

An angular analysis of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays at the LHCb experiment.

by

Christopher Everett

Thesis

Submitted to the University of Warwick

For the degree of

Masters by Research

Department of Physics

September 2016



WARWICK
THE UNIVERSITY OF WARWICK



Abstract

This work presents an angular analysis of the suppressed $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay, performed using proton-proton collision data collected by the LHCb experiment during 2011 and 2012. The distribution of moments describing the decay with a non-zero transverse Λ_b^0 baryon polarisation production are extracted for the first time, and used to determine the 5-dimensional angular distribution in the 15 to 20 GeV^2/c^4 dimuon invariant mass squared interval for the first time. The forward-backward asymmetry of the Λ baryon and $\mu^+ \mu^-$ produced from the Λ_b^0 baryon (A_{FB}^{lh}) is measured, along with the previously published f_L , A_{FB}^l and A_{FB}^h . These values are found to be $f_L = 0.546 \pm 0.132$, $A_{\text{FB}}^l = -0.004 \pm 0.077$, $A_{\text{FB}}^h = -0.312 \pm 0.150$ and $A_{\text{FB}}^{lh} = -0.007 \pm 0.071$, where uncertainties are statistical, and account for correlation between the moments used to determine them.

Acknowledgements

I'm going to go through thing chronologically, so I'd like to start by thanking Tim Gershon for offering me a place in the LHCb group in the first place. I came in with no prior experience with computer programming, and a limited knowledge of particle physics, but I was given the chance to prove myself here. I feel like I've gained invaluable experience in working both with computer languages and project groups, something that no doubt serve me well in the future (whatever that may contain).

Next, my supervisors Michal Kreps and Tom Blake. Both have been more than happy to help me overcome the roadblocks I've hit during my project, even those that I really should have known how to breakthrough myself. They've kept me on track, helped refine some of the more elaborate coding ideas I've had, and more importantly got me to the end of the project in one piece. Michal has helped keep me focused on the task at hand, and always provided clear steps on the progression of the work, while Tom was able to explain the more challenging concepts in terms I could easily understand. To them I give my heartfelt thanks, as while they could have completed this analysis without me, I'd likely not have finished it without them.

For their contributions to the project, I'd like to thank both Nigel Watson and Georgios Chatzikonstantinidis of the University of Birmingham. The amount of work associated to this project was much larger than I would have originally presumed, and much more that I'd have been able to complete successfully on my own. Their contributions were vital to the successful completion of the analysis, and I sincerely thank them for this.

I'd like to also thank Tom Latham for his on-the-ground support with some of the more interesting computer quirks I've come across over the year. His input both in the office and teaching during programming courses has helped this project progress smoothly. The brief cricketing banter, and not-so-brief discussions on the state of the country also helped take the edge off the more frustrating days.

I should also thank Ron, Cayo, Chris and Andy for keeping me sane outside of the office, for appreciating the joys of free coffee with me, and often providing completely surreal conversation to enjoy/tolerate. Also, for their assistance and humour, I'd like to thank everyone in the LHCb office at Warwick over the last academic year.

Last, but of course not least, I should thank my mother for helping bankroll this whole misguided venture. She's helped keep me in food and shelter for longer than either of us care to remember, so hopefully this qualification will go some way to helping me be able to repay her, and to get out of her hair.

Declaration

This thesis contains primarily my own work, however, as the analysis was carried out by a group of LHCb physicists from the Universities of Warwick and Birmingham, for completeness I also include work produced by other members. The weighting functions and differential decay rate equations in Section 2.2, the work presented in Section 5.8 on efficiencies, and the simulation validation studies in Sections 6.2 and 6.4 were performed by the group, and are therefore not entirely my own work. Work on the selection requirements applied to the dataset, training and optimisation of the neural network, extraction of event yields, and subtraction of background events in the moment calculation is my own. Additionally, the final results published here were extracted by myself. Any and all programming packages used during the analysis are mentioned when used. This work has not been submitted in any form to any other university for any other qualification.

Christopher Everett

Contents

1	Physics of Heavy Flavour Processes	12
1.1	The Standard Model of Particle Physics	12
1.2	Flavour Changing Neutral Currents and the GIM mechanism	14
1.3	The Cabibbo-Kobayashi-Maskawa Matrix	15
1.4	Beyond the Standard Model	17
2	The Suppressed $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ Decay	20
2.1	The $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ Decay, and the q^2 Spectrum	20
2.2	Distribution of Angles and Definition of Moments	25
2.3	Analysis Strategy and Goals	29
3	The LHC and the LHCb Experiment	31
3.1	The LHC	31
3.2	The LHCb Experiment	32
3.3	The Bending Magnet	34
3.4	The Tracking System	36
3.5	Ring Imaging Cherenkov Detector	39
3.6	Calorimeter System	42
3.7	Muon Detectors	44
3.8	Trigger	45
4	Analysis Software	47
4.1	Introduction to NeuroBayes	47
4.2	ROOT, RooFit, and <i>sPlot</i>	48
5	Analysis Method	50
5.1	Dataset, Stripping and Trigger Requirements	50
5.2	Preselection Requirements and Neural Network Training	53
5.3	Signal Modelling	59
5.4	Background Sources	61
5.5	Fits to $\Lambda_b^0 \rightarrow J/\psi \Lambda$ Data	63
5.6	Optimisation of Neural Network Selection	66
5.7	Fits to $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ Data	69
5.8	Signal Efficiency Correction	75
5.9	Background Removal	79
5.10	Determination of Uncertainties	81
6	Simulation Studies	82
6.1	Generation of Events	82
6.2	Validation with Simulated Events	83
6.3	Reconfirmation of Neural Network Optimisation	87
6.4	Validation with $\Lambda_b^0 \rightarrow J/\psi \Lambda$ Events	89

7	Results	95
7.1	Distributions of Moments and Angles, and Measurement of Angular Ob- servables	95
7.2	Determination of Systematic Uncertainties	101
8	Conclusions and the Future	103
8.1	Discussion of Analysis Method	103
8.2	Discussion of Angular Distributions and Observables	104
8.3	The Future and Final Remarks	106
	Appendices	107
A	Definition of Variables used for Selection and Training	107
B	Comparison of Signal and Background Events for Neural Network Training Variables	108
C	Distributions of Pull Plots used in Moment Validation	117
9	Figure References	121
10	Literature References	121

List of Figures

1	Flavour changing vertices in $\Lambda_b^0 \rightarrow J/\psi \Lambda$ (top) and $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ (bottom) decays. In the top diagram, the bottom quark decays into a charm in a flavour changing charged current process, which then undergoes a second change to the strange. The bottom quark cannot do this in $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ however, and must decay into an intermediate virtual quark state, before this becomes the strange quark.	16
2	Recent NASA results on the predicted make up of the mass-energy in the universe, measured with WMAP. Figure and values published by the NASA/WMAP Science Team.	18
3	A BSM $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay diagram, where Y is a new neutral boson mediating the flavour change.	18
4	Feynman diagrams of contributing Standard Model $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays. Each proceeds via an intermediate quark of the opposite type to the bottom (u, c, t), with the particles mediating this flavour change then producing the dimuon.	21
5	A sketch of the distribution of the q^2 spectrum for the decay $B^0 \rightarrow K^* \mu^+ \mu^-$. The narrow $c\bar{c}$ resonances are the J/ψ meson and $\psi(2S)$ resonances respectively, while the pole at low q^2 is symptomatic of dimuons produced from photons. Figure is taken from [6]	22
6	Branching fractions of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ in different q^2 bins. The black data points are data, while the purple are SM prediction. Values from [28], figure taken from [31].	23
7	A Feynman diagram of the $\Lambda_b^0 \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) \Lambda$ decay used to normalise the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ differential branching fraction.	24
8	Definition of angular variables measurable in the $\Lambda_b^0 \rightarrow \Lambda (\rightarrow p \pi^-) \mu^+ \mu^-$ decay, assuming a non-zero Λ_b^0 baryon transverse production polarisation. Definitions of the angles and axes displayed are given in the text.	25
9	Layout of the experiments at the Large Hadron Collider. Figure taken from the CERN Document Server.	31
10	A schematic diagram of the LHCb experiment. Figure taken from [43] . . .	33
11	Definition of the co-ordinate system used in the LHCb. The solid black arrows are the proton beams, and the red circle is the interaction point. . .	33
12	Schematic view of the LHCb Bending Magnet. Units are in millimetres. Figure taken from [43]	34
13	Magnetic field strength as a function of distance travelled downstream. The two lines represent the field strength in both the up and down magnet configurations. Figure taken from [43]	35
14	Cross section of the VELO in its closed state. Figure taken from [43] . . .	37

15	Definition of various track types as they pass through the tracking systems. Track types are defined by which parts of the experiment they are recorded in. For this analysis we split our dataset into Long and Downstream tracks. This is done as the two track types have different resolutions, with Long tracks better measured than Downstream. Figure taken from [43]	37
16	The cyan unit is the OT, while the purple is the whole of the ST. The TT is on the left, and the IT is on the right. Figure taken from [43]	38
17	Left - RICH-1 layout. Right - RICH-2 layout. Figures taken from [43]	40
18	A schematic view and photograph of the Hybrid Photon Detectors used in RICH-1 and RICH-2. Figure taken from [43]	40
19	Expected Cherenkov angles (θ_c) for each radiator used in RICH by particle species. The use of aerogel was only present during the 2011 data taking. Figure taken from [43]	41
20	Segmentation of the SPD/PS and ECAL (left) and HCAL (right). The dimensions presented in the left figure are for the ECAL. Quarter of each face is shown. Figures taken from [43]	42
21	Schematic view of a unit used in the HCAL. Figure taken from [43]	43
22	Schematic view of the LHCb muon system. Figure taken from [43]	44
23	Nodes involved in a typical neural network. In this analysis, there are many more input nodes, corresponding to the training variables, with a single output. It is the job of the hidden layer to transform the inputs in to something that may be used by the output. Figure taken from Wikimedia Commons.	48
24	A $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay. Solid arrows are particle momenta vectors, and dashed lines are those vectors traced back. The red circle is the interaction point. The double ended arrows represent the distance measured to calculate impact parameter χ^2 values.	52
25	Left - Presence of the peak in the Λ baryon decay vertex in the real dataset. Right - positions of Λ baryon decays in the xy plane ± 20 mm about the peak at $z = 2350$ mm.	53
26	Signal (green) and background (red) event yields as a function of neural network output.	56
27	Purity by neural network output. The black points represent the final training, while the grey are those produced by preprocessing.	57
28	Left - Average neural network value from the signal region in simulated events. Right - Average neural network value in the background region from real data.	57
29	Variable correlation matrix from NN training. The white regions arise from the separation of LL and DD tracks, as an event is either one or the other, no correlation exists between these variables.	58

30	Separation of the Crystal Ball shapes needed for the total fit to 15 to 20 GeV^2/c^4 q^2 interval $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ simulation. The solid green line is the total Crystal Ball shape, while the dashed red and blue lines represent the individual CB shapes from which it is formed. The parameters describing this fit are given in Table 13 in Section 5.7.	61
31	Left - an exponential function (red dashed line) modelling the combinatorial background in the side band of the data sample after preselection (black points) used to train background in the neural network. Right - fit (black dashed line) to extract parameters for events misreconstructed as $B^0 \rightarrow J/\psi K_S^0$ from simulation data (black points). Both plots are presented with goodness-of-fit values.	62
32	Fits to $\Lambda_b^0 \rightarrow J/\psi \Lambda$ simulation to obtain PDF parameters, along with goodness-of-fit values. Top - full dataset. Bottom left - Long candidates. Bottom right - Downstream candidates.	64
33	Fits to the full $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset after preselection, with an invariant dimuon mass window applied, with linear and log scales, along with goodness-of-fit values. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively. The dashed red and black lines represent the combinatorial background and $B^0 \rightarrow J/\psi K_S^0$ contributions. The B^0 meson background parameters are listed in Table 8, and the signal fit parameters are given in Table 9.	65
34	$\Lambda_b^0 \rightarrow J/\psi \Lambda$ events separated into Long (left) and Downstream (right) tracks. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively, along with goodness-of-fit values. The dashed red and black lines represent the combinatorial background and $B^0 \rightarrow J/\psi K_S^0$ contributions. The B^0 meson background parameters are listed in Table 8, and the signal fit parameters are given in Table 9.	65
35	Fits to the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset after applying steps to remove the $B^0 \rightarrow J/\psi K_S^0$ background. Comparison between this Figure, and Figure 33 shows the success of these restrictions. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively, along with goodness-of-fit values. The dashed red and black lines represent the combinatorial background and $B^0 \rightarrow J/\psi K_S^0$ contributions.	66
36	Left - 15 to 20 GeV^2/c^4 q^2 $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ signal significance as a function of neural network output. The significance is found to maximise at 0.24. Right - Efficiency of the neural network requirement selection over the full q^2 range, for events with a output value of greater than 0.24.	68
37	ROC curves produced in the 15 to 20 GeV^2/c^4 q^2 interval. The black line is a generically trained network, while the red represents a specific training for this q^2 interval. No preference can be seen for either method.	69

38	Simulated events in the 15 to 20 GeV^2/c^4 q^2 interval, for the Mu (left) and Psi (right) fit models, along with goodness-of-fit values. Top - Full dataset. Middle row - Long candidates. Bottom row - Downstream candidates. The black points are the data, and the green line is the signal shape.	71
39	Fits to $\Lambda\mu^+\mu^-$ invariant mass distribution using the Mu fit model. Top - Both Long and Downstream tracks. Bottom left - Long tracks. Bottom right - Downstream tracks. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively, along with goodness-of-fit values. The dashed red line represents the combinatorial background contributions.	72
40	Fits to $\Lambda\mu^+\mu^-$ invariant mass distribution using the Psi fit model. Top - Both Long and Downstream tracks. Bottom left - Long tracks. Bottom right - Downstream tracks. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively, along with goodness-of-fit values. The dashed red line represents the combinatorial background contributions.	73
41	Profile log-Likelihood against number of $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ events in the 15 to 20 GeV^2/c^4 q^2 region.	74
42	One dimensional plots of the 6-dimensional efficiencies of $\cos\theta_l, \cos\theta_h, \cos\theta, \phi_l, \phi_h$ and q^2 , for LL events, with q^2 values between 11 and 20 GeV^2/c^4 , compared to the real $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ dataset, after preselection. Black points are real data after being corrected with the efficiency model. The blue distribution is the angular PDF derived from the efficiency model. Distributions are normalised to 1, giving probability per bin on the y-axis.	77
43	One dimensional plots of the 6-dimensional efficiencies of $\cos\theta_l, \cos\theta_h, \cos\theta, \phi_l, \phi_h$ and q^2 , for DD events, with q^2 values between 11 and 20 GeV^2/c^4 , compared to the real $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ dataset, after preselection. Black points are real data after being corrected with the efficiency model. The blue distribution is the angular PDF derived from the efficiency model. Distributions are normalised to 1, giving probability per bin on the y-axis.	78
44	2-Dimensional plots of reconstructed Λ_b^0 baryon mass by angular variables and q^2 for Long candidates. Plots created using the final analysis dataset, without q^2 limits. No correlation to the Λ_b^0 mass can be seen.	80
45	2-Dimensional plots of reconstructed Λ_b^0 baryon mass by angular variables and q^2 for Downstream candidates. Plots created using the final analysis dataset, without q^2 limits. No correlation to the Λ_b^0 mass can be seen. . . .	80
46	Comparison of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ (top) and $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ (bottom) moments generated from a 6-D distribution with uniform efficiency (red) to those generated from an efficiency weighted Monte Carlo simulated dataset (black). Left - Long candidates. Right - Downstream candidates.	84

47	Difference between the moment values determined from the two aforementioned $\Lambda_b^0 \rightarrow J/\psi \Lambda$ (top) and $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ (bottom) datasets. Left - Long candidates. Right - Downstream candidates.	85
48	Means and widths of the Gaussian distributions of pull plots for the moments calculated in LL (left) and DD (right) simulation studies, using their respective efficiencies. For unbiased determination of the values of M_i , the mean value of the Gaussian is required to be 0, with a width of 1, and this is seen in both datasets for all moments.	86
49	Mean uncertainties by neural network output for the moments associated to unpolarised Λ_b^0 baryon production in $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$, produced via simulation studies.	88
50	Comparison of our calculated expected values of specific moments of the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decay (black) to 2011 data (red) and 2011 + 2012 (blue). The values of M_i sensitive to decay amplitudes do not have expected values plotted. Good agreement is seen between both extracted and expected values for each moment.	92
51	Angular distribution of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decays from data (black points) superimposed with the angular distribution as obtained by setting angular observables in the simulation studies to the values obtained by extracting moments for LL events (top 5) and DD events (bottom 5) using the 2011 dataset.	93
52	Angular distribution of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decays from data (black points) superimposed with the angular distribution as obtained by setting angular observables in the simulation studies to the values obtained by extracting moments for LL events (top 5) and DD events (bottom 5) using the 2011 + 2012 datasets.	94
53	Distribution of the combined Long and Downstream moments determined from the separate datasets assuming non-zero Λ_b^0 baryon transverse polarisation production. Values measured with events taken from the 15 to 20 GeV^2/c^4 q^2 interval.	97
54	Comparison of the distributions of moments extracted from LL (red circles) and DD (blue squares) datasets independently. Values determined from events taken from the 15 to 20 GeV^2/c^4 q^2 interval.	97
55	Angular distributions of Long events from data (black points) over the 15 to 20 GeV^2/c^4 q^2 range, after all selection requirements and application of optimal neural network working point. Events are sWeighted, and taken from a 5350 to 5850 MeV/c^2 invariant $\Lambda \mu^+ \mu^-$ mass range. The blue histogram is the angular PDF generated by simulation studies.	98
56	Angular distributions of Downstream events from data (black points) over the 15 to 20 GeV^2/c^4 q^2 range, after all selection requirements and application of optimal neural network working point. Events are sWeighted, and taken from a 5350 to 5850 MeV/c^2 invariant $\Lambda \mu^+ \mu^-$ mass range. The blue histogram is the angular PDF generated by simulation studies.	99

57	Histograms produced to determine systematic uncertainty on fixing shape parameters. Top left - Distribution in the yields from $B^0 \rightarrow J/\psi K_S^0$. Top right - distribution of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ event yields. Bottom - distribution of $\Lambda_b \rightarrow \Lambda \mu^+ \mu^-$ event yields.	102
58	Theoretical values for observables associated to $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays. Black data points are the results measured in the previous LHCb angular analysis, while the purple regions are theory. The lepton-hadron asymmetry is previously unmeasured by the LHCb. Figures taken from [31].	105
59	Lifetime of the Λ_b^0 baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	108
60	Goodness-of-fit of measured lifetime of the Λ_b^0 baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	109
61	Transverse momentum of the Λ_b^0 baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	109
62	Goodness-of-fit of the decay vertex of the Λ_b^0 baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	110
63	Cosine of the direction angle of the Λ_b^0 baryon momentum vector from its creation vertex. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	110
64	Measured asymmetry of event to background transverse momenta of all tracks in a cone about the Λ_b^0 . The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	111
65	Lifetime of the Λ baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	111
66	Transverse momentum of the Λ baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background. The structure just above 500 MeV/c arises due to a combination of the tight Λ baryon mass restriction and the trigger requirements we apply, which kills these low p_T events.	112
67	Invariant $p\pi^-$ mass distribution used to reconstruct the Λ baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	112
68	Cosine of the direction angle of the Λ baryon momentum vector from its creation vertex. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	113

69	Flight distance fit goodness of the Λ baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	113
70	Z-component of the Λ baryon decay vertex. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background. The features at 250 and 650 mm correspond to sources of background events caused by material boundaries in the detector [79]. The minima occur as these events are removed via preselection and trigger requirements.	114
71	$\text{IP}\chi^2$ of the proton. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	114
72	$\text{IP}\chi^2$ of the pion. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	115
73	$\text{IP}\chi^2$ of μ^+ . The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	115
74	$\text{IP}\chi^2$ of μ^- . The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.	116
75	Pull distributions of Moments 1 - 16 produced from 500 simulation studies, each with 108 LL signal events, and the relevant signal efficiency model. . .	117
76	Pull distributions of Moments 17 - 34 produced from 500 simulation studies, each with 108 LL signal events, and the relevant signal efficiency model. . .	118
77	Pull distributions of Moments 1 - 16 produced from 500 simulation studies, each with 162 DD signal events, and the relevant signal efficiency model. . .	119
78	Pull distributions of Moments 17 - 34 produced from 500 simulation studies, each with 162 DD signal events, and the relevant signal efficiency model. . .	120

List of Tables

1	Particles/Antiparticles in the current Standard Model, with properties published by the Particle Data Group [1]. Note that there are 3 types of each quark, and 8 individual gluons.	12
2	Differential branching fraction ratios of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ relative to $\Lambda_b^0 \rightarrow J/\psi \Lambda$. Uncertainties are statistical and systematic, respectively. Values are taken from [28].	24
3	Weighting functions for calculation of moments, and determination of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ angular observables. Polarised and unpolarised functions given, respectively. Functions for M_{11} through M_{34} become 0 when the Λ_b^0 baryon has no transverse polarisation, as they are products of P_b and f_i	28
4	Stripping requirements applied to the dataset. Where two values are present, the first is applicable to Long candidates, and the value in brackets to Downstream.	50
5	Trigger requirements applied to the data, all Trigger on Signal. L0 is applied to the dimuon, and the HLT to the reconstructed Λ_b^0 baryon.	52
6	Preselection cuts applied to both the real and simulated datasets before training of the neural network.	54
7	Variables used in NeuroBayes training, in order of their final ranking. ‘Added significance’ is the amount of information a variable adds to all above. ‘Only this’ is the variable on its own. ‘Loss without’ is the information lost by removing solely that variable.	56
8	Fit parameters for the $B^0 \rightarrow J/\psi K_S^0$ mass shape. Values extracted from simulation data, with preselection requirements, but with no restriction on q^2 or neural network output.	62
9	Fit parameters for the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ mass shape. Values are extracted from simulation data, with selection, trigger and truth requirements applied, but with no restriction on q^2 or neural network output.	63
10	Yields extracted from modelling to the $\Lambda \mu^+ \mu^-$ invariant mass spectrum for $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events.	64
11	Species yields after the application of the neural network working point and proton probability to the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset. Note, that the signal yields presented here are not those used to optimise the network, but instead to show the effect of the restrictions.	66
12	Neural network optimum working points, as determined by maximising the signal significance in each q^2 region.	68
13	Fit parameters extracted from MC simulation with their associated errors for both fit shapes. The widths listed here are taken straight from simulation, and have not been corrected for MC/data discrepancies.	70
14	$\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ yields extracted from the $\Lambda \mu^+ \mu^-$ invariant mass distribution for both the Mu ($\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$) and Psi ($\Lambda_b^0 \rightarrow J/\psi \Lambda$) mass shape fit models, using parameters determined from simulation data.	74

15	Orders required to describe the Legendre polynomials used for the efficiency parametrisation of LL and DD track type events.	76
16	Values of moments expected using weighing functions to extract $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ angular observables in terms of the physics parameters associated with $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decays.	90
17	Combined moments for the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decays using DD and LL events from 2011 and 2011 + 2012 datasets, and with expected values.	91
18	Values of the moments determined from $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events in the 15 to 20 GeV^2/c^4 q^2 interval, assuming that the Λ_b^0 baryon has a non-zero transverse production polarisation. All values are produced from the Run 1 dataset. Uncertainties are the standard deviation of the bootstrap samples. The moments measurable in the decay of a Λ_b^0 baryon with non-zero transverse production polarisation P_b are found to be consistent with zero.	96
19	Angular observables measured from extraction of the moments associated to the decay $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$. These measurements are specific to the 15 to 20 GeV^2/c^4 q^2 interval assuming that the Λ_b^0 baryon has a non-zero transverse production polarisation. Values presented for the total dataset and for the individual LL and DD datasets.	100
20	Comparison of observables determined in this work to the previous analysis, and to theoretical values determined from lattice QCD. Theoretical values of moments are those predicted not to be zero in non-polarised Λ_b^0 baryon production. All values are from the 15 to 20 GeV^2/c^4 q^2 interval. Uncertainties are statistical and systematic, respectively. Values in brackets are confidence on the last digit.	100
21	Sources and values of the individual systematic uncertainties on yields arising from the analysis method. The total uncertainty is determined by adding the values in quadrature.	102

1 Physics of Heavy Flavour Processes

1.1 The Standard Model of Particle Physics

The Standard Model provides a framework to aid understanding into the interactions of the matter and energy that make up the universe, and an insight into the forces between them. The current model describes three of the four known fundamental forces of nature - electromagnetism, and the strong and weak forces, along with the bosons mediating them, and the elementary fermions which form the building blocks of the matter we see around us. A list of the elementary particles in the Standard Model is shown in Table 1.

Particle Name	Type	Label	Mass (GeV/ c^2)	Electric charge (e)
Up Quark	Fermion	$u/\bar{u}$	2.3×10^{-3}	$\pm \frac{2}{3}$
Down Quark	Fermion	$d/\bar{d}$	4.8×10^{-3}	$\mp \frac{1}{3}$
Charm Quark	Fermion	$c/\bar{c}$	1.28	$\pm \frac{2}{3}$
Strange Quark	Fermion	$s/\bar{s}$	95×10^{-3}	$\mp \frac{1}{3}$
Top Quark	Fermion	$t/\bar{t}$	173.21	$\pm \frac{2}{3}$
Bottom Quark	Fermion	$b/\bar{b}$	4.18	$\mp \frac{1}{3}$
Electron	Fermion	$e/\bar{e}$	0.51×10^{-3}	∓ 1
Electron Neutrino	Fermion	$\nu_e/\bar{\nu}_e$	$< 0.225 \times 10^{-3}$	0
Muon	Fermion	$\mu/\bar{\mu}$	105.68×10^{-3}	∓ 1
Muon Neutrino	Fermion	$\nu_\mu/\bar{\nu}_\mu$	$< 0.19 \times 10^{-3}$	0
Tauon	Fermion	$\tau/\bar{\tau}$	1.78	∓ 1
Tau Neutrino	Fermion	$\nu_\tau/\bar{\nu}_\tau$	$< 18.2 \times 10^{-3}$	0
Photon	Boson	γ	0.00	0
Gluon	Boson	g	0.00	0
W Boson	Boson	W^+/W^-	80.39	± 1
Z Boson	Boson	Z^0	91.19	0
Higgs Boson	Boson	H^0	125.09	0

Table 1: Particles/Antiparticles in the current Standard Model, with properties published by the Particle Data Group [1]. Note that there are 3 types of each quark, and 8 individual gluons.

The matter we see around us in the universe is made up of quarks and leptons, primarily the up and down quarks, and the electron. These particles are held together by the bosons, with quarks being bound by gluons to form baryons and mesons. The other, heavier fermions are only found in more extreme regimes, be that inside neutron stars, as a result of cosmic rays striking the atmosphere, or inside particle accelerator experiments, as they very rapidly decay into their lighter rel-

atives. As such, high energy experiments are the only reliable way to study their properties.

The creation of the current Standard Model was seeded by Sheldon Glashow in 1961 [2], when he found that electromagnetic and weak processes could be described by a single model, aptly named the electroweak interaction. Later, the current and more familiar iteration was produced via addition of the Higgs mechanism [3] by Steven Weinberg [4] and Abdus Salam. The fundamental forces in nature are expected to have a field and force carrying bosonic particle associated to it, and the Standard Model uses these to describe the interactions between particles. It can therefore be used as a framework to understand the processes we see in high energy particle physics.

The Λ_b^0 baryon decays solely by the weak force, which is the only fundamental force of nature which allows flavour changes in decays, causing heavy quarks to become lighter, more stable bodies, as is seen in the decay of free neutrons. Weak processes are mediated by the $W^\pm$ and Z^0 bosons, the heavy masses of which reduce the rates at which these processes will occur when compared to strong or electromagnetic interactions, as well as limiting the range of interactions to $\mathcal{O}(10^{-18})$ metres. Using leptonic processes [5], the coupling constant relation for weak interactions is given as

$$\frac{G_F}{\sqrt{2}} = \frac{g_W^2}{M_W^2} \quad (1)$$

where G_F is the Fermi coupling constant, and g_W is the coupling associated with the $W^\pm$ boson-lepton vertices, and is necessary to calculate probabilities of flavour changing processes. The Fermi coupling constant is equal to $1.166 \times 10^{-5} \text{ GeV}^{-2}$. As can be seen, the coupling is proportional to the inverse of the mass of the mediator boson squared, greatly increasing the life times of particles that decay weakly. As such, a particle that decays exclusively via the weak force will exist longer than a similar particle that may decay strongly.

While the Standard Model may be the best framework we have to explain the processes and interactions seen between forces and particles, the model is flawed. It is unable to describe 95% of the mass-energy of the universe, for example, and nor is it able to explain gravitation at the quantum level. Processes which are suppressed in the Standard Model are a great place to search for hints of new physics, and with the LHCb experiment [6, 7], it's now possible to produce b-hadrons and their rare decays at quantities out of reach of previous experiments, as well as being able reconstruct and analyse them accurately. Production is increased due to higher luminosities, and vast centre of mass energies increasing particle creation cross sections. Similar b-hadron decays have helped refine parameters necessary to describe particle interactions, as well as finding significant deviations from the theory used to describe them [8, 9], suggesting shortcomings in our knowledge [10]. Recent measurements have found tensions in the measured branching fractions of the decays of $B_S^0 \rightarrow \phi \mu^+ \mu^-$ and $B \rightarrow K^{(*)} \mu^+ \mu^-$, which deviate from theoretical expectations [11, 12]. Both were discovered using the Run 1, 3 fb^{-1}

data collected at the LHCb, indicating that the experiment is able to detect these variations.

1.2 Flavour Changing Neutral Currents and the GIM Mechanism

Currently in nature, the weak force is the only physical process known not to conserve quark flavour in its interactions, and thus allows decay of heavy quarks into those in lower generations. In the case of $b \rightarrow s$ however, while the flavour changes, the net electric charge difference is 0. This means that for a tree-level decay the mediating boson must also be neutrally charged, and this property belongs to the Z^0 boson. The Z^0 , however, exists as its own antiparticle, and as such, its flavour quantum numbers and charges are all zero [1]. Due to this, the only properties exchangeable via its interactions are energy and momentum, and thus quark flavour may not be altered. Despite this, bottom to strange transitions - and other similar processes - have been observed, meaning that that may proceed in different, non-tree level manners. The $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay proceeds via loop and box paths, and these are the methods in which all FCNC processes are expected to proceed in the Standard Model. Instead of a single-stage decay from one quark to another of the same family, there is an intermediate state in which the decaying quark becomes a member of the opposite quark type ($u, c, t \leftrightarrow d, s, b$), before undergoing a second flavour change into a quark of the same charge as the original. In $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$, a down-type bottom quark decays into a strange quark, via a preferred intermediate top quark.

Decays proceeding via the weak force must obey certain selection rules. Most importantly, flavour can only change by a single unit, and most proceed via a charged current process. These require FCNC processes to occur at least at the second order, as seen already with the increased number of flavour changing vertices. In 1970, Glashow, Iliopoulos and Maiani published their seminal paper on what became known as the GIM mechanism [13]. This work describes charged current processes, and also the suppression of flavour changing neutral currents. Originally, the paper offered an extension to Cabibbo's work on the weak charged current processes of the three quarks known at the time, up, down and strange [14]. The GIM mechanism added a fourth quark to this system, in order to explain the low rate of FCNC kaon decays and restore symmetry, this becoming known as the charm quark. This addition expanded Cabibbo's original interaction current

$$J_\mu(x) = u'(x)\gamma_\mu(1 + \gamma_5)[\cos\theta d(x) + \sin\theta s(x)] \quad (2)$$

to

$$J_\mu(x) = \begin{bmatrix} u'(x) & c'(x) \end{bmatrix} \gamma_\mu(1 + \gamma_5) \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} d(x) \\ s(x) \end{bmatrix} \quad (3)$$

where θ is the Cabibbo angle, and γ_μ are the Dirac matrices. With the existence of 4 quarks, only doublets were necessary. The GIM mechanism allowed for description of charged current interactions, with previous models not being able to successfully predict

decays rates. The loop level decay $K_L \rightarrow \mu^+ \mu^-$, for example, was observed less often experimentally than predictions would suggest. The GIM mechanism found this to be due to strange and charm contributing much less than would be expected to these processes, and thus could explain the suppression the FCNCs experienced in loop level decays.

1.3 The Cabibbo-Kobayashi-Maskawa Matrix

Originally, the likelihoods of transition between the first to quark families were determined via the previously mentioned combination of Nicola Cabibbo's work with the up, down and strange quarks, and the GIM mechanism's addition of the charm quark. After the 2 generation model could not explain violations of charge-parity symmetry measured in neutral kaon decays [15], Makoto Kobayashi and Toshihide Maskawa extended the model to predict the unseen 3rd generation quarks [8]. The CKM matrix contains information on the probabilities of flavour changes in charged current weak processes, or to put it simply, a measure of the likelihood of a change between an up type (u, c, t) or down type (d, s, b) quark into another of the opposite type. The matrix elements come into play at every flavour changing vertex in a Feynman diagram, and are necessary to predict the rates associated with the interactions. The flavour changes go as

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \times \begin{bmatrix} d \\ s \\ b \end{bmatrix},$$

where (d, s, b) are the initial state quarks, V_{ij} are the matrix elements describing likelihoods, and (d', s', b') are the post transformation quarks. It is required that d', s' and b' are single quark states, and not eigenstates of multiple flavours. The coupling strengths at each flavour changing vertex is calculated as the matrix element describing the flavour change multiplied by g_W . The vertices and respective couplings for both the non-suppressed and rare decays of the Λ_b^0 baryon are displayed in Figure 1, where they are referenced as $g_{i,j}$. These values are given by

$$g_{cb} = V_{cb} \cdot g_W, \quad g_{cs} = V_{cs} \cdot g_W, \quad g_{tb} = V_{tb} \cdot g_W \quad \text{and} \quad g_{ts} = V_{ts} \cdot g_W. \quad (4)$$

The probability of a given flavour change is equal to the square of the coupling constant, and as such, proportional to the square of the associated CKM matrix component. The magnitudes of the nine components of the matrix have been measured experimentally, with each being determinable via different decays.

The elements relevant to $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ are $|V_{tb}|$ and $|V_{ts}|$. The latter is not accurately measurable in tree level diagrams, and as such is determined from $B - \bar{B}$ oscillations proceeding via box Feynman diagrams, or rare loop level kaon decays [1]. $|V_{tb}|$ is measured in electroweak top quark production, a tree level process. Measured independently of one another, the individual published [1] CKM elements are

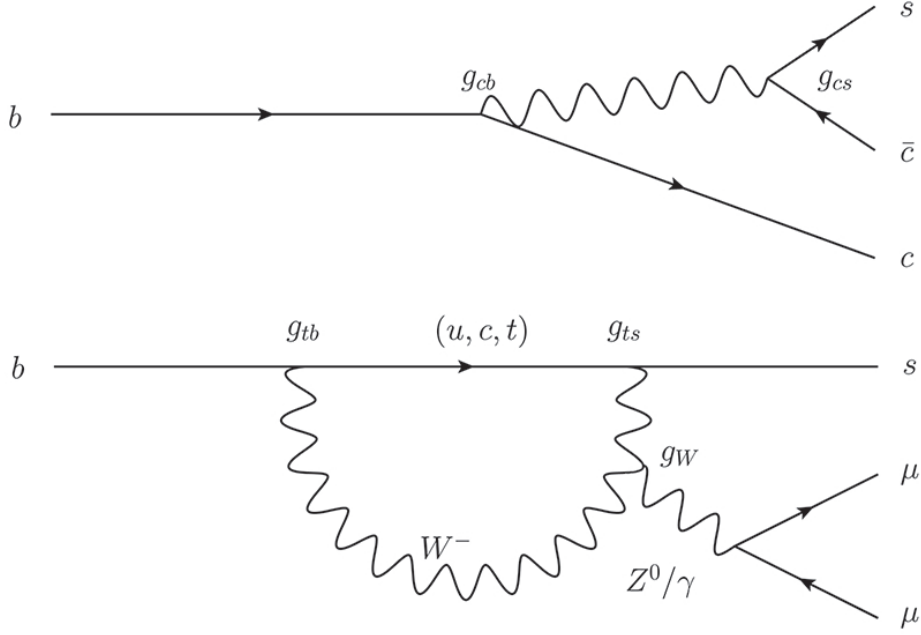


Figure 1: Flavour changing vertices in $\Lambda_b^0 \rightarrow J/\psi \Lambda$ (top) and $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ (bottom) decays. In the top diagram, the bottom quark decays into a charm in a flavour changing charged current process, which then undergoes a second change to the strange. The bottom quark cannot do this in $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ however, and must decay into an intermediate virtual quark state, before this becomes the strange quark.

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97425 \pm 0.0002 & 0.2253 \pm 0.0008 & 0.00413 \pm 0.00049 \\ 0.225 \pm 0.008 & 0.986 \pm 0.016 & 0.0411 \pm 0.0013 \\ 0.0084 \pm 0.0006 & 0.040 \pm 0.0022 & 1.021 \pm 0.032 \end{bmatrix}.$$

With these individual magnitudes, the unitarity of the CKM matrix can be checked. To be unitary, the sum of the squares across a row or down a column are required to be 1. By enforcing this over all three quark generations, and modelling the values with a global fit over the available measurements, the matrix components can be refined. The PDG [1] publishes best fit values of the parameters to be

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886^{+0.00033}_{-0.00032} & 0.0405^{+0.00011}_{-0.00012} & 0.99914 \pm 0.00005 \end{bmatrix}.$$

With the rate of a decay being proportional to the square of g_W and the CKM matrix element, it is easy to see that a Feynman diagram containing many flavour changing

matrices will be less probable than one containing fewer. Therefore, as FCNC's contain at least two - with one or more off-diagonal CKM elements, and a loop suppression factor, their reduced rates may be explained. While both $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi \Lambda$ contain two flavour changing vertices with similarly valued matrix elements, suppression arises from the loop present in the rare decay diagram. This reduces the rate by a factor of $1/4\pi^2$, as well as extra suppression from the additional vertex produced in creating the virtual Z^0/γ particle. It can be seen that the $b \rightarrow t$ flavour change is the most preferred in the rare decay, but not the only valid intermediate.

1.4 Beyond the Standard Model

A somewhat major failure of the Standard Model is that it only describes 5% of the make up of the universe [16], as shown in Figure 2. The remaining 95% is theorised to comprise of dark energy and dark matter, neither of which have been directly observed. Dark matter is only seen in huge quantities via its gravitational effects on light. The dark matter particles however are thought to interact via the weak force. Currently, popular candidates are weakly interacting massive particles - WIMPs - theorised to be around $100 \text{ GeV}/c^2$ in mass [17]. Searches at the LHC are already under way in attempts to find the unknown particles [18]. In addition to this, no explanation to the existence of a matter dominated universe may be found in the model. During the Big Bang, an equal amount of matter and antimatter are expected to have been created, however the universe has since evolved to be matter dominated. The conditions which lead to this asymmetry are understood [19], however the Standard Model cannot explain the levels of asymmetry we see. It is one of the LHCb's main goals to understand whether matter and antimatter behave differently in high energy interactions. Another caveat is the lack of a model of gravitation which is compatible with the other 3 forces described by the Standard Model. As such, one of the most sought after prizes in physics currently is to create the "Theory of Everything", one model to describe all four forces of nature using the same set of equations. On the topic of gravity, another issue called the Hierarchy Problem [20] exists. There is a huge difference in the Planck and electroweak mass scales, with the weak force carriers $W^\pm$, Z^0 are much, much smaller than would be expected with current measured properties of the Higgs field. This implies a massive difference in the mass scales of the weak and gravitational forces, and an unstable Higgs field, with a Higgs mass expected to be much, much larger than measured.

