the distribution of the pseudo experiment's pulls for the signal channel fit.

The value of the mean of the pull distribution is 0.080 ± 0.011 and the value of the standard deviation is 0.986 ± 0.008 . The standard deviation therefore is consistent with unity at the level of 1.8σ . The mean however has a significant deviation from zero, meaning the fit has a slight bias towards lower signal yields. This bias corresponds to 8% of the error on the signal yield from the signal fit and this is taken as a systematic uncertainty. To gauge the size of this error, assuming that the error from the signal fit will be roughly of order $\sim 30\%$, 8% of 30% corresponds to an error of size 2.5%. This fit bias is not corrected for, instead it is taken into account with an additional systematic uncertainty.

8.2.9 Verifying the fit coverage and bias for the normalisation channel

To verify the fit coverage and bias for the normalisation channel, the same procedure as carried out in [subsection 8.2.8](#) is repeated for the case of the fit to $A_b^0 \rightarrow J/\psi p\pi^-$ data, again using 10000 pseudo experiments but taking $n_{sig} = 1017 \pm 41$ for the number of signal events. The fit to the distribution of the pulls is shown in [Figure 8.14](#). The mean from the fit to the pull distribution is consistent with zero at the 0.6σ level and the standard deviation from the fit is consistent with unity at the 0.3σ level.

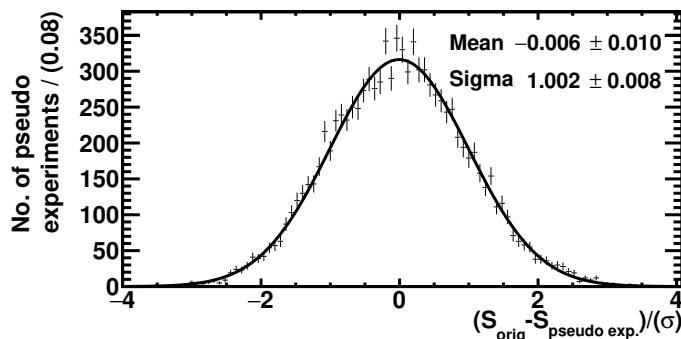


Figure 8.14: The fit to the distribution of the pseudo experiment's pulls for the normalisation channel fit.

8.2.10 Errors on the $\Lambda_b^0 \rightarrow p\pi^- J/\psi (\rightarrow \mu^+\mu^-)$ branching fraction measurement

The branching fraction for $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ is measured relative to that of the $\Lambda_b^0 \rightarrow J/\psi pK^-$ decay, by combining the measurements in Ref. [70] and Ref. [69]. The $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ branching fraction is given as

$$(2.61 \pm 0.09 \pm 0.13^{+0.47}_{-0.37}) \times 10^{-5}, \quad (8.3)$$

where the first uncertainty is statistical, the second is due to the systematic uncertainty on $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-)/\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi pK^-)$, and the third is due to the systematic uncertainty on $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi pK^-)$.

In turn, the $\Lambda_b^0 \rightarrow J/\psi pK^-$ branching fraction is measured relevant to $B^0 \rightarrow J/\psi K^{*0}$ in Ref. [70] and is given as

$$(3.17 \pm 0.04 \pm 0.07 \pm 0.34^{+0.45}_{-0.28}) \times 10^{-4}, \quad (8.4)$$

where the first uncertainty is statistical, the second is due to the systematic uncertainty, the third is due to the uncertainty on the branching fraction of $B^0 \rightarrow J/\psi K^{*0}$, and the fourth is due to the knowledge of the fraction of Λ_b^0 baryons produced within LHCb by the number of B^0 mesons produced, $f_{\Lambda_b^0}/f_d$. The error on $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi pK^-)$ is dominated by the uncertainty on $f_{\Lambda_b^0}/f_d$ and on $\mathcal{B}(B^0 \rightarrow J/\psi K^{*0})$, which in turn dominates the error on the $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ branching fraction.

The total error on $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-)$ is $^{+19\%}_{-15\%}$, making it the dominant systematic on the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction.