In the context of this work however, we search for the possibility of new particles in BSM sensitive decays, while the LHCb on the whole works on CP-sensitive decays. Should a statistically significant deviation be seen in the angular distributions of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ or similar decays when compared to theoretical models, it could be theorised that an unknown particle was present in the process. This may be a new intermediate quark in a loop level or box Feynman diagram, or a new uncharged boson which allows FCNCs to proceed at the tree level, as shown in Figure 3. Due to the reduction in the total number of flavour changing vertices, it could be seen that this would yield an altered branching

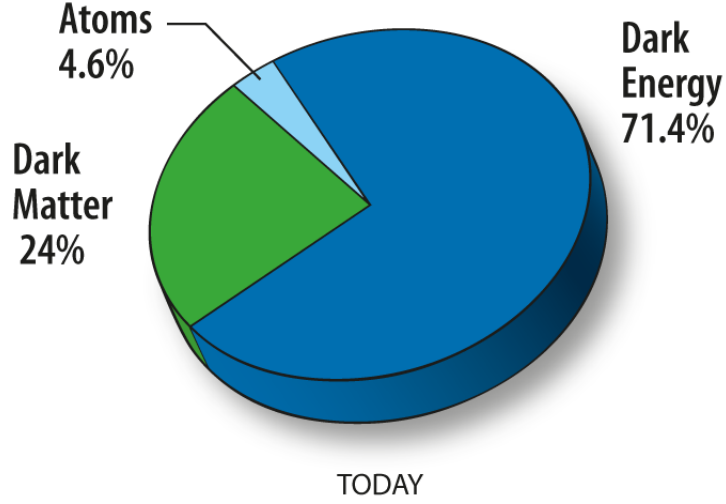


Figure 2: Recent NASA results on the predicted make up of the mass-energy in the universe, measured with WMAP. Figure and values published by the NASA/WMAP Science Team.

fraction than the current Standard Model process. It would be fair to assume that this new boson is very massive, and as such a decay proceeding via this virtual intermediate state would only been seen in high luminosity collisions. As previously stated, work has already been done to produce theoretical models on the effects of new quarks and bosons on both the branching fractions of suppressed decays, and the properties of the final state leptons in other $b \rightarrow s\ell\ell$ transitions. These, coupled with the new higher energies achievable at the LHCb, and the future high luminosity upgrade will allow the LHCb to look for hints of these new particles, be they bosons, quarks, or leptons.

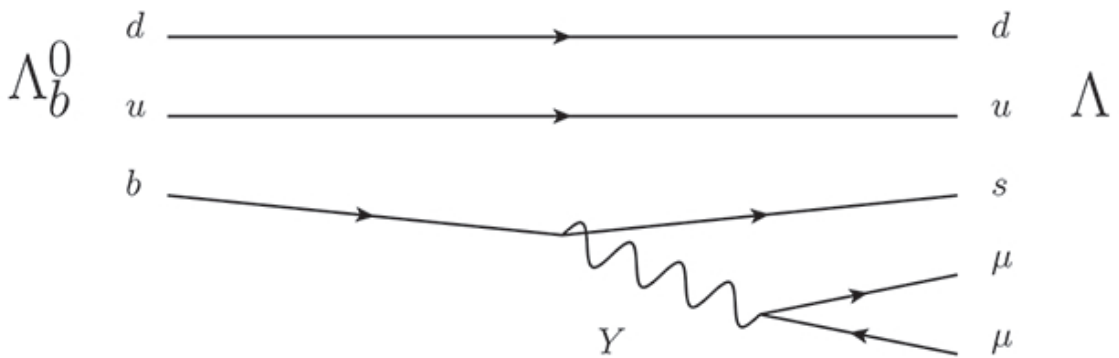


Figure 3: A BSM $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay diagram, where Y is a new neutral boson mediating the flavour change.

While new bosons are required in order to describe the nature of gravity in the

Standard Model, and other new particles are needed to describe the missing mass-energy in the universe, there is no such immediate need for a new generation of quarks and leptons. This being said, searches are ongoing in high energy collisions at various accelerators around the world [1], in order to narrow down the window in which new particles are likely to be seen. With the Higgs Boson mass at its measured value, it has been theorised that the maximum quark mass lies at about $300 \text{ GeV}/c^2$, however this contradicts with already defined limitations [21]. Additionally, other works have been produced to theorise the maximum masses of the leptons and neutrinos in the 4th generation that are required to maintain the lepton-hadron symmetry [22]. The LHCb is not a general purpose experiment, and does not provide the full coverage of detectors around the interaction point necessary to find new particles. As such, the LHCb would only be able to observe their effects on known matter, with the results narrowing the search intervals for the other more general purpose detectors to use. [23]. Dark matter particles are not expected to be directly observable in any of the experiments at the Large Hadron Collider, as they would not interact with matter in detectable ways. Instead, decays involving these particles would appear to have missing energy, as they escape the experiment without detection. To infer the presence of dark matter particles, it is therefore necessary to have precise decay tracking and reconstruction to ensure the energy is genuinely missing, and not only appearing to be after bad event construction.

A potential BSM theory is supersymmetry (SUSY) [1, 24] - a model in which every boson and fermion has a superpartner, with a spin that differs by $\frac{1}{2}$. In a model with a perfectly unbroken symmetry, apart from spin, these superpartners would have identical quantum numbers and masses. Under this assumption however, supersymmetrical particles would have been discovered already - and as such the theory must be a spontaneously broken symmetry, such that mass differences may arise [25]. Supersymmetry is touted as a solution to the hierarchy problem and also as a source of WIMPs [26] - the properties of supersymmetric particles are in agreement to those expected for this form of dark matter, known as the WIMP miracle. No experimental evidence has been found however [27], as results from the LHC yield no supersymmetric particles in the expected mass ranges. Should SUSY be proven however, the number of particles in the current Standard Model (See Table 1) would be doubled, greatly increasing the number of possible interactions.

2 The Suppressed $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ Decay

2.1 The $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ Decay, and the q^2 Spectrum

The $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay is an example of a suppressed $b \rightarrow s$ FCNC process. The decay may only occur via electroweak loop or box level processes, with the contributing Standard Model Feynman diagrams displayed in Figure 4. Suppressed processes provide a way to probe for Beyond the Standard Model contributions, as measurable effects on branching fractions and distributions of the final state particles may be seen. Importantly, the scarcity of the suppressed decays means new processes are not drowned out by known physics. The decays are mediated by virtual particles, allowing experimental facilities to probe interactions of particles that cannot not be created on mass shell at current centre of mass energies. It has been found previously that some of the measured parameters of the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay show deviation from Standard Model predictions [28]. It is therefore of interest to measure the angular distribution of the process, as it is sensitive to the effects of new physics. Frameworks describing the theory behind angular distributions of unpolarised $\Lambda_b^0 \rightarrow \Lambda(\rightarrow N\pi)\ell\ell$ [29–31] are used for this analysis.

The decay was first discovered in a sample of 24 events with a 5.8σ significance by the CDF II experiment at the Tevatron in 2011 [32], with further studies by the LHCb [28] experiment finding larger yields. The world average branching fraction [1] of the decay is published as

$$\mathcal{B}(\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-) = (1.08 \pm 0.28) \times 10^{-6}. \quad (5)$$

We reconstruct $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events from a Λ baryon and a muon pair originating from a common vertex, with the Λ baryon reformed from $p\pi^-$. The same is true for the more common $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decay, and due to the similarities, they are not easily distinguishable. With dimuon final states, both the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ and $\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\Lambda$ decays exist on the q^2 spectra - a distribution produced with the square of the invariant mass of the dimuon in these and other similar processes. This distribution allows for rare decay to be found, as the dimuon mass of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ may be found at any value, while $\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\Lambda$ is confined to specific regions of the spectra. By omitting the regions associated with $c\bar{c} \rightarrow \mu^+ \mu^-$ decays, we are able to remove $\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\Lambda$ events. A theoretical sketch of the q^2 spectra is shown as Figure 5, produced for $B^0 \rightarrow K^* \mu^+ \mu^-$ decays, however the distribution is similar to $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ - the lower the energy of the hadron, the greater the invariant mass of the dimuon. In the hard minimum value of the distribution is the square of the dimuon rest mass, while the maximum is dependant on the mass difference between the original particle and the hadron bodies in the final state of a given decay. For $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$, the invariant mass difference between the Λ_b^0 and Λ baryons determines the upper limit, corresponding to a q^2 value of $20.29 \text{ GeV}^2/c^4$, with events expected to be seen across this range. The muons produced from a virtual photon will have a lower q^2 , as it is favourable for the photon to have as close to its real mass as possible, with dimuons from the off shell Z^0 being of higher invariant mass. The spectra is dominated by the dimuon decays of J/ψ and $\psi(2S)$ charmonia meson

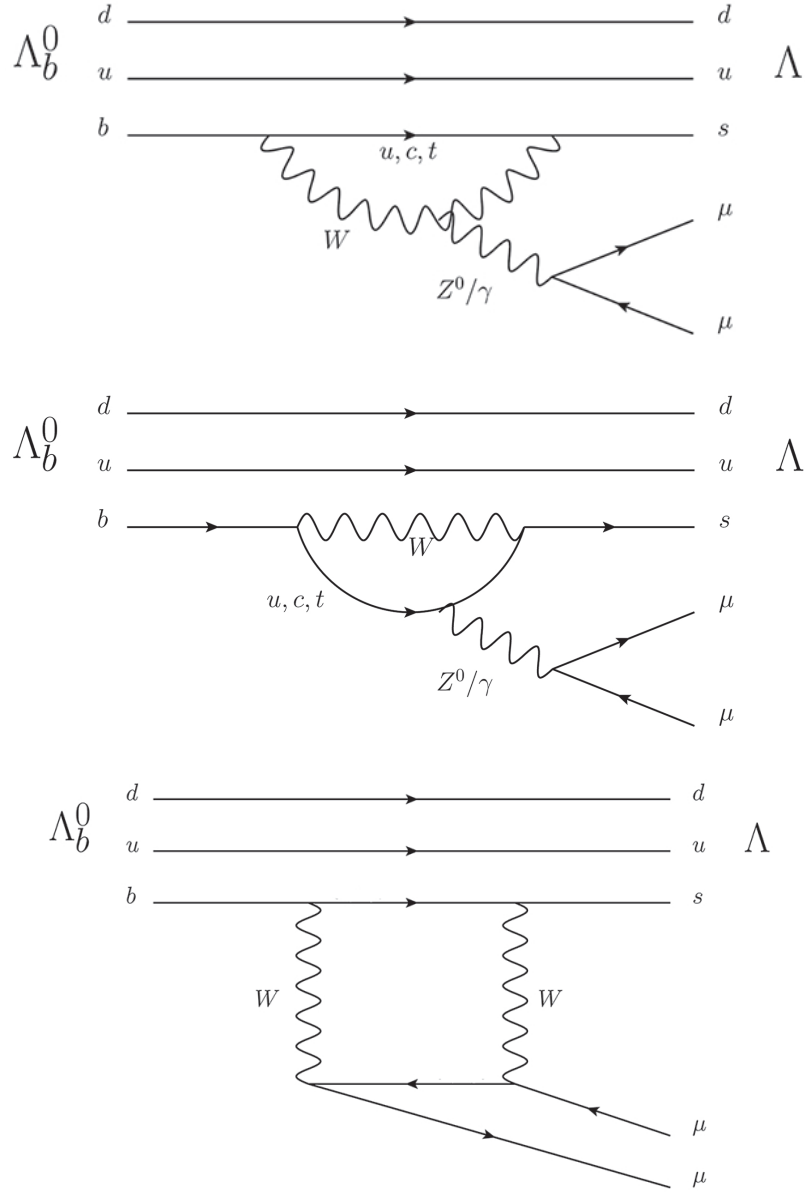


Figure 4: Feynman diagrams of contributing Standard Model $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays. Each proceeds via an intermediate quark of the opposite type to the bottom (u, c, t), with the particles mediating this flavour change then producing the dimuon.

resonances - peaking at approximately $9.59 \text{ GeV}^2/c^4$ and $13.59 \text{ GeV}^2/c^4$ respectively. In order to prevent the events in these regions swamping suppressed decays, they are omitted during searches for rare $b \rightarrow s \ell \ell$ processes. While other charmonium resonances exist, their contributions are negligible in comparison to the two mentioned. Despite this, the contributions cover a large q^2 range, making them hard to remove, meaning that their effects must be modelled in predictions.

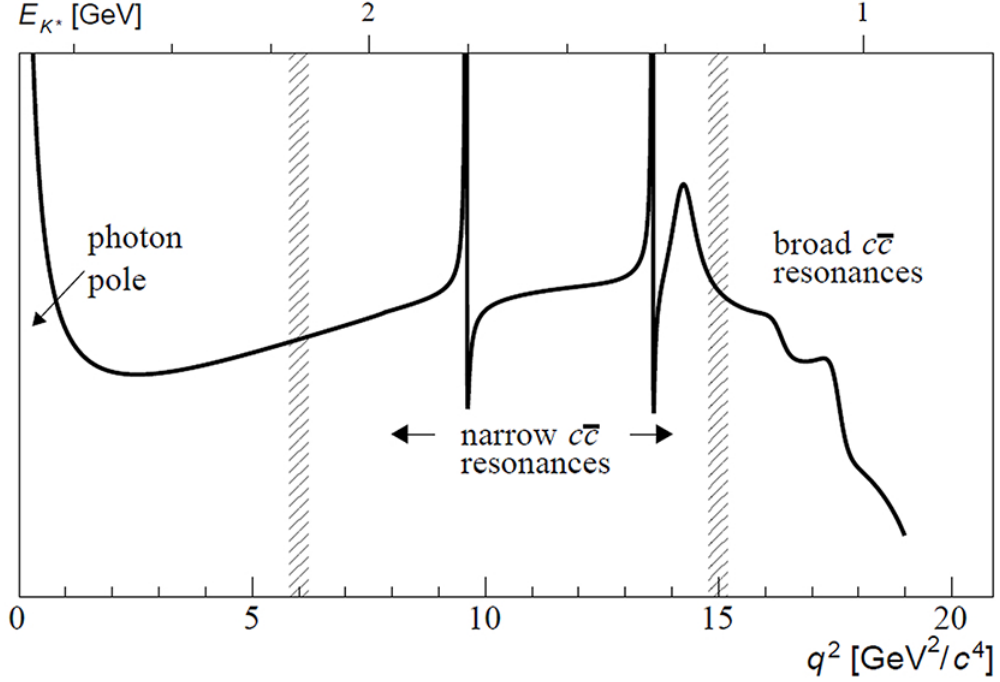


Figure 5: A sketch of the distribution of the q^2 spectrum for the decay $B^0 \rightarrow K^* \mu^+ \mu^-$. The narrow $c\bar{c}$ resonances are the J/ψ meson and $\psi(2S)$ resonances respectively, while the pole at low q^2 is symptomatic of dimuons produced from photons. Figure is taken from [6]

Previous searches and analyses of the decay $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ have only found a significant number of signal events in the 15 to 20 GeV^2/c^4 region [28, 33] of the q^2 distribution. In this limit, the differential branching fraction ratio is measured to be

$$\frac{d\mathcal{B}(\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-)}{dq^2} = \left(1.18_{-0.08}^{+0.09} \pm 0.03 \pm 0.27\right) \times 10^{-7} (\text{GeV}^2/c^4)^{-1}, \quad (6)$$

where uncertainties are statistical, systematic, and due to branching fraction of $\Lambda_b^0 \rightarrow J/\psi \Lambda$, the decay mode which was used for normalisation. As this angular analysis is using the same 3 fb^{-1} dataset as [28], we focus on this q^2 region, due to the expected higher yield in this region with respect to lower q^2 bins. This region is also sufficiently far away from the $\psi(2S)$ peak to not contain contributions which may affect the distribution of our moments. We choose to use a wide range as opposed to finer bins for our search in order to compare to the theory, as these predictions are better over large q^2 bins. This is due to the previously discussed broad charmonia resonances providing quark level contributions in the smaller q^2 bins, while the effect is negated in larger regions [34]. The previous LHCb angular analysis of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ was not limited to the 15 to 20 GeV^2/c^4 q^2 region, but also performed in the 1.1 to 2.0 and 11 to 12.5 GeV^2/c^4 range. Despite finding 4.4σ and 6.5σ significances of events in these bins respectively, there was not a large enough

quantity of events found to accurately determine the distribution of moments, and so our search was not repeated there. Also described in [28] is a search for events in the 1.1 to 6.0 GeV^2/c^4 q^2 interval, as this region is clean of both interference from the J/ψ peak, and potential contributions from $\Lambda_b^0 \rightarrow \Lambda \phi (\rightarrow \mu^+ \mu^-)$ present at low q^2 values. There was not a significant $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ yield however, and the decay is still undiscovered in this q^2 region. Due to the expected low statistics here, no angular analysis is performed in the interval. Figure 6 shows differential branching fraction of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ by q^2 interval, with the numerical values of these differential branch fractions are given in Table 2, as ratios relative to the branching fraction of $\Lambda_b^0 \rightarrow J/\psi \Lambda$. It can be seen that the preference is for $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events to occur with high invariant mass dimuons.

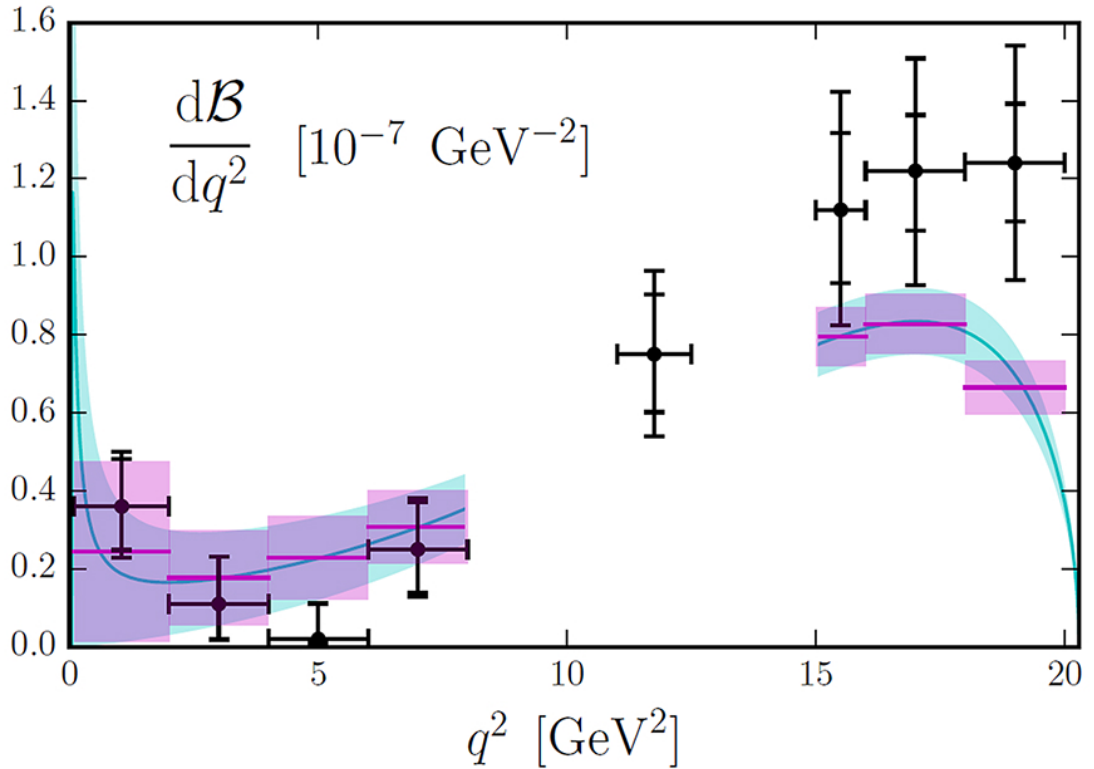


Figure 6: Branching fractions of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ in different q^2 bins. The black data points are data, while the purple are SM prediction. Values from [28], figure taken from [31].

While not contributing to the source of events used in the angular analysis, we use $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events in order to optimise the output of the neural network used to find $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events. As a previous angular analysis has been performed on the decay [35], we also use these events in validating the method used to determine the moments of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$. The diagram describing the decay of $\Lambda_b^0 \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) \Lambda$ is shown in Figure 7. By imposing a 3030 to 3150 MeV/c^2 dimuon invariant mass limit (or a q^2 range of 9.18 to 9.92 GeV^2/c^4), the signal peak in the distribution of $M(\Lambda \mu^+ \mu^-)$ after preselection

q^2 Region (GeV^2/c^4)	$\frac{d\mathcal{B}(\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-)/dq^2}{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi \Lambda)} \cdot 10^{-4} [(\text{GeV}^2/c^4)^{-1}]$
0.1 - 2.0	$0.56^{+0.20}_{-0.17}^{+0.03}_{-0.03}$
1.1 - 6.0	$0.14^{+0.10}_{-0.09}^{+0.01}_{-0.01}$
11.0 - 12.5	$1.19^{+0.24}_{-0.23}^{+0.04}_{-0.07}$
15.0 - 20.0	$1.90^{+0.14}_{-0.14}^{+0.04}_{-0.06}$

Table 2: Differential branching fraction ratios of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ relative to $\Lambda_b^0 \rightarrow J/\psi \Lambda$. Uncertainties are statistical and systematic, respectively. Values are taken from [28].

requirements is expected to be primarily due to $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events.

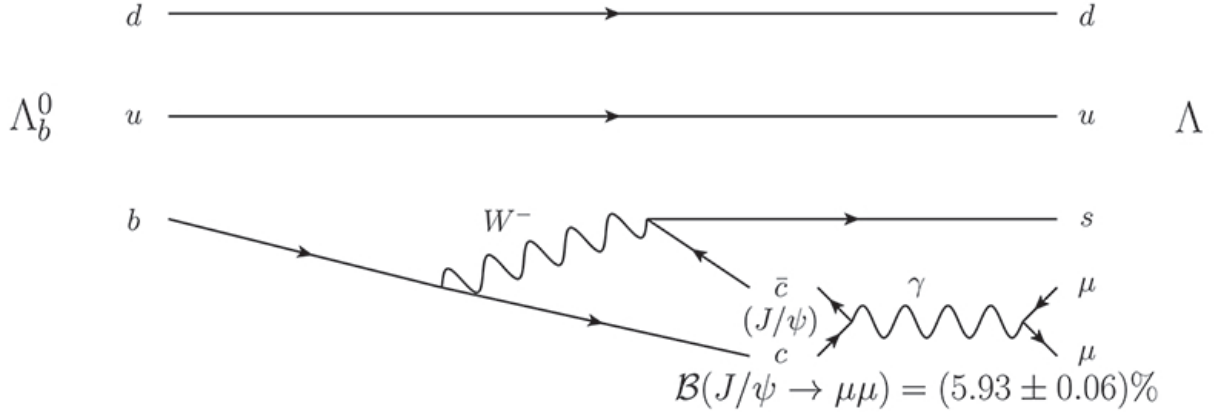


Figure 7: A Feynman diagram of the $\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \Lambda$ decay used to normalise the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ differential branching fraction.

Theories are already in place for describing possible new physics processes that may be found in rare $b \rightarrow s \ell \ell$ transitions, including 4th generation quark effects on similar decays [36, 37]. Like all $b \rightarrow s \ell \ell$ the decay of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ occurs via two intermediate virtual states, leading to increased possibility of effects by particles of much higher masses than available in the centre-of-mass frame being seen. Current estimates predict that the LHCb will be able to accurately probe a coupling-to-mass ratio of approximately 200 TeV in rare decays [38], much greater than the 13 TeV available in centre of mass collisions for on mass shell particle creation. To probe these levels, it is necessary to precisely know the hadronic physics associated, and as such the work currently being performed by the LHCb and other experiments is vital to this.

2.2 Distribution of Angles and Definition of Moments

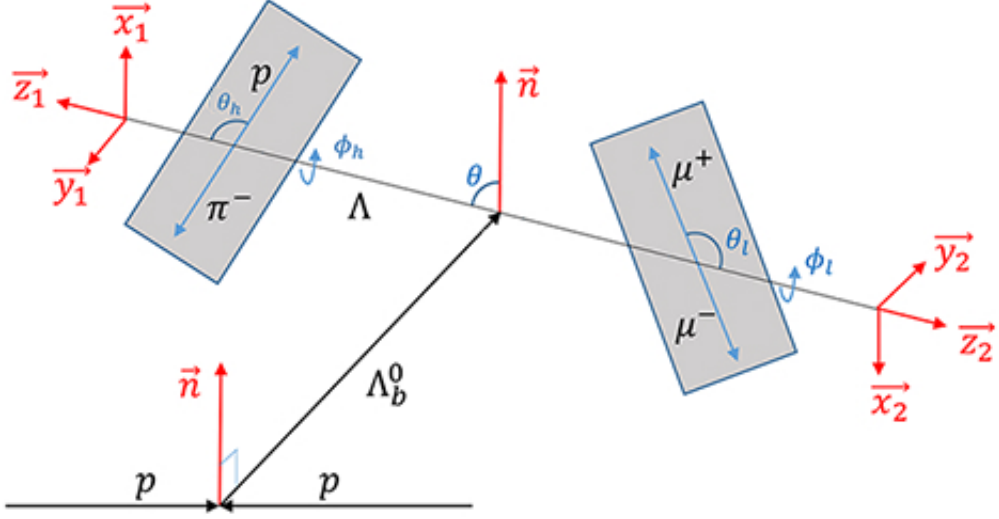


Figure 8: Definition of angular variables measurable in the $\Lambda_b^0 \rightarrow \Lambda(\rightarrow p\pi^-)\mu^+\mu^-$ decay, assuming a non-zero Λ_b^0 baryon transverse production polarisation. Definitions of the angles and axes displayed are given in the text.

Like all baryons, the Λ_b^0 baryon is a fermion, meaning it has a non-integer spin and may be produced with non-zero transverse polarisation relative to the beamline. Polarisation is a measure of the preferential alignment of the baryon spin to the LHC beamline. The Λ_b^0 baryon has a measured transverse production polarisation (P_b) of 0.06 ± 0.07 [35], as determined by the LHCb experiment. If we assume $P_b \neq 0$, the 5 angles $\theta, \theta_l, \theta_h, \phi_l, \phi_h$ displayed in Figure 8 are necessary to describe the decay. The definition of these axes follows from an angular analysis on $\Lambda_b^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\Lambda(\rightarrow p\pi^-)$ [35]. These are given as $\vec{n} = (\vec{p}_{\Lambda_b^0} \times \vec{p}_{beam})/|\vec{p}_{\Lambda_b^0} \times \vec{p}_{beam}|$, with $\vec{x}_1$ and $\vec{x}_2$ parallel to $\vec{n}$, defined as perpendicular to the beamline and the Λ_b^0 momentum in the lab frame. Axes $\vec{y}_1$ and $\vec{y}_2$ are perpendicular to $\vec{n}$, and are given by $(\vec{n} \times \vec{p}_\Lambda)/|\vec{n} \times \vec{p}_\Lambda|$ and $(\vec{n} \times \vec{p}_{\mu^+\mu^-})/|\vec{n} \times \vec{p}_{\mu^+\mu^-}|$. Lastly, axes $\vec{z}_1$ and $\vec{z}_2$ simply go as $\vec{p}_\Lambda/|\vec{p}_\Lambda|$ and $\vec{p}_{\mu^+\mu^-}/|\vec{p}_{\mu^+\mu^-}|$, respectively. The angles θ_h and θ_l are defined as those between the momentum vector of the p or μ^+ and $\vec{z}_1$ or $\vec{z}_2$ respectively, measured in the frame of reference of the Λ baryon and dimuon respectively. The angle θ is related to the production polarisation, and is defined as the angle between $\vec{n}$ and $\vec{z}_1$, with respect to the Λ_b^0 baryon rest frame. Angles ϕ_l and ϕ_h are measured between the decay planes of $\Lambda \rightarrow p\pi$ and $\mu^+\mu^-$ against the $\vec{n}$ axis, and are again measured in the reference frame of the particle from which they are produced.

If P_b is assumed to be 0 - meaning the Λ_b^0 baryon is produced unpolarised - then only 3 angles describe the distribution. With no preferred spin alignment to the beamline, the axis $\vec{n}$ is no longer fixed, as it was in the polarised case. Now, θ, ϕ_l and ϕ_h are not determined against a fixed axis; but a random one. All three of these are still measurable,

however will contain no physical information. In the place of ϕ_l and ϕ_h , the angle between the $p\pi^-$ and $\mu^+\mu^-$ decay planes ϕ is measured instead. The angle θ is not replaced by other measurements, and now not physically relevant.

In terms of the angles described here, the differential decay rate of $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$, assuming a non-zero transverse polarisation production is

$$\begin{aligned} \frac{32\pi}{3} \frac{d^6\Gamma}{dq^2 d\cos\theta d\cos\theta_l d\cos\theta_h d\phi_l d\phi_h} = & (J_1 \sin^2 \theta_l + J_2 \cos^2 \theta_l + J_3 \cos \theta_l) + \\ & (J_4 \sin^2 \theta_l + J_5 \cos^2 \theta_l + J_6 \cos \theta_l) \cos \theta_h + \\ & (J_7 \sin \theta_l \cos \theta_l + J_8 \sin \theta_l) \sin \theta_h \cos(\phi_h + \phi_l) + \\ & (J_9 \sin \theta_l \cos \theta_l + J_{10} \sin \theta_l) \sin \theta_h \sin(\phi_h + \phi_l) + \\ & \cos \theta \{ (J_{11} \sin^2 \theta_l + J_{12} \cos^2 \theta_l + J_{13} \cos \theta_l) + \\ & (J_{14} \sin^2 \theta_l + J_{15} \cos^2 \theta_l + J_{16} \cos \theta_l) \cos \theta_h + \\ & (J_{17} \sin \theta_l \cos \theta_l + J_{18} \sin \theta_l) \sin \theta_h \cos(\phi_h + \phi_l) + \\ & (J_{19} \sin \theta_l \cos \theta_l + J_{20} \sin \theta_l) \sin \theta_h \sin(\phi_h + \phi_l) \} + \\ & \sin \theta \{ (J_{21} \cos \theta_l \sin \theta_l + J_{22} \sin \theta_l) \sin \phi_l + \\ & (J_{23} \cos \theta_l \sin \theta_l + J_{24} \sin \theta_l) \cos \phi_l + \\ & (J_{25} \cos \theta_l \sin \theta_l + J_{26} \sin \theta_l) \sin \phi_l \cos \theta_h + \\ & (J_{27} \cos \theta_l \sin \theta_l + J_{28} \sin \theta_l) \cos \phi_l \cos \theta_h + \\ & (J_{29} \cos^2 \theta_l + J_{30} \sin^2 \theta_l) \sin \theta_h \sin \phi_h + \\ & (J_{31} \cos^2 \theta_l + J_{32} \sin^2 \theta_l) \sin \theta_h \cos \phi_h + \\ & J_{33} \sin^2 \theta_l \sin \theta_h \cos(2\phi_l + \phi_h) + \\ & J_{34} \sin^2 \theta_l \sin \theta_h \sin(2\phi_l + \phi_h) \}. \end{aligned} \tag{7}$$

The observables J_i are extracted via an angular analysis in individual q^2 regions, as the values are dependant on this interval. All J_i values are extractable via the moments of the distribution. The angular distributions consist of functions which are orthogonal, such that fitting is not required to measure them. The normalised moments are defined as

$$\begin{aligned} M_i = \frac{1}{N} \cdot \int \frac{d^6\Gamma}{dq^2 d\cos\theta d\cos\theta_l d\cos\theta_h d\phi_l d\phi_h} \\ \times f_i(\cos\theta, \cos\theta_l, \cos\theta_h, \phi_l, \phi_h) d\cos\theta d\cos\theta_l d\cos\theta_h d\phi_l d\phi_h, \end{aligned} \tag{8}$$

where

$$N = \int \frac{d^6\Gamma}{dq^2 d\cos\theta d\cos\theta_l d\cos\theta_h d\phi_l d\phi_h} d\cos\theta d\cos\theta_l d\cos\theta_h d\phi_l d\phi_h. \tag{9}$$

In these equations, f_i represents weighting functions unique to the moment in question. The normalisation factor N is proportional to the differential branching fraction in a given q^2 interval, and in terms of J_i , it is equal to

$$N = 2J_1 + J_2 \quad (10)$$

and with this, normalised moments can be described as

$$M_i = \frac{J_i}{2J_1 + J_2}. \quad (11)$$

The functions f_i are listed in Table 3, along with the observable they correspond to. All 34 functions are necessary for the polarised decay, while only the first 10 are needed to describe the unpolarised case. In this case, the two angles ϕ_l and ϕ_h are now combined as ϕ . When it comes to the actual measurement of moments from data, they are calculated as

$$M_i = \frac{1}{N_{\text{Sig}}} \sum_e^{N_{\text{Sig}}} f_i(\cos \theta^e, \cos \theta_l^e, \cos \theta_h^e, \phi_l^e, \phi_h^e), \quad (12)$$

where this runs over all signal events N_{Sig} , and assumes perfect efficiency, and no background contributions. To correct for this, each event is multiplied by ${}_s\mathcal{P}_n(y_e)/\epsilon$, where ${}_s\mathcal{P}_n(y_e)$ is an event weight to remove background, and ϵ is the efficiency - these will be described in Sections 5.8 and 5.9. This equation is derived from a Monte Carlo integration [39] of Equation 8.

The previous analysis [28] extracted θ_h and θ_l from data, and used these to measure the forward-backward asymmetries of the hadron and lepton systems. These values measure the preference of the p/π^- decaying from Λ baryon or the muons to be produced in a specific direction with respect to the origin particle's momentum. Using Figure 8 as a basis, in the case that $\cos \theta_h$ is greater than 0, the decay is forward - the z -component of the proton momentum is the same sign as the Λ baryon, while a backward decay occurs when $\cos \theta_h$ is less than zero and the components are of opposite sign. The same holds true for θ_l , measured using the μ^+ momentum vector. Also determined was the fraction of longitudinally polarised dimuons f_L . An additional forward-backward asymmetry, that of the production of the Λ baryon and the dimuon from the Λ_b^0 baryon - A_{FB}^{lh} - may also be measured [31], however this has not previously been determined by the LHCb. These observables can be expressed in terms of the normalised moments M_i [31], as

$$\begin{aligned} f_L &= 2M_1 - M_2, \\ A_{\text{FB}}^l &= \frac{3}{2}M_3, \\ A_{\text{FB}}^h &= \frac{2M_4 + M_5}{2}, \\ A_{FB}^{lh} &= \frac{3}{4}M_6. \end{aligned} \quad (13)$$

Observable	$f_i(\cos \theta, \cos \theta_l, \cos \theta_h, \phi_l, \phi_h)$	$f_i(\cos \theta_l, \cos \theta_h, \phi)$
M_1	$1/4(3 - 5 \cos^2 \theta_l)$	$1/4(3 - 5 \cos^2 \theta_l)$
M_2	$1/2(5 \cos^2 \theta_l - 1)$	$1/2(5 \cos^2 \theta_l - 1)$
M_3	$\cos \theta_l$	$\cos \theta_l$
M_4	$16 \sin \theta_l \cos \theta_h / \pi - 3 \cos \theta_h$	$16 \sin \theta_l \cos \theta_h / \pi - 3 \cos \theta_h$
M_5	$-32 \sin \theta_l \cos \theta_h / \pi + 9 \cos \theta_h$	$-32 \sin \theta_l \cos \theta_h / \pi + 9 \cos \theta_h$
M_6	$3 \cos \theta_l \cos \theta_h$	$3 \cos \theta_l \cos \theta_h$
M_7	$128/(3\pi^2 \cos \theta_l \cos(\phi_l + \phi_h))$	$128/(3\pi^2 \cos \theta_l \cos \phi)$
M_8	$32/(3\pi^2 \cos(\phi_l + \phi_h))$	$32/(3\pi^2 \cos \phi)$
M_9	$128/(3\pi^2 \cos \theta_l \sin(\phi_l + \phi_h))$	$128/(3\pi^2 \cos \theta_l \sin \phi)$
M_{10}	$32/(3\pi^2 \sin(\phi_l + \phi_h))$	$32/(3\pi^2 \sin \phi)$
M_{11}	$3/4(3 - 5 \cos^2 \theta_l) \cos \theta$	0
M_{12}	$3/2(5 \cos^2 \theta_l - 1) \cos \theta$	0
M_{13}	$3 \cos \theta_l \cos \theta$	0
M_{14}	$3 \cos \theta (16 \sin \theta_l \cos \theta_h / \pi - 3 \cos \theta_h)$	0
M_{15}	$3 \cos \theta (-32 \sin \theta_l \cos \theta_h / \pi + 9 \cos \theta_h)$	0
M_{16}	$9 \cos \theta \cos \theta_l \cos \theta_h$	0
M_{17}	$128/\pi^2 \cos \theta \cos \theta_l \cos(\phi_l + \phi_h)$	0
M_{18}	$32/\pi^2 \cos \theta \cos(\phi_l + \phi_h)$	0
M_{19}	$128/\pi^2 \cos \theta \cos \theta_l \sin(\phi_l + \phi_h)$	0
M_{20}	$32/\pi^2 \cos \theta \sin(\phi_l + \phi_h)$	0
M_{21}	$16/\pi \sin \theta \cos \theta_l \sin \phi_l$	0
M_{22}	$3/2 \sin \theta \sin \theta_l \sin \phi_l$	0
M_{23}	$16/\pi \sin \theta \cos \theta_l \cos \phi_l$	0
M_{24}	$3/2 \sin \theta \sin \theta_l \cos \phi_l$	0
M_{25}	$48/\pi \sin \theta \cos \theta_l \cos \theta_h \sin \phi_l$	0
M_{26}	$9/2 \sin \theta \sin \theta_l \cos \theta_h \sin \phi_l$	0
M_{27}	$48/\pi \sin \theta \cos \theta_l \cos \theta_h \cos \phi_l$	0
M_{28}	$9/2 \sin \theta \sin \theta_l \cos \theta_h \cos \phi_l$	0
M_{29}	$8/3(9 \sin \phi_h + 15 \sin \phi_h \cos(2\theta_l))/\pi^2$	0
M_{30}	$4/3(3 \sin \phi_h - 15 \sin \phi_h \cos(2\theta_l))/\pi^2$	0
M_{31}	$8/3(9 \cos \phi_h + 15 \cos \phi_h \cos(2\theta_l))/\pi^2$	0
M_{32}	$4/3(3 \cos \phi_h - 15 \cos \phi_h \cos(2\theta_l))/\pi^2$	0
M_{33}	$6/\pi \sin \theta \cos(\phi_h + 2\phi_l)$	0
M_{34}	$6/\pi \sin \theta \sin(\phi_h + 2\phi_l)$	0

Table 3: Weighting functions for calculation of moments, and determination of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ angular observables. Polarised and unpolarised functions given, respectively. Functions for M_{11} through M_{34} become 0 when the Λ_b^0 baryon has no transverse polarisation, as they are products of P_b and f_i .

For completeness, when considering the unpolarised case with the angles θ_h, θ_l, ϕ , the differential decay rate instead becomes:

$$\begin{aligned} \frac{8\pi}{3} \frac{d^4\Gamma}{dq^2 d\cos\theta_l d\cos\theta_h d\phi} = & (J_1 \sin^2\theta_l + J_2 \cos^2\theta_l + J_3 \cos\theta_l) + \\ & (J_4 \sin^2\theta_l + J_5 \cos^2\theta_l + J_6 \cos\theta_l) \cos\theta_h + \\ & (J_7 \sin\theta_l \cos\theta_l + J_8 \sin\theta_l) \sin\theta_h \cos\phi + \\ & (J_9 \sin\theta_l \cos\theta_l + J_{10} \sin\theta_l) \sin\theta_h \sin\phi, \end{aligned} \quad (14)$$

and the normalised moment distribution is

$$\begin{aligned} M_i = \frac{1}{N} \cdot \int \frac{d^4\Gamma}{dq^2 d\cos\theta_l d\cos\theta_h d\phi} \\ \times f_i(\cos\theta_l, \cos\theta_h, \phi) d\cos\theta_l d\cos\theta_h d\phi, \end{aligned} \quad (15)$$

with the normalisation factor calculated as

$$N = 2J_1 + J_2 = \int \frac{d^4\Gamma}{dq^2 d\cos\theta_l d\cos\theta_h d\phi} d\cos\theta_l d\cos\theta_h d\phi. \quad (16)$$

Thus, moments for events with the Λ_b^0 baryon being assumed to have zero transverse polarisation are calculated as

$$M_i = \frac{1}{N_{\text{Sig}}} \sum_e^{N_{\text{Sig}}} f_i(\cos\theta_l^e, \cos\theta_h^e, \phi^e). \quad (17)$$

In this case, the weighting function $f_i(\cos\theta_l, \cos\theta_h, \phi)$ only involves the three angles in the decay with physical information. These functions are also given in Table 3.