8.2.11 Summary of systematic uncertainties

A summary of the systematic uncertainties can be found below in Table 8.3

Although the resultant 16% systematic error, when disregarding the systematic uncertainty on $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-)$, is large, the statistical error is of order 30% and thus the

Error source	Error Assigned/%
Error on $\mathcal{B} (\Lambda_b^0 \rightarrow J/\psi p\pi^-)$	$^{+19}_{-15}$
Error on $\mathcal{B} (J/\psi \rightarrow \mu^+ \mu^-)$	0.5
Error on efficiency ratio	
q^2 distribution	7.9
$p\pi$ weighting	7.7
Choice of BDT efficiency proxy	5.6
Error due to simulation statistics	4.4
Trigger	3.4
Error due to the removal of last q^2 bin	2.5
Kinematic weighting	1.3
PID	1.0
Error on normalisation and signal yields	
Effect of the fit shape	6.9
Error on number of $\Lambda_b^0 \rightarrow J/\psi p\pi^-$ events	4.0
Fit bias	2.2
Effect of the $\Lambda_b^0 \rightarrow pK^- \mu^+ \mu^-$ component	1.6
Without $\mathcal{B} (\Lambda_b^0 \rightarrow J/\psi p\pi^-)$ error	16.1
Total	$^{+25}_{-22}$

Table 8.3: The different sources of systematic uncertainty and the total systematic uncertainty assigned.

statistical error is still dominant. The dominant systematic is from the uncertainty of the phase space simulation's ability to successfully model the kinematics of the decay, as reflected in the 7.9% and 7.7% uncertainties on the effect of the dimuon and dihadron mass spectrum modelling on the efficiency respectively.

Chapter 9

Conclusions and outlook

This thesis presents the first observation of a $b \rightarrow d$ transition in the baryon sector, via the decay $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$. This comes 29 years after the first observation of a $b \rightarrow d$ transition in the meson sector by the ARGUS experiment which measured $B^0 - \overline{B^0}$ mixing in 1987 [79]. The observation of $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ opens up the possibility of using baryonic decays to investigate the modest tension observed [15] between the value of the CKM element ratio $|V_{td}/V_{ts}|$ measured via either tree- or loop-level processes.

In the future, by combining the measurement of $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)$ with that of $\mathcal{B}(\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-)$ ¹, the value of $|V_{td}/V_{ts}|$ will be extracted using

$$\frac{\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)}{\mathcal{B}(\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-)} = \left| \frac{V_{td}}{V_{ts}} \right|^2 f^2, \quad (9.1)$$

where f^2 is the ratio of the relevant form factors and Wilson coefficients, integrated over the relevant phase space. In order to extract the value of $|V_{td}/V_{ts}|$, the value of f must be calculated by theorists. New measurements of $|V_{td}/V_{ts}|$ via different channels are important in resolving the modest tension observed between tree and loop measurements of $|V_{td}/V_{ts}|$ as the value calculated from neutral B mixing is currently theory-limited.

The $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ decay was observed with a significance of 5.5σ using 3 fb^{-1} of

¹The $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ branching fraction is also being measured by LHCb but is not a public result at the time of writing.

LHCb data. The value of the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction is found to be

$$\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-) = (6.9 \pm 1.9 \pm 1.1^{+1.3}_{-1.0}) \times 10^{-8}, \quad (9.2)$$

where the first error is the statistical uncertainty, the second is the systematic uncertainty and the third is the uncertainty on $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-)$. The fitted $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ yield is 22 ± 6 .

There is no theory prediction for the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction. However, assuming that the $\Lambda_b^0 \rightarrow p\pi^-J/\psi (\rightarrow\mu^+\mu^-)$ branching fraction is ~ 100 times larger than the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ branching fraction, as discussed in section 5.1, a value of $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)$ of order 10^{-8} is expected.

This thesis also presents a novel technique to measure the downstream tracking efficiency of the process $K_s^0 \rightarrow \pi^+\pi^-$ in both data and simulation. This technique has been important for analyses which feature K_s^0 decays to help calibrate the difference between simulation and data. Given this, the same technique can also be applied to decays from other long-lived particles such as Λ^0 baryons.