2.3 Analysis Strategy and Goals

In this analysis, it is not assumed that $P_b = 0$, and therefore the goal is to extract the 5 angles that describe $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$, using them to determine the distribution of moments for the first time. Due to the similar topology of the rare decay and $\Lambda_b^0 \rightarrow J/\psi \Lambda$, the decay is used to optimise and validate the analysis on the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay. This method expands on the last angular analysis, using all 5 dimensions in comparison to the 2 used previously [28]. We are able to use the moments to remeasure the previously published observables A_{FB}^l , A_{FB}^h and f_L , while making a first measurement of A_{FB}^{lh} , in addition to the 34 moments associated to the decay. These new methods of determining the observables allow a cross check of those previously measured, as well as to theoretical predictions. Additionally, we are able to compare our experimental moments to prediction. By determining these moments, it is easy to extract a lot of information about the decay which would otherwise be difficult to obtain. Along with the comparisons to

experiment and theory, we may also use the distribution of moments to confirm whether this analysis agrees that the Λ_b^0 baryon is produced unpolarised, as current findings imply. After the discussion of heavy flavour physics and the decay of interest, this thesis begins with an introduction to the LHCb experiment in Section 3. With the scientific and experimental groundwork in place, a brief discussion on the software used is given in Section 4. The variables used through this thesis are defined in Appendix A.

The work is carried out on the 3 fb^{-1} data collected during 2011 and 2012 by the LHCb experiment at the Large Hadron Collider. The data is subject to stripping, trigger and preselection requirements, as described in Section 5.1, removing decays of the wrong topography and those which are not well reconstructed by the experiment. This process leaves us with a much smaller, and better quality source of events to work with. The $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ signal events are found within this data using multivariate analysis (MVA), with a neural network being trained and optimised to maximise signal yield, while rejecting background. Optimisation of the network is performed by maximising signal significance, and to do this it is also necessary to model the decay of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ to obtain an event yield. The steps detailing these processes are discussed in Sections 5.2, 5.3, 5.4, 5.5 and 5.6. Using parameters extracted from simulated data, Section 5.7 details the invariant mass distribution of $\Lambda \mu^+ \mu^-$ being modelled with probability density functions, allowing for event yields to be extracted. To determine the moments from the final dataset, an analysis code was produced in order to compensate for detector effects, and to remove the final background events. Efficiency corrections are performed on the data, in order to compensate for detector effects on the dataset, as detailed in Section 5.8. Background subtraction is performed with event weighting, leaving only the signal events to contribute to the measurement of the moments, and this will be discussed in Section 5.9. Finally, uncertainties on our results are produced, as 5.10 describes.

The output of the analysis code was validated before allowing it to handle the final $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$, in order to ensure correct results. Simulated data was used to confirm the efficiency corrections, background subtractions, moment calculations and uncertainty measurements were performed successfully. A dataset of real $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events being used as a final verification, as the angular observables associated to this decay are already published. The steps used to validate the code are given in Section 6.

With the analysis code's output verified and validated, the angular and moment distributions for the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay were able to be determined, allowing for a new way to measure the previously published observables A_{FB}^l , A_{FB}^h , and f_L , as well as the previously unmeasured A_{FB}^{lh} , with relative ease. This method provides a complete set of the 34 moments that describe a non-zero transverse polarisation production $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay, along with the distributions of $\theta, \theta_l, \theta_h, \phi_l$ and ϕ_h , as well as the aforementioned observables. The results are presented and discussed in Section 7, with my final conclusions and discussions in Section 8.

3 The LHC and the LHCb Experiment

3.1 The LHC

The Large Hadron Collider [40] is a particle accelerator 27 km in circumference located under the Franco-Swiss border, and is the largest scientific experiment on the planet. Bunches of charged particles are accelerated in opposite directions around the beamline using radio-frequency cavities, while they're kept in the accelerator by superconducting magnets. The two beams reach energies of 6.5 TeV per beam as of 2015, resulting in a centre of mass (CoM) energy ($\sqrt{s}$) of 13 TeV. Occupying the tunnel originally inhabited by the LEP collider, construction was approved in 1995 and the first proton beams were accelerated in 2008.

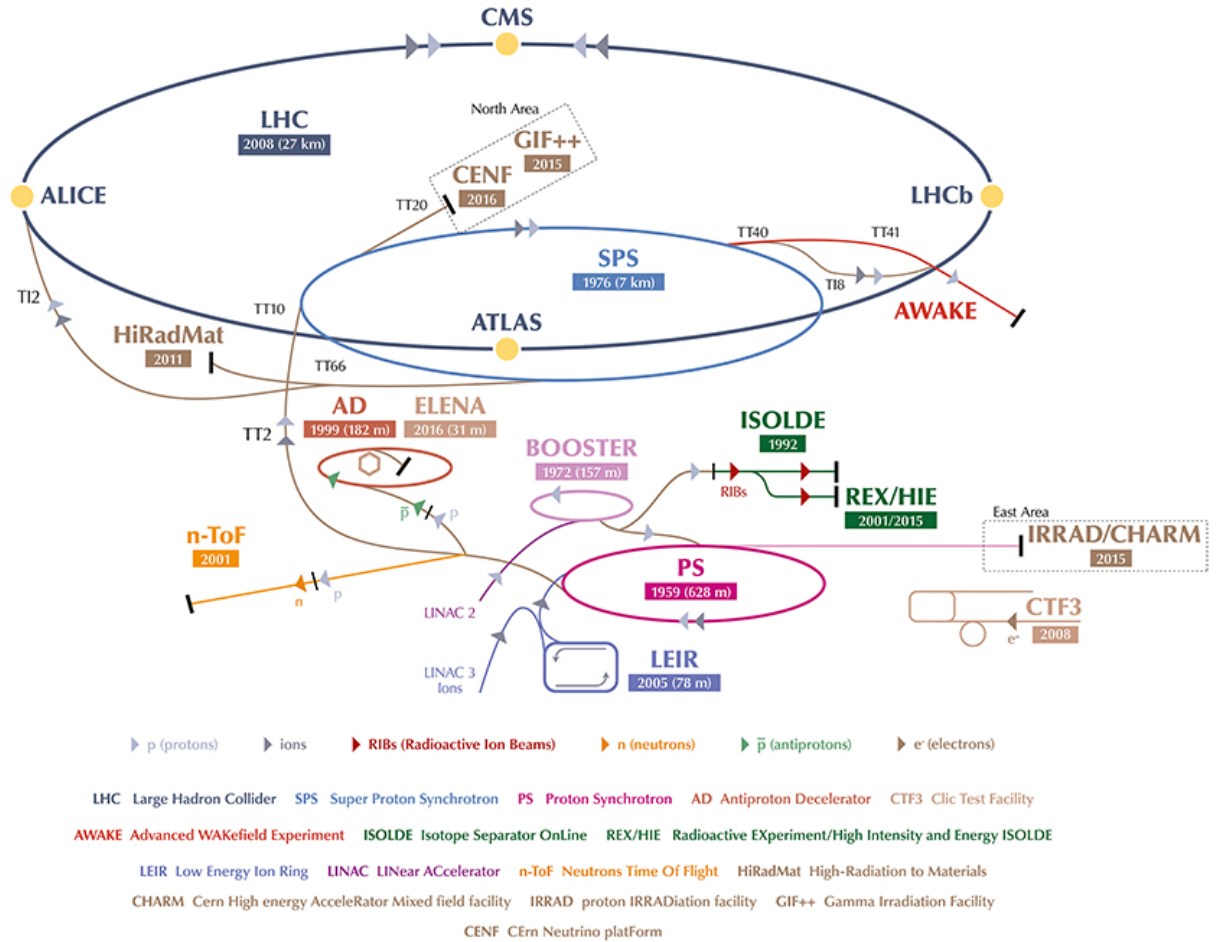


Figure 9: Layout of the experiments at the Large Hadron Collider. Figure taken from the CERN Document Server.

Protons are sourced by ionising hydrogen gas then accelerated by large potential differences in several stages [41]. The process starts with the linear accelerator Linac 2, into the Proton Synchrotron Booster, then cycling 3-4 times in the Proton Synchrotron, before they complete 12 cycles of the Super Proton Synchrotron. At this point, the protons have an average energy of 450 GeV, and are injected into the main LHC beamline in bunches. The layout of the experiments, and the accelerator complex is displayed by Figure 9. A total fill of the two beamlines takes about 260 seconds, before accelerating up to 6.5 TeV per beam over 20 minutes, and kept in orbit for hours under normal conditions. Proton beams can be held in storage in the main ring until collisions are required, with cryogenics keeping the machine cold against the synchrotron radiation produced by the accelerating particles [40, 41]. Without these cryogenics, the superconducting magnets would heat, resulting in high currents being passed through resistive components - destroying them.

While being home to many experiments [42], the most prominent at the LHC are ALICE (**A Large Ion Collider Experiment**), ATLAS (**A Toroidal LHC ApparatuS**), CMS (**Compact Muon Solenoid**), and LHCb (**Large Hadron Collider beauty**). Each experiment is specifically designed for individual physics purposes.

3.2 The LHCb Experiment

The LHCb detector [7, 43] - shown in Figure 10 - is a single arm forward spectrometer, covering the $2 < \eta < 5$ pseudorapidity range, where

$$\eta \equiv -\ln \left[\tan \frac{\theta}{2} \right], \quad (18)$$

and θ is measured against the beam. The detector is specifically designed to study the decays of hadrons containing charm or bottom quarks. As heavy flavour baryons and mesons tend to be produced at small angles relative to the beamline, the detector is designed to only cover these regions, minimising space and material costs. The LHCb experiment is designed to work at a low luminosity compared to the other experiments, with a design of one interaction per bunch crossing, in order to reduce the pile-up of events that would arise otherwise. Additionally, the sensitive VELO detector is designed to handle low particle flux, and an increased luminosity would damage its components.

In 2012, the maximum instantaneous luminosity was about 4×10^{32} events $\text{cm}^{-2}\text{s}^{-1}$; some 20 times lower than ATLAS/CMS. Despite this comparatively low rate, during Run 1 in 2011 and 2012, 3 fb^{-1} of data was collected. Using the fragmentation ratio of Λ_b^0 to B^0 meson [44], and the B^0 production cross section [45], it can be estimated that roughly 5 trillion Λ_b^0 baryons were created via pp collisions during Run 1. At a 100% efficiency, this would result in a maximum extractable yield of 5400 $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events from the dataset, assuming the PDG branching fraction [1]. In 2018, the LHCb detector will undergo an upgrade in order to increase the luminosity at which it can

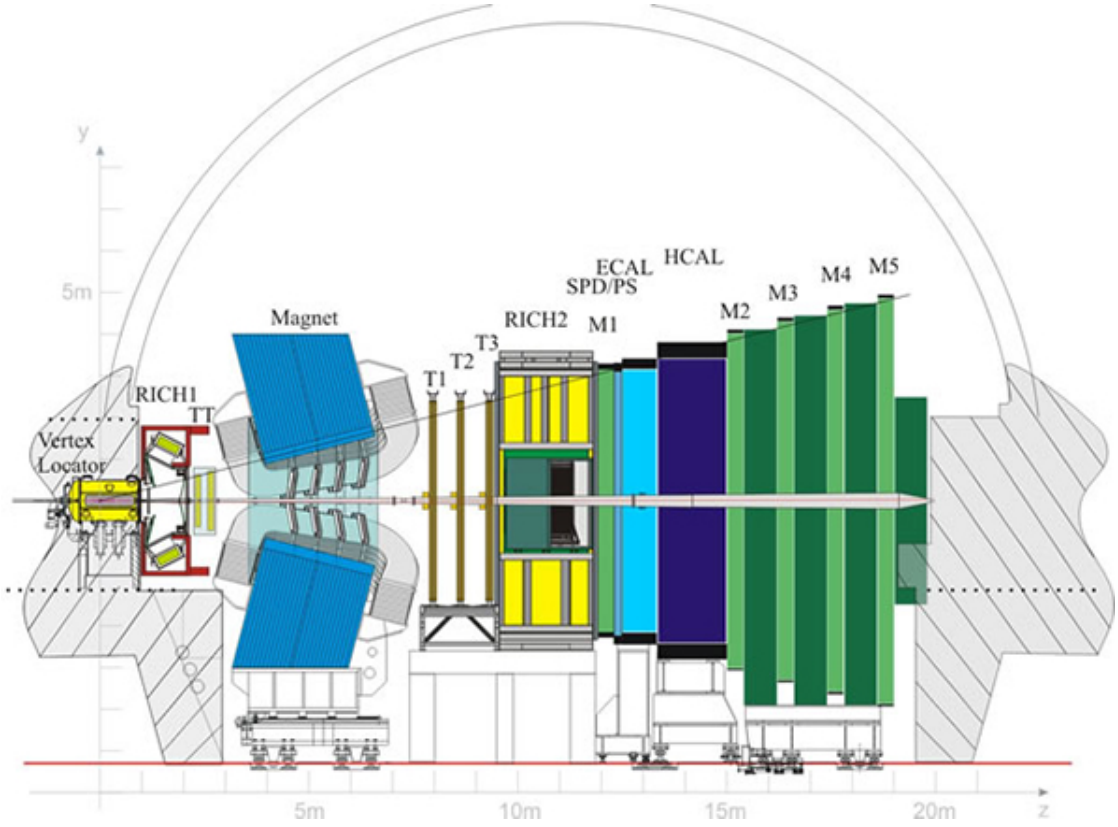


Figure 10: A schematic diagram of the LHCb experiment. Figure taken from [43]

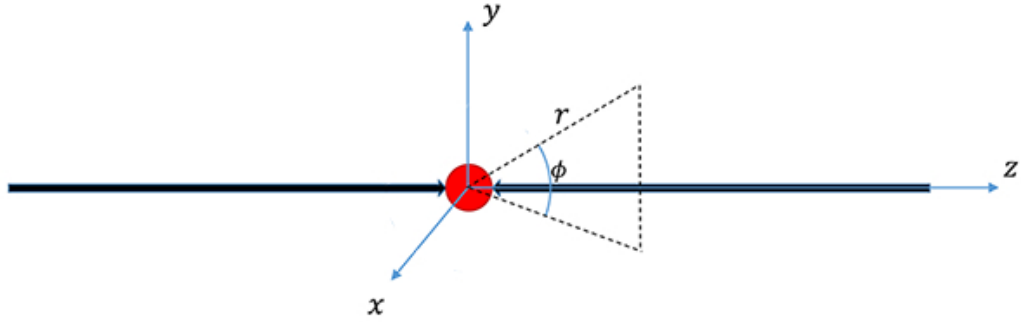


Figure 11: Definition of the co-ordinate system used in the LHCb. The solid black arrows are the proton beams, and the red circle is the interaction point.

operate [46]. Currently, the aim is to record an integrated luminosity of 50 fb^{-1} over a 10 year operating period, a huge improvement of the aforementioned 3 fb^{-1} in 2011 and 2012.

The LHCb is a specialised ‘beauty factory’, an experiment specially designed for decays of bottom quark containing hadrons [7, 43]. To this end, the detector is designed

with systems to allow for hadron and lepton identification, meaning the LHCb is able to measure the properties of B-meson decays, as well as those of b-baryons. Additionally, the LHCb experiment has tracking systems throughout the length of the detector, meaning mass and momentum resolution is better than achievable in ATLAS/CMS, due to increased tracking length. In its relatively short operational period, analyses performed on data taken by the LHCb experiment have returned the first significant indications of pentaquarks [47], the measurement of the branching fraction of the heavily suppressed $B_s \rightarrow \mu^+ \mu^-$ decay [48], as well as many papers on flavour changing neutral current decays.

3.3 The Bending Magnet

The LHCb detector utilises a large dipole magnet operating in the y axis [43, 49] in order to deflect charged particles in the $x - z$ plane, with the curvature of these paths being used to estimate their momenta. The magnet is chosen to be warm, as this allows for the polarisation to be flipped as necessary, without having to allow the equipment to warm up and then be cooled back down again. Full use is made of this, with data being collected for both magnetic field directions.

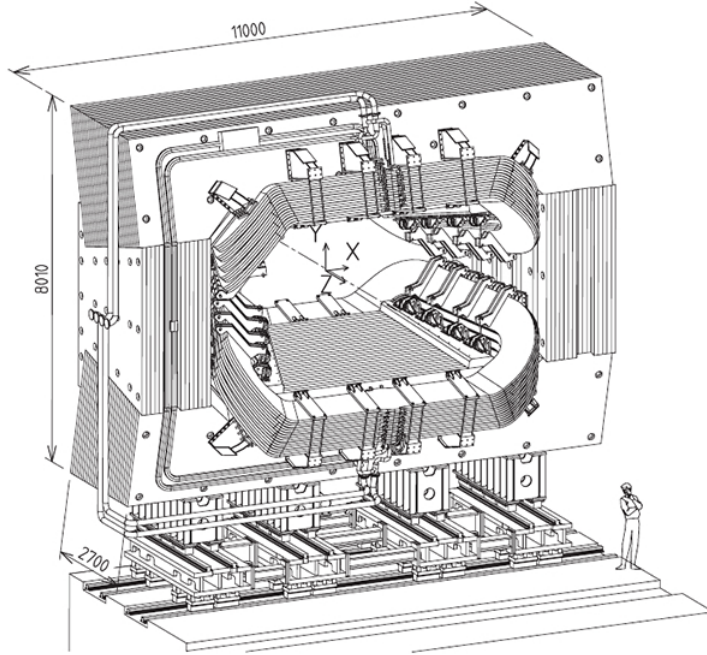


Figure 12: Schematic view of the LHCb Bending Magnet. Units are in millimetres. Figure taken from [43]

The magnet itself is comprised of 2 identical coils, each comprised of 225 turns of wire. These are placed in a magnet yoke, mirror symmetrical to one another. In total, this

arrangement has a mass of about 1600 tons, and is designed to handle a current of 5850 Amps. The resultant magnetic field has a bending power of 4 Tm [49]. The complete assembly is 11.0 by 8.01 by 2.7 metres, with a schematic view being displayed in Figure 12

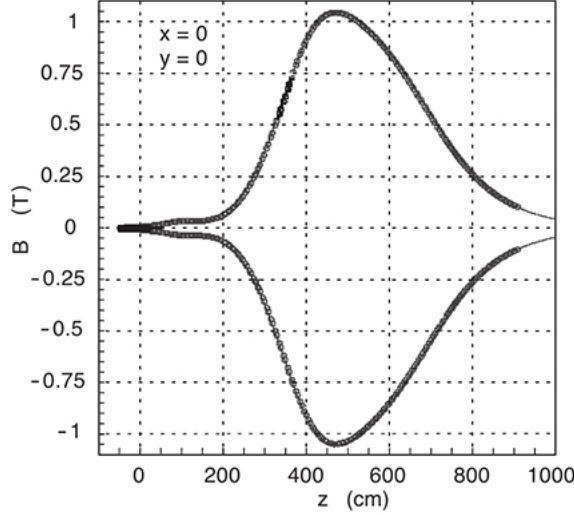


Figure 13: Magnetic field strength as a function of distance travelled downstream. The two lines represent the field strength in both the up and down magnet configurations. Figure taken from [43]

The motion of a charged particle moving perpendicular to a constant magnetic field is given as

$$r = \frac{mv}{Bq}, \quad (19)$$

where m is particle mass, v is particle velocity, B is the magnetic field strength, q is particle charge, and r is the radius of curvature - i.e the particle will follow a circular path. As can be seen, curvature is proportional to momenta and it follows that high momentum particles will have a greater radius of curvature, and as such are less deflected. As the sign of the particle charge is determined by the direction of its deflection in the magnetic field, and dB/dl is known, the momentum of the particle can be determined via measurement of the track curvature within the magnetic field, using data from the Tracking System. With this, the sign of the charge on the particle can be determined, and its magnitude is assumed to be 1. The magnetic field has influence over a large distance, as Figure 13 shows, with $z = 0$ representing the interaction point. The strength of the field present in the VELO is small enough however that charged particles can still be thought to be moving in straight lines; allowing for easier calculations of impact parameters, and other similar quantities.

3.4 The Tracking System

The tracking system at the LHCb experiment is comprised of the Vertex Locator (VELO), the Turicensis Tracker (TT), and the T1, T2 and T3 trackers. The VELO and the TT use silicon strip detectors, with the T1, T2, and T3 utilising the silicon strip detectors around the beamline only, and straw drift tubes elsewhere.

The VELO [43, 50] is a silicon strip detector surrounding the pp interaction region. Its function is to separate primary and secondary decay particle vertices, and measure the distance travelled from the primary vertex r and the angle in the $x - y$ plane ϕ , allowing the trajectory of charged particles to be mapped. Figure 11 shows the co-ordinate systems in the LHCb. A Λ baryon exists for over 180 times longer than a Λ_b^0 baryon [1], resulting in good separation between primary and secondary vertices. These displaced secondary vertices are characteristic of b and c hadron decays. When charged particles pass through the strips, electron-hole pairs are created, which drift to electrodes to allow determination of position of the hit, with these allowing for tracks to be reconstructed. During injection, the VELO is moved away from the beamline in order to protect it from damage caused by the unstabilised proton beam, as intense radiation would quickly damage them. Once the fill is complete, the detector returns to its ‘Physics’ position, 7 millimetres from the beamline. The detector is designed to cover a pseudorapidity range of $1.6 < \eta < 4.9$ - approximately 14.9 to 398.4 milliradians. Within the LHCb acceptance range of ± 250 mrad vertically, and ± 300 mrad horizontally; all tracks pass through at least 3 modules.

As displayed in Figure 14, there is smaller spacing of the silicon detector chips is greater closer to the pp interaction point, as this minimises the flight distance between particle decay and first measurement, and allows for accurate track reconstruction for particles which decay shortly after production. The coordinates measured of particle hits by the VELO are vitally important in track reconstruction. Before passing through the bending magnet, particles pass through the TT [43] - another silicon strip detector. The TT provides a measurement just before the particle enters the magnetic field and gets deflected. This aids calculation of particle momentum.

Immediately succeeding the magnet lies T1, T2 and T3, and the outer regions of these are the Outer Tracker (OT). The OT uses drift tubes filled with an argon/carbon dioxide gas at a 70% to 30% mix. The tubes also contain a thin wire anode in the centre of each. As particles pass through, they ionise the gas, and the resulting free electrons are detected by the anode. By measuring the drift time of these electrons, the hits are recorded, with drift time and straw position proving a coordinate of the hit in the detector. The chosen gas mixture provides very short drift times under 50 ns, comparable to the proton bunch spacing in the beamline during Run 1. In total, the Outer Tracker contains about 55000 drift tubes, and has a total active area of 5971×4850 mm². In the centre of the OT, surrounding the beamline is the Inner Tracker, another silicon strip setup, which along with the TT detector before the bending magnet, forms the Silicon

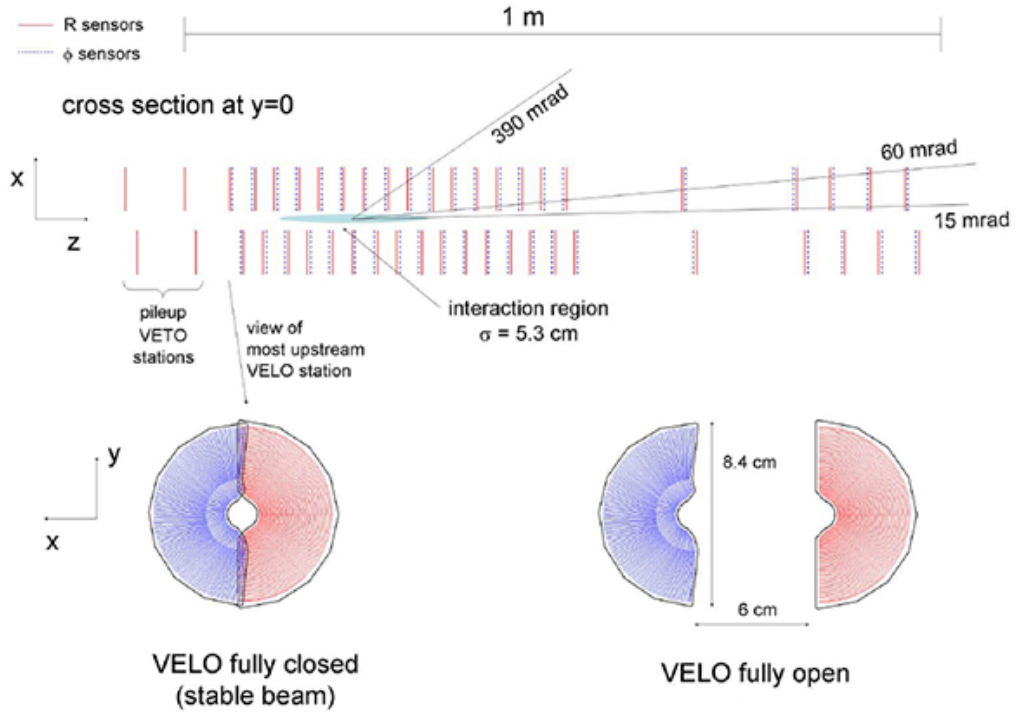


Figure 14: Cross section of the VELO in its closed state. Figure taken from [43]

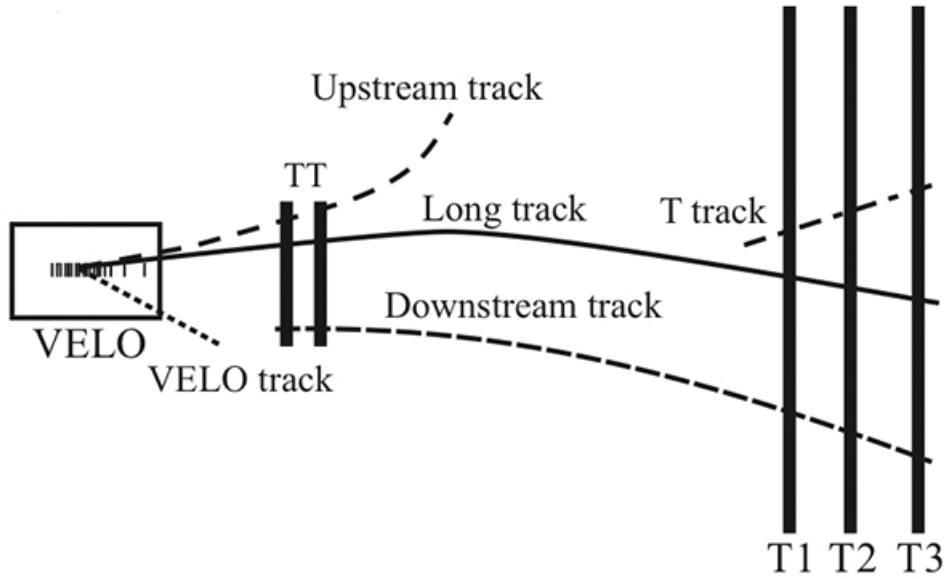


Figure 15: Definition of various track types as they pass through the tracking systems. Track types are defined by which parts of the experiment they are recorded in. For this analysis we split our dataset into Long and Downstream tracks. This is done as the two track types have different resolutions, with Long tracks better measured than Downstream. Figure taken from [43]

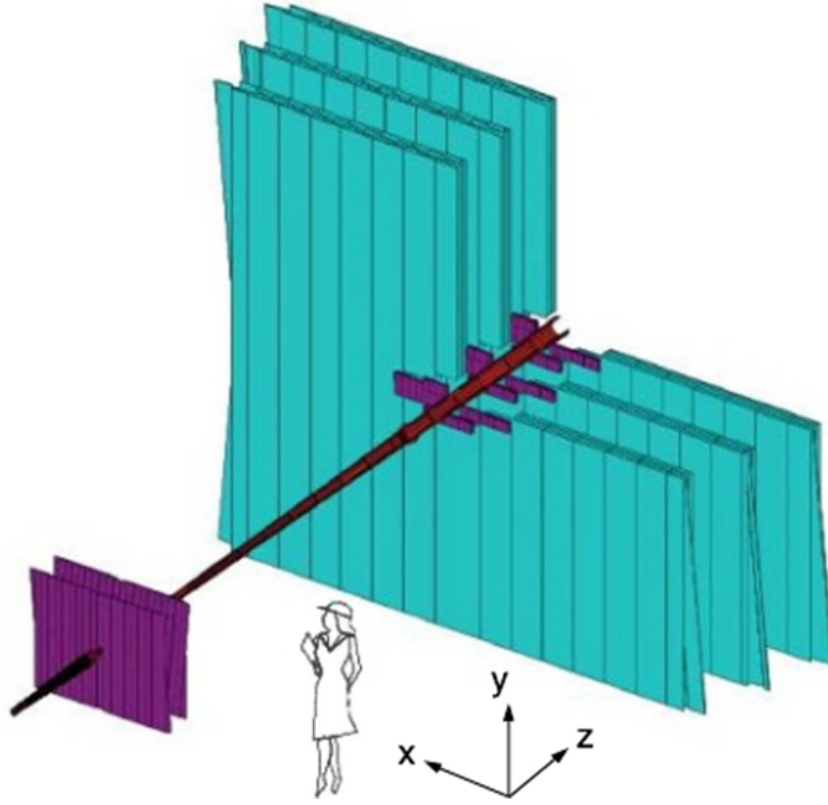


Figure 16: The cyan unit is the OT, while the purple is the whole of the ST. The TT is on the left, and the IT is on the right. Figure taken from [43]

Tracker. The Inner Tracker is necessary to detect higher momentum particles that do not have their paths deflected greatly in the magnetic field. Due to the higher particle flux surrounding the beamline, the silicon detectors are chosen for high resolution in order to separate and identify the individual tracks, as well as fast readout speeds. Due to the high cost of these however, in the lower flux areas the drift straws provide ideal coverage.

Tracks with hits in the VELO and either the ST or OT are labelled Long (LL), while those only recorded in the ST and OT are Downstream, as displayed in Figure 15. As they've passed through additional tracking systems, the Long tracks have better momentum resolution, and therefore Λ baryons formed from Long track protons and pions have better defined masses when compared to those reconstructed from Downstream tracks. While other tracks categories are present, this analysis focuses on LL and DD candidate differentiation only, as these are the only tracks that pass through a great enough proportion of the tracking system to be completely reconstructed. The Tracking System and the RICH detectors both detect only charged particles, via ionisation processes or emission of Cherenkov photons. The neutral Λ baryon is subsequently reconstructed from these charged tracks, while the Λ_b^0 baryon is reconstructed from Λ and dimuon tracks originating from a common vertex. Due to the long life time of the Λ baryon, many will

not decay until after the TT, and the majority of tracks used to reconstruct them will be Downstream. Due to this, it is advantageous to separate both the tracks of the Λ baryon and its decay products by their type, as to improve resolution on the Long tracks.

3.5 Ring Imaging Cherenkov Detector

Cherenkov radiation is produced as a result of a charged particle travelling through a medium at a velocity greater than the speed of light in that medium. While the speed of light in a vacuum is fixed and unbreakable, in other materials, its velocity will be reduced by an amount dependant on the refractive index of the material. Due to this, it is possible for particles to exceed the speed of light in the medium, and emit Cherenkov light. The angle (θ_c) of this emission of this radiation is dependant on both the refractive index of the medium (n), and the velocity of the charged particle (v), given as

$$\cos \theta_c = \frac{c}{nv}. \quad (20)$$

The LHCb experiment uses two Cherenkov detectors, RICH-1 and RICH-2 [43, 51]. RICH-1 is located immediately after the VELO, and is designed to identify particles with low track momentum, using a perfluorobutane (C_4F_{10}) gaseous radiator with a refractive index of 1.0014 [51]. Further down the beamline is RICH-2, which identifies particles with high momentum tracks. This detector instead uses tetrafluoromethane (CF_4), which works for a 15 to 100 GeV/ c momentum range and has a refractive index of 1.0005 [51]. RICH-2 also has an acceptance range in the low angular region, where the high momenta particles reside. The layout of the detectors are displayed in Figure 17. The aerogel feature of RICH-1 was only present in 2011 data taking.

As the particles travel through the media and emit Cherenkov radiation, the photons are reflected via a spherical and plane mirror onto a plane of hybrid photon detectors (HPD), where they are detected and recorded. Figure 18 describes the layout of the HPD. For a low momentum particle, the average number of photons detected is given by

$$N = \frac{N_c q^2 \sin^2 \theta_c}{(1 - \frac{1}{n^2})}, \quad (21)$$

where q is the particle charge and N_c is the optical response of the detector, or a measure of how quickly photons can be detected and recorded - the main limiting factor. From this, the statistical error (σ_m) on the mean value of the distribution of emitted Cherenkov photons (σ) is given as

$$\sigma_m = \frac{\sigma}{\sqrt{N}}. \quad (22)$$

The HPD units work by transforming incident photons into photoelectrons in a photocathode, and these are then accelerated by a 20 kV voltage onto the rear silicon detector. With the known radiator refractive index, and the momentum of the particle, the expected value of θ_c for each particle type may be determined, and the measured angle

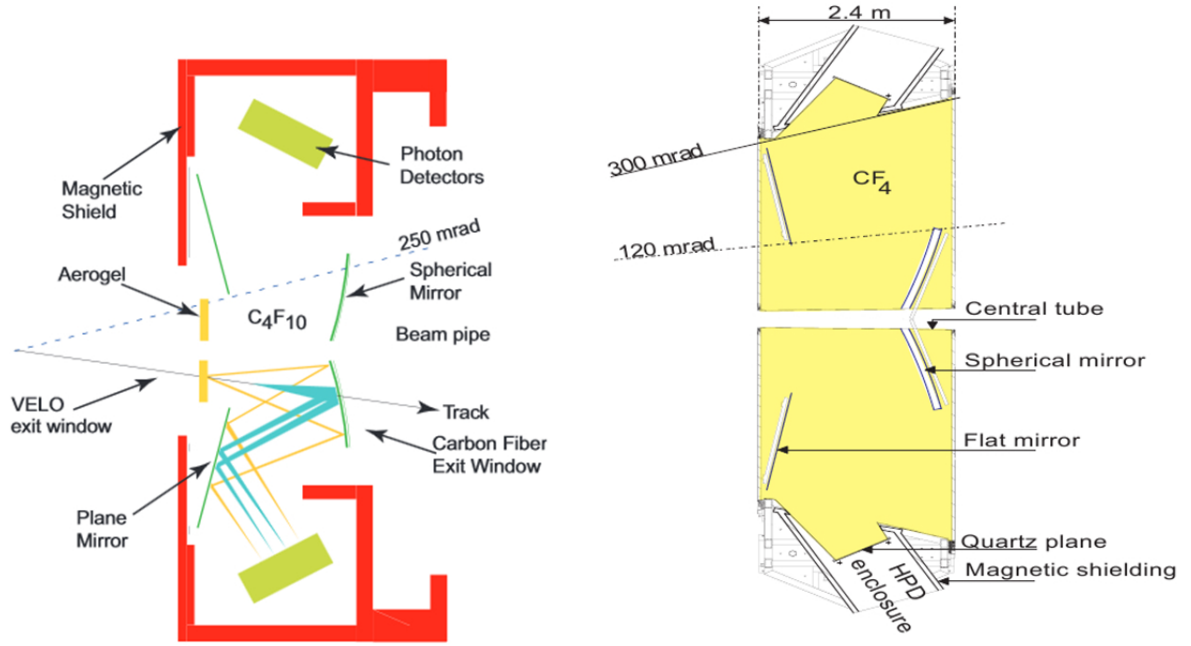


Figure 17: Left - RICH-1 layout. Right - RICH-2 layout. Figures taken from [43]

is compared for consistency with the expectation for different particles. From this, the likelihood of a track being produced by a specific particle is determined. The predicted Cherenkov angles for various charged particles passing through either RICH detector are presented in Figure 19.

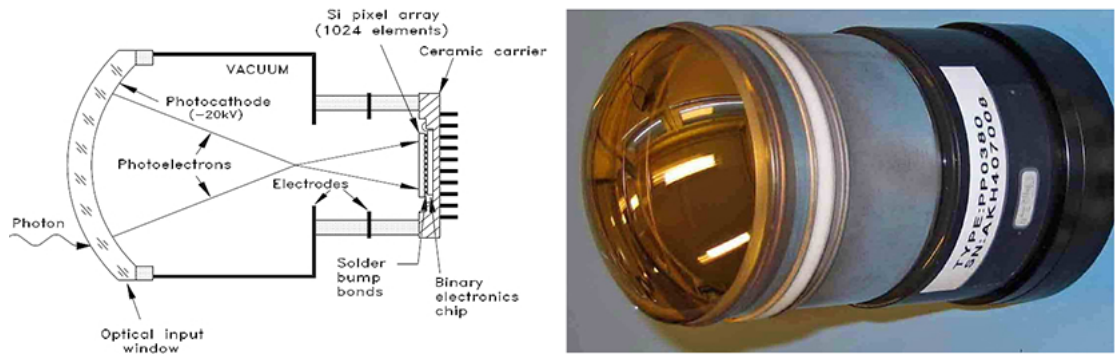


Figure 18: A schematic view and photograph of the Hybrid Photon Detectors used in RICH-1 and RICH-2. Figure taken from [43]

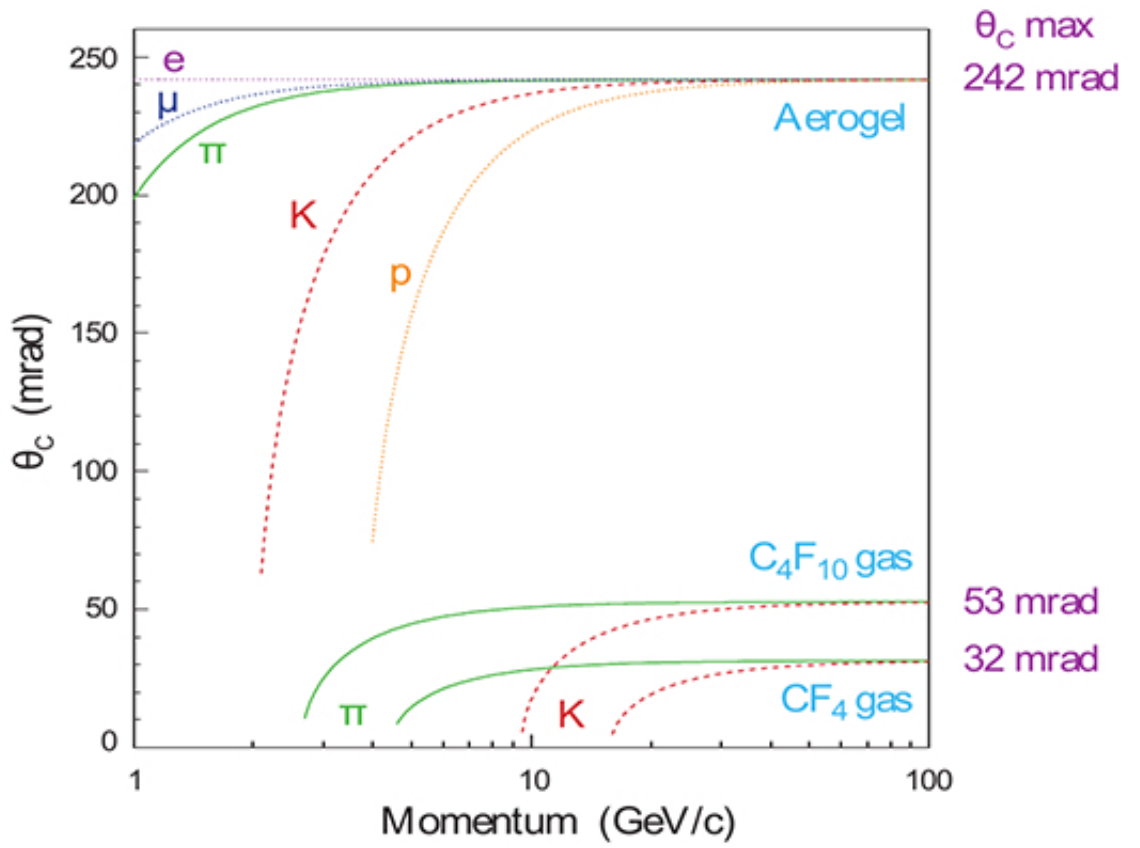


Figure 19: Expected Cherenkov angles (θ_c) for each radiator used in RICH by particle species. The use of aerogel was only present during the 2011 data taking. Figure taken from [43]

3.6 Calorimeter System

The LHCb experiment utilises a multi-layer calorimeter system [43, 52], comprising of the Scintillation Pad Detector (SPD), the Pre-Shower Detector (PS), the Electronic Calorimeter (ECAL), and the Hadronic Calorimeter (HCAL). The system is designed to bring particles to a halt as they pass through, sampling their energy, as well as providing a form of particle identification. As part of the trigger system, the readout of hits in the calorimeter is required to be fast, with decisions being made $4 \mu\text{s}$ after interaction. The layout of the detectors is shown in Figure 20.

The SPD and PS are used as particle identification, with the SPD separating charged and neutral particles, with the PS then aiming to determine what those particles are. Both consist of scintillation pads, separated by sheets of lead approximately 1.4 cm thick - or two and a half radiation lengths. As particles pass through transparent material, light is produced via interactions with the media, slowing said particle, and allowing sampling of the energy. Scintillation light produced by particles travelling through these sheets are captured with wavelength shifting fibres. While particles pass through the lead sheets, further energy is lost via physical interactions [1], including pair production, ionisation processes and scattering effects, which are again measured by scintillation.

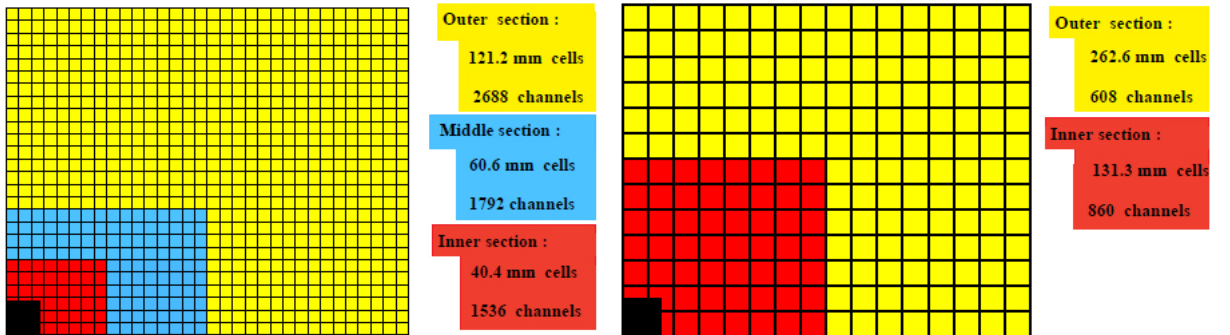


Figure 20: Segmentation of the SPD/PS and ECAL (left) and HCAL (right). The dimensions presented in the left figure are for the ECAL. Quarter of each face is shown. Figures taken from [43]

Following identification, the particle reaches the ECAL, made up again of alternating scintillating tiles and lead plates, with smaller units surrounding the beam for higher resolution, and larger ones at greater angles. The ECAL specialises in detection of energy losses from electronically interacting particles, as these lose energy in shorter distances through the calorimeter than hadronic interactions. It is thus necessary to have the ECAL closer to the interaction point than the HCAL. The HCAL therefore sits down the beamline, and is comprised of layers of scintillators and iron sheets, running parallel to the proton beams, as shown in Figure 21. For optimum resolution, the LHCb experiment requires full measurement of high energy photons, and as such the ECAL stretches 25 radiation

lengths to ensure this. The HCAL is much shorter however, as there is no requirement to fully contain any of the particles interacting here, and this also saves space. Both these calorimeters work in a similar way to the SPD and PS, scintillation light is produced and measured when the incident body passes through, or when secondary particles produced via interactions in the metal plates they travel through. Each interaction deposits a little energy into the detector, and by following paths particles make through the calorimeters, the total energy is determinable.

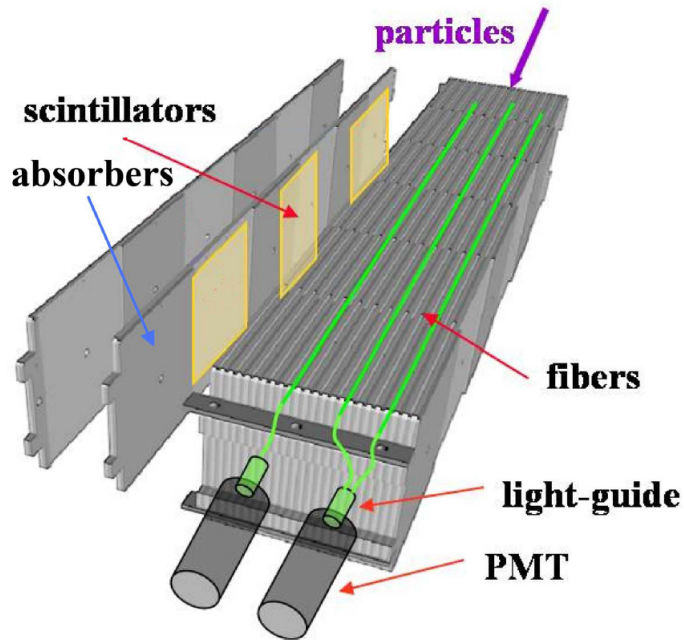


Figure 21: Schematic view of a unit used in the HCAL. Figure taken from [43]

3.7 Muon Detectors

The last stations of the experiment comprise of 5 muon detectors - M1 through M5, each separated by sheets of iron. These are needed to measure the energy and momenta of muons, which are not detectable by the calorimeters [43]. The iron sheets only allow energetic muons to pass through, with the lower energy ones being stopped. The muon detector is another gas based system, utilising carbon dioxide, argon and tetrafluoromethane, with the resultant ionisation of these gases via muon interactions detected by electrodes.

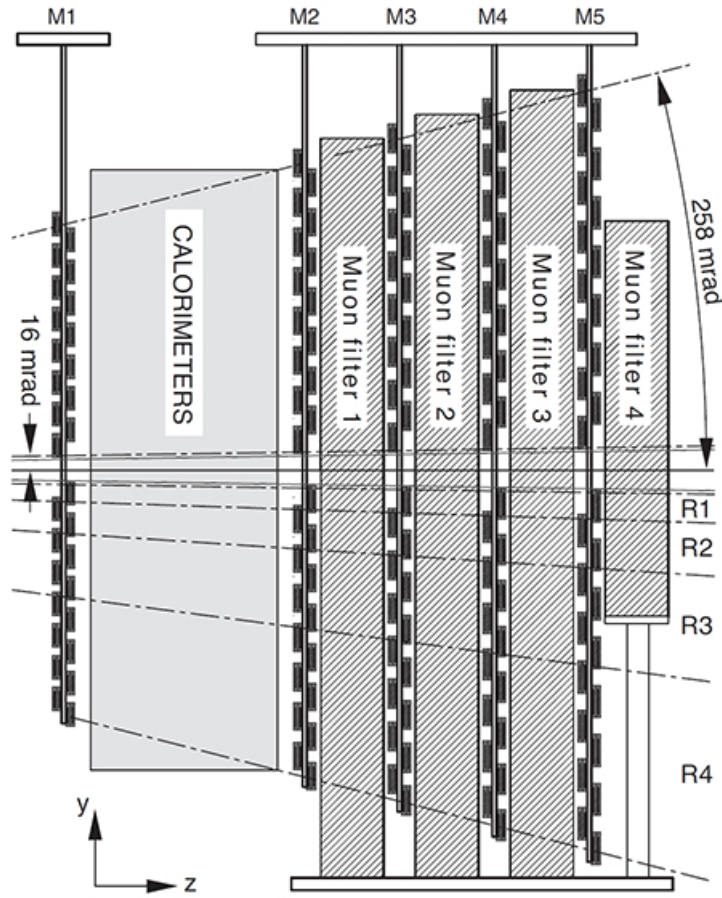


Figure 22: Schematic view of the LHCb muon system. Figure taken from [43]

In addition to the 5 stations, the whole detector is split into the regions R1 through R4, as displayed in Figure 22. These regions are at increasing angular distance from the beamline, but have approximately the same acceptance in each. Each region has a granularity that is appropriate to the particle density through it, which is intended to keep occupancy the same over the whole detector. Granularity is increased in the

horizontal plane, as this gives more accurate determination of the muon momentum.

The detector covers ± 300 millirad horizontally, and ± 250 millirad vertically, with the first of the stations located before the SPD in the calorimeter set up, and the remaining 4 located after HCAL. The muon system is one of the detectors used in the first level of the trigger system, and so information from events in the detector must be measured quickly - within 20 nanoseconds to be precise. The detectors are required to provide information on muon momentum for the first level of the trigger, and muon identification for the successive levels.

Muons require specialised detection, as their increased mass relative to the electron means that they do not lose energy as readily as electrons, and also appear to be charged hadrons in the tracking system. Additionally, as muons only interact electromagnetically, they very rarely lose energy from interactions with nuclear matter. As muons are so minimally ionising, they are the only particles expected to make impacts in the muon stations, and so tracks here are assumed to be due to them. Due to all these factors, the energy loss per unit of distance travelled in matter is much less than other particles seen in the experiment, including charged pions, which have a similar mass to charge ratio, but may be stopped via hadronic processes. Specialised detectors are developed for the sole purpose of sampling the energy and measuring the properties of these penetrative leptons, despite still not stopping them.

3.8 Trigger

The trigger at the LHCb utilises both hardware and software based systems, these being the Level 0 (L0) and High Level Trigger (HLT), respectively [43, 53]. Due to the huge number of collisions occurring every second, it is impossible to record each and every event, due to limits on how fast the data can be transmitted and written, and by actual available storage space. Assuming the luminosity is chosen to provide one interaction per beam crossing, and one crossing occurring every 50 ns during run 1, and every 25 ns in Run 2 - roughly 20 million/40 million events happen each second. Less than 1 in every 100 inelastic events contain b-quarks [54] however, and so the L0 trigger is required to reject 99% of inelastic pp collisions. Thus, the Level 0 trigger looks for evidence of high transverse momentum (p_T) events, with the HLT trigger partially recombining vertices, and analysing tracks through the detector, and only records those of meaningful scientific worth. This greatly reducing the data output, importantly to under current computational limits.

The L0 trigger uses information from the calorimeters and muon systems as a means of making its decisions, with the L0 Muon trigger only accepting events with muons with p_T greater than 1.48 GeV/ c in the $\sqrt{s} = 7$ TeV data or 1.76 GeV/ c for $\sqrt{s} = 8$ TeV. The L0 is designed to output events at a rate of 1 MHz, and the muon

system accounts for 400 kHz of that. The remaining output comes from the hardware trigger on the electronic and hadronic calorimeters, with a threshold requirement in the $\sqrt{s} = 8$ TeV data of 3 GeV of transverse energy (E_T) - the energy associated to a particle's transverse momentum - being deposited into the ECAL, while the HCAL needs an E_T greater than 3.68 GeV. These contribute an output of 150 kHz and 450 kHz, respectively. The hadron calorimeter L0 trigger is designed to reject high p_T neutral pions - as these represent a large source of background. This is achieved via selection of events with high transverse momentum electrons, as b-hadron decays will tend to produce particles with significantly higher transverse momentum than other decays. The hardware calorimeter only looks for the highest energy tracks, and by making this selection on electrons, events containing the more massive neutral pions can be rejected.

The software based HLT trigger works in 2 stages, HLT1 and HLT2, and further reduces the rate of events that pass L0. HLT1 works by performing a partial reconstruction of the event. The events are first reconstructed from their tracks in the VELO, before events with large impact parameters are traced back to the main tracking system, along with events that can be matched to hits in the muon detectors, and are consistent with a high transverse momentum event. In this analysis, we require one of the final state particles to have a transverse momentum of 1.4 or 1.6 GeV/ c^2 , for 7 and 8 TeV data, respectively. Should the particle be a muon however, the minimum p_T is instead 1.0 GeV/ c^2 . These tracks are required to have impact parameter greater than 100 μ m with respect to each primary vertex in the event. The HLT1 reduces output from 1 MHz to 70 kHz.

HLT2 reconstructs tracks that pass HLT1 with a minimum p_T of 300 MeV/ c , from at least two particles coming from a displaced vertex, consistent with b-hadron decays. This reduced the output to 3/5 kHz for the 2011/2012 data taking, and rising to about 12.5 kHz [55] for Run 2. After the HLT2, the events are written to disk. Additionally to this, heavy flavour decays with final state hadrons, such as $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ are selected with a multivariate classifier that identifies decays with two, three and four track vertices. This topological trigger is based on a Boosted Decision Tree [56], and has a typical efficiency of 60%. The trigger cuts are also split into two more types, Trigger On Signal (TOS) or Trigger Independent of Signal (TIS). If a signal candidate is sufficient to active the trigger itself, then it is classed as a TOS event. A candidate where something else in the event activates the trigger is classified as TIS.

4 Analysis Software

4.1 Introduction to NeuroBayes

NeuroBayes [57] is a Multi-Variate Analysis (MVA) program used to aid the separation of signal from background events. Neural network programs are designed to work in a similar way to a brain, with neurons firing when they're sufficiently stimulated by neighbouring neurons. Nodes are used in place of neurons, with input and output layers connected by hidden nodes, and the information is stored in the connections between these nodes - an example is given in Figure 23. They can exploit correlation between the input variables, and do not require all input variables to be filled for every event. As a multi-variate program, NeuroBayes can handle many individual variables, allowing for a model of the process of interest to be described and learned by the network. The training process provides an output of the ranking of each variable, giving a measure of how significant they are in describing the process, and how strongly they correlate with other inputs. The output of the neural network provides a simple and efficient way to further reduce the background events in the dataset, which would be difficult to do efficiently by hand. The neural network is created with a series of nodes, of which the value of a given node (j) in a given layer (n) is given by the weighted sum of the nodes in the previous layer. This is calculated as

$$x_j^n = g \left(\sum_k w_{jk}^n \cdot x_k^{n-1} + \mu_j^n \right), \quad (23)$$

where $g(t)$ is a sigmoid function, and μ_j^n is a bias node, which effectively acts like a neuron in a brain, "firing" when its input reaches a certain threshold. The training process adjusts the weights w_{jk}^n to minimise a loss function, the measure of the agreement of outputs against the desired values. This function is usually a sum of quadratic deviations, or entropy. NeuroBayes consists of a Teacher and Expert program, in which the former is run over existing data to learn the relationships and correlations between variables to train the neural network, and the latter uses the result of the training on new data. The Expert assigns each event in the dataset with a value between -1 and 1, corresponding to how background or signal like it believes it to be. The requirement on neural network output is then optimised, to maximise acceptance of wanted data. These abilities of the program make it a very popular choice in not only particle physics, but also in industry, such as banking and insurance applications [57].

The program makes use of Bayesian statistics, and rates the events in the dataset based on their likelihood of being signal or background. Like all neural networks, NeuroBayes makes use of theoretical information. By implementing Bayesian statistics to make use of this, *a priori*, unphysical results may be avoided, as the software has been provided with a source of physically correct solutions. Bayes' Theorem states

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (24)$$

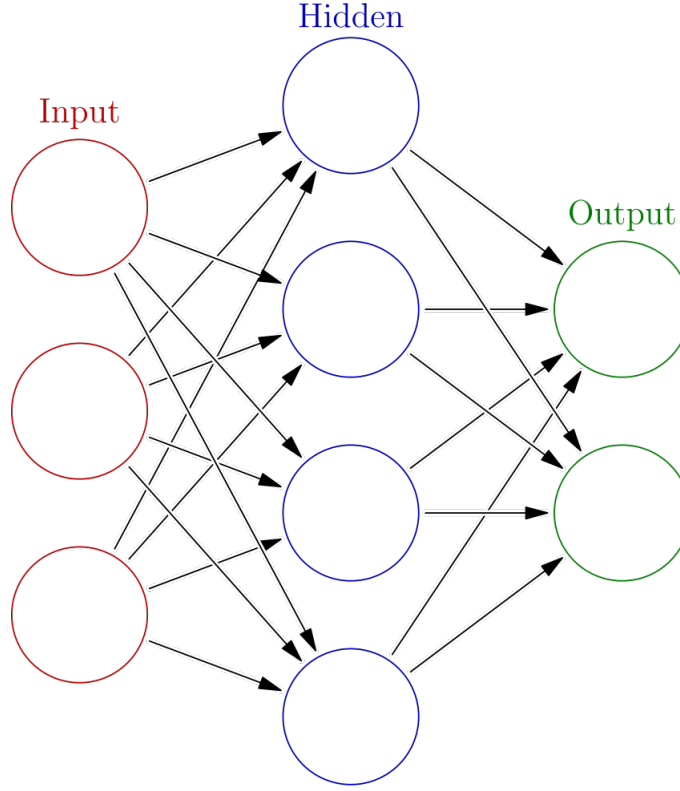


Figure 23: Nodes involved in a typical neural network. In this analysis, there are many more input nodes, corresponding to the training variables, with a single output. It is the job of the hidden layer to transform the inputs in to something that may be used by the output. Figure taken from Wikimedia Commons.

where $P(A)$ and $P(B)$ are individual probabilities, with $P(A|B)$ and $P(B|A)$ being the probabilities of observing one after having already seen the other. In this case, A represents theory, and B is data. It is the goal of NeuroBayes to use the Bayesian theorem to make a statement on the posterior probability of an event being signal - that being the probability of a random event occurring after previously obtained information is examined, as well as relative background and evidence is taken into account.

4.2 ROOT, RooFit, and $gPlot$

ROOT [58] is a data processing framework developed at CERN, and is the brainchild of Rene Brun and Fons Rademakers. It allows for easy analysis of vast amounts of information at once, for simulations to be produced, plots to be easily created, and can be worked into existing code very simply, providing a huge scope of possibility for analysis.

Within ROOT exists RooFit [59, 60], a toolkit for modelling data. The framework provides a way to model the distribution of observables in the data with respect to parameters of interest, allowing for easy creation of fit models. Users are able to create complex probability density functions (PDF) using simple single functions. RooFit is then able to perform a maximum likelihood fit, or χ^2 fit using MINUIT [61], producing the values necessary to describe the function afterwards. For this analysis, extended maximum likelihood fits were used in order to model the invariant mass distribution and extract yields. The specifics of this process are detailed in Section ???. RooFit allows the user to easily create PDFs - multiple functions can be combined into the total fit. RooFit was used in this analysis to create models using the simulated data, and then apply them to the real dataset. This fitting is discussed in detail in Section 5.7.

The *sPlot* package [62] is a data analysis program that statistically discriminates components in data, and provides the distribution of these. The goal of the program is to infer the behaviour of individual event sources with respect to a control variable, without the need for prior theoretically obtained knowledge. *sPlot* is able to use the probability density functions obtained from the extended maximum likelihood fit. With these, the dataset can be weighted against the likelihood of an event belonging to a certain species. These covariance-weighted values are called sWeights, ${}_s\mathcal{P}_n(y_e)$. An inverse covariance matrix ($\mathbf{V}_{nj}^{-1}$) is produced within *sPlot*, and contains the information on the PDF values for given events, as well as the yields of the various species. Here, as detailed in Section ??, three are present - signal, peaking background and combinatorial background. With this, the matrix is 3×3 , and is calculated as

$$\mathbf{V}_{nj}^{-1} = \sum_{e=1}^N \frac{f_n(y_e) f_j(y_e)}{(\sum_{k=1}^{N_s} N_k f_k(y_e))^2}. \quad (25)$$

Here, N and N_s are the number of events and species respectively, $f_n(y_e)$ and $f_j(y_e)$ are the PDF values of an event being a given species, and N_k is the yield of species k extracted from the extended maximum likelihood fit. The non-inverted covariance matrix allows for events to be weighted on their likelihood of belonging to either species. Using the previous definitions, sWeights are calculated as

$${}_s\mathcal{P}_n(y_e) = \frac{\sum_{j=1}^{N_s} \mathbf{V}_{nj} f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}. \quad (26)$$

The sum of the sWeights for a given species is equal to the number of events within it. Their use is described in Section 6, and are necessary to determine the distribution of moments for $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays. The sWeighting process used in *sPlot* is only valid while the control and discriminating variables of the dataset are uncorrelated. These checks are made and displayed later in Section 5.9, when the use of sWeights is discussed further.

5 Analysis Method

5.1 Dataset, Stripping and Trigger Requirements

In this analysis, the dimuon dataset from pp collision data collected in 2011 and 2012, corresponding to an integrated luminosity of 3fb^{-1} is used. Candidate reconstruction occurs via stripping [63], in which the full dataset is reduced using a series of selection requirements to select only decays of the required topology and final state particles. Here, we choose a stripping line which reduces the data to contain events with a hadron and dimuon final state. Specifically, version **21r1** [64] is used for data taken in 2011 and version **21** [65] for 2012. The selection requirements involved are listed in Table 4. Dimuon candidates are first formed from pairs of oppositely charged muons. The Λ baryon is reconstructed from p and π^- tracks with a well reconstructed vertex, separated from the primary vertex. The dimuon and Λ candidates are combined to form $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ events. As we apply no requirement to the invariant mass of the dimuon, $\Lambda_b^0 \rightarrow J/\psi\Lambda$ decays are also reconstructed. Appendix A contains definitions of all the variables used in the analysis, as well as how they are measured.

Particle	Variable	Requirement
Λ_b^0	Mass	$4.9 < M < 7.0 \text{ GeV}/c^2$
	$\cos(\theta_D)$	0.9999
	Impact Parameter χ^2	< 16.0
	Flight Distance χ^2	< 121.0
	Vertex $\chi^2/\text{Deg. Of Freedom}$	< 8.0
Λ	Mass - $M_{\text{WorldAverage}}(\Lambda)$	$\pm 30 \text{ MeV}/c^2$
	Vertex χ^2	$< 30.0 \text{ (25.0)}$
	τ	$> 2 \text{ ps}$
p/π	Momentum	$> 2 \text{ GeV}/c$
	Impact Parameter χ^2	$> 9 \text{ (4)}$
Muons	$\chi^2/\text{No. Deg. Freedom}$	< 5
	Track Ghost Prob	< 0.35
	Particle ID Muon	> -3.0
	Impact Parameter χ^2	> 9.0
Dimuon	Mass	$< 7.1 \text{ GeV}/c^2$
	Vertex $\chi^2/\text{Deg. Of Freedom}$	< 12.0

Table 4: Stripping requirements applied to the dataset. Where two values are present, the first is applicable to Long candidates, and the value in brackets to Downstream.

The Λ_b^0 baryons are formed from $b\bar{b}$ pairs, which hadronise quickly, so the impact parameter χ^2 associated to them will be small, as they will effectively be produced

at the interaction point. This also means that the pointing angle (θ_D) will be close to zero. The $\text{IP}\chi^2$ is a measure of the change in χ^2 of the PV with and without the considered particle. A small value indicates the particle was produced close to the primary vertex. Conversely, the $\text{IP}\chi^2$ values of the proton and pion used to form the Λ baryon are required to be large, in order to remove those created by pp interactions. Enforcing a large minimum value in the stripping will ensure that only those which are produced away from the proton interaction point are accepted. Similarly, we are able to restrict the inclusion of protons and pions that were not produced from a decay by introducing the minimum momentum limit. Figure 24 shows the measurement of $\text{IP}\chi^2$ values within a $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ decay. The decay shown would pass our stripping requirements, due to the large final state particles $\text{IP}\chi^2$ values. While this process may reject legitimate decays, the restriction removes many more background events.

The flight distance χ^2 is the difference in χ^2 values of a primary vertex fit with and without the relevant particle. The χ^2 value associated with the vertex is a simple goodness-of-fit check, and ensures well determined origin and decay co-ordinates. It is advantageous to select only well defined tracks and vertices. As particles leave hits in the detectors, misreconstruction is possible by incorrectly joining impacts left by multiple particles to form a single track. Information from the tracking systems, calorimeters and Cherenkov detectors helps to reduce this possibility, though there is still a chance of a reconstructed event being fake. This ghost track probability is applied to the individual muons in order to limit the amount of these fake events from the final dataset. A minimum is also imposed on the particle identification (PID) associated to the individual muons, as determined by the PID systems in the LHCb. Using information from the muon stations and the RICH detectors to determine the nature of the particles, this restriction allows tracks from similar charged bodies, such as kaons and pions to be reduced. It is the goal of these cuts to remove the easily recognisable background events and incorrect decay topologies.

In addition to the stripping, it is required that the events trigger a specific set of requirements at both the software and hardware levels. The trigger decision must be due to the candidate itself, and to successfully pass the requirements, an event must pass at least one criteria at each of the levels. The requirements chosen for this analysis are detailed in Table 5. The L0 hardware trigger level is applied to the dimuon, with the HLT levels used on the reconstructed Λ_b^0 baryon. The L0 trigger accepts tracks which are deemed to contain either one or two muons of sufficient transverse momentum. In the first level of HLT, tracks are required to have triggered on both the muon and dimuon decisions. The topological boosted decision tree decisions in the HLT2 trigger level were discussed in Section 3.8, and only allow events with a number of bodies in the final state greater than or equal to a specific value to pass. As a result of the stripping and trigger cuts, the remaining data set contains only events of a similar topology to that which we have an interest in. The selection and application of the trigger requirements before the neural network selection means that training does not occur with events that would

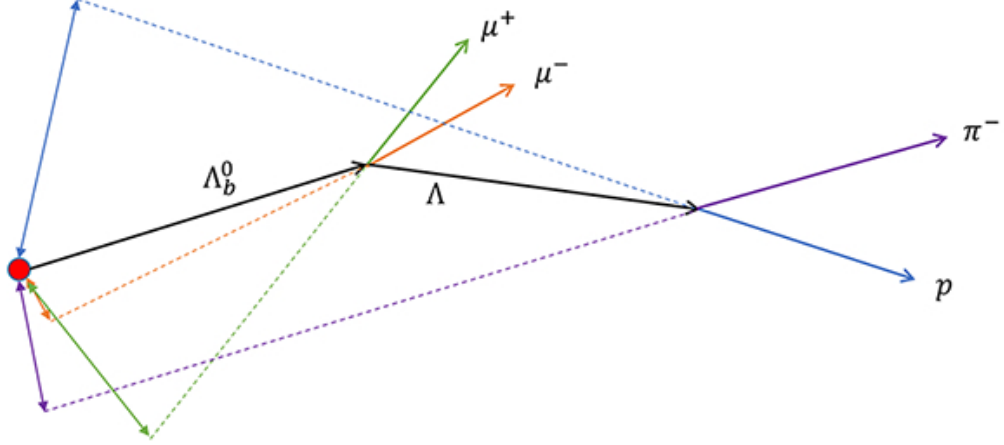


Figure 24: A $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay. Solid arrows are particle momenta vectors, and dashed lines are those vectors traced back. The red circle is the interaction point. The double ended arrows represent the distance measured to calculate impact parameter χ^2 values.

eventually have been removed anyway.

Trigger Level	Requirement
L0	MuonDecision
	DiMuonDecision
HLT1	DiMuonHighMassDecision
	TrackAllL0Decision
	TrackMuonDecision
HLT2	Topo2BodyBBDTDecision
	Topo3BodyBBDTDecision
	Topo4BodyBBDTDecision
	TopoMu2BodyBBDTDecision
	TopoMu3BodyBBDTDecision
	TopoMu4BodyBBDTDecision
	DiMuonDetatchedDecision

Table 5: Trigger requirements applied to the data, all Trigger on Signal. L0 is applied to the dimuon, and the HLT to the reconstructed Λ_b^0 baryon.

Simulated proton-proton collisions are modelled with the **PYTHIA** program [66], using specific parameters to describe the LHCb detector [67]. The decays of the hadronic particles are simulated using **EVTGEN** [68], with final state radiation modelled by **PHOTOS** [69]. Finally, the interactions of all particles with the detector are described by **GEANT4** [70, 71]. This simulated data is vital in analysis work, as it allows unbiased Neural Network training,

and provides description of the different shapes we see in the invariant mass distributions in real data. Simulated datasets are used for determination of efficiencies, and are subject to the same trigger requirements as the real event data.

5.2 Preselection Requirements and Neural Network Training

Our neural network was trained to recognise signal and background events with training samples of both. The $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ signal is taken from an invariant $\Lambda \mu^+ \mu^-$ mass distribution of 5550 to 5700 MeV/c^2 from a Monte Carlo simulation dataset. The background sample contains events with an invariant mass between 5900 and 7000 MeV/c^2 in the real dataset - this region is deemed to contain a true source of combinatorial background events. No background contributions from the lower $\Lambda \mu^+ \mu^-$ mass sideband were included, as these events are more likely to be real physics background contributions. These peaking backgrounds are much easier to identify and model than combinatorial sources, and will likely appear similar to our signal decay. In addition to the stripping and trigger selections, further restrictions are applied to the samples before training to remove other sources of misreconstructed events. A large quantity of events were found to have a Λ baryon decay vertex around 2350 mm downstream of the interaction point. The peak was found to be due to the material boundary between the RICH-1 detector and the beam pipe [72]. This excess is displayed in Figure 25, along with the density in the xy plane about the peak. The source of events was found to contain no real Λ_b^0 baryons, and so only candidates with Λ baryon end vertices displaced between 0 and 2330 mm from the pp interaction point were removed accepted in the training sample.

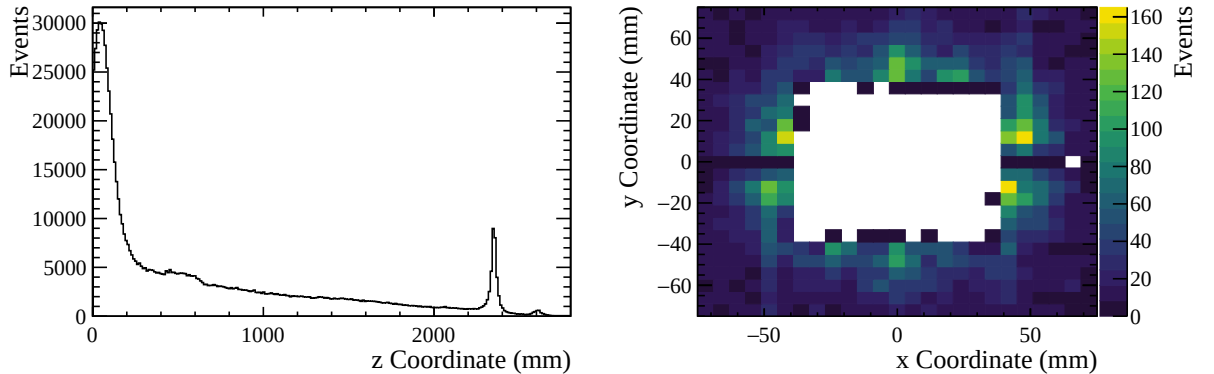


Figure 25: Left - Presence of the peak in the Λ baryon decay vertex in the real dataset. Right - positions of Λ baryon decays in the xy plane ± 20 mm about the peak at $z = 2350$ mm.

It was found that both the Long and Downstream $p\pi^-$ invariant mass distributions peaked between 1105 to 1125 MeV/c^2 , so a ± 10 MeV/c^2 window about the world average Λ baryon mass was imposed. A maximum lifetime of 2 nanoseconds for the Λ baryon was applied to limit the number of incorrectly identified particles included in the training sample. Finally, only events with positive values for the direction goodness-of-fit of the

flight duration associated to the Λ baryon are used, as well as those with a pointing angle greater than $\pi/2$ radians. The Monte Carlo simulated dataset contains signal events, as well as various backgrounds. The different decay types are assigned a Background Category value [73], which allows for them to be identified and removed if needed. To best represent the real data sample, the simulated data was stripped using this background category to contain signal and signal-like events, and low mass background. The simulated data associated to $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events is reconstructed as the $\Lambda_b^0 \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) \Lambda$ decay. The dataset is reduced to contain candidates that involve all specified particles, those with a missing intermediate (such as the J/ψ meson in the rare decay), and those with an additional final state body - such as missing energy radiating away as a gamma photon. The low mass background contains well reconstructed decays not originating from simulated Λ_b^0 baryons, but instead particles with masses $\pm 100 \text{ MeV}/c^2$ about $M(\Lambda_b^0)$. The applied restrictions, plus the particles they were performed on are given in Table 6.

Variable	Minimum	Maximum	Notes
Λ Mass	1105 MeV/c^2	1125 MeV/c^2	-
Λ τ	0 ns	2 ns	-
Λ Dira Own PV	0	None	-
Λ FD χ^2 Own PV	0	None	-
Λ Decay Vertex z	0 mm	2330 mm	-
Λ_b^0 Bkg Category	Between 0 and 11 inclusive	= 50	Simulation only
$\mu^+ \mu^-$ Invariant Mass	3030 MeV/c^2	3150 MeV/c^2	Omitted from
$\mu^+ \mu^-$ Invariant Mass	3635 MeV/c^2	3735 MeV/c^2	training only

Table 6: Preselection cuts applied to both the real and simulated datasets before training of the neural network.

Further omissions are made to the data sample after preselection cuts, but these are not permanently removed, just prevented from being used in training. Specifically, events with invariant dimuon masses that correspond to either the J/ψ meson, or $\psi(2S)$ resonance are not included in the background sample, as they also do not provide a true representation of combinatorial events. The omitted dimuon invariant mass windows were chosen to be 3030 to 3150 MeV/c^2 for the J/ψ meson, and 3635 and 3735 MeV/c^2 for the $\psi(2S)$ resonance peak. Apart from the charmonia restrictions, no requirements are made on the q^2 variable.

With the final training sample selections made, variables which strongly describe the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decay were chosen, such that the neural network could learn the signal and background decays. Sixteen input variables are used, containing information about kinematics, track quality and vertices. Variables associated with the Λ baryon and its decay products are split into Long and Downstream categories, and treated as separate through the training process, due to differing resolutions. Flags are applied to these

Long and Downstream variables to prevent correlation checks between them. Within the training, only one value of the impact parameter χ^2 from each of the $p\pi^-$ and $\mu^+\mu^-$ pair were accepted per event, with this being determined by which hadron and lepton had the highest transverse momentum in that event. Distributions of the chosen variables for both the signal and background samples are compared in Appendix B. It can be seen that due to the overlaps of distributions, there is no way to efficiently separate the candidates via 1-D selection requirements. In order to maintain the validity of the analysis, any variables that were to be used later in the process were excluded from training. These include the angular variables, and the q^2 parameter. These omissions are vital in order not to skew the results. After splitting the variables into independent Long and Downstream track types, 21 inputs are used in total. NeuroBayes was run with 15 training iterations in order to better close in on the optimum network, with each successive iteration attempting to improve the previous. The variables are listed along with their final training ranking and contributing significance in Table 7.

The Λ_b^0 and Λ baryons both decay via the weak force, and therefore have long decay times when compared to similar mass bodies. This means both will travel deep into the detector, with their tracks and decay times able to be cleanly measured. The lifetime of the Λ baryon means that many do not decay until after travelling through the VELO, and will be recorded as Downstream tracks. The chosen training variables allow for events with misidentified baryons to be rejected easily. The most powerful variable in this case is the Λ_b^0 baryon decay vertex χ^2 , the goodness-of-fit of the vertex from which $\Lambda\mu^+\mu^-$ originate. For Downstream tracks, the most important variable is the transverse momentum of the Λ baryon, which allows rejection of random combinations of tracks, which have lower p_T values associated to them. In the Long candidate case, the variable with highest significance is the χ^2 of the Λ baryon flight distance from its own primary vertex, measured by determining the uncertainty on the vertex after considering other tracks.

Final signal background event distributions, and network purity ($N_{\text{Sig}}/(N_{\text{Sig}} + N_{\text{Bkg}})$) as a function of NN output are shown in Figures 26 and 27. The distributions of the selected training samples are overlaid, and have a similar shape. They are found to fluctuate differently, which would imply that the network has not been overtrained. The purity plot of the network is seen to be linear, suggesting that there is no output value at which the optimisation cuts too harshly or loosely on the signal or background samples. As no bias may be seen at any NN value, again we may say there has been no overtraining of the network. The network is therefore deemed to have successfully learned how to separate signal and background events. Average values of the neural network output were calculated over the signal and background sideband regions in order to confirm there is no strong correlation between neural network output and Λ_b^0 baryon mass. These values are plotted in Figure 28, and no real tendency to create spurious peaks can be seen, as those in the signal sample are due to low statistics in the fringes.

Variable	Matrix ID	Added significance	Only this	Loss without
Λ_b^0 Decay Vertex χ^2	4	161.46	164.46	67.25
$\Lambda_b^0 \tau \chi^2$	3	151.26	66.63	57.0
$\Lambda_{DD} p_T$	19	83.72	143.73	54.68
$\Lambda_{LL} \text{FD } \chi^2$	9	63.13	96.76	13.74
Max p_T^{lepton} IP χ^2	22	50.86	106.42	32.19
$\Lambda_b^0 p_T$	5	49.91	106.57	29.60
Λ_{LL} Dira Own PV	12	36.75	57.49	32.10
Λ_b^0 Cone p_T asymmetry	6	28.61	97.65	24.16
Λ_{LL} Mass	11	23.37	87.16	21.29
$\Lambda_b^0 \tau$	2	21.33	63.98	16.13
Λ_{DD} Decay Vertex z	16	17.98	65.52	9.27
$\Lambda_{LL} p_T$	13	17.62	78.72	15.45
Λ_{DD} Mass	17	12.85	60.77	13.97
$\Lambda_{DD} \tau$	14	9.09	57.12	8.91
Max p_T^{hadronLL} IP χ^2	20	9.51	83.49	8.72
Λ_{DD} Dira Own PV	18	9.42	78.84	9.02
Λ_b^0 Dira Own PV	7	8.77	114.79	8.18
$\Lambda_{DD} \text{FD } \chi^2$	15	5.78	55.27	4.83
Max p_T^{hadronDD} IP χ^2	21	1.99	18.81	1.98
$\Lambda_{LL} \tau$	8	1.17	37.24	1.29
Λ_{LL} Decay Vertex z	10	0.61	35.48	0.61

Table 7: Variables used in NeuroBayes training, in order of their final ranking. ‘Added significance’ is the amount of information a variable adds to all above. ‘Only this’ is the variable on its own. ‘Loss without’ is the information lost by removing solely that variable.

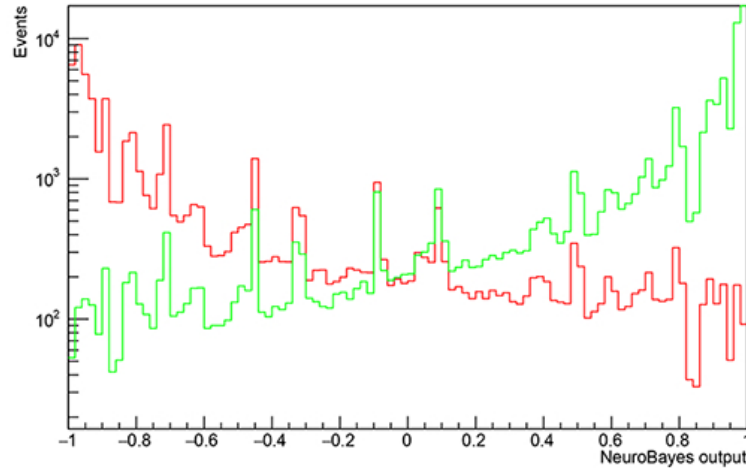


Figure 26: Signal (green) and background (red) event yields as a function of neural network output.

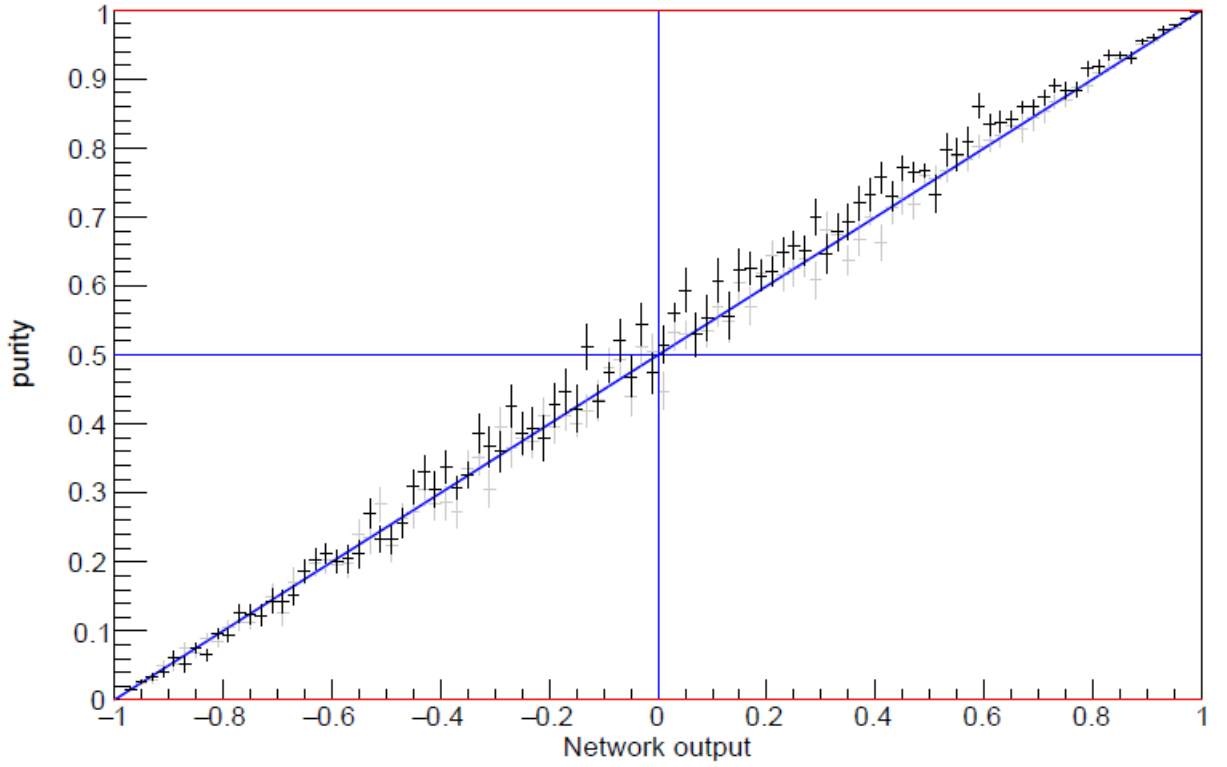


Figure 27: Purity by neural network output. The black points represent the final training, while the grey are those produced by preprocessing.