Bibliography

- [1] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, *Tenth-Order QED Contribution to the Electron $g-2$ and an Improved Value of the Fine Structure Constant*, *Phys. Rev. Lett.* **109** (2012) 111807.
- [2] D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse, *Cavity control of a single-electron quantum cyclotron: Measuring the electron magnetic moment*, *Phys. Rev. A* **83** (2011) 052122.
- [3] M. Roos, *Dark Matter: The evidence from astronomy, astrophysics and cosmology*, ArXiv e-prints (2010) [arXiv:1001.0316](https://arxiv.org/abs/1001.0316).
- [4] O. Hartmann *et al.*, *Proceedings of the International Conference on Exotic Atoms and Related Topics, EXA 2011* (Vienna 5-9 September 2011).
- [5] M. Schumann, *Dark Matter 2014*, *EPJ Web of Conferences* **96** (2015) 01027.
- [6] G. Isidori, Y. Nir, and G. Perez, *Flavor Physics Constraints for Physics Beyond the Standard Model*, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 355, [arXiv:1002.0900](https://arxiv.org/abs/1002.0900).
- [7] K.A. Olive et al. (Particle Data Group), *Review of particle physics*, *Chin. Phys. C*, **38**, 090001 (2014) and 2015 update.
- [8] Y. Ahn, H.-Y. Cheng, and S. Oh, *Wolfenstein parametrization at higher order: Seeming discrepancies and their resolution*, *Physics Letters B* **703** (2011), no. 5 571 .
- [9] The LHCb collaboration, Taken from the Lectures on Particle Physics series, Nikhef, *The Standard Model Higgs Bosons*, 2013-2014. P.P. notes, Nikhef, Acquired 2016.

- [10] W. Detmold, C.-J. D. Lin, S. Meinel, and M. Wingate, *$\Lambda_b \rightarrow \Lambda \ell^+ \ell^-$ form factors and differential branching fraction from lattice QCD*, *Phys. Rev. D* **87** (2013) 074502.
- [11] S. Dürr *et al.*, *Ab Initio Determination of Light Hadron Masses*, *Science* **322** (2008), no. 5905 1224.
- [12] UTfit, M. Bona *et al.*, *Model-independent constraints on $\Delta F = 2$ operators and the scale of new physics*, *JHEP* **03** (2008) 049, [arXiv:0707.0636](https://arxiv.org/abs/0707.0636).
- [13] CKMfitter Group, J. Charles *et al.*, *Eur. Phys. J. C41*, 1-131 (2005) [[hep-ph/0406184](https://arxiv.org/abs/hep-ph/0406184)], updated results and plots available at: <http://ckmfitter.in2p3.fr>.
- [14] The LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the $B_s^0 - \bar{B}_s^0$ oscillation frequency Δm_s in the decay $B_s^0 \rightarrow D_s^+ \pi^-$* , *New J. Phys.* **15** (2013) 053021, [arXiv:1304.4741](https://arxiv.org/abs/1304.4741).
- [15] Fermilab Lattice, MILC, A. Bazavov *et al.*, *$B_{(s)}^0$ -mixing matrix elements from lattice QCD for the Standard Model and beyond*, , [arXiv:1602.03560v1](https://arxiv.org/abs/1602.03560v1).
- [16] The LHCb collaboration, R. Aaij *et al.*, *First measurement of the differential branching fraction and CP asymmetry of the $B^\pm \rightarrow \pi^\pm \mu^+ \mu^-$ decay*, *JHEP* **10** (2015) 034, [arXiv:1509.0041](https://arxiv.org/abs/1509.0041).
- [17] The LHCb collaboration, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \rightarrow K^{(*)} \mu^+ \mu^-$ decays*, *JHEP* **06** (2014) 133, [arXiv:1403.8044](https://arxiv.org/abs/1403.8044).
- [18] A. Ali, A. Ya. Parkhomenko, and A. V. Rusov, *Precise Calculation of the Dilepton Invariant-Mass Spectrum and the Decay Rate in $B^\pm \rightarrow \pi^\pm \mu^+ \mu^-$ in the SM*, *Phys. Rev. D* **89** (2014), no. 9 094021, [arXiv:1312.2523](https://arxiv.org/abs/1312.2523).
- [19] M. Bartsch, M. Beylich, G. Buchalla, and D.-N. Gao, *Precision flavour physics with $B \rightarrow K \nu \bar{\nu}$ and $B \rightarrow K \ell^+ \ell^-$* , *JHEP* **2009** (2009), no. 11 011.