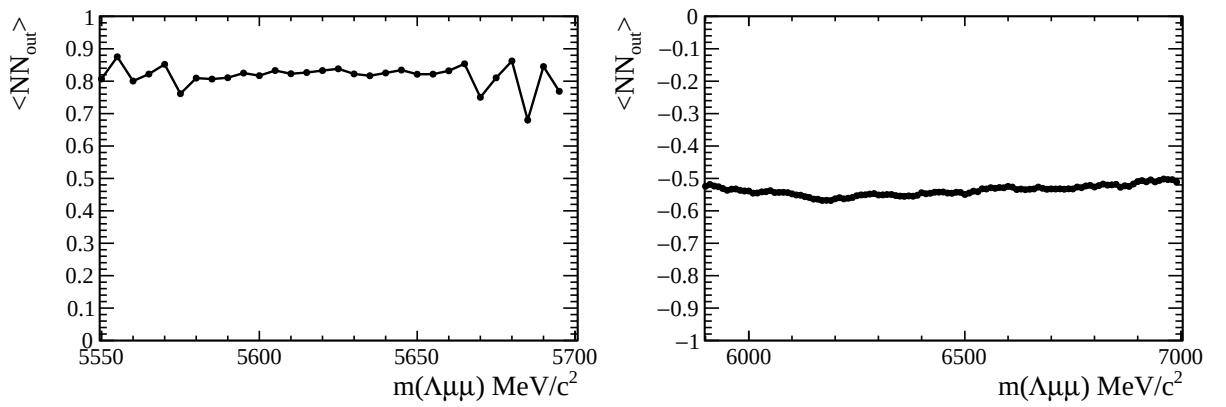


Figure 28: Left - Average neural network value from the signal region in simulated events. Right - Average neural network value in the background region from real data.

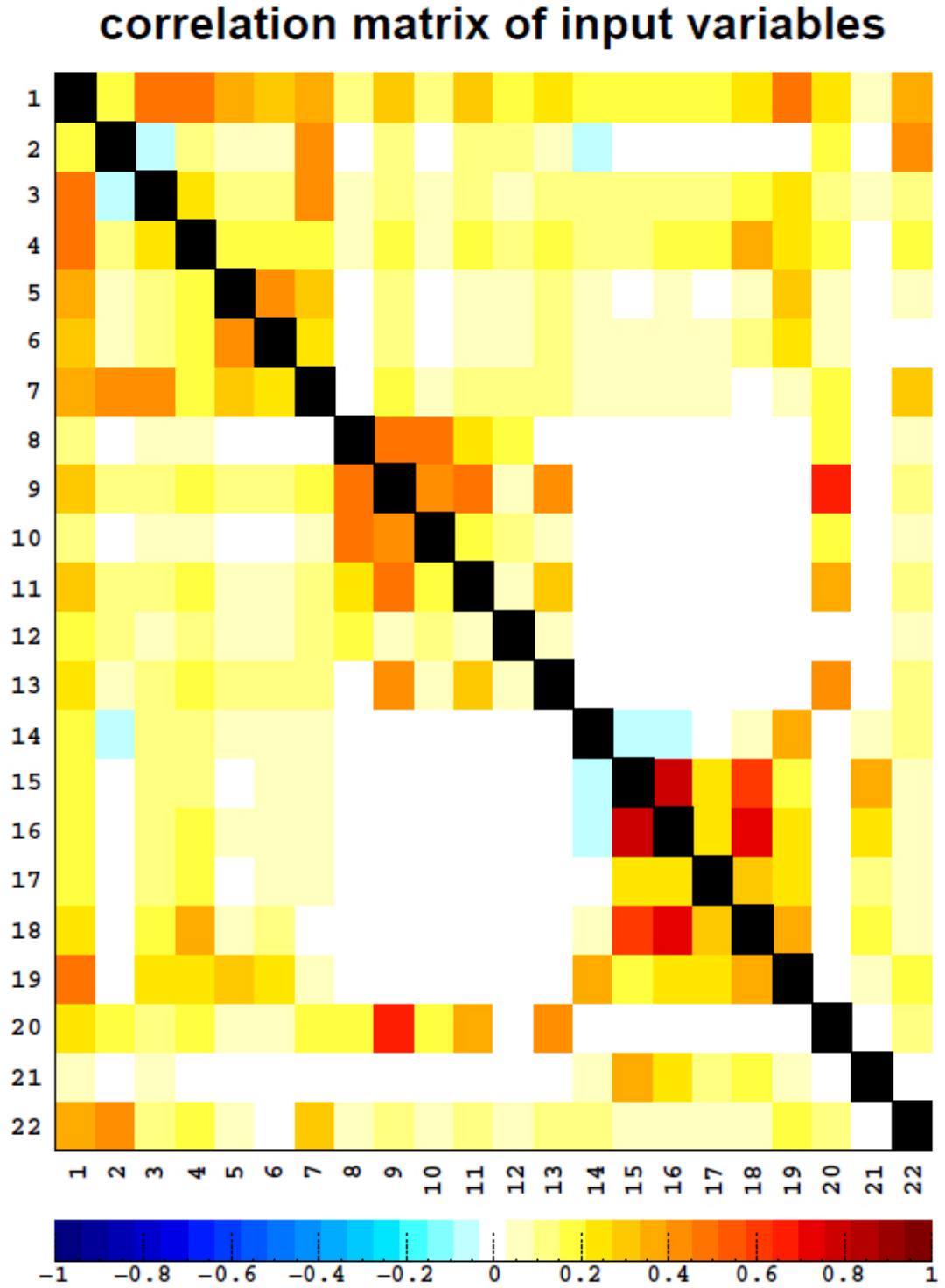


Figure 29: Variable correlation matrix from NN training. The white regions arise from the separation of LL and DD tracks, as an event is either one or the other, no correlation exists between these variables.

Correlations between the training parameters are shown in Figure 29. Variable ID = 1 is missing from the table, and present in the matrix, as this is the measure of whether the target is signal (+1) or background (-1). Heavy correlation can be seen between variables 15, 16 and 18 - the Λ baryon's flight distance χ^2 , decay vertex z co-ordinate, and θ_D . This is expected, as all three measure very similar things. Correlation can also be seen between the lifetime of the Λ baryon and its flight distance χ^2 , and this is due to the proportionality of the flight duration to how long it exists for in the detector. The flight distance χ^2 of the Λ is also found to correlated with the impact parameters of the proton and pion, as the decay vertex of the Λ baryon and the primary vertex of $p\pi^-$ are identical. It can also be seen that the impact parameter χ^2 of the leptons is dependant on the decay time variables of the Λ_b^0 baryon. The IP of the muons is dependant on the location of decay vertex of the Λ_b^0 baryon, so its lifetime will affect these co-ordinates. The output of the matrix was used while selecting the final variables for training, as it was possible to remove those that heavily correlated with others. It was advantageous to do this, as they were found to contribute little to the final training, and slowed the overall process.

With the neural network trained, the each event in the real and simulation datasets was ranked on how signal or background-like it appeared to be. With this new variable, the requirement on the output of the neural network could be optimised such that maximum signal was retained while rejecting the greatest amount of background. As the decay of $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ is previously observed, we can optimise the output of the neural network by maximising the expected signal yield significance. The calculation of an expected $\Lambda\mu^+\mu^-$ event count requires a measure of the yield of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events in the dataset after our preselection requirements. It is therefore necessary to model the signal and background contributions of the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decay before the network may be optimised.

5.3 Signal Modelling

Yields are extracted from data via the maximum likelihood fits to data. Signal and background models are created using simple functions, then combined to make a total function to describe the probabilities of all events in the sample. Naively, the signal shapes required in this analysis would appear to be Gaussian functions, however closer examination reveals final state radiation in the fringes of the peak, meaning the model requires alteration. Instead, the Crystal Ball function is used [74], which employs a Gaussian core, but switches to a power law decay tail at a specific distance (α) from the mean value (M_0). The Crystal Ball function produces the power-law tail to the left of the mean value. This has the form

$$C_l(M_i; \alpha_l, n_l, M_0, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(M_i - M_0)^2}{2\sigma^2}\right) & \text{if } \frac{(M_i - M_0)}{\sigma} \geq -|\alpha_l|, \\ A \left(B - \frac{(M_i - M_0)^2}{\sigma^2}\right)^{-n_l} & \text{if } \frac{(M_i - M_0)}{\sigma} < -|\alpha_l|, \end{cases} \quad (27)$$

where M_i is the mass of event i the function is calculating the probability for, σ is the width of the decay peak, and n is the power law coefficient. For normalisation and continuity,

$$A = \left(\frac{n_l}{|\alpha_l|} \right)^{n_l} \cdot \exp \left(-\frac{\alpha_l^2}{2} \right), \quad (28)$$

and

$$B = \frac{n_l}{|\alpha_l|} - |\alpha_l|. \quad (29)$$

In this analysis, signal peaks are found to require a double Crystal Ball Shape - a Gaussian core with power law tails on both sides. This is modelled with second values of both α and n , while keeping a common mean and width. The new value of α exists on the opposite side of the mean, and is the opposite sign. To account for this, the form of the Crystal Ball with a tail slope to the right of the central value becomes

$$C_r(M_i; \alpha_r, n_r, M_0, \sigma) = N \cdot \begin{cases} \exp \left(-\frac{(M_i - M_0)^2}{2\sigma^2} \right) & \text{if } \frac{-(M_i - M_0)}{\sigma} \geq -|\alpha_r|, \\ A \left(B - \frac{(M_i - M_0)^2}{\sigma} \right)^{-n_r} & \text{if } \frac{-(M_i - M_0)}{\sigma} < -|\alpha_r|, \end{cases} \quad (30)$$

with

$$A = \left(\frac{n_r}{|\alpha_r|} \right)^{n_r} \cdot \exp \left(-\frac{\alpha_r^2}{2} \right), \quad (31)$$

and

$$B = \frac{n_r}{|\alpha_r|} - |\alpha_r|. \quad (32)$$

With these definitions, the form describing the double Crystal Ball function is given as

$$C_{\text{double}}(M_i) = f C_l(M_i - M_0, \sigma, \alpha_l, n_l) + (1 - f) C_r(M_i - M_0, \sigma, \alpha_r, n_r). \quad (33)$$

This equation requires a value for f , which is a measure of how much an individual Crystal Ball function contributes to the total - referred to from now as the CB Ratio. Figure 30 displays the requirement for this number in the modelling of simulated $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events.

Due to the correlations between the values of α and n , and also to avoid biasing, the fit parameters of the double Crystal Ball functions are determined by modelling on simulation events before being applied to the real data invariant mass distribution. To fit the final PDF to the simulated dataset, an extended unbinned maximum likelihood fit is used. The likelihood has the form

$$\mathcal{L} = e^{-(N_S + N_B + N_{\text{phsbg}})} \times \prod_{i=1}^N [N_S C_S(M_i) + N_B C_B(M_i) + N_{\text{phsbg}} C_{\text{phsbg}}(M_i)] \quad (34)$$

where

- N is the number of events in the sample,

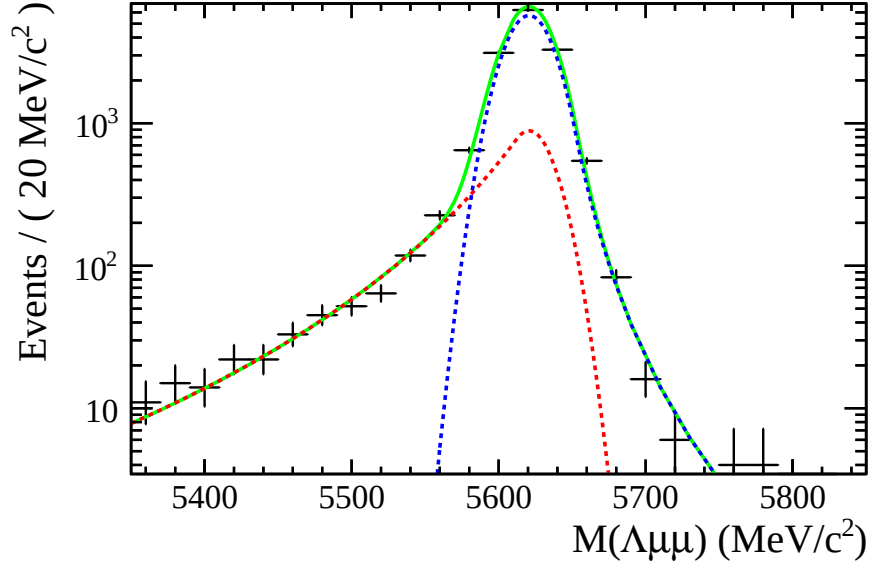


Figure 30: Separation of the Crystal Ball shapes needed for the total fit to 15 to 20 GeV^2/c^4 q^2 interval $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ simulation. The solid green line is the total Crystal Ball shape, while the dashed red and blue lines represent the individual CB shapes from which it is formed. The parameters describing this fit are given in Table 13 in Section 5.7.

- N_S is the number of signal events,
- N_B is the background events,
- N_{phsbg} is the number of peaking background events,
- $C_k(M_i)$ is the corresponding value of the probability density function.

The parameters found to maximise this likelihood with simulated data are then recorded and fixed, before being applied to the real data, along with functions describing the background sources present in the distribution.

5.4 Background Sources

Along with the signal, it is necessary to model the background contributions present in our invariant mass distribution. The primary background in the dataset is combinatorial - events reconstructed within the detector from random tracks as an attempt to create b -hadron decays. Probability density functions are fitted to this source of background in every fit involving the real data, and we find that an exponential function models this well and the coefficient describing this is the only mass shape parameter not fixed in the final yield extractions. Figure 31 displays the modelling of an exponential function to the dataset used as our background sample in the neural network training.

An additional source of background is found to be present in the invariant mass distribution of $\Lambda_b^0 \rightarrow J/\psi \Lambda$, peaking to the left of the Λ_b^0 baryon peak. This source of events is from misreconstructed $B^0 \rightarrow J/\psi K_S^0$ decays. The decay is visible as its similar topology and final state particles allows it to survive preselection requirements, while the reconstructed B^0 meson mass is in our window of interest. The contribution requires modelling, and this was performed on simulated data, with the fit displayed in Figure 31, and the parameters given in Table 8. The mass distribution can clearly be seen to not be Gaussian, and is modelled with a double Crystal Ball Shape. The $B^0 \rightarrow J/\psi K_S^0$ background may be reduced using a proton ID cut and applying a loose NN restriction, as discussed in Section 5.5. This peaking background does not contribute in the high q^2 interval in which this analysis is performed, as the muons are produced by the J/ψ meson decay, and are restricted to the q^2 intervals associated with charmonia decays, as mentioned in Section 2.1. The rare decay $B^0 \rightarrow K_S^0 \mu^+ \mu^-$ may occur at high q^2 , however a previous search for $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ determined yields of this decay are negligible [33].

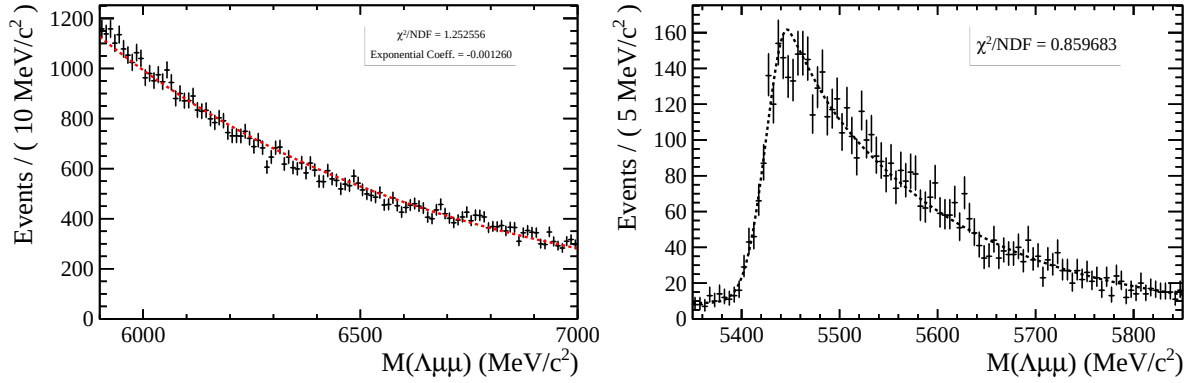


Figure 31: Left - an exponential function (red dashed line) modelling the combinatorial background in the side band of the data sample after preselection (black points) used to train background in the neural network. Right - fit (black dashed line) to extract parameters for events misreconstructed as $B^0 \rightarrow J/\psi K_S^0$ from simulation data (black points). Both plots are presented with goodness-of-fit values.

Decay Mode	M_0 (MeV/ c^2)	σ (MeV/ c^2)	α_l	α_r	n_l	n_r	CB Ratio
$B^0 \rightarrow J/\psi K_S^0$	5446.56 ± 2.12	21.36 ± 1.62	9.61×10^{-7} $\pm 2.09 \times 10^{-7}$	-0.13 ± 0.01	18.67 ± 4.20	50.00 8.52	0.04 ± 0.004

Table 8: Fit parameters for the $B^0 \rightarrow J/\psi K_S^0$ mass shape. Values extracted from simulation data, with preselection requirements, but with no restriction on q^2 or neural network output.

5.5 Fits to $\Lambda_b^0 \rightarrow J/\psi \Lambda$ Data

As mentioned at the end of Section 5.2, modelling of the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ invariant mass peak is necessary so we may extract a yield from data to optimise the neural network. In order to model the decay, the dataset after preselection requirements was restricted to only contain events with an dimuon invariant mass between 3030 and 3150 MeV/ c^2 . The total $\Lambda_b^0 \rightarrow J/\psi \Lambda$ probability density function consists of the two aforementioned background sources, plus a signal shape using parameters taken from a $\Lambda_b^0 \rightarrow J/\psi \Lambda$ simulated dataset. The signal shape was fitted for the full dataset, in addition to separate Long and Downstream candidates models. Fits to simulation data are shown as Figure 32, while the fit parameters are given in Table 9.

Decay Mode (Track Type)	M_0 (MeV/ c^2)	σ (MeV/ c^2)	α_l	α_r	n_l	n_r	CB Ratio
$\Lambda_b^0 \rightarrow$ $J/\psi \Lambda$ (LL + DD)	5620.66 ± 0.12	15.49 ± 0.14	1.35 ± 0.11	-1.47 ± 0.13	1.89 ± 0.16	8.24 ± 2.59	0.57 ± 0.08
$\Lambda_b^0 \rightarrow$ $J/\psi \Lambda$ (LL)	5620.67 ± 0.20	15.69 ± 0.20	1.49 ± 0.10	-1.40 ± 0.16	1.59 ± 0.11	4.95 ± 0.97	0.68 ± 0.09
$\Lambda_b^0 \rightarrow$ $J/\psi \Lambda$ (DD)	5620.66 ± 0.11	15.70 ± 0.09	1.35 ± 0.01	-1.83 ± 0.03	1.61 ± 0.03	3.70 ± 0.21	0.49 ± 0.01

Table 9: Fit parameters for the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ mass shape. Values are extracted from simulation data, with selection, trigger and truth requirements applied, but with no restriction on q^2 or neural network output.

Combining the signal shape with the combinatorial and B^0 meson backgrounds produces the fits required for $J/\psi \Lambda$ yield extraction. Figure 33 shows the results of the fit to data for the complete dataset, as well as a plot with a logarithmic scale to better show the contributions. Figure 34 presents the data split into the separate LL and DD track categories. From the individual Long and Downstream yields given in Table 10, it can be seen that roughly two-thirds of the events are in the DD category, as would be expected with the long Λ baryon lifetime.

Mass distributions in Monte Carlo simulated datasets are generated assuming a perfect detector set up. The magnetic field is assumed to be perfectly known, and the detector is perfectly aligned. This gives accurate and unrealistic resolutions on particle masses, resulting in finer peaks. In reality however, the lack of this perfection results in a reduced resolution, and wider peaks in the real data when compared to the simulation. A discrepancy was not a problem in the fit to $\Lambda_b^0 \rightarrow J/\psi \Lambda$, as the mean and width of the peak were not fixed when modelling the data. When considering the later fits to $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events, it is necessary to compensate for this deviation. To do so, the ratio of the widths of the $J/\psi \Lambda$ peaks in data and simulation are determined, and a

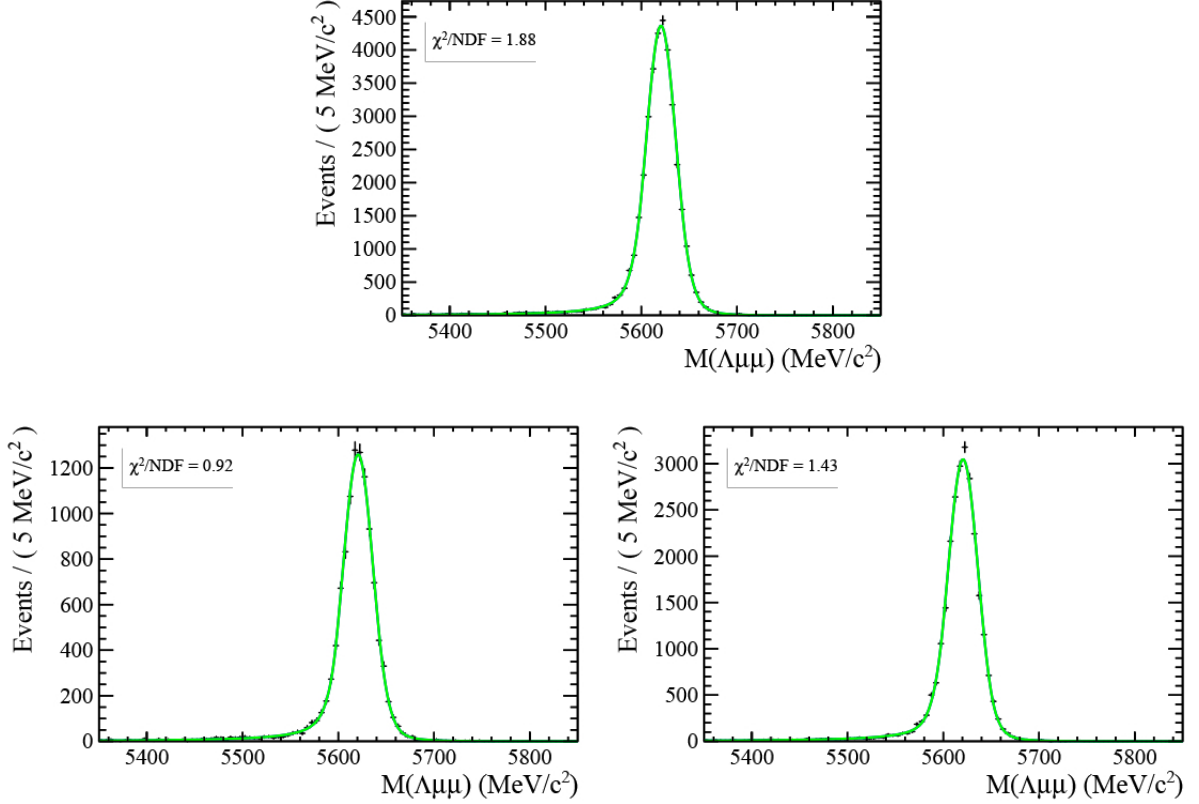


Figure 32: Fits to $\Lambda_b^0 \rightarrow J/\psi \Lambda$ simulation to obtain PDF parameters, along with goodness-of-fit values. Top - full dataset. Bottom left - Long candidates. Bottom right - Downstream candidates.

compensation factor is calculated as $\sigma_{\text{Data}}/\sigma_{\text{Simulation}}$. It is found that the width of this decay in the real dataset is 1.11 times larger than in simulation, and this value is used to correct $\sigma_{\text{Simulation}}$ before application to data.

Track Type	$N_{\Lambda_b^0 \rightarrow J/\psi \Lambda}$	$N_{B^0 \rightarrow J/\psi K_S^0}$	$N_{\text{Combinatorial Bkg}}$
LL + DD	18885 ± 185	5017 ± 392	64835 ± 894
LL	6719 ± 115	1230 ± 249	26553 ± 552
DD	12273 ± 146	3751 ± 303	38308 ± 707

Table 10: Yields extracted from modelling to the $\Lambda\mu^+\mu^-$ invariant mass spectrum for $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events.

The PDFs used to model the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset were then used to produce sWeights, as the events were required to validate the analysis method used to determine the moments of the rare $\Lambda\mu^+\mu^-$ decay. This was possible, as a previous angular analysis of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ provided results we could use to produce theoretical moments for comparison [35]. The steps

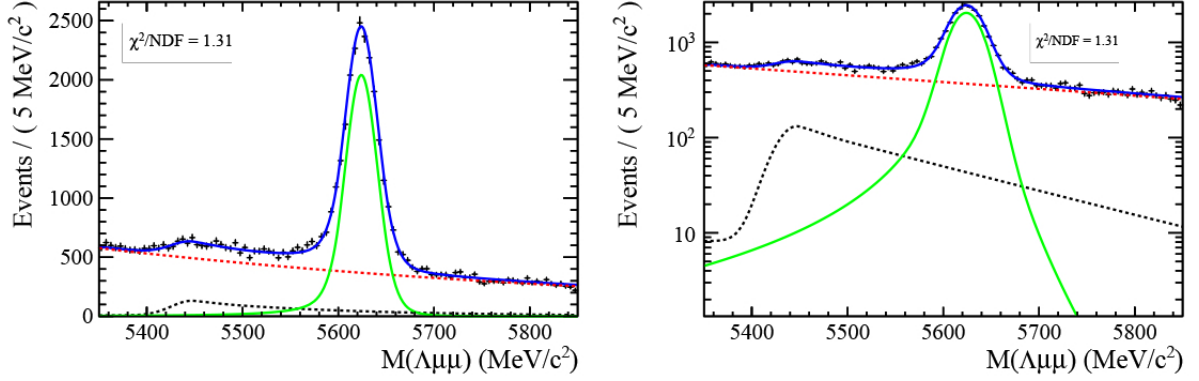


Figure 33: Fits to the full $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset after preselection, with an invariant dimuon mass window applied, with linear and log scales, along with goodness-of-fit values. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively. The dashed red and black lines represent the combinatorial background and $B^0 \rightarrow J/\psi K_S^0$ contributions. The B^0 meson background parameters are listed in Table 8, and the signal fit parameters are given in Table 9.

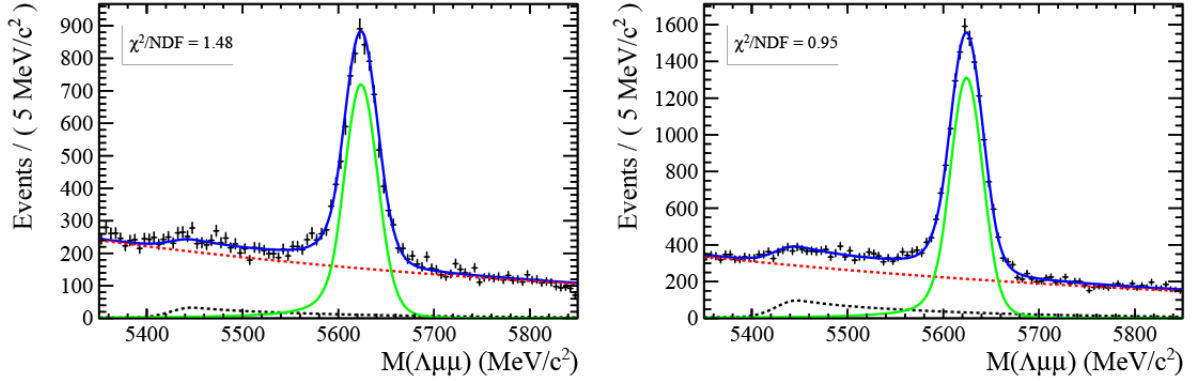


Figure 34: $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events separated into Long (left) and Downstream (right) tracks. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively, along with goodness-of-fit values. The dashed red and black lines represent the combinatorial background and $B^0 \rightarrow J/\psi K_S^0$ contributions. The B^0 meson background parameters are listed in Table 8, and the signal fit parameters are given in Table 9.

describing this will be detailed in Section 6.4, however it was discovered that contributions from the $B^0 \rightarrow J/\psi K_S^0$ background were affecting this process. To compensate, a neural network restriction of 0.0 was applied, along with a probability of the proton being a real proton with a minimum value of 0.1 imposed on the dataset before the sWeighting process. This successfully removed the majority of the $B^0 \rightarrow J/\psi K_S^0$ events, with the new yields being given in Table 11, with a logarithmic plot of the full LL + DD dataset

Track Type	$N_{\Lambda_b^0 \rightarrow J/\psi \Lambda}$	$N_{B^0 \rightarrow J/\psi K_S^0}$	$N_{\text{Combinatorial Bkg}}$
LL + DD	14802 ± 130	197 ± 126	4898 ± 144
LL	5467 ± 78	0 ± 9	1261 ± 43
DD	9417 ± 104	401 ± 108	3353 ± 121

Table 11: Species yields after the application of the neural network working point and proton probability to the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset. Note, that the signal yields presented here are not those used to optimise the network, but instead to show the effect of the restrictions.

given in Figure 35. The total $J/\psi \Lambda$ yield has also decreased, but there is still enough events for validation. The optimisation of the neural network was not repeated after these changes, as the value used from the original method had already been reconfirmed by an independent process, as will be discussed in Section 6.3.

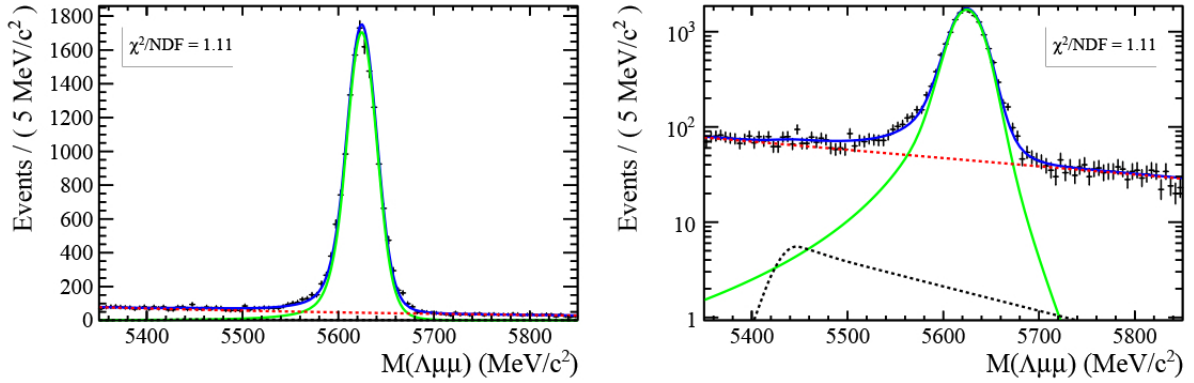


Figure 35: Fits to the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset after applying steps to remove the $B^0 \rightarrow J/\psi K_S^0$ background. Comparison between this Figure, and Figure 33 shows the success of these restrictions. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively, along with goodness-of-fit values. The dashed red and black lines represent the combinatorial background and $B^0 \rightarrow J/\psi K_S^0$ contributions.

5.6 Optimisation of Neural Network Selection

The $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ signal is previously observed in the 15 to 20 GeV^2/c^4 q^2 region, so the optimal neural network cut may be determined by maximising signal significance in this interval. Significance is calculated as $N_{\text{Sig}}/\sqrt{N_{\text{Sig}} + N_{\text{Bkg}}}$, where N_{Sig} and N_{Bkg} are the expected signal and background yields in a given q^2 bin. The value of N_{Sig} is calculated as

$$N_{\text{Sig}} = N_{\Lambda_b^0 \rightarrow J/\psi \Lambda} \cdot \frac{d\mathcal{B}(\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-)/dq^2}{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi \Lambda)} \cdot \frac{\Delta q^2}{\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)} \cdot \frac{N_{MC}^{\text{Cuts}}}{N_{MC}^{\text{Stripping}}}, \quad (35)$$

where

- $N_{\Lambda_b^0 \rightarrow J/\psi \Lambda}$ is the number of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events extracted from data after preselection given in Table 10,
- $\frac{d\mathcal{B}(\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-)/dq^2}{\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi \Lambda)}$ is the differential branching fraction ratio for the necessary q^2 interval from Table 2,
- Δq^2 is the width of the q^2 interval in units of GeV^2/c^4 ,
- $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$ is equal to $5.93 \pm 0.06 \times 10^{-2}$ [1],
- N_{MC}^{Cuts} is the number of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events from a simulated dataset in the q^2 interval with preselection requirements, for a specific neural network cut,
- $N_{MC}^{\text{Stripping}}$ is the number of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events from a simulated dataset in the same q^2 interval with preselection requirements, and no restriction on neural network output.

The value of $N_{\Lambda_b^0 \rightarrow J/\psi \Lambda}$ is extracted from data as described in Section 5.5, with the exact number used given in Table 10. The number of background events - N_{Bkg} is the expected yield under the 5580 to 5660 MeV/c^2 invariant mass region which tightly confines the signal peak. We measure the value from the real data, with a fit to the background in the 5900 to 7000 MeV/c^2 rare sample $\Lambda \mu^+ \mu^-$ invariant mass distribution extrapolated back over the signal range. Modelling the background within the signal window directly results in RooFit attempting to maximise signal yield, giving a measured N_{Bkg} less than the actual value. To find the optimal value for the neural network, the signal significance is measured for 50 working points between -1 and +1. The background yield is extracted for each neural network working point, along with N_{MC}^{Cuts} . Signal significance as a function of neural network output requirement is shown in Figure 36, giving an optimum value of 0.24. The efficiency of the chosen working point is also shown in Figure 36, where the neural network requirement is applied to the full q^2 interval, regardless of from which it was produced. The optimisation is found to show reduced efficiency towards the 15 to 20 GeV^2/c^4 q^2 region, and this is likely due to low statistics in this range.

In Section 2.1, the reasons behind only performing the angular analysis on the 15 to 20 GeV^2/c^4 q^2 region were detailed. Despite this, the neural network was optimised in three other q^2 bins, 0.1 to 2.0, 1.1 to 6.0 and 11.0 to 12.5 GeV^2/c^4 . The resulting optimum working points from all the optimised q^2 bins are listed in Table 12. For simplicity, the working points were obtained in the same way as previously detailed, using the branching fraction ratios from Table 2, despite the decay not being discovered in the 1.1 to 6.0 GeV^2/c^4 bin. As they were not used as part of the analysis, no fits will be shown, nor event yields given. Instead, these working points were only used to model detector efficiencies with simulated data in other q^2 bins, as will be discussed later.

q^2 region (GeV^2/c^4)	Optimal NN Working Point
0.1 - 2.0	0.90
1.1 - 6.0	0.98
11.0 - 12.5	0.65
15.0 - 20.0	0.24

Table 12: Neural network optimum working points, as determined by maximising the signal significance in each q^2 region.

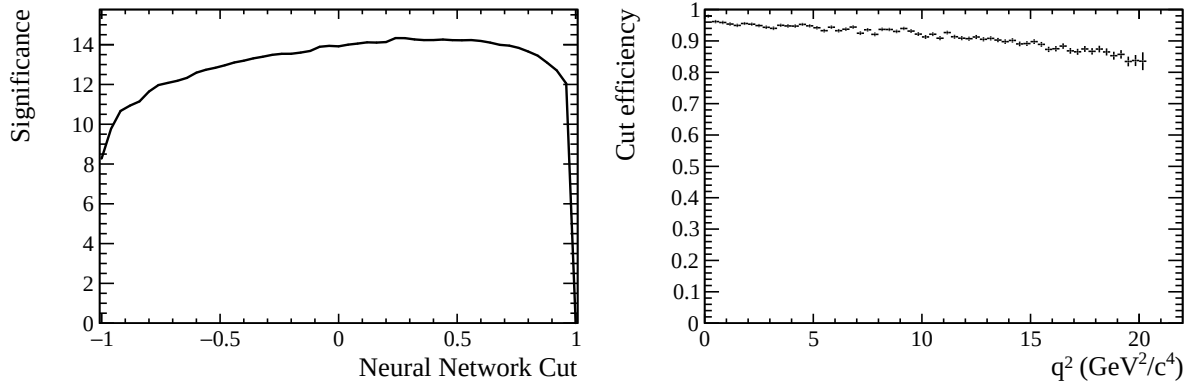


Figure 36: Left - 15 to 20 GeV^2/c^4 q^2 $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ signal significance as a function of neural network output. The significance is found to maximise at 0.24. Right - Efficiency of the neural network requirement selection over the full q^2 range, for events with a output value of greater than 0.24.

The neural network was trained using data from the whole q^2 spectrum, while the angular analysis is only performed in the 15 to 20 GeV^2/c^4 q^2 interval. To check this was sensible, a q^2 specific neural network training was produced with the same variables and cuts. To ensure sufficient events remained for training, the Λ baryon mass window was increased to $\pm 30 \text{ MeV}/c^2$. Receiver operating characteristic (ROC) plots were produced from both networks to compare the efficiency of background rejection to signal acceptance as a check of training success. The ROC plots produced from the generic and specific training methods are overlaid in Figure 37, and both show a clean, reliable training was achieved - but no real variation between them can be seen. As a result, the generic, full q^2 neural network training is used.

After the optimisation of the neural network, we restrict our dataset to only contain events with a NN output value of 0.24 or greater. This provides the most efficient method of retaining $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ decays while losing background events. The remaining dataset, with preselection, q^2 and neural network restrictions applied is now able to be modelled in order to measure $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ yields.

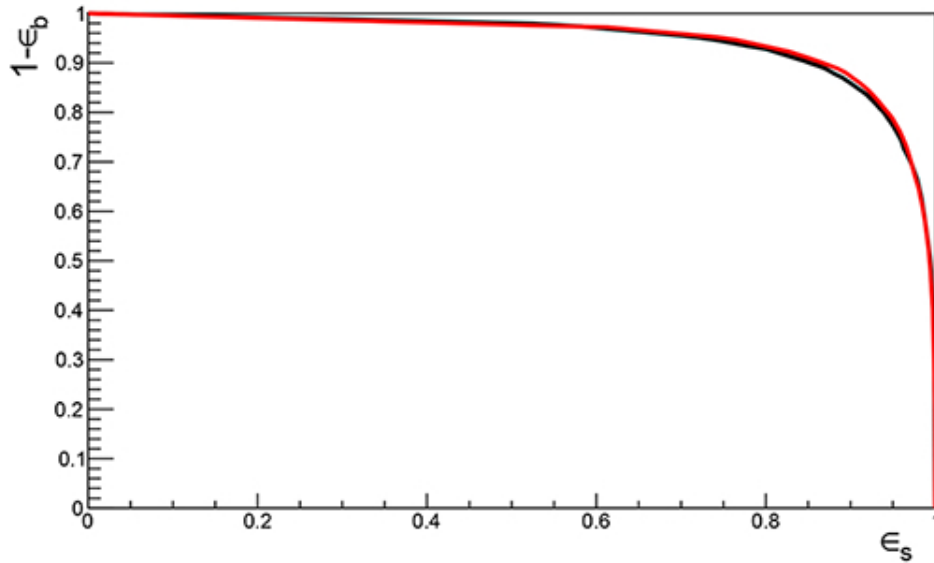


Figure 37: ROC curves produced in the 15 to 20 GeV^2/c^4 q^2 interval. The black line is a generically trained network, while the red represents a specific training for this q^2 interval. No preference can be seen for either method.

5.7 Fits to $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ Data

The dataset after the preselection, neural network working point and q^2 window requirements were applied was found to contain no visible peaking backgrounds, so was treated as containing only combinatorial background events. The combinatorial background is again modelled as an exponential, with the coefficient defining it not being fixed during fitting. The contributions of $B^0 \rightarrow J/\psi K_S^0$ are removed by applying the q^2 window, and so are not seen. The suppressed $B^0 \rightarrow K_S \mu^+ \mu^-$ decay is also unseen, as its branching fraction is less than $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ [75] in this q^2 interval, even when considering the fragmentation ratio of B mesons to Λ_b^0 baryons [44]. A discussion of the peaking backgrounds from a previous search for $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ provides a total expected yield of no more than 10 events for $B^0 \rightarrow K_S \mu^+ \mu^-$ and $B^0 \rightarrow K_S \mu^+ \mu^-$ when integrated over q^2 [33]. The $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ mass peak is modelled with two different probability density functions as a cross check method. First, the $\Lambda \mu^+ \mu^-$ simulation data is subjected to preselection requirements and truth matching before the q^2 and neural network limitations are applied. The parameters required to describe the simulated $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ are taken and applied to the data. The second set of parameters are obtained by applying the same preselection and truth requirements to $\Lambda_b^0 \rightarrow J/\psi \Lambda$ simulation data. This time no neural network or q^2 restrictions are applied, resulting in a $\Lambda_b^0 \rightarrow J/\psi \Lambda$ mass peak, from which we take the fit parameters. The fits will be described from here on as the Mu and Psi respectively to avoid confusion. Both sets of fits are shown in Figure 38. The numerical values defining the probability density functions are given in Table 13, however these widths are before the previously mentioned correction factor of 1.11 has been applied.

Fit Model (Track Type)	M_0 (MeV/ c^2)	σ (MeV/ c^2)	α_l	α_r	n_l	n_r	CB Ratio
Mu (LL + DD)	5620.46 ± 0.14	16.32 ± 0.12	1.28 ± 0.04	-1.71 ± 0.04	2.02 ± 0.40	4.82 ± 0.97	0.47 ± 0.02
Mu (LL)	5620.39 ± 0.01	16.21 ± 0.01	1.45 ± 0.001	-1.44 ± 0.002	2.00 ± 0.003	12.50 ± 0.17	0.62 ± 0.001
Mu (DD)	5620.50 ± 0.37	16.31 ± 0.32	1.12 ± 0.68	-1.77 ± 0.23	2.10 ± 0.76	4.22 ± 0.70	0.38 ± 0.28
Psi (LL + DD)	5620.82 ± 0.33	15.85 ± 0.37	1.31 ± 0.21	-1.74 ± 0.34	1.93 ± 0.38	4.45 ± 1.98	0.56 ± 0.18
Psi (LL)	5620.43 ± 0.05	15.20 ± 0.05	1.01 ± 0.01	-1.33 ± 0.01	2.61 ± 0.03	5.81 ± 0.14	0.49 ± 0.004
Psi (DD)	5620.98 ± 0.38	16.17 ± 0.33	1.27 ± 0.26	-2.46 ± 0.28	1.60 ± 0.41	2.61 ± 1.02	0.44 ± 0.004

Table 13: Fit parameters extracted from MC simulation with their associated errors for both fit shapes. The widths listed here are taken straight from simulation, and have not been corrected for MC/data discrepancies.

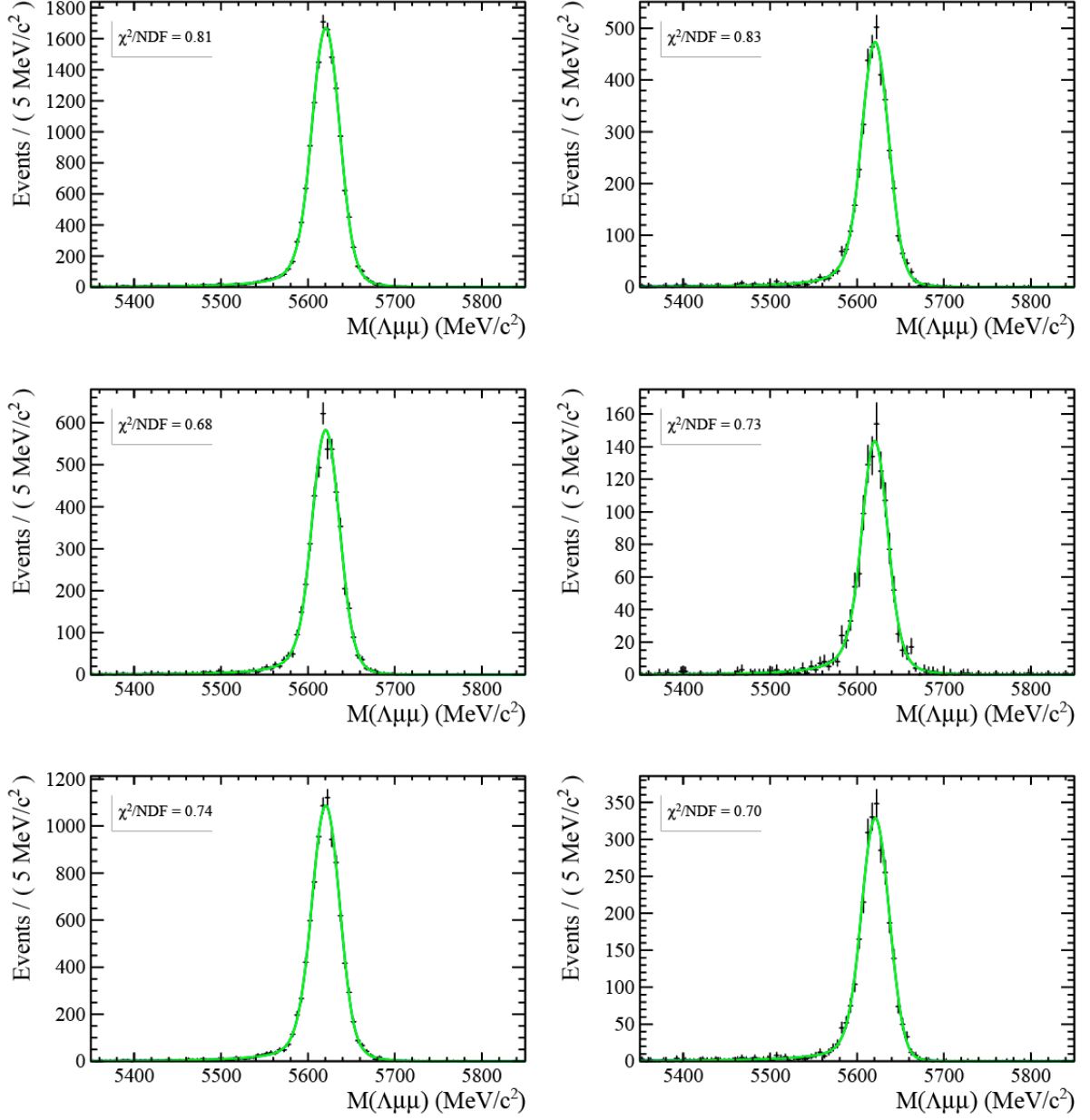


Figure 38: Simulated events in the 15 to 20 GeV^2/c^4 q^2 interval, for the Mu (left) and Psi (right) fit models, along with goodness-of-fit values. Top - Full dataset. Middle row - Long candidates. Bottom row - Downstream candidates. The black points are the data, and the green line is the signal shape.

The full probability density function required to fit to the $\Lambda\mu^+\mu^-$ dataset are now described, and both models are used in order to extract yields. Figures 39 and 40 show the application of the Mu and Psi fit models to the data respectively.

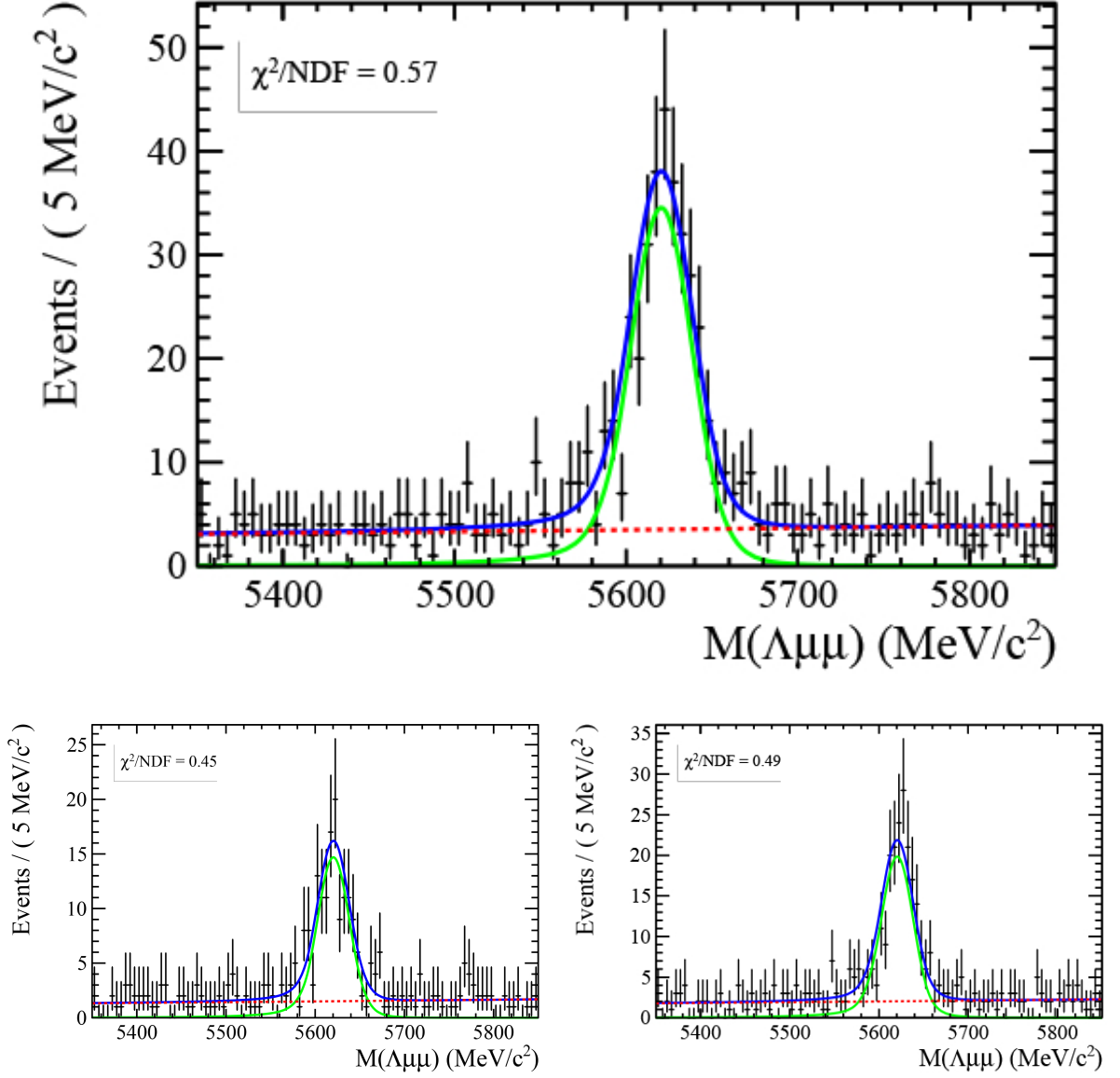


Figure 39: Fits to $\Lambda\mu^+\mu^-$ invariant mass distribution using the Mu fit model. Top - Both Long and Downstream tracks. Bottom left - Long tracks. Bottom right - Downstream tracks. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively, along with goodness-of-fit values. The dashed red line represents the combinatorial background contributions.

Final event yields determined by the fits to data presented in Figures 39 and 40 are

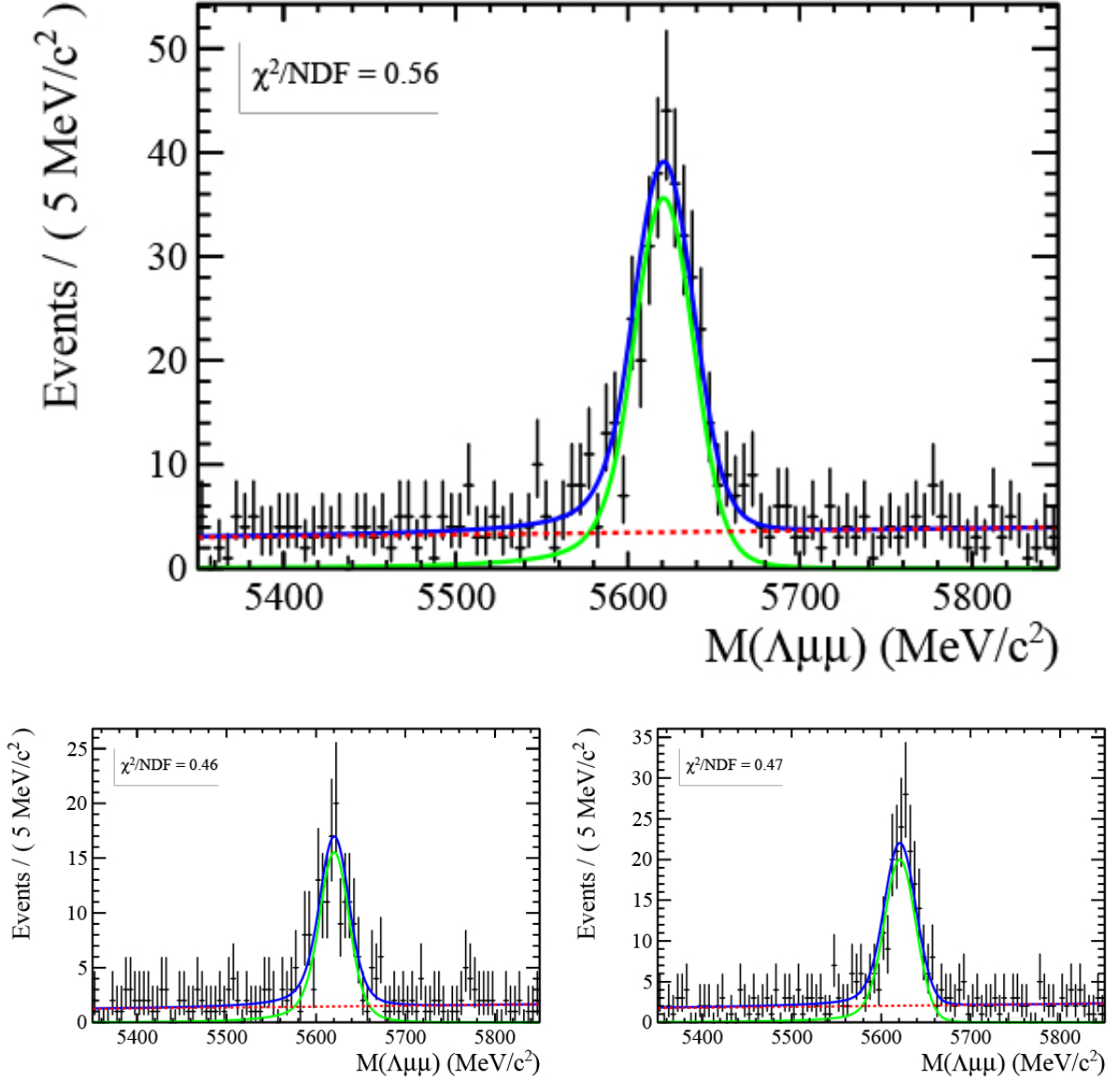


Figure 40: Fits to $\Lambda\mu^+\mu^-$ invariant mass distribution using the Psi fit model. Top - Both Long and Downstream tracks. Bottom left - Long tracks. Bottom right - Downstream tracks. Black points are data, the blue and green solid lines are the total fit and signal shape, respectively, along with goodness-of-fit values. The dashed red line represents the combinatorial background contributions.

listed in Table 14. Significances are measured using Wilks' theorem,

$$\sigma = \sqrt{2 \times (\log \mathcal{L}^{\text{Bkg}} - \log \mathcal{L}^{\text{Sig,Bkg}})}. \quad (36)$$

Here, $\log \mathcal{L}^{\text{Sig,Bkg}}$ is the log-Likelihood value for the original fit to the dataset, while $\log \mathcal{L}^{\text{Bkg}}$ is produced by fitting with the signal yield set to be zero. Wilks' theorem is only

valid if the amount of degrees of freedom are equal in both fits. By determining and fixing the values of $M_0, \sigma, \alpha_l, \alpha_r, n_l, n_r$ and CB Ratio from simulation, this condition is met.

Fit model	q^2 interval (GeV ² /c ⁴)	Track Type	Signal Yield	Significance
Mu	15.0 - 20.0	LL & DD	326 ± 21	23.4
		LL	131 ± 13	14.5
		DD	195 ± 16	18.3
Psi	15.0 - 20.0	LL & DD	326 ± 21	23.4
		LL	134 ± 14	14.5
		DD	193 ± 16	18.3

Table 14: $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ yields extracted from the $\Lambda \mu^+ \mu^-$ invariant mass distribution for both the Mu ($\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$) and Psi ($\Lambda_b^0 \rightarrow J/\psi \Lambda$) mass shape fit models, using parameters determined from simulation data.

The fits give a $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ yield of 326 ± 21 in the 15 to 20 GeV²/c⁴ q^2 interval, compared to 276 ± 20 from the previous analysis using the same 3 fb⁻¹ dataset [28]. We find the Mu and Psi fit shapes are found to provide yields in good agreement with each other. As seen in the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ yields in Table 10, the majority of events are found to be Downstream candidates, again due to the long-lived Λ baryons.

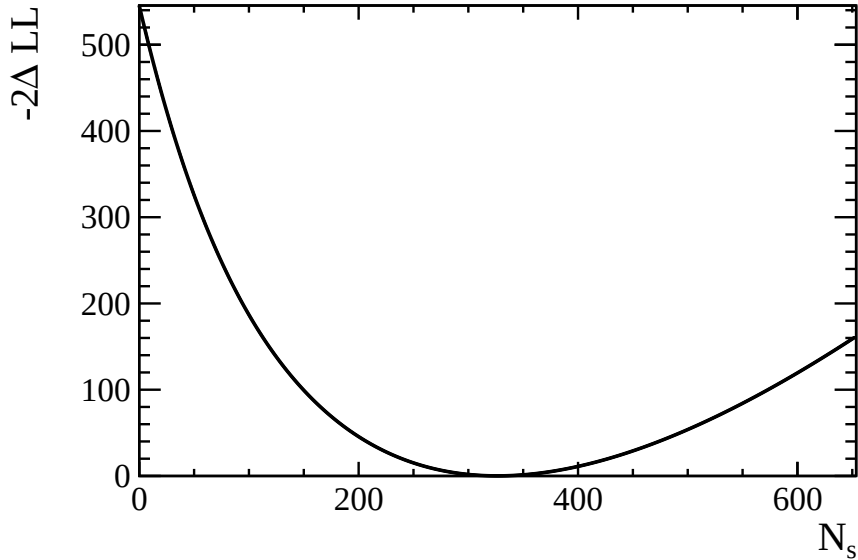


Figure 41: Profile log-Likelihood against number of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events in the 15 to 20 GeV²/c⁴ q^2 region.

As confirmation of the maximum likelihood fit, plots of the profile likelihood were produced by repeating the fit by fixing the signal at zero, and increasing in steps of +1 to

a maximum of twice the value given in Table 14. The values are calculated as

$$-2 \times (\log \mathcal{L}^{\text{Sig}', \text{Bkg}} - \log \mathcal{L}^{\text{Sig}, \text{Bkg}}), \quad (37)$$

where $\log \mathcal{L}^{\text{Sig}, \text{Bkg}}$ is the log-Likelihood value produced from the original fit, and $\log \mathcal{L}^{\text{Sig}', \text{Bkg}}$ is the value with the altered signal yield. The obtained values were then plotted by the signal yield which produced them. Figure 41 shows the plot to be a parabola, with the minimum at the previously given yield.

Using the Mu fit parameters, sWeights are calculated for the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ dataset. As will be described in Section 5.9, these weights are required to remove contributions from background events in the measurement of the distribution of moments. This is necessary as both the polarised and unpolarised moment calculations are otherwise unable to distinguish between signal and background while running over the dataset.

5.8 Signal Efficiency Correction

Due to preferential selection of specific regions of the phase space, the signal events in the final dataset are distorted during reconstruction and selection. It is necessary to weight events to repair this before we can extract moments. The required weights are defined as the inverse of the efficiency function, which in this case is 6-dimensional $(\cos \theta, \cos \theta_l, \cos \theta_h, \phi_h, \phi_l, q^2)$, and is modelled using Legendre polynomials. The acceptance function is calculated as

$$\epsilon(\Omega_5, q^2) = \sum_{i,j,k,l,m,n} c_{i,j,k,l,m,n} \times P_i(\cos \theta) P_j(\cos \theta_l) P_k(\cos \theta_h) P_l(\phi_l) P_m(\phi_h) P_n(q^2),$$

where $\Omega_5 \equiv (\cos \theta, \cos \theta_l, \cos \theta_h, \phi_l, \phi_h)$, and $P_i(x)$ represents the Legendre polynomial necessary to model the dimension, with x being the order of that polynomial, with those found to best model the efficiency function given in Table 15. As Legendre polynomials are defined over a range of -1 to +1, the variables ϕ_l , ϕ_h and q^2 have their values rescaled to accommodate this, as reflected in the integral limits. The coefficients $c_{i,j,k,l,m,n}$ are extracted using simulation data, after the preselection cuts are made, and the neural network working point has been applied. It is possible to determine the moments using the Legendre polynomials. These are equivalent to the normalised moments obtained using Equation 12, and are calculated as

$$M_{i,j,k,l,m,n} = \frac{1}{N_{\text{events}}} \sum_e^{N_{\text{events}}} P_i(\cos \theta^e) P_j(\cos \theta_l^e) P_k(\cos \theta_h^e) P_l(\phi_l^e) P_m(\phi_h^e) P_n(q^{2^e}), \quad (38)$$

where N_{events} is the number of events in the simulated dataset, and e is the given event. As previously stated in Section 2.2, these normalised moments are equivalent to the theoretical ones,

$$M_{i,j,k,l,m,n} = \frac{1}{N} \int_{-1}^{+1} \int_{-1}^{+1} \int_{-1}^{+1} \int_{-1}^{+1} \int_{-1}^{+1} \int_{-1}^{+1} \epsilon(\Omega_5, q^2) \times P_i(\cos \theta) P_j(\cos \theta_l) P_k(\cos \theta_h) P_l(\phi_l) P_m(\phi_h) P_n(q^2) dV, \quad (39)$$

where

$$N \equiv \int_{-1}^{+1} \int_{-1}^{+1} \int_{-1}^{+1} \int_{-1}^{+1} \int_{-1}^{+1} \int_{-1}^{+1} \epsilon(\Omega_5, q^2) dV \quad (40)$$

and

$$dV \equiv d \cos \theta d \cos \theta_l d \cos \theta_h d \phi_l d \phi_h d q^2. \quad (41)$$

Using these, and the orthogonal nature of Legendre polynomials,

$$\int_{-1}^{+1} P_n(x) P_m(x) dx = \frac{2}{2n+1} \delta_{nm}, \quad (42)$$

an expression for the coefficients is given as

$$c_{i,j,k,l,m,n} = c_{000000} \cdot M_{i,j,k,l,m,n} (2i+1)(2j+1)(2k+1)(2l+1)(2m+1)(2n+1). \quad (43)$$

Variable	Orders (LL)	Orders (DD)
$\cos \theta$	6	6
$\cos \theta_l$	6	6
$\cos \theta_h$	6	6
ϕ_l	8	8
ϕ_h	8	8
q^2	6	6

Table 15: Orders required to describe the Legendre polynomials used for the efficiency parametrisation of LL and DD track type events.

One caveat with this determination is that each coefficient is multiplied by an unmeasurable constant, c_{000000} . This is equal to $= 2N$, and has no effect on final moments, so is set to equal 1. Using the efficiency models, 1-D angular projections are produced for each of the 6 observables, by generating a dataset of only signal candidates. These distributions are then overlaid with the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ dataset after preselection, q^2 and NN restrictions are applied, to confirm that our efficiency model is successful in describing the distortions arising within the analysis. Figures 42 and 43 show the comparison of distributions in the real dataset to those produced from the efficiency models. Good agreement between the data points and distributions generated from the efficiency models is seen, implying the model successfully corrects for the distortions that arise through the analysis.

The shape of $\cos \theta_l$ is symmetric about 0, as the masses of the two leptons are the same. Conversely, the distribution of $\cos \theta_h$ is asymmetric, the proton is more likely to be found to be travelling backwards relative to the Λ baryon. This arises from the mass difference between the proton and the pion, with the p roughly 7 times more massive than the π^- . With the Λ baryon only decaying weakly, the effect arises from vector

and axial-vector contributions [76], and this causes the breakdown of parity within the decay. As the azimuthal distribution of the Λ baryon is uniform, $\cos \theta$ is also seen to be symmetric about 0. The distribution of ϕ_l shows two minima at about 1.57 and -1.57 rad, corresponding to 90° and 270° respectively. At these values, the momentum vectors of the muons are parallel to that of the Λ_b^0 baryon, with one pointing back towards the primary vertex. As discussed in Section 5.1, muon tracks with small impact parameter values are removed in stripping, so these events will be reduced. The same trend can be seen in ϕ_h , however as the majority of Λ baryons are reconstructed from Downstream tracks - which have a lower IP minimum in our stripping requirements than the leptons - this effect is less pronounced. In the Downstream ϕ_h distribution, an asymmetric minima is visible at about -1.57 rad, which would correspond to the momentum vector of the proton being perpendicular to the Λ_b^0 baryon direction, pointing towards the interaction point. The asymmetry arises from the proton-pion mass difference, with the proton likely to be pointed towards the interaction point, and therefore removed from the dataset by stripping requirements.

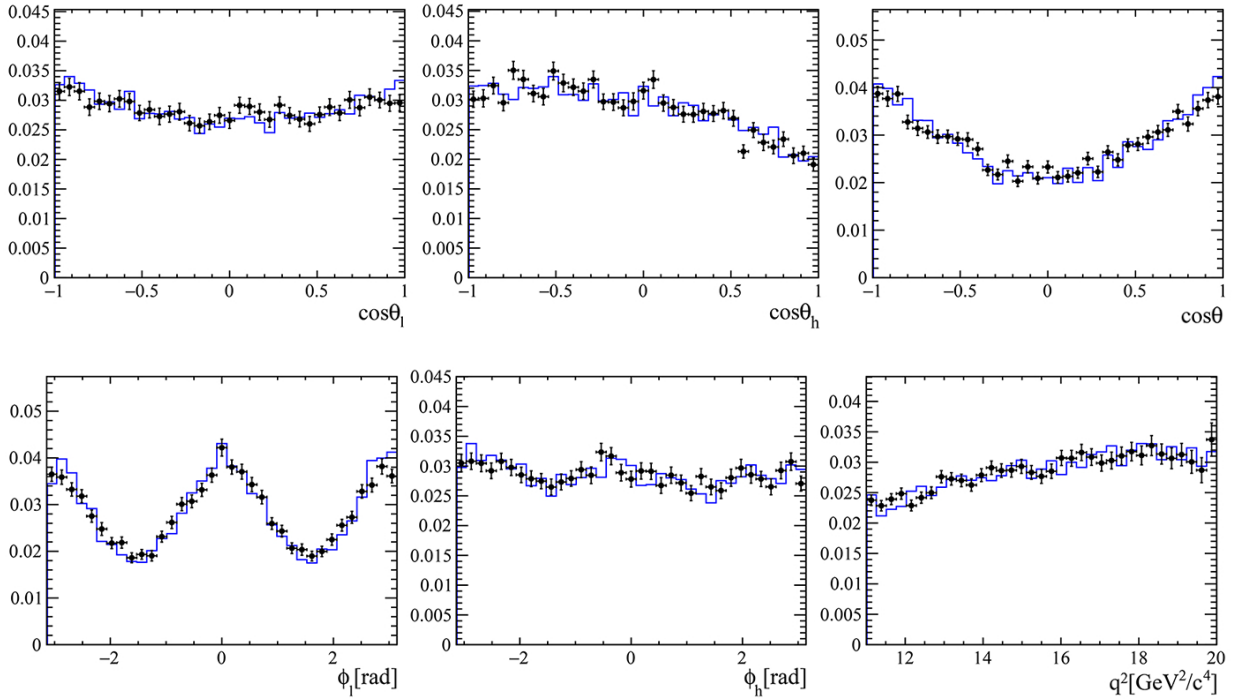


Figure 42: One dimensional plots of the 6-dimensional efficiencies of $\cos \theta_l$, $\cos \theta_h$, $\cos \theta$, ϕ_l , ϕ_h and q^2 , for LL events, with q^2 values between 11 and 20 GeV²/c⁴, compared to the real $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ dataset, after preselection. Black points are real data after being corrected with the efficiency model. The blue distribution is the angular PDF derived from the efficiency model. Distributions are normalised to 1, giving probability per bin on the y-axis.

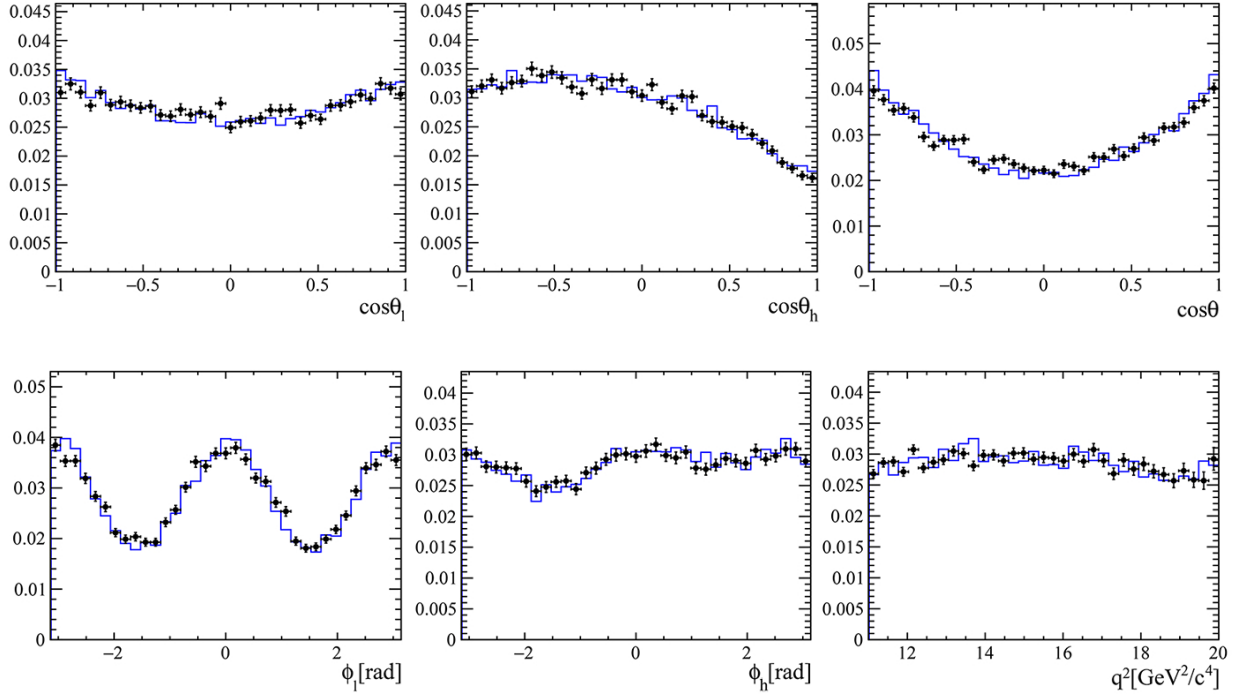


Figure 43: One dimensional plots of the 6-dimensional efficiencies of $\cos \theta_l$, $\cos \theta_h$, $\cos \theta$, ϕ_l , ϕ_h and q^2 , for DD events, with q^2 values between 11 and 20 GeV²/c⁴, compared to the real $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ dataset, after preselection. Black points are real data after being corrected with the efficiency model. The blue distribution is the angular PDF derived from the efficiency model. Distributions are normalised to 1, giving probability per bin on the y-axis.

5.9 Background Removal

After preselection requirements, the application of the q^2 interval, and neural network selection, the dataset still contains background events. As Equation 12 runs over every event given to it, it is necessary to weight the distribution such that only $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays contribute to the results. In order to do this, a probability density function is fitted to the data with a maximum likelihood fit within RooFit, before being sWeighted [62] against the Λ_b^0 baryon mass shape. These signal weights provide the means to correct the distribution. The fits, as previously described in Section 5.3, are performed by fixing the values of $\alpha_l, \alpha_r, n_l, n_r$ and CB Ratio while allowing M_0 and σ to vary in the signal shape. The exponential coefficient describing the combinatorial background was free to be determined by the code, then fixed for the sWeighting process. Although not contributing in the high q^2 interval, the parameters associated with the $B^0 \rightarrow J/\psi K_S^0$ distribution were included in the modelling process. This is required, as despite the additional restrictions applied to the $J/\psi \Lambda$ dataset, Figure 35 shows there is still contribution from these misreconstructed events that requires modelling. Inclusion of the sWeights means that the equation for calculation of moments after modification for signal efficiency and background correction becomes

$$M_i = \left(\sum_e^{N_{\text{events}}} {}_s\mathcal{P}_n(y_e) \right)^{-1} \cdot \sum_e^{N_{\text{events}}} \frac{{}_s\mathcal{P}_n(y_e)}{\epsilon} f_i(\Omega_5^e), \quad (44)$$

with $f_i(\Omega_5^e)$ representing the relevant weighting function from Table 3 for event e , and ${}_s\mathcal{P}_n(y_e)$ being the sWeight for that event. The sum of the sWeights for a given species is equal to the number of events of that species. As previously discussed, sWeighting is only valid as long as the control and discriminating variables are uncorrelated. To confirm this was the case, correlation plots of the reconstructed mass of the Λ_b^0 baryon and $\cos \theta_l, \cos \theta_h, \cos \theta, \phi_l, \phi_h$ and q^2 were made. The dataset after application of stripping, trigger, preselection, and after neural network requirements was used, without a q^2 limit. The plots are shown in Figures 44 and 45. The structures in these plots are the same as discussed at the end of Section 5.8, with those in q^2 due to $J/\psi \rightarrow \mu^+ \mu^-$ decays.

Background may alternatively be removed by sideband correction. The distribution of moments was determined using events in a given signal window and in a sideband containing only background. The background yields are required to be equal in both windows, and this is achieved by multiplying the sideband moment by the ratio of these. The alternative determination of moments then becomes

$$M_i^{\text{SB Corrected}} = \frac{1}{N_{\text{sig}}} \cdot \left(\left(\sum_e^{N_{\text{sig}} + N_{\text{bkg}}} \frac{f_i(\Omega_5^e)}{\epsilon} \right) - \left(\frac{N_{\text{bkg}}}{N_{\text{bkg}'}} \cdot \sum_e^{N_{\text{bkg}'}} \frac{f_i(\Omega_5^e)}{\epsilon} \right) \right), \quad (45)$$

where N_{bkg} is the yield in the signal window, and $N_{\text{bkg}'}$ is in the sideband. Both run independently to allow a confirmation of the processes. Both methods are successful in removing background contributions from the final determination of the moments.

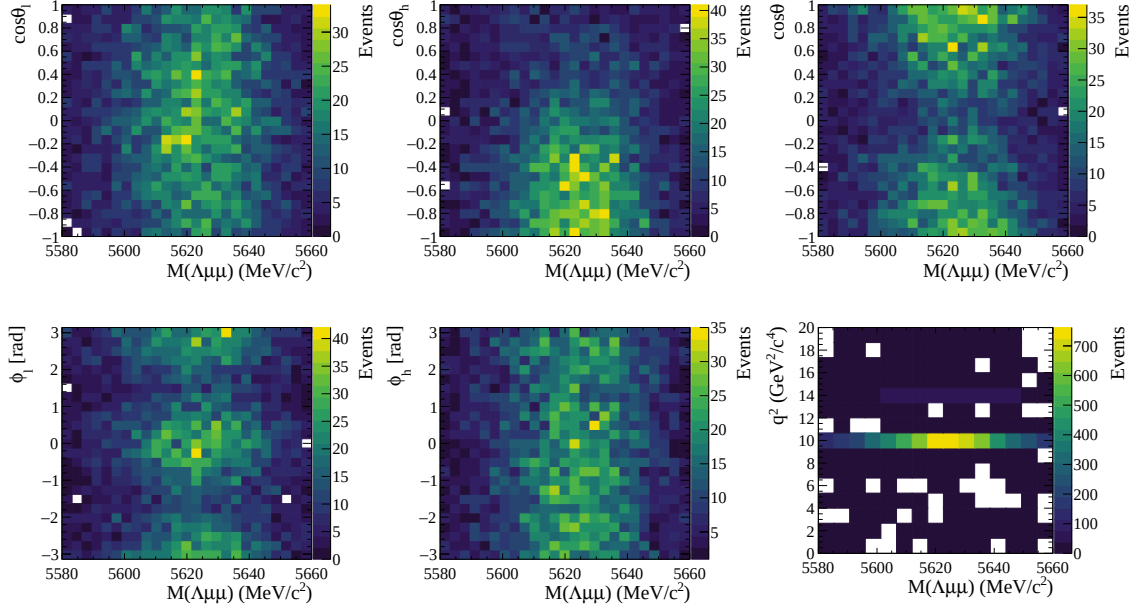


Figure 44: 2-Dimensional plots of reconstructed Λ_b^0 baryon mass by angular variables and q^2 for Long candidates. Plots created using the final analysis dataset, without q^2 limits. No correlation to the Λ_b^0 mass can be seen.

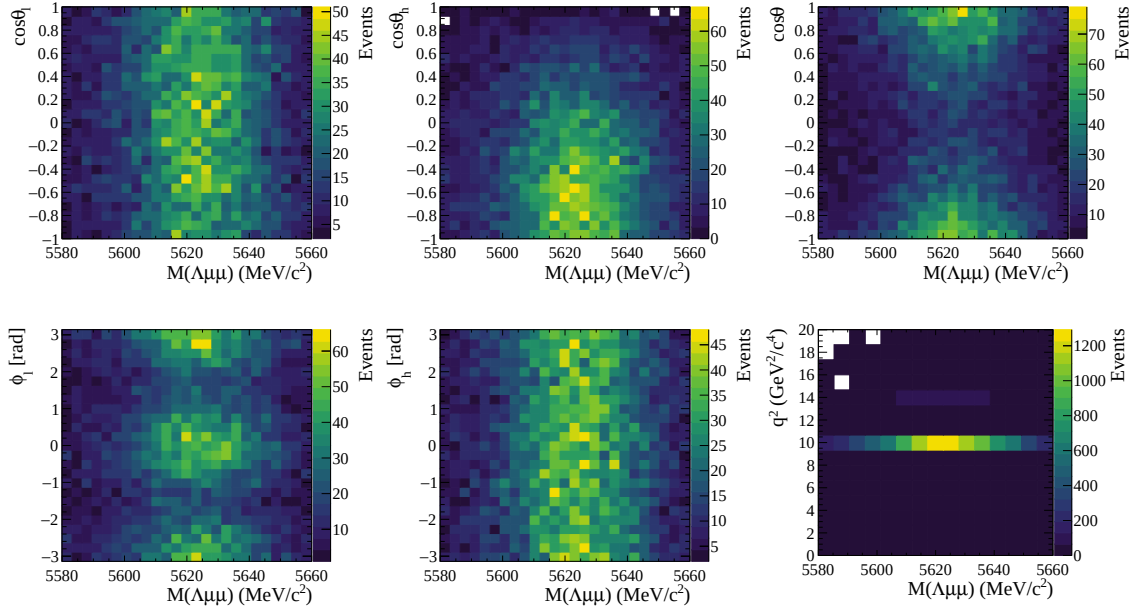


Figure 45: 2-Dimensional plots of reconstructed Λ_b^0 baryon mass by angular variables and q^2 for Downstream candidates. Plots created using the final analysis dataset, without q^2 limits. No correlation to the Λ_b^0 mass can be seen.

5.10 Determination of Uncertainties

Uncertainties are measured using bootstrapping [77], in which the dataset is re-sampled in order to determine statistical uncertainties. A set number of new datasets are created with a random selection of events from the original set. Each then undergoes signal efficiency and background subtraction corrections before producing a distribution of moments. A preselected number of bootstrap samples are produced, while the number of events in a given sample is generated from a Poisson distribution with the total event yield as its mean. The width of the distribution of the bootstrapped moments is then taken as the uncertainty on the value produced with the full dataset. The widths are measured using root mean square (RMS) values, and are determined from the moment distributions as

$$\sigma(M_i) = \sqrt{\frac{\langle M_i^2 \rangle - \langle M_i \rangle^2}{N}}, \quad (46)$$

where

$$\langle M_i^2 \rangle = \sum_s^N (M_i^s)^2, \quad (47)$$

and

$$\langle M_i \rangle^2 = \left(\sum_s^N M_i^s \right)^2. \quad (48)$$

The value of N is the total number of samples taken, s is the current sample, and M_i is the moment of interest. By resampling all 34 moments, we may produce uncertainties for each value of M_i . This also helps determine that the values of any moments do not change wildly depending on the events chosen from a given study.