- [20] C. Hambrock, A. Khodjamirian, and A. Rusov, *Hadronic effects and observables in $B \rightarrow \pi\ell^+\ell^-$ decay at large recoil*, Phys. Rev. **D92** (2015), no. 7 074020, [arXiv:1506.0776](#).
- [21] Z.-H. Li, Z.-G. Si, Y. Wang, and N. Zhu, *$B \rightarrow \pi\ell^+\ell^-$ decays revisited in the standard model*, [arXiv:1411.0466](#).
- [22] I. Sentitemsu Imsong, A. Khodjamirian, T. Mannel, and D. van Dyk, *Extrapolation and unitarity bounds for the $B \rightarrow \pi$ form factor*, JHEP **02** (2015) 126, [arXiv:1409.7816](#).
- [23] P. Ball and R. Zwicky, *New results on $B \rightarrow \pi, K, \eta$ decay form factors from light-cone sum rules*, Phys. Rev. **D71** (2005) 014015, [arXiv:hep-ph/0406232](#).
- [24] W. Detmold and S. Meinel, *$\Lambda_b \rightarrow \Lambda\ell^+\ell^-$ form factors, differential branching fraction, and angular observables from lattice QCD with relativistic b quarks*, Phys. Rev. **D93** (2016), no. 7 074501, [arXiv:1602.0139](#).
- [25] The LHCb collaboration, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.
- [26] The LHCb VELO group, R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) 09007, [arXiv:1405.7808](#).
- [27] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, Tech. Rep. LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, May, 2014.
- [28] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, J. Inst. , Volume 8 (2013), [arXiv:1211.3055v2](#).
- [29] The LHCb collaboration, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30** (2014) 1530022. 82 p.

- [30] LHCb HLT project, J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, *J. Phys. Conf. Ser.* **513** (2014) 012001, [arXiv:1310.8544](https://arxiv.org/abs/1310.8544).
- [31] B. P. Roe *et al.*, *Boosted decision trees, an alternative to artificial neural networks*, *Nucl. Instrum. Meth.* **A543** (2005), no. 2-3 577, [arXiv:physics/0408124](https://arxiv.org/abs/physics/0408124).
- [32] V. Gligorov *et al.*, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, 2012. [arXiv:1210.6861](https://arxiv.org/abs/1210.6861) 2013 *JINST* 8 P02013.
- [33] V. V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, Tech. Rep. LHCb-PUB-2011-016. CERN-LHCb-PUB-2011-016. LHCb-INT-2011-030, CERN, Geneva, Sep, 2011. LHCb-INT-2011-030.
- [34] A. Puig, *The LHCb trigger in 2011 and 2012*, Tech. Rep. LHCb-PUB-2014-046. CERN-LHCb-PUB-2014-046, CERN, Geneva, Nov, 2014.
- [35] Sam Cunliffe, *The scalar component of $B^0 \rightarrow K^+ \pi^- \mu^+ \mu^-$* , 2015. PhD Thesis, Imperial College London.
- [36] *Image of the Outer Tracker taken from the LHCb public web page*, Acquired 2016.
- [37] J. Tilburg, *Track reconstruction in LHCb*, 2013. **LHCb UK student seminar 29 Oct. 2016**.
- [38] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and Performance Studies*, Tech. Rep. LHCb-2007-015. CERN-LHCb-2007-015, CERN, Geneva, May, 2007.
- [39] O. Callot and M. Schiller, *PatSeeding: A Standalone Track Reconstruction Algorithm*, Tech. Rep. LHCb-2008-042. CERN-LHCb-2008-042, CERN, Geneva, Aug, 2008.
- [40] The LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, 2013. [arXiv:1408.1251](https://arxiv.org/abs/1408.1251) *JINST* 10 (2015) P02007.

- [41] O. Callot, *Downstream Pattern Recognition*, Tech. Rep. LHCb-2007-026. CERN-LHCb-2007-026, CERN, Geneva, Mar, 2007.
- [42] *Image of the rich detector taken from the LHCb public web page*, Acquired 2016.
- [43] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, *Eur. Phys. J.* **C73** (2013) 2431, [arXiv:1211.6759](https://arxiv.org/abs/1211.6759).
- [44] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, *Nucl. Instrum. Meth.* **A555** (2005) 356, [arXiv:physics/0402083](https://arxiv.org/abs/physics/0402083).
- [45] A. A. Alves *et al.*, *Performance of the LHCb muon system*, *J. Inst.* **8** (2013), no. 02 P02022.
- [46] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, *JINST* **8** (2013) P10020, [arXiv:1306.0249](https://arxiv.org/abs/1306.0249).
- [47] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [arXiv:hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175).
- [48] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, [arXiv:0710.3820](https://arxiv.org/abs/0710.3820).
- [49] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152.
- [50] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, *Eur. Phys. J.* **C45** (2006) 97, [arXiv:hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026).
- [51] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, *IEEE Trans. Nucl. Sci.* **53** (2006) 270.
- [52] The LHCb collaboration, R. Aaij *et al.*, *Prompt K_S^0 production in pp collisions at $\sqrt{s} = 0.9 \text{ TeV}$* , *Phys. Lett.* **B693** (2010) 69, [arXiv:1008.3105](https://arxiv.org/abs/1008.3105).