6 Simulation Studies

Before moments could be extracted from our final dataset, it was necessary to validate the signal efficiency model and background subtraction methods mentioned in Sections 5.8 and 5.9, respectively. To avoid biasing, studies were performed using simulated events, which underwent the same treatment as the real data would do - as discussed in Section 6.1. The results from these studies allowed us to validate the machinery involved in calculating the moments and uncertainties on signal-only data, and these steps are discussed in Section 6.2. The simulation studies also allowed for a reconfirmation of the optimal working point of the neural network, and this is detailed in Section 6.3. Comparison between moments extracted for $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events to those previously published allowed for validation of the efficiency and background removal models, as described in Section 6.4. The goal of the simulation is to test the analytical tools working in the background to calculate the distribution of moments and the uncertainties. With the output validated, the studies are performed using the final dataset in place of simulated events, after preselection, neural network and q^2 requirements are applied, allowing determination of the distribution of moments of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$.

6.1 Generation of Events

The simulation studies used in this analysis create events with angular distributions and masses determined by pre-defined probability density functions. Polarised events are created with random number generators which produce values between -1 and 1 for $\cos \theta$, $\cos \theta_h$, $\cos \theta_l$, and between $-\pi$ and π for ϕ_h and ϕ_l . For unpolarised studies, the value of $\cos \theta$ is fixed at 1, while ϕ_h is set to be 0, with ϕ_l then being treated as the angle ϕ . The probability for an event with this angular distribution is calculated with Equation 7 or 14, dependant on chosen polarisation, and accept-rejected against a randomly generated number. Candidates which pass this stage are then assigned a random $\Lambda \mu^+ \mu^-$ invariant mass value, in the range 5350 to 5850 MeV/ c^2 . As the efficiency models differ, the studies are required to be able to produce either only Long or Downstream candidates, in addition to both. With a mass assigned to the event, t is generated as a random number between 0 and 1, then used to determine the track type of the event. The ratio of Long to Downstream stream events is determined from the $J/\psi \Lambda$ yields given in Table 10. Specifically, we calculate the value as

$$\frac{N_{\Lambda_b^0 \rightarrow J/\psi \Lambda}^{\text{LL}}}{N_{\Lambda_b^0 \rightarrow J/\psi \Lambda}^{\text{LL}} + N_{\Lambda_b^0 \rightarrow J/\psi \Lambda}^{\text{DD}}} \approx 0.354. \quad (49)$$

If t is less than or equal to 0.354, the event is considered LL, and DD if larger. The mass is then accept-rejected against the Long or Downstream PDF, modelled by the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ (Mu) fit parameters in Table 13.

Background is added to the simulated datasets with a uniform angular distribution, with only the mass variable being accept-rejected. Combinatorial background is modelled

by an exponential function with a coefficient of -1.56×10^{-3} , determined from simulation data. This is the only background present in the generation of simulated $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events. The simulation may also generate a mass distribution based on the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ fit parameters in Table 9, while using the same angular PDF and LL/DD candidate fraction. In this validation channel, misreconstructed $B^0 \rightarrow J/\psi K_S^0$ decays is also be modelled, with a fixed percentage of the total generated background being produced this way. Similar to the method used to split events into Long and Downstream track types, the fraction of the total background contributed by $B^0 \rightarrow J/\psi K_S^0$ is calculated as

$$\frac{N_{B^0 \rightarrow J/\psi K_S^0}}{N_{B^0 \rightarrow J/\psi K_S^0} + N_{\text{Combinatorial Bkg}}} = B_{\text{Frac}}^0. \quad (50)$$

The value of this fraction depends on whether the simulation is producing $J/\psi \Lambda$ events with or without neural network and proton ID restrictions. The yields for both cases were given in Tables 10 and 11 respectively. In either case, a random number is again generated between 0 and 1, and should this be less than B_{Frac}^0 , the event is treated as being due to $B^0 \rightarrow J/\psi K_S^0$. Like the LL and DD signal events, the background is then accept/rejected against the required PDF. This allows us to generate a simulation sample of the $J/\psi \Lambda$ dataset without the restrictions applied to remove $J/\psi K_S^0$ from data, giving us opportunity to test their removal.

The analysis code will continue to run until a preset number of signal and background events have been accepted. With these candidates produced, the dataset is corrected for detector efficiencies, and sWeighted in order to remove the background events. The distributions of moments are then determined, using either Equation 12 or 17, depending on polarisation. Uncertainties are measured by taking a predetermined number of bootstrap samples from the simulated dataset, in order to obtain RMS values. The process of generating events, applying efficiency and background corrections, calculating values for each moment, and then determining uncertainties with the specified number of bootstrap samples is referred to as a study. When multiple studies are performed, the process is repeated from the beginning, creating a new dataset of the same number of signal and background events, and then applying the same corrections to these, before determining new uncertainties with the same number of bootstrap samples, and the new distribution of moments. These multiple studies are then combined into a single output dataset, creating distributions of the calculated values.

6.2 Validation with Simulated Events

We validate the simulated studies by comparing the moments generated from two datasets. The first is produced from Monte Carlo simulated events after preselection requirements, that are weighted by the q^2 phase space of generator level MC events. This dataset the same as that used to determine the efficiency functions, and we weight the events with the inverse of the model. The dataset we compare to is a distribution of events generated with

a uniform efficiency model, as would be expected in a situation with no detector effects or distortions from the analysis process. Should the efficiency parametrisation be valid, the moments produced from these two sets of 6 dimensional ($\cos\theta, \cos\theta_l, \cos\theta_h, \phi_l, \phi_h, q^2$) distributions should be in good agreement, as the model would correct for distortions. This test is performed on both $\Lambda_b^0 \rightarrow J/\psi \Lambda$ and $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events, using the signal and background yields extracted from fits to data. The validation checks were performed separately on Long and Downstream track type events. Figure 46 shows direct comparison between the values of the moments for each dataset for both the $J/\psi \Lambda$ and $\Lambda \mu^+ \mu^-$ decays, while Figure 47 shows the variations between each value. No significant deviations between the calculated values of M_i are found in either dataset for either track type, and the model therefore successfully accounts for the detector effects in simulated events.

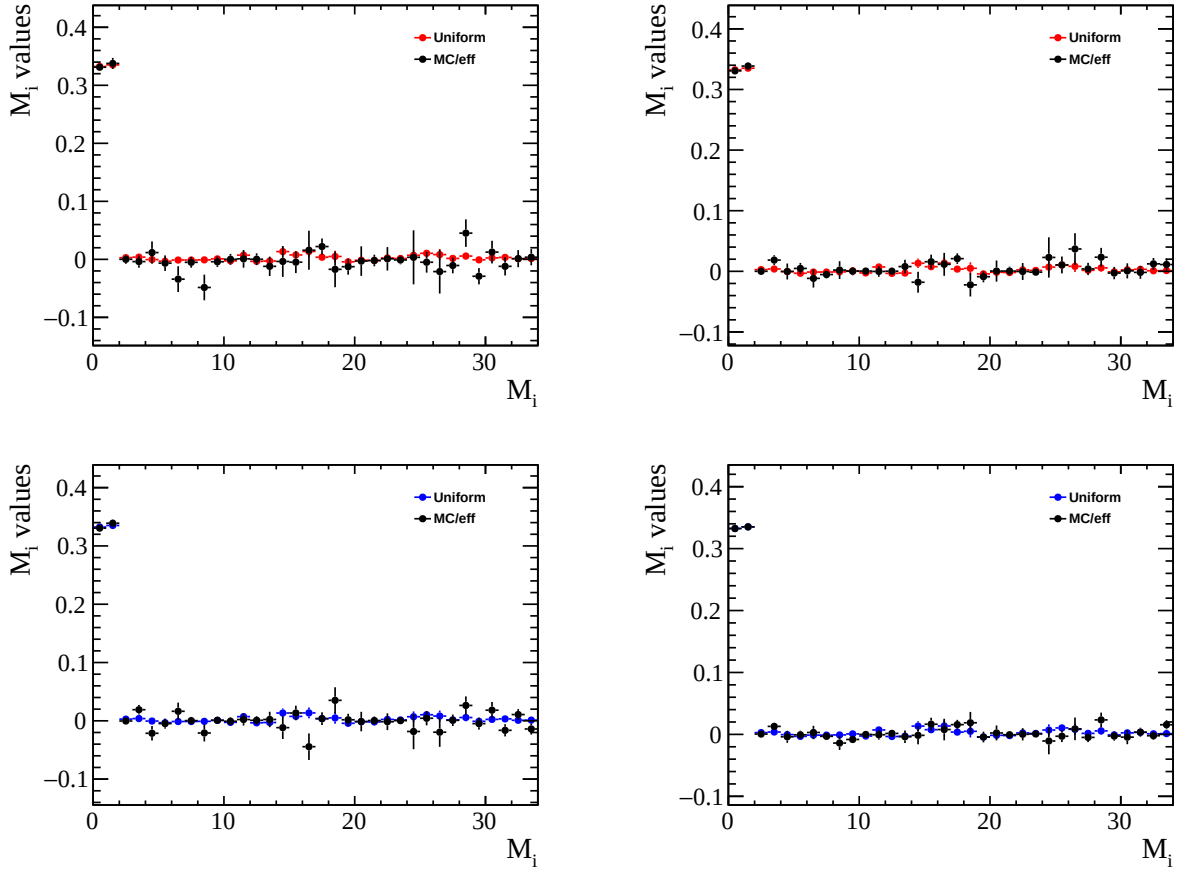


Figure 46: Comparison of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ (top) and $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ (bottom) moments generated from a 6-D distribution with uniform efficiency (red) to those generated from an efficiency weighted Monte Carlo simulated dataset (black). Left - Long candidates. Right - Downstream candidates.

After validating output with alternating efficiency models, the determination of moments and uncertainties generated for simulated events was tested. Datasets for Long and

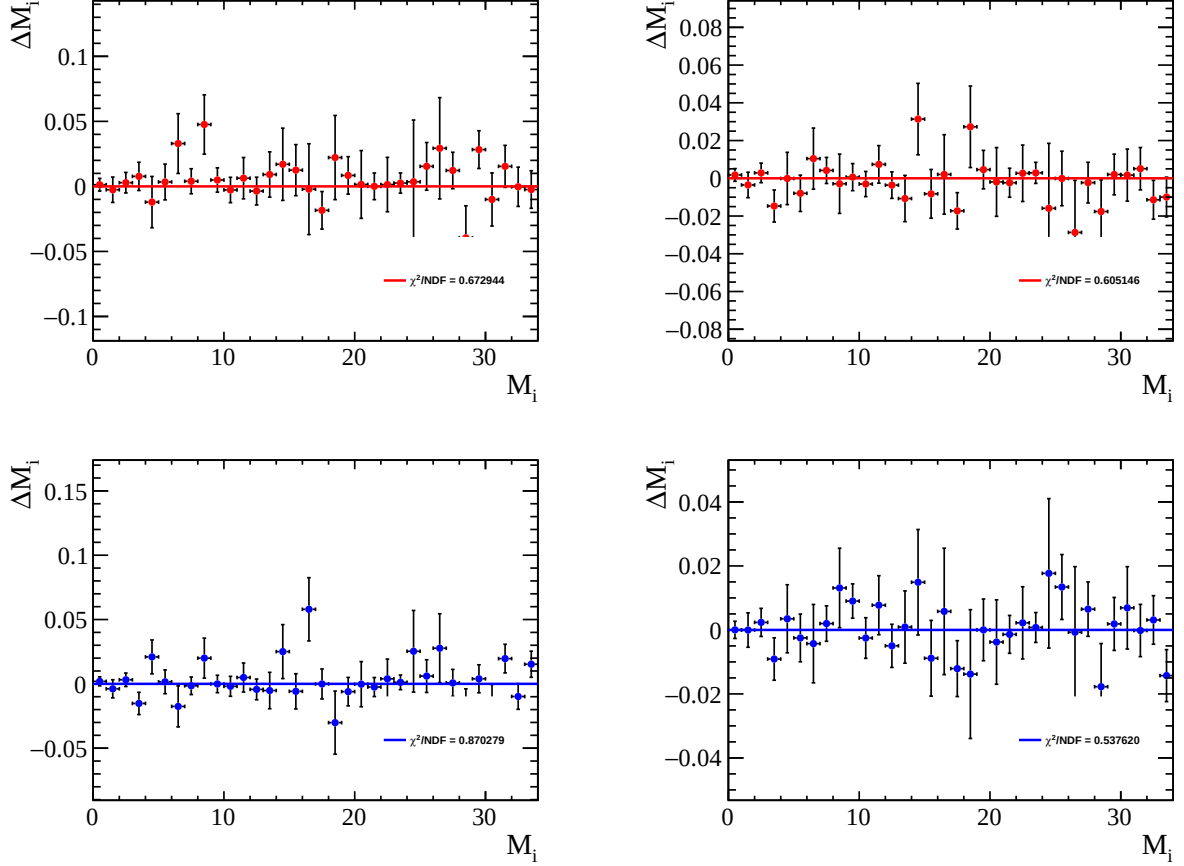


Figure 47: Difference between the moment values determined from the two aforementioned $\Lambda_b^0 \rightarrow J/\psi \Lambda$ (top) and $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ (bottom) datasets. Left - Long candidates. Right - Downstream candidates.

Downstream candidates were produced with their respective efficiency models, each containing 500 studies of signal polarised $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays. Studies were performed using expected Long and Downstream event yields, instead of extracted values. Equation 35 was used to calculate N_{Sig} for a 0.24 neural network restriction, which was then treated with the LL/DD ratio determined in Equation 49 to give 108 Long and 162 Downstream events. The individual studies were then produced using these values. Pull is calculated as

$$\text{Pull} = \frac{(M_i^{\text{Extracted}} - M_i^{\text{True}})}{\sigma(M_i)}, \quad (51)$$

where the values of $M_i^{\text{Extracted}}$ and M_i^{True} are taken as the means of the Gaussian distributions generated from multiple studies. The value of $\sigma(M_i)$ is determined by Equation 46, using the distribution of uncertainties on M_i , generated via bootstrapping over multiple simulation studies. For perfectly unbiased moment determination, and correct error calculation, the pull distributions will be Gaussian in shape, with a mean of 0, and a width

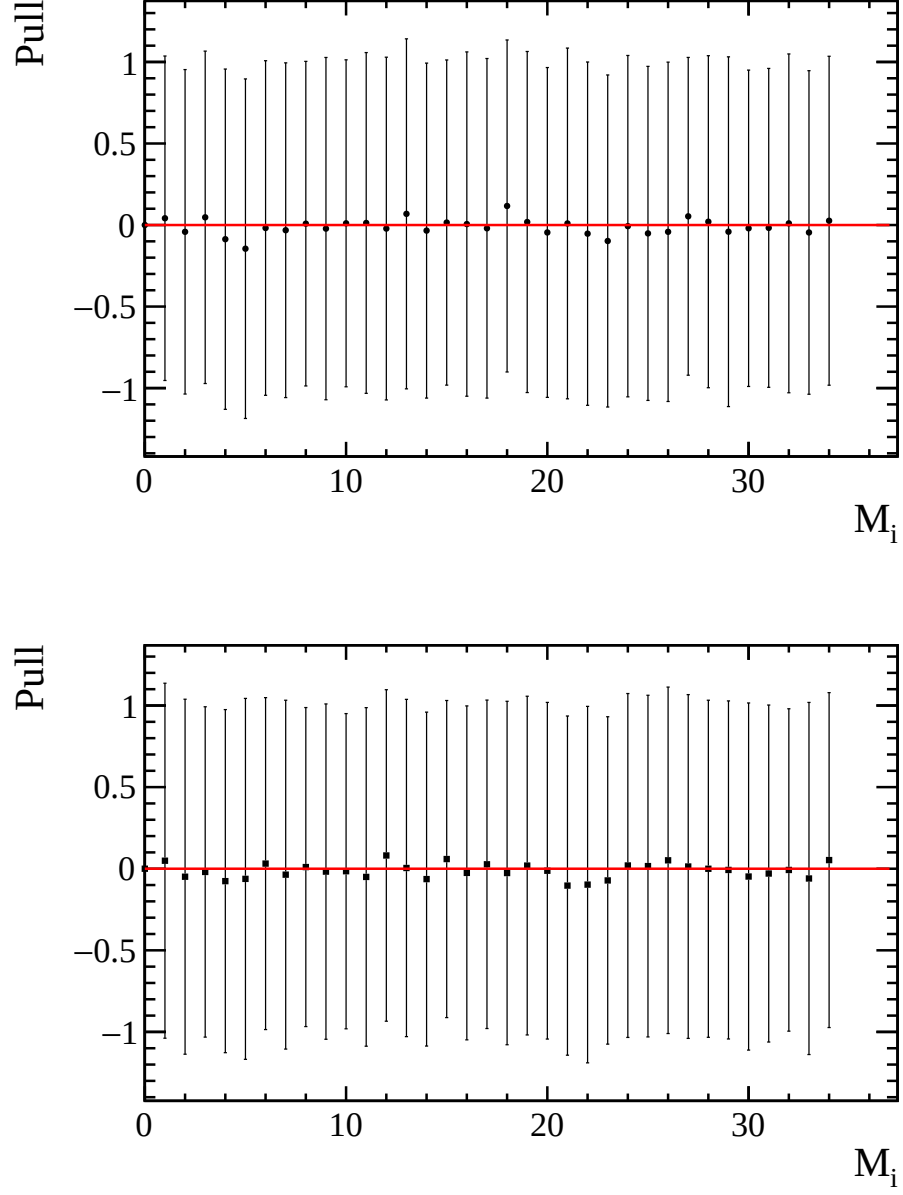


Figure 48: Means and widths of the Gaussian distributions of pull plots for the moments calculated in LL (left) and DD (right) simulation studies, using their respective efficiencies. For unbiased determination of the values of M_i , the mean value of the Gaussian is required to be 0, with a width of 1, and this is seen in both datasets for all moments.

of 1. The means and widths from each distribution created in these checks are shown in Figure 48, and they show no bias present in the determination of the distribution of moments. The individual distributions of the pull plots generated in this validation study are displayed in Appendix C.

6.3 Reconfirmation of Neural Network Optimisation

Maximising significance to optimise the neural network is slightly crude, as the background and signal yields are dependant on arbitrarily selected mass intervals. Too wide of a window about the signal peak will greatly increase the expected background yield, with no similar improvement in signal yield to counteract this. Simulation studies offer a more accurate method to optimise of the neural network, using the expected values of signal and background yields in the 5350 to 5850 MeV/ c^2 $\Lambda\mu^+\mu^-$ invariant mass window of the dataset. The steps detailed in Section 5.6 were repeated to extract signal and background yields for 40 neural network cuts between -1 and 1. The value of N_{Sig} was calculated with Equation 35, and N_{Bkg} is determined by integrating the PDF fitted to the 5900 to 7000 MeV/ c^2 sideband over the 5350 to 5850 MeV/ c^2 range. By generating 500 uniform efficiency simulation studies with these signal and background values each with 200 bootstrap samples, the distribution of unpolarised moments was determined. The distribution of uncertainties associated to these moments produced by studies were Gaussian, and the mean values of each were taken. Plotting the means by the neural network cut used to extract the yields which produced them allowed for reconfirmation of the optimisation. The optimum neural network cut is the value at which moment uncertainty is lowest. The plots are shown as Figure 49, and show the same trend as the significance plot produced originally, with minimum mean uncertainties at neural network cuts consistent with the value of 0.24 used to optimise the neural network to find $\Lambda_b^0 \rightarrow \Lambda\mu^+\mu^-$ events. This process confirms the neural network optimisation performed in Section 5.6. All 10 plots have minimum uncertainties at similar working points, and they all agree with the value of 0.24 determined from maximising the signal significance.

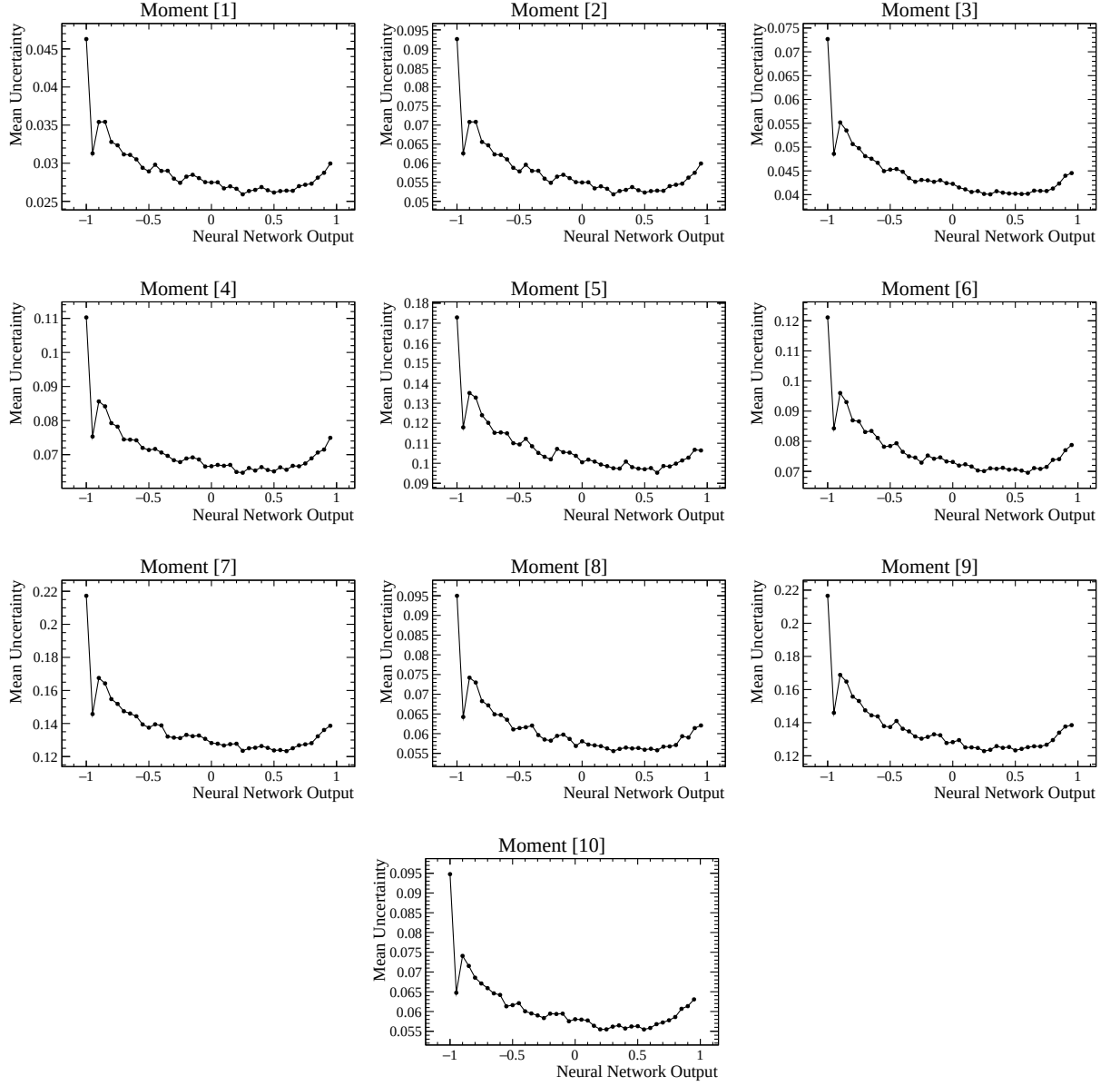


Figure 49: Mean uncertainties by neural network output for the moments associated to unpolarised Λ_b^0 baryon production in $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$, produced via simulation studies.

6.4 Validation with $\Lambda_b^0 \rightarrow J/\psi \Lambda$ Events

Further validation of the output was possible by comparing results extracted from real data to previously published values. Using the parameters listed in Table 16, we are able to determine the values of M_i we would expect should the analysis code be provided with a $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset. As an angular analysis on $\Lambda_b^0 \rightarrow J/\psi \Lambda$ has been performed [35], we may use these results to determine expected values for the moments of the decay. The four amplitudes a_+, a_-, b_+, b_- and the Λ_b^0 baryon transverse polarisation production are measured by the original $\Lambda_b^0 \rightarrow J/\psi \Lambda$ angular analysis, while the Λ baryon decay asymmetry α_Λ is published as 0.642 ± 0.013 [1]. Using these values and our expressions, we produce the expectation values in Table 16. The phases of the decay amplitudes are unmeasured, and we cannot determine any expectation values for the moments sensitive to them. By comparing expected values to those measured by passing the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ dataset after neural network and proton ID restrictions to our analysis code, we are able to validate with real results. In this case, the mass shape of $M(\Lambda_b^0)$ is determined by fitting to data, and not simulation. Comparison of our calculated expected values to those measured from data allows us to test the sWeighting and efficiency model. The results taken from the analysis code are given in Table 17, for both the 3 fb^{-1} dataset, but also using only the events collected during 2011, as this is what the original analysis was performed on. Reasonable agreement between the expected and generated non-zero values were found, and those which we determined to be zero were measured to be consistent with this value. The values are overlaid in the plot shown in Figure 50, in order to better see the comparisons.

Our analysis code is developed to also work in reverse - by providing it with values for moments, it may generate angular distributions corresponding to those numbers. With the extracted moments, simulation studies were performed using the values of M_i as a basis for generation of a large signal-only dataset. By overlaying these angular distribution produced with those in the sWeighted $\Lambda_b^0 \rightarrow J/\psi \Lambda$ data, we are able to confirm that the real angular distributions and those used to calculate the moments by the simulation studies are in good agreement. This allows us to confirm the weighting background removal is successful in removing peaking and combinatorial events. Plots are produced for both the 2011 and full Run 1 datasets, for all 5 angular variables, split into both LL and DD track types. The plots are presented in Figure 51 for 2011, with Figure 52 for the full Run 1 dataset. We find that the predicted angular distributions are in good agreement with sWeighted events in the dataset.

Observable	Physics parameters	Expectation
M_1	$1/4 \times (2 a_+ ^2 + 2 a_- ^2 + b_+ ^2 + b_- ^2)$	0.395 ± 0.005
M_2	$1/2 \times (b_+ ^2 + b_- ^2)$	0.210 ± 0.010
M_3	0	0
M_4	$\alpha/4 \times (b_- ^2 - b_+ ^2 + 2 a_+ ^2 - 2 a_- ^2)$	-0.278 ± 0.024
M_5	$\alpha/2 \times (b_- ^2 - b_+ ^2)$	-0.196 ± 0.026
M_6	0	0
M_7	$\alpha/\sqrt{2} \times \Re(-b_+^* a_+ + b_- a_-^*)$	n/a
M_8	0	0
M_9	$\alpha/\sqrt{2} \times \Im(b_+^* a_+ - b_- a_-^*)$	n/a
M_{10}	0	0
M_{11}	$P_b/4 \times (b_+ ^2 - b_- ^2 + 2 a_+ ^2 - 2 a_- ^2)$	-0.008 ± 0.009
M_{12}	$P_b/2 \times (b_+ ^2 - b_- ^2)$	0.018 ± 0.022
M_{13}	0	0
M_{14}	$P_b \alpha/4 \times (- b_- ^2 - b_+ ^2 + 2 a_+ ^2 + 2 a_- ^2)$	0.009 ± 0.011
M_{15}	$-P_b \alpha/2 \times (b_+ ^2 + b_- ^2)$	-0.008 ± 0.010
M_{16}	0	0
M_{17}	$-P_b \alpha/\sqrt{2} \times \Re(b_+^* a_+ + b_- a_-^*)$	n/a
M_{18}	0	0
M_{19}	$P_b \alpha/\sqrt{2} \times \Im(b_+^* a_+ + b_- a_-^*)$	n/a
M_{20}	0	0
M_{21}	$-P_b/\sqrt{2} \times \Im(b_+^* a_- - b_- a_+^*)$	n/a
M_{22}	0	0
M_{23}	$P_b/\sqrt{2} \times \Re(b_+^* a_- - b_- a_+^*)$	n/a
M_{24}	0	0
M_{25}	$P_b \alpha/\sqrt{2} \times \Im(b_+^* a_- + b_- a_+^*)$	n/a
M_{26}	0	0
M_{27}	$-P_b \alpha/\sqrt{2} \times \Re(b_+^* a_- + b_- a_+^*)$	n/a
M_{28}	0	0
M_{29}	0	0
M_{30}	$P_b \alpha \times \Im(a_+ a_-^*)$	n/a
M_{31}	0	0
M_{32}	$-P_b \alpha \times \Re(a_+ a_-^*)$	n/a
M_{33}	$-P_b \alpha/2 \times \Re(b_+^* b_-)$	n/a
M_{34}	$P_b \alpha/2 \times \Im(b_+^* b_-)$	n/a

Table 16: Values of moments expected using weighing functions to extract $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ angular observables in terms of the physics parameters associated with $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decays.

Observable	Expectation	Numerical Value (2011)	Numerical Value (2011 + 2012)
M_1	0.395 ± 0.005	0.383 ± 0.007	0.372 ± 0.004
M_2	0.210 ± 0.01	0.235 ± 0.014	0.255 ± 0.008
M_3	0	-0.014 ± 0.011	-0.010 ± 0.006
M_4	-0.278 ± 0.024	-0.268 ± 0.021	-0.248 ± 0.011
M_5	-0.196 ± 0.026	-0.152 ± 0.028	-0.185 ± 0.016
M_6	0	-0.027 ± 0.021	-0.019 ± 0.011
M_7	n/a	0.048 ± 0.035	0.022 ± 0.018
M_8	0	-0.010 ± 0.016	-0.020 ± 0.008
M_9	n/a	0.004 ± 0.035	-0.004 ± 0.020
M_{10}	0	-0.011 ± 0.016	-0.004 ± 0.009
M_{11}	-0.008 ± 0.009	0.015 ± 0.015	0.004 ± 0.008
M_{12}	0.018 ± 0.022	-0.024 ± 0.021	-0.010 ± 0.011
M_{13}	0	0.001 ± 0.014	-0.003 ± 0.008
M_{14}	0.009 ± 0.001	0.017 ± 0.028	-0.004 ± 0.015
M_{15}	-0.008 ± 0.010	-0.008 ± 0.036	0.013 ± 0.019
M_{16}	0	-0.015 ± 0.027	-0.014 ± 0.014
M_{17}	n/a	0.059 ± 0.046	-0.003 ± 0.025
M_{18}	0	-0.008 ± 0.022	0.001 ± 0.012
M_{19}	n/a	0.089 ± 0.046	0.035 ± 0.023
M_{20}	0	-0.003 ± 0.022	0.014 ± 0.012
M_{21}	n/a	-0.025 ± 0.041	-0.005 ± 0.023
M_{22}	0	-0.003 ± 0.017	0.002 ± 0.009
M_{23}	n/a	0.005 ± 0.034	-0.000 ± 0.017
M_{24}	0	0.003 ± 0.015	0.014 ± 0.008
M_{25}	n/a	-0.032 ± 0.074	-0.009 ± 0.040
M_{26}	0	0.019 ± 0.032	-0.004 ± 0.017
M_{27}	n/a	0.000 ± 0.060	-0.007 ± 0.033
M_{28}	0	0.007 ± 0.029	-0.013 ± 0.015
M_{29}	0	0.008 ± 0.038	-0.018 ± 0.020
M_{30}	n/a	-0.043 ± 0.025	-0.014 ± 0.013
M_{31}	0	-0.060 ± 0.032	0.002 ± 0.016
M_{32}	n/a	0.030 ± 0.028	-0.005 ± 0.015
M_{33}	n/a	-0.029 ± 0.024	-0.020 ± 0.013
M_{34}	n/a	-0.042 ± 0.025	-0.014 ± 0.013

Table 17: Combined moments for the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decays using DD and LL events from 2011 and 2011 + 2012 datasets, and with expected values.

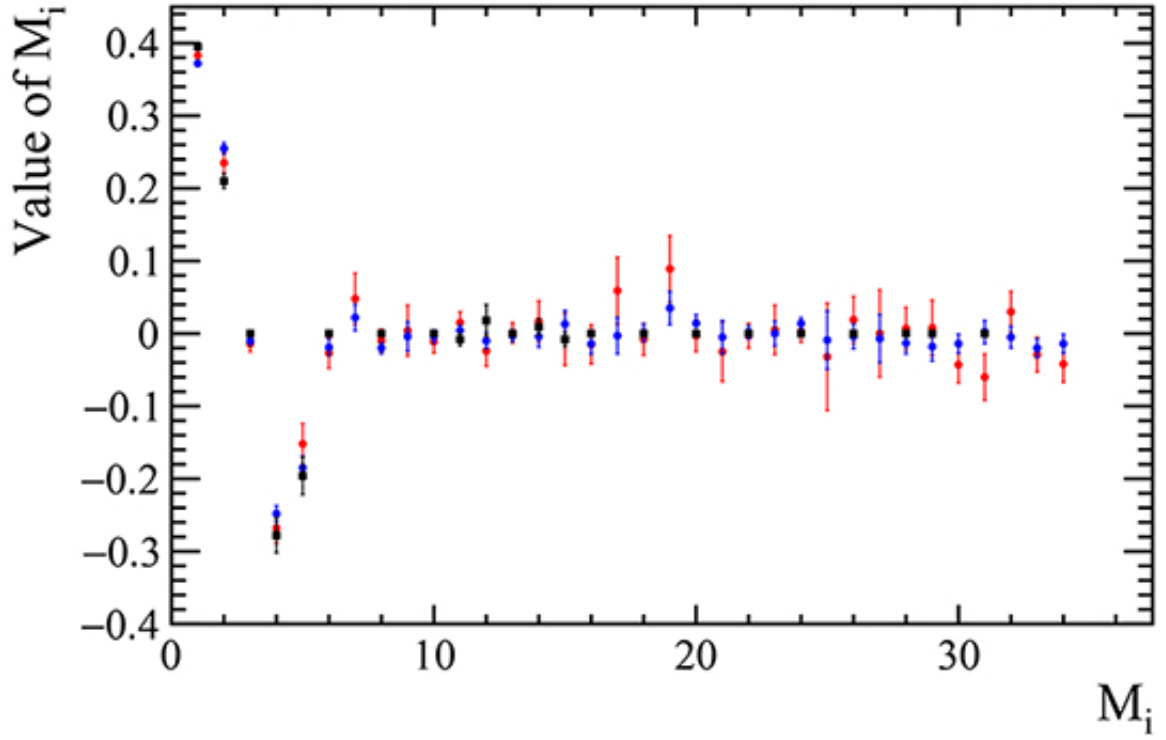


Figure 50: Comparison of our calculated expected values of specific moments of the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decay (black) to 2011 data (red) and 2011 + 2012 (blue). The values of M_i sensitive to decay amplitudes do not have expected values plotted. Good agreement is seen between both extracted and expected values for each moment.

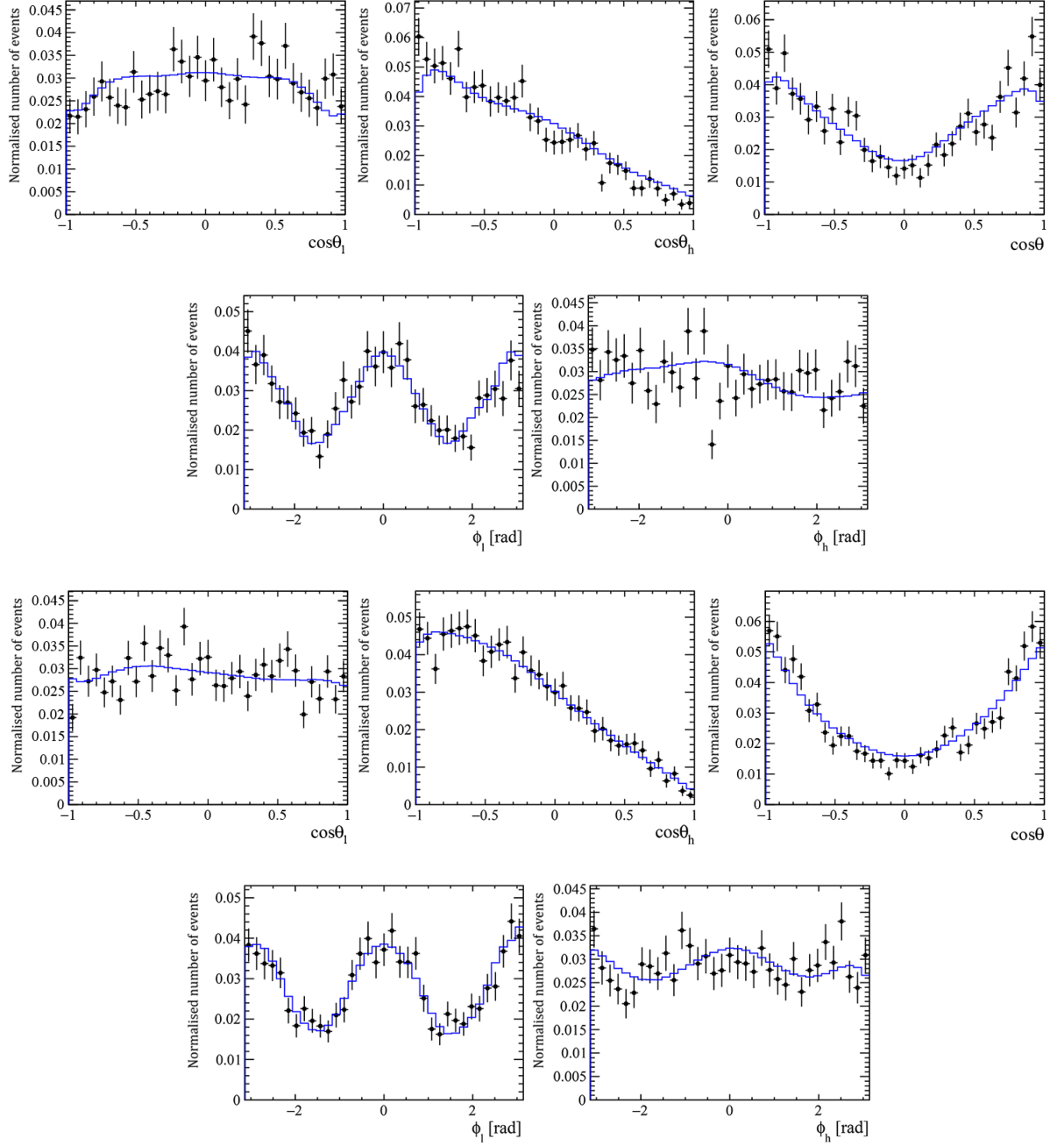


Figure 51: Angular distribution of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decays from data (black points) superimposed with the angular distribution as obtained by setting angular observables in the simulation studies to the values obtained by extracting moments for LL events (top 5) and DD events (bottom 5) using the 2011 dataset.