- [53] The LHCb collaboration, R. Aaij *et al.*, *Measurement of V^0 production ratios in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV*, **JHEP** **08** (2011) 034, [arXiv:1107.0882](https://arxiv.org/abs/1107.0882).
- [54] The LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \rightarrow J/\psi K_S^0$ branching fraction*, **Phys. Lett.** **B713** (2012) 172, [arXiv:1205.0934](https://arxiv.org/abs/1205.0934).
- [55] The LHCb collaboration, R. Aaij *et al.*, *Measurement of the isospin asymmetry in $B \rightarrow K^{(*)}\mu^+\mu^-$ decays*, **JHEP** **07** (2012) 133, [arXiv:1205.3422](https://arxiv.org/abs/1205.3422).
- [56] The LHCb collaboration, R. Aaij *et al.*, *Study of D_{sJ} decays to $D^+K_S^0$ and D^0K^+ final states in pp collisions*, **JHEP** **10** (2012) 151, [arXiv:1207.6016](https://arxiv.org/abs/1207.6016).
- [57] The LHCb collaboration, R. Aaij *et al.*, *Search for the rare decay $K_S^0 \rightarrow \mu^+\mu^-$* , **JHEP** **01** (2013) 090, [arXiv:1209.4029](https://arxiv.org/abs/1209.4029).
- [58] The LHCb collaboration, R. Aaij *et al.*, *A model-independent Dalitz plot analysis of $B^\pm \rightarrow DK^\pm$ with $D \rightarrow K_S^0 h^+h^-$ ($h = \pi, K$) decays and constraints on the CKM angle γ* , **Phys. Lett.** **B718** (2012) 43, [arXiv:1209.5869](https://arxiv.org/abs/1209.5869).
- [59] The LHCb collaboration, R. Aaij *et al.*, *Measurement of the time-dependent CP asymmetry in $B^0 \rightarrow J/\psi K_S^0$ decays*, **Phys. Lett.** **B721** (2013) 24, [arXiv:1211.6093](https://arxiv.org/abs/1211.6093).
- [60] The LHCb collaboration, R. Aaij *et al.*, *Searches for CP violation in the $D^+ \rightarrow \phi\pi^+$ and $D_s^+ \rightarrow K_S^0\pi^+$ decays*, **JHEP** **06** (2013) 112, [arXiv:1303.4906](https://arxiv.org/abs/1303.4906).
- [61] The LHCb collaboration, R. Aaij *et al.*, *Measurements of the $\Lambda_b^0 \rightarrow J/\psi\Lambda$ decay amplitudes and the Λ_b^0 polarisation in pp collisions at $\sqrt{s} = 7$ TeV*, **Phys. Lett.** **B724** (2013) 27, [arXiv:1302.5578](https://arxiv.org/abs/1302.5578).
- [62] The LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \rightarrow J/\psi K_S^0$ effective lifetime*, **Nucl. Phys.** **B873** (2013) 275, [arXiv:1304.4500](https://arxiv.org/abs/1304.4500).
- [63] The LHCb collaboration, R. Aaij *et al.*, *Measurement of the differential branching fraction of the decay $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$* , **Phys. Lett.** **B725** (2013) 25, [arXiv:1306.2577](https://arxiv.org/abs/1306.2577).

- [64] The LHCb collaboration, R. Aaij *et al.*, *Branching fraction ratio and CP asymmetry measurements of the $B^+ \rightarrow K^0\pi^+$ and $B^+ \rightarrow \bar{K}^0K^+$ decays*, Phys. Lett. **B726** (2013) 646, [arXiv:1308.1277](#).
- [65] The LHCb collaboration, R. Aaij *et al.*, *Study of $B_{(s)}^0 \rightarrow K_S^0 h^+ h^-$ decays with first observation of $B_s^0 \rightarrow K_S^0 K^\pm \pi^\mp$ and $B_s^0 \rightarrow K_S^0 \pi^+ \pi^-$* , JHEP **10** (2013) 143, [arXiv:1307.7648](#).
- [66] P. H. Owen, *Measurement of branching fractions, isospin asymmetries and angular observables in exclusive electroweak penguin decays*, 2014. CERN thesis 2014-057.
- [67] The LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015), no. 02 P02007, [arXiv:1408.1251](#).
- [68] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.
- [69] The LHCb collaboration, R. Aaij *et al.*, *Observation of the $\Lambda_b^0 \rightarrow J/\psi p \pi^-$ decay*, JHEP **07** (2014) 103, [arXiv:1406.0755](#).
- [70] The LHCb collaboration, R. Aaij *et al.*, *Study of the production of Λ_b^0 and \bar{B}^0 hadrons in pp collisions and first measurement of the $\Lambda_b^0 \rightarrow J/\psi p K^-$ branching fraction*, Chin. Phys. **C40** (2016), no. 1 011001, [arXiv:1509.0029](#).
- [71] I. Belyaev *et al.*, *Handling of the generation of primary events in GAUSS, the LHCb simulation framework*, Nuclear Science Symposium Conference Record (NSS/MIC) IEEE (2010) 1155.
- [72] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.
- [73] M. Clemencic *et al.*, *The LHCb simulation application, GAUSS: design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

- [74] LHCb, R. Aaij *et al.*, *Measurement of matter-antimatter differences in beauty baryon decays*, [arXiv:1609.0521](https://arxiv.org/abs/1609.0521).
- [75] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- [76] The LHCb collaboration, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, *Nature Phys.* **11** (2015) 743, [arXiv:1504.0156](https://arxiv.org/abs/1504.0156).
- [77] A. Blum *et al.*, *Beating the hold-out: Bounds for k-fold and progressive cross-validation*, in *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, , COLT 1999 (New York USA) pp. 203-208.
- [78] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, [arXiv:physics/0308063](https://arxiv.org/abs/physics/0308063).
- [79] The ARGUS collaboration, H. Albrecht *et al.*, *Observation of $B^0 - \overline{B}{}^0$ mixing*, *Physics Letters B* (1987), *Physics Letters B Volume 192 Issue 2, Pages 245-252*.

Appendices

Appendix A

Boosted Decision Trees

Many rare decay analyses make extensive use of BDTs and they are important in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis. Firstly, the concept of a decision tree is introduced followed by a brief explanation of boosted decision trees.

A decision tree, in the context of data mining, is a supervised machine learning method which allows for the prediction of the value of a target variable based on several input variables. In particle physics, the purpose of the decision tree is to classify an event as being either signal or background, based on the event's input variables. The input variables, $\{x_i\}$, are various physics parameters. Each cut point in the tree is referred to as a node and the final nodes are referred to as leaves. A very simple example is shown in [Figure A.1](#). The purity, P , of a leaf refers to the fraction of the weight of a leaf due to signal events, e.g. if a leaf had 20 signal events and 15 background events it would have a purity of 0.75. If a leaf has a purity larger than 0.5 it is deemed to correspond to signal and if lower, to background.

A decision tree is constructed by a process called training. For this, samples of known signal and background events are used. These samples could be either simulation or data. For each x_i the best dividing point is decided, that is, the cut that gives the best separation between signal and background. This optimum point is decided by using the Gini index defined as

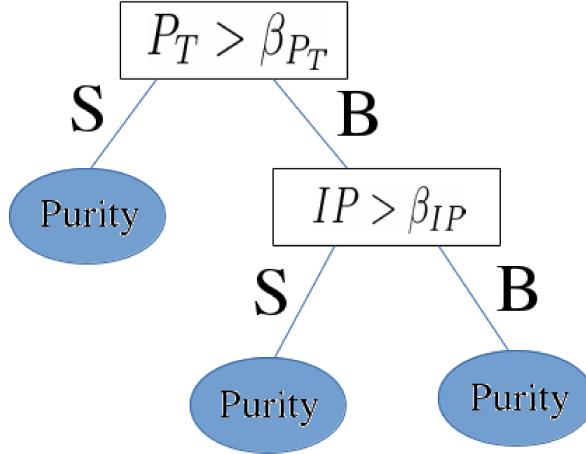


Figure A.1: An example decision tree. The S and B stand for ‘Signal-like’ and ‘Background-like’. The β_i variables refer to the cut values chosen by the machine learning algorithm after the tree has been trained on signal and background samples. The blue ovals represent final nodes called leafs, which each leaf having an associated purity, i.e. the fraction of the weight of a leaf due to signal events.

$$Gini = \sum_{i=1}^n W_i P(1 - P), \quad (\text{A.1})$$

where W_i is weight of the i^{th} event, which would generally be unity for the case of a non-boosted decision tree. The cutting point is then found by maximising the separation, Δ , between the Gini index of the parent node and the combined Gini index of the child nodes, as given in [Equation A.2](#)

$$\Delta = Gini_{parent} - Gini_{child_1} - Gini_{child_2}. \quad (\text{A.2})$$

The depth of a tree (the maximum number of cuts or nodes) is normally a number specified before the training begins.

Boosting a decision tree involves training many trees ($\mathcal{O} \sim 1000$) and giving misclassified events a higher weight. A misclassified event is defined as a known signal event being

placed on a background leaf and vice versa. By giving the events which are difficult to classify more weight, the next tree to be trained will effectively have to work harder in order to classify events correctly.

The total score on an event is deduced by following an event through from tree to tree and, for the algorithms used in this thesis, is simply given by the weighted sum of the scores over the individual trees.

Data sets are split into two (or more) sub samples, where one half is used for training the tree and the other is used for testing the tree, and the distributions of the event scores (the BDT output) for training and testing samples are compared for signal and background. Cases where the training sample performs better than the testing sample are referred to as over-trained trees, which is often due to the BDT becoming sensitive to the statistical fluctuations of the training sample.

The distribution of events scores for a given dataset can then be cut on in order to increase the fraction of signal events.

Appendix B

The *sPlot* technique

The *sPlot* technique is used extensively throughout this thesis. It is used in cases when there is a merged dataset which consists of data from different sources of data species, namely background and signal. These datasets are assumed to have two different sets of variables associated with the events they contain. Discriminating variables are those whose distributions are known for background and signal. Control variables are those whose distributions are unknown, or are assumed to be unknown.

The *sPlot* technique allows the distribution of the control variables for each data species to be deduced by using the species discriminating variable. This method relies on the assumption that there is no correlation between the discriminating variable and the control variable. The discriminating variable used in this thesis is always the mass distribution. The full mathematical description of the *sPlot* technique can be found in Ref [44], the key points are outlined here.

An unbinned extended maximum likelihood analysis of a data sample of several species is considered. The log-likelihood is expressed as

$$\mathcal{L} = \sum_{e=1}^N \left\{ \ln \sum_{i=1}^{N_s} N_i f_i(y_e) \right\} - \sum_{i=1}^{N_s} N_i, \quad (\text{B.1})$$

where N is the total number of events considered, N_s is the number of species of event (i.e. two - background and signal), N_i is the average number of expected events for the i^{th}

species, y represents the set of discriminating variables, $f_i(y_e)$ is the value of the Probability Density Function (PDF) of y for event e for the i^{th} species and the control variable, x , does not appear in the expression of \mathcal{L} by definition.

For the simple (and not particularly practical) case of the control variable x being a function of y , i.e. completely correlated, one could naively assume that the probability of a given event of the discriminating variable y being of the species n would be given by

$$\mathcal{P}_n(y_e) = \frac{N_n f_n(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}. \quad (\text{B.2})$$

The distribution for a control variable x for the n^{th} species, $M_n(x)$, can be deduced by histogramming in x and applying $\mathcal{P}_n(y_e)$ as a weight to event e . In this scenario the probability, $\mathcal{P}_n(y_e)$, would run from 0 to 1.

In the case considered in this thesis, where x is entirely uncorrelated with y , it can be shown that $\mathcal{P}_n(y_e)$ can be written as

$$\mathcal{P}_n(y_e) = \frac{\sum_{j=1}^{N_s} V_{nj} f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}, \quad (\text{B.3})$$

where V_{nj} is the covariance matrix between the species n and the j^{th} species. The inverse of this covariance matrix is given by the second derivative of $-\mathcal{L}$ in [Equation B.1](#).

The quantity in [Equation B.3](#) is denoted as the sWeight. In this thesis the species, n , in [Equation B.3](#) is always the signal. Because of the presence of the covariant derivative the sWeight of an event can be both positive and negative. The more negative an event is, the more likely it is to be background and vice versa for positive sWeights. The signal distribution for the control variable x , $M_s(x)$, can again be deduced by histogramming events in x , applying the sWeight to each event.