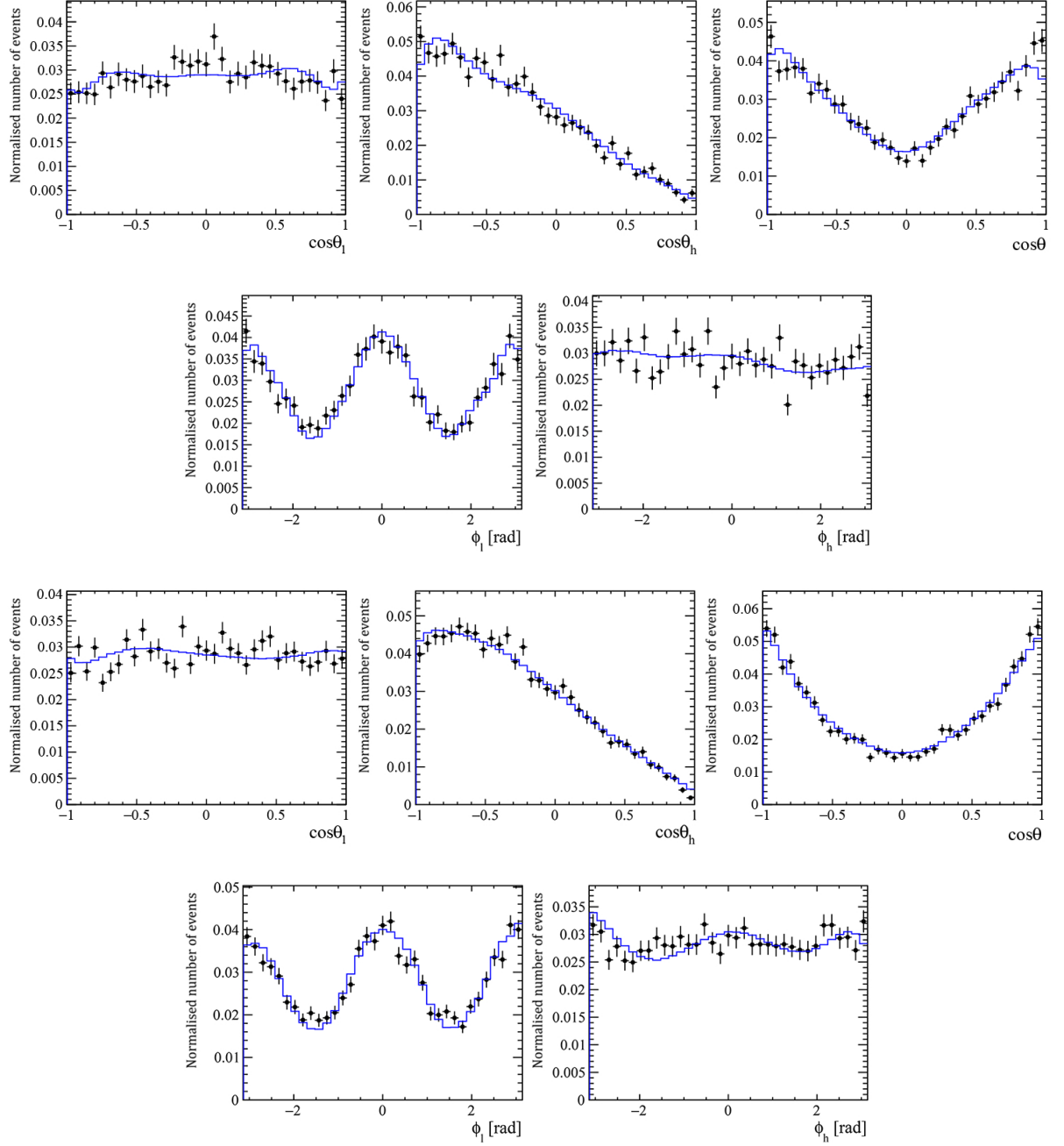


Figure 52: Angular distribution of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ decays from data (black points) superimposed with the angular distribution as obtained by setting angular observables in the simulation studies to the values obtained by extracting moments for LL events (top 5) and DD events (bottom 5) using the 2011 + 2012 datasets.

7 Results

7.1 Distributions of Moments and Angles, and Measurement of Angular Observables

With the analysis code validated, moments were extracted from the $\Lambda\mu^+\mu^-$ dataset after preselection, neural network and q^2 requirements were applied, and the distribution had been sWeighted. Individual values of the extracted moments assuming $P_b \neq 0$, are shown in Table 18. Due to the differing efficiency models required for LL and DD events, the values were extracted independently. The results from the complete dataset are determined by taking a weighted average of the separate values. The value of each moment in the full dataset is calculated as

$$M_i^{\text{Tot}} \pm \delta M_i^{\text{Tot}} = \frac{w_{\text{LL}} M_i^{\text{LL}} + w_{\text{DD}} M_i^{\text{DD}}}{w_{\text{LL}} + w_{\text{DD}}} \pm (w_{\text{LL}} + w_{\text{DD}})^{-1/2}, \quad (52)$$

where

$$w_{\text{LL}} = \frac{1}{(\delta M_i^{\text{LL}})^2}, \quad (53)$$

and

$$w_{\text{DD}} = \frac{1}{(\delta M_i^{\text{DD}})^2}. \quad (54)$$

M_i^x represents the measured value of moment i for track type x , and δM_i^x is its statistical uncertainty. Moments 11 to 34 - those associated to polarised Λ_b^0 production - are found to be consistent with zero, which is expected with unpolarised Λ_b^0 baryon production. This method is not as precise as the determination of the polarisation from the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ analysis [35], but does show agreement. Figure 53 gives the distribution of the moments extracted from data, while Figure 54 compares the values determined from the individual Long and Downstream track datasets. Both Table 18 and Figure 54 show larger uncertainties on values of M_i determined from Long track events when compared to those of Downstream moments. This is caused by the use of bootstrapping to determine them, and the resultant distribution not being Gaussian. The small data sample leads to large RMS variations, with the same event appearing multiple times in the different bootstrap datasets, causing large uncertainties that do not scale as expected. Using our extracted moments, simulation studies were performed to produce angular PDFs for $\cos\theta$, $\cos\theta_h$, $\cos\theta_l$, ϕ_h and ϕ_l in the 5350 to 5850 MeV/ c^2 $\Lambda\mu^+\mu^-$ invariant mass range, separated into Long and Downstream events. These were overlaid with the sWeighted real data distributions, and are shown in Figures 55 and 56.

Observable	Numerical Value (LL + DD)	Numerical Value (LL)	Numerical Value (DD)
M_1	0.387 ± 0.033	0.377 ± 0.110	0.388 ± 0.035
M_2	0.226 ± 0.067	0.245 ± 0.220	0.224 ± 0.070
M_3	-0.003 ± 0.051	0.041 ± 0.170	-0.007 ± 0.054
M_4	-0.181 ± 0.072	-0.146 ± 0.123	-0.198 ± 0.089
M_5	-0.269 ± 0.143	0.032 ± 0.632	-0.286 ± 0.147
M_6	-0.009 ± 0.095	0.053 ± 0.399	-0.013 ± 0.098
M_7	0.217 ± 0.144	-0.189 ± 0.282	0.360 ± 0.167
M_8	-0.120 ± 0.058	-0.200 ± 0.092	-0.068 ± 0.075
M_9	-0.117 ± 0.160	0.397 ± 0.692	-0.146 ± 0.165
M_{10}	0.006 ± 0.066	0.009 ± 0.196	0.005 ± 0.070
M_{11}	-0.020 ± 0.067	-0.014 ± 0.124	-0.023 ± 0.079
M_{12}	-0.039 ± 0.104	-0.202 ± 0.735	-0.035 ± 0.105
M_{13}	0.017 ± 0.076	0.128 ± 0.447	0.014 ± 0.077
M_{14}	0.075 ± 0.139	0.185 ± 0.289	0.043 ± 0.158
M_{15}	0.090 ± 0.174	-0.182 ± 1.755	0.092 ± 0.175
M_{16}	0.185 ± 0.131	0.203 ± 1.237	0.184 ± 0.132
M_{17}	0.125 ± 0.225	-0.246 ± 0.573	0.192 ± 0.245
M_{18}	-0.041 ± 0.101	-0.053 ± 0.187	-0.037 ± 0.121
M_{19}	0.379 ± 0.232	0.039 ± 1.823	0.385 ± 0.234
M_{20}	-0.201 ± 0.108	-0.099 ± 0.566	-0.205 ± 0.110
M_{21}	-0.192 ± 0.156	-0.144 ± 0.264	-0.218 ± 0.194
M_{22}	-0.012 ± 0.056	-0.070 ± 0.101	0.013 ± 0.067
M_{23}	0.116 ± 0.115	0.288 ± 0.176	-0.011 ± 0.152
M_{24}	0.000 ± 0.049	-0.028 ± 0.073	0.022 ± 0.065
M_{25}	-0.162 ± 0.287	-0.742 ± 0.494	0.133 ± 0.353
M_{26}	-0.128 ± 0.100	0.070 ± 0.193	-0.200 ± 0.117
M_{27}	-0.211 ± 0.214	-0.391 ± 0.374	-0.124 ± 0.260
M_{28}	0.016 ± 0.090	0.115 ± 0.130	-0.077 ± 0.126
M_{29}	-0.109 ± 0.155	0.154 ± 0.376	-0.163 ± 0.170
M_{30}	-0.091 ± 0.086	0.023 ± 0.129	-0.184 ± 0.116
M_{31}	0.048 ± 0.144	0.017 ± 1.023	0.049 ± 0.146
M_{32}	-0.065 ± 0.100	-0.015 ± 0.284	-0.072 ± 0.106
M_{33}	0.016 ± 0.085	-0.027 ± 0.156	0.034 ± 0.102
M_{34}	-0.082 ± 0.088	0.120 ± 0.151	-0.186 ± 0.109

Table 18: Values of the moments determined from $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events in the 15 to 20 GeV^2/c^4 q^2 interval, assuming that the Λ_b^0 baryon has a non-zero transverse production polarisation. All values are produced from the Run 1 dataset. Uncertainties are the standard deviation of the bootstrap samples. The moments measurable in the decay of a Λ_b^0 baryon with non-zero transverse production polarisation P_b are found to be consistent with zero.

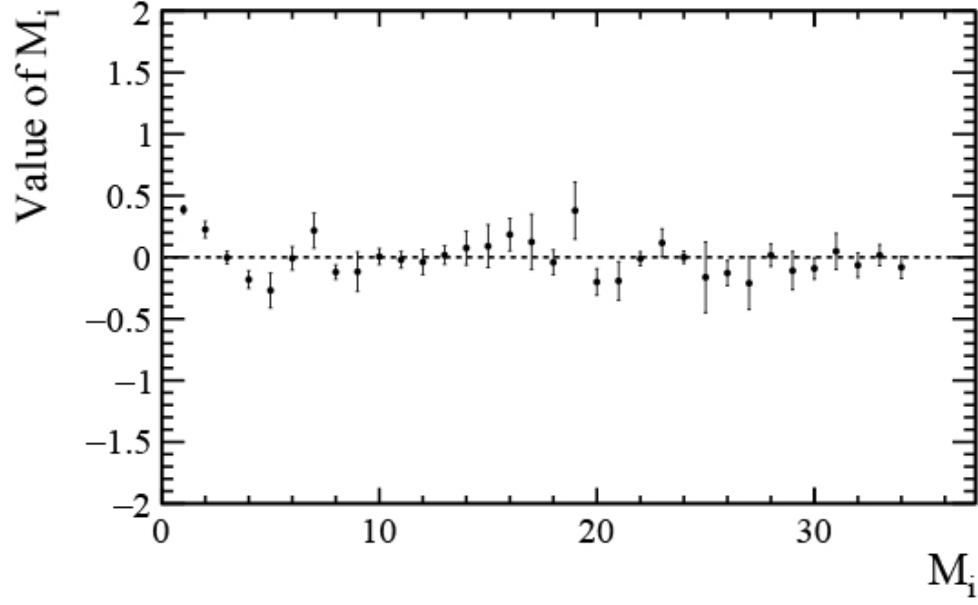


Figure 53: Distribution of the combined Long and Downstream moments determined from the separate datasets assuming non-zero Λ_b^0 baryon transverse polarisation production. Values measured with events taken from the 15 to 20 GeV^2/c^4 q^2 interval.

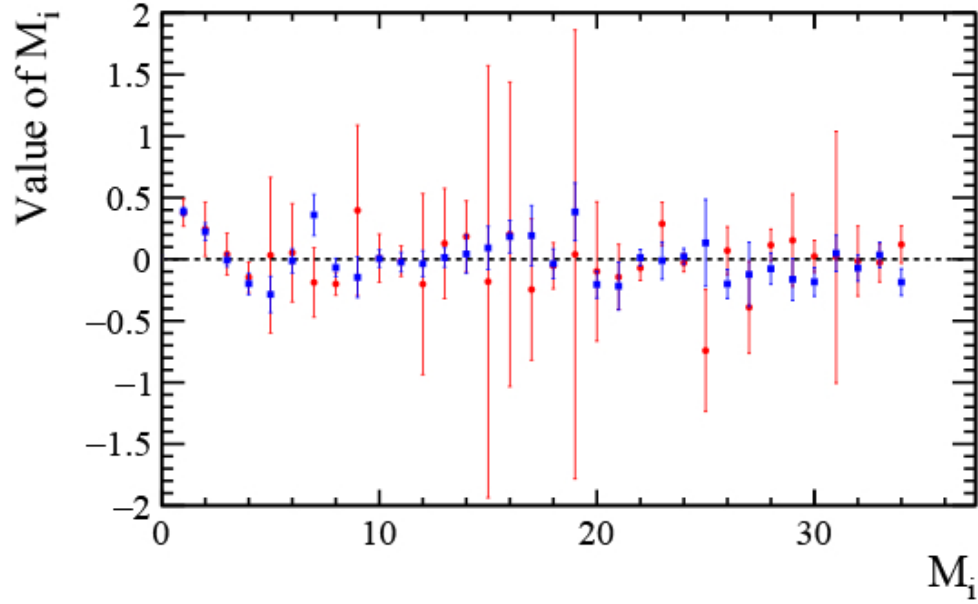


Figure 54: Comparison of the distributions of moments extracted from LL (red circles) and DD (blue squares) datasets independently. Values determined from events taken from the 15 to 20 GeV^2/c^4 q^2 interval.

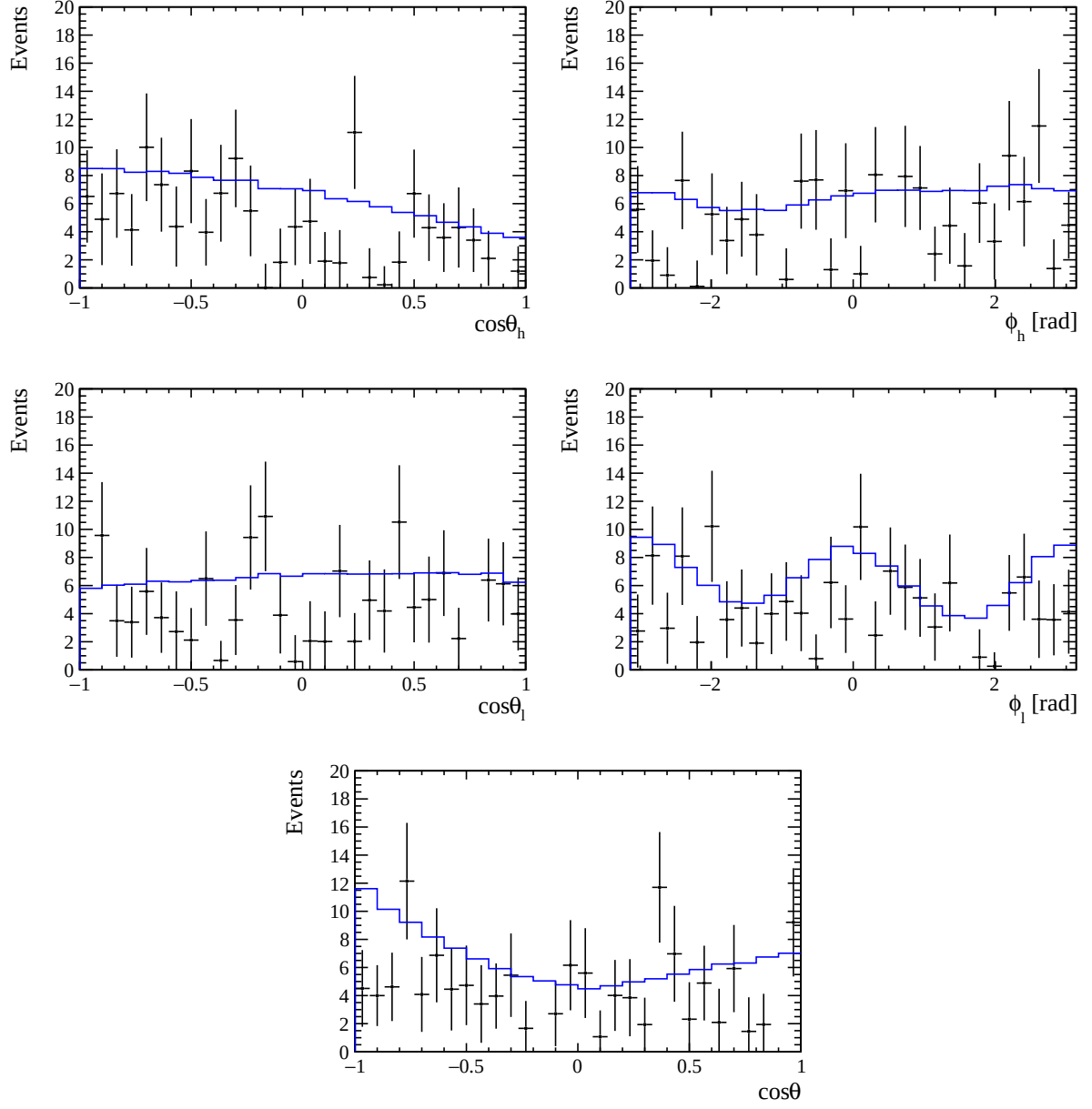


Figure 55: Angular distributions of Long events from data (black points) over the 15 to 20 GeV^2/c^4 q^2 range, after all selection requirements and application of optimal neural network working point. Events are sWeighted, and taken from a 5350 to 5850 MeV/c^2 invariant $\Lambda\mu^+\mu^-$ mass range. The blue histogram is the angular PDF generated by simulation studies.

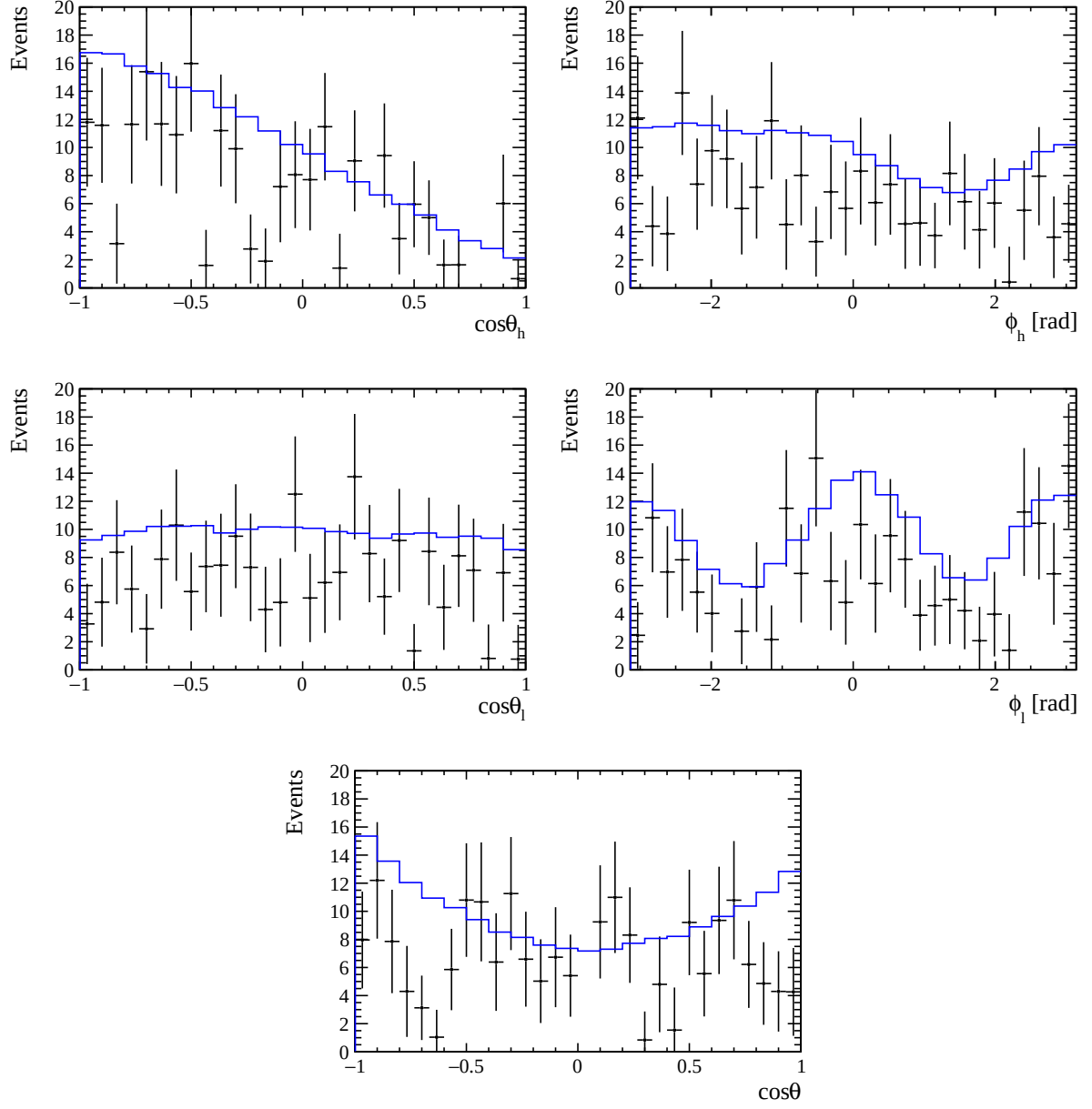


Figure 56: Angular distributions of Downstream events from data (black points) over the 15 to 20 GeV^2/c^4 q^2 range, after all selection requirements and application of optimal neural network working point. Events are sWeighted, and taken from a 5350 to 5850 MeV/c^2 invariant $\Lambda\mu^+\mu^-$ mass range. The blue histogram is the angular PDF generated by simulation studies.

Using the extracted values and functions given in Equation 13, measurements of f_L , A_{FB}^l , A_{FB}^h and A_{FB}^{lh} are made, as shown in Table 19. Where observables are calculated from two moments, uncertainties are corrected to account for correlation with use of covariances. The values for the full dataset observables are determined using the same weighted average calculation used to measure moments. Observables determined by moments are comparable to those measured in [28], and these are given in Table 20, along with the theoretical predictions to these and additional measured observables.

Observable	Numerical Value (LL + DD)	Numerical Value (LL)	Numerical Value (DD)
f_L	0.546 ± 0.132	0.509 ± 0.376	0.552 ± 0.141
A_{FB}^l	-0.004 ± 0.077	0.062 ± 0.255	-0.010 ± 0.081
A_{FB}^h	-0.312 ± 0.150	-0.130 ± 0.407	-0.341 ± 0.162
A_{FB}^{lh}	-0.007 ± 0.071	0.040 ± 0.299	-0.009 ± 0.073

Table 19: Angular observables measured from extraction of the moments associated to the decay $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$. These measurements are specific to the 15 to 20 GeV^2/c^4 q^2 interval assuming that the Λ_b^0 baryon has a non-zero transverse production polarisation. Values presented for the total dataset and for the individual LL and DD datasets.

Observable	This Work	Theoretical Values [31]	Previous LHCb Analysis [28]
f_L	0.546 ± 0.132	0.409 ± 0.013	$0.61_{-0.14}^{+0.11} \pm 0.03$
A_{FB}^l	-0.004 ± 0.077	-0.350 ± 0.013	$-0.05 \pm 0.09 \pm 0.03$
A_{FB}^h	-0.312 ± 0.150	-0.2710 ± 0.0092	$-0.29 \pm 0.07 \pm 0.03$
A_{FB}^{lh}	-0.007 ± 0.071	$0.1398(43)$	-
M_4	-0.181 ± 0.072	$-0.1947(68)$	-
M_5	-0.269 ± 0.143	$-0.1526(65)$	-
M_9	0.117 ± 0.160	$-0.0196(55)$	-
M_{10}	0.006 ± 0.066	$-0.1031(97)$	-

Table 20: Comparison of observables determined in this work to the previous analysis, and to theoretical values determined from lattice QCD. Theoretical values of moments are those predicted not to be zero in non-polarised Λ_b^0 baryon production. All values are from the 15 to 20 GeV^2/c^4 q^2 interval. Uncertainties are statistical and systematic, respectively. Values in brackets are confidence on the last digit.

7.2 Determination of Systematic Uncertainties

Systematic uncertainties arose in the process where choices were made on how to perform the analysis, and needed to be measured. These were mostly found to be in the models used in the fitting of probability density functions to the data, both in terms of the parameters describing them, and the shapes chosen. Systematic uncertainties associated to fixing PDF values from simulation data to the real invariant mass distributions were determined by taking a selection of values from the final variance matrix, then using these to recalculate signal yields. These yields were plotted as a histogram, and produced Gaussian distributions with a mean at the optimum value, and a width representing the uncertainty. As the yield extracted for $\Lambda_b^0 \rightarrow J/\psi \Lambda$ events is used to optimise the neural network, the systematic uncertainties on fixing the PDF parameters associated to this mass shape were also determined in the same way as the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ fits. With the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ plots also containing misreconstructed $B^0 \rightarrow J/\psi K_S^0$ events, these parameters are varied too. 500 additional fits for each were produced, and the histograms of the result yields are shown in Figure 57, with systematics as percentages determined by the mean and RMS values given in the plots.

Uncertainties on the functions used to define the signal and background PDFs are determined by switching to an alternative, then comparing resulting yields. For $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$, the double Crystal Ball modelling the signal was switched to a Gaussian, and the background exponential was changed to a 2nd order polynomial. In the case of the $\Lambda_b^0 \rightarrow J/\psi \Lambda$, the double Crystal Ball signal shape was again switched for a Gaussian, with the background exponential then separately replaced with a 5th order polynomial, as this was the lowest order found to accurately model the data. Differences between the selected and alternate fit yields were used to determine the uncertainties. The systematic uncertainty of the neural network was measured using the difference in yields between the generic and specific q^2 training, as previously discussed in Section 5.2. Each training was optimised independently to produce two working points. The differences in the yields of the decay modes modelled are given in Table 21 as percentage uncertainties. Values associated to the fixing of PDF shapes are measured as $\frac{\text{Std Dev} \times 100}{\text{Mean}}$. All other values are calculated by comparing final decay mode yields to those measured with the final analysis method, to give percentage differences between them. Total uncertainty on the $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ yield is measured by adding the individual values in quadrature.

Systematic uncertainties on the moments themselves are not published here, but they are dependant on the signal efficiency model and use of sWeighting to remove the background events. The same processes involved in fitting PDFs to invariant mass distributions are used in the background extraction in the simulation studies, however.

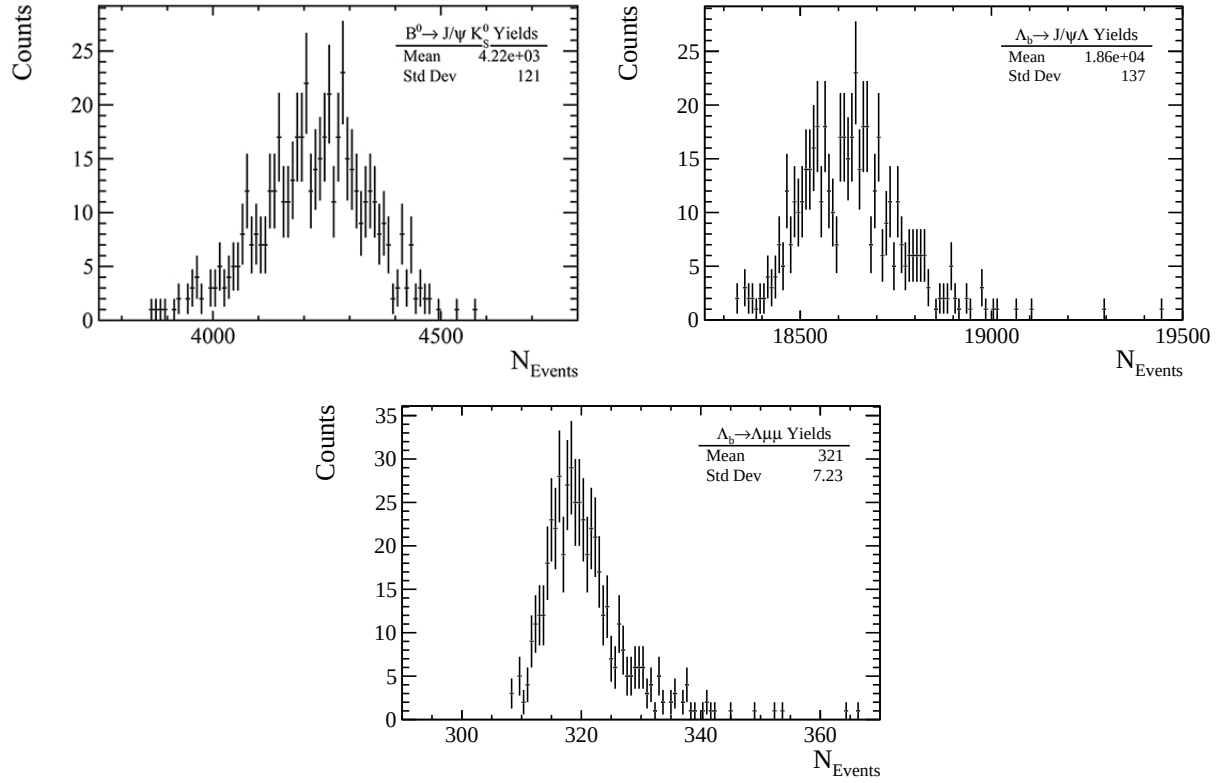


Figure 57: Histograms produced to determine systematic uncertainty on fixing shape parameters. Top left - Distribution in the yields from $B^0 \rightarrow J/\psi K_S^0$. Top right - distribution of $\Lambda_b^0 \rightarrow J/\psi \Lambda$ event yields. Bottom - distribution of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ event yields.

Source	Uncertainty on Decay Mode Yield
Network Training	1.84%
Fixing $B^0 \rightarrow J/\psi K_S^0$ parameters	2.86%
Fixing $\Lambda_b^0 \rightarrow J/\psi \Lambda$ parameters	0.74%
Fixing $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ parameters	2.20%
$J/\psi \Lambda$ Signal Shape	5.36%
$J/\psi \Lambda$ Background Shape	1.90%
$\Lambda \mu^+ \mu^-$ Signal Shape	4.08%
$\Lambda \mu^+ \mu^-$ Background Shape	0.07%
Total $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ Yield Uncertainty	8.12%

Table 21: Sources and values of the individual systematic uncertainties on yields arising from the analysis method. The total uncertainty is determined by adding the values in quadrature.

8 Conclusions and the Future

8.1 Discussion of Analysis Method

Despite a previous angular analysis of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ being performed using the same dataset, it was decided that a new search method should be used this time. The changes included new stripping line, trigger requirements, preselection restrictions, and a new selection of variables for training the neural network. The invariant $\Lambda \mu^+ \mu^-$ dataset after preselection was found to contain only one peaking background in addition to the combinatorial, which proved easy to identify and model. This suggests the trigger, stripping and selection cuts were successful in removing unwanted events from the data. The search performed in this work gave an 18% improvement in $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ yield. The biggest appreciable change was the reduction of the $\pm 30 \text{ MeV}/c^2$ on the Λ baryon mass to $\pm 10 \text{ MeV}/c^2$. This new neural network had a looser optimal working point than the previous, which may explain our improved signal yield. The signal shape, combinatorial and peaking backgrounds in the $\Lambda \mu^+ \mu^-$ invariant mass distribution were easy to describe and model. The use of two fit models when applied to the $\Lambda \mu^+ \mu^-$ distribution was an easy method to confirm the extracted signal yields, with additional checks also confirming final values. It is also found that in this case at least, specifically training the neural network to only contain events in the q^2 region of interest was not advantageous.

While not given here, the yields measured in the 0.1 to 2.0, 1.1 to 6.0 and 11.0 to 12.5 GeV^2/c^4 q^2 range, were used as extra statistics to validate the output and efficiencies of the simulation studies. Despite these regions not having significant $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ yields, the same method was reused to optimise the neural network output on these intervals. By assuming that the Λ_b^0 baryon was produced with non-zero transverse polarisation, we were able to develop our simulation to handle all 5 angles necessary to describe the decay. Overall, use of the simulation code was a successful and powerful tool in determination of efficiencies, while also allowing the processes used to determine the observables from data to be checked for flaws and oversights. With the simulations able to provide a validation of themselves, the majority of functions within it could be verified, while the use of the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ moments allowed us to confirm the final results from data were valid. Our distribution of moments associated with $\Lambda_b^0 \rightarrow J/\psi \Lambda$ were found to agree with those previously published.

With the above points, the analysis has proved to be successful in both improving on the previous $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ angular analysis, as switching from maximum-Likelihood fits to 1-dimensional angular projections in order to extract moments allowed easy measurement of observables which were not previously visible to us. Repeating this analysis with the extra data collected during Run 2 will allow for further reduction in uncertainties and better precisions on the central measurements.

8.2 Discussion of Angular Distributions and Observables

The results published here would seem to agree that Λ_b^0 baryon production has no preferred alignment to the beamline. We find that the moments measurable when $P_b \neq 0$ are undefined. Repeating the analysis on a larger number of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events would help to confirm this, and better define the unpolarised moments M_1 to M_{10} . Due to low statistics, the uncertainties on some individual moments for LL and DD do not scale accordingly to their statistics, as the bootstrapping process used to measure them does not give a Gaussian distribution. Data distributions displayed in Figures 55 and 56 show good agreement to the overlaid PDFs, confirming the effectiveness of the sWeighting and efficiency correction models.

We are able to compare our values of A_{FB}^l , A_{FB}^h and f_L to both previous measurement [28] and theory [31]. The theoretical predictions are all made on the assumption that $P_b = 0$. We calculate these observables with moments M_1 to M_5 , none of which are dependant on angles only measurable in polarised Λ_b^0 baryon production. Our values are found to be in good agreement with published measurements, suggesting this determination is as valid as the method used in [28]. We also find uncertainties on A_{FB}^l and f_L to be similar to previous measurement. Our uncertainty on A_{FB}^h is more than twice as large, likely due to the large uncertainty we have on M_5 , the only contributing variable. This is a result of M_5^{LL} having a large uncertainty due to low sample size and the bootstrapping method. While we agree with experimental results, those were previously found to not be consistent with theory. In the case of the f_L we find some of the tension to the SM theory to be relieved, while still consistent with the previous experimental value. Our result determined from moments is much closer to the published lattice QCD prediction, and almost within uncertainties. The forward-backward asymmetry of the muons is consistent with 0, and as the leptons are of identical mass, this does not indicate any preferential treatment between matter and antimatter. This is still in tension with theory, however we are consistent with the previous experimental value. This would lead to the assumption that the experimentally obtained observables are the physically correct ones. A non-zero asymmetry is found in A_{FB}^h as before, this may be explained by the mass difference between the proton and pion, as well as the asymmetrical Λ decay parameter. Our measured value of this asymmetry shows good agreement to both the theory and previous work. The previously unmeasured A_{FB}^{lh} maybe be determined with moments via Equation 13. As A_{FB}^{lh} is measured with moment M_6 , which is only dependant on θ_l and θ_h , comparisons may be drawn. We find the value to be about 1.5σ from theory, but not be all that well defined in this analysis. Values of observables in the 15 to 20 GeV^2/c^4 q^2 bin are listed in Table 20.

Theoretical predictions for M_4 , M_5 , M_9 and M_{10} are also published [31], while those not presented are predicted to be 0. Experimental values were not measured previously, but our this analysis allows the values of M_4 , M_5 , M_9 and M_{10} to be compared directly, while A_{FB}^{lh} may be determined and compared also. Moments M_9 and M_{10} are

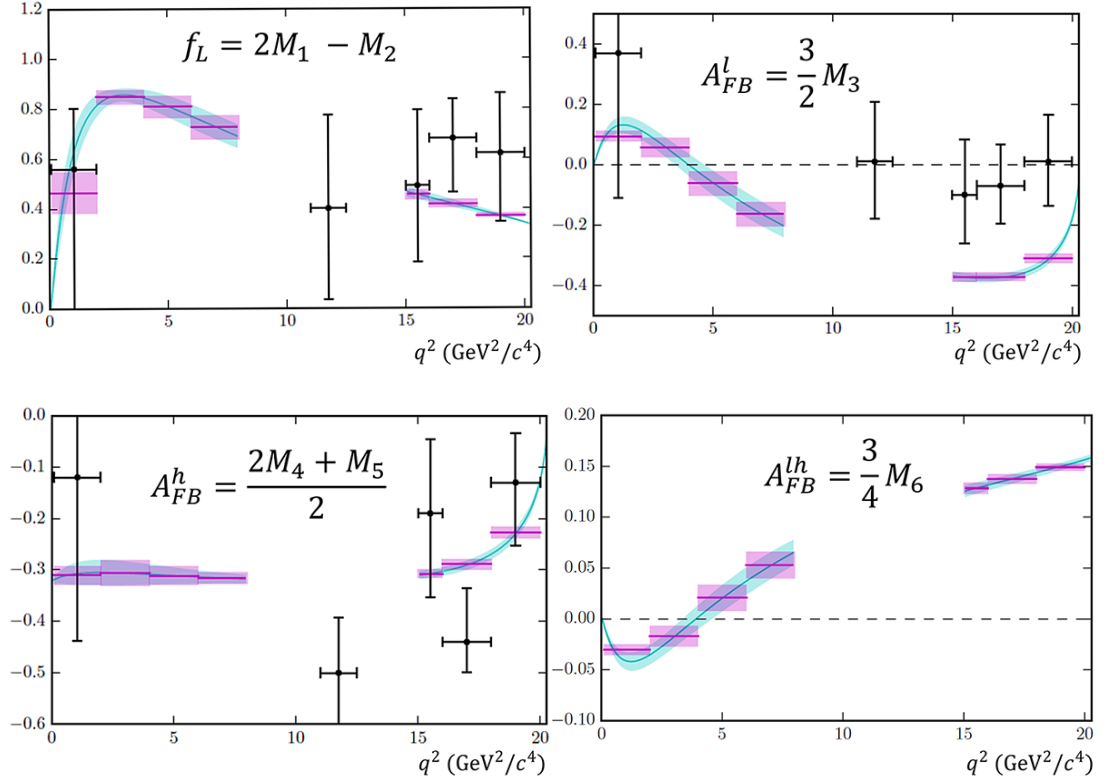


Figure 58: Theoretical values for observables associated to $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays. Black data points are the results measured in the previous LHCb angular analysis, while the purple regions are theory. The lepton-hadron asymmetry is previously unmeasured by the LHCb. Figures taken from [31].

calculated from ϕ_l and ϕ_h , which require non-zero Λ_b^0 polarisation production to be measured independently, but the weighting functions uses the sum of the two, which is the angle ϕ measured when $P_b = 0$. Our values of M_4 , M_5 agree with predictions within error margins, though M_5 is not accurately known. The value of M_9 is not defined enough to draw valid conclusions at this point, either. Moment M_{10} does not show good agreement - our value is consistent with 0, while the theoretical is not. Similarly, moments M_1 , M_2 , M_3 , M_6 , M_7 and M_8 are all assumed to be zero in theory, while we find this to not be the case. We show comparison for our moments values, as these are used to extract the J_i coefficients in the differential decay rate equation and it is of interest to contrast these to theory. The lattice QCD predictions [31] give a differential decay rate in the 15 to 20 GeV^2/c^4 bin of 0.756×10^{-7} , less than two-thirds of the measured value, implying a tension between experiment and theory. Predictions of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays with polarised Λ_b^0 production are not made, as the value of P_b is dependant on the detector, and centre of mass energies. Also, the measured value of P_b is consistent with zero, so no work has been done to theorise values which are likely to not have physical relevance.

8.3 The Future and Final Remarks

The second data taking run is already under way at the LHC, and will continue until 2017. Particle production cross-sections are approximately linearly proportional to $\sqrt{s}$, and with the increase from 8 TeV to 13 TeV between Runs 1 and 2, a factor of ≈ 1.6 should be seen in particle creation cross-section. With analyses already being carried out on this data, it will not be long until this angular analysis can be repeated with new data, and the results here can be improved upon. As stated in Section 3.2, approximately 5 trillion Λ_b^0 baryons were produced during Run 1. Using the Λ_b^0 to B^0 fragmentation ratio, the increased particle production cross sections, and accounting for the higher luminosity gives a yield of 135 trillion Λ_b^0 baryons. This would return 9000 $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ events in the high q^2 region with the same search efficiency - 28 times more than in the 3 fb^{-1} dataset. This allows for greatly reduced uncertainties on repeat measurements of angular observables and moments, and a better insight to the potential BSM physics at work in the suppressed decay. Additionally, the insight provided to us by the data collected during Runs 1 and 2 of the LHC will improve knowledge of the processes observed, helping to refine the theories used to describe them. Eventually, a larger hadron collider will be necessary to probe the more extreme energy regimes, with some already planned to be capable of reaching the $\sqrt{s} = 100 \text{ TeV}$ mark [78], allowing similar analyses to be performed with much greater event yields. Currently no average values over the q^2 spectrum have been measured due to low yields, so these upgrades would finally allow this.

The method of determining the moments of a decay has proved itself to be a powerful way to perform an angular analysis. With the right theory in place, this could be used to model many other rare decays, and would allow easy extraction of the observables associated to them. We have been able to make first measurements of several observables, and compare these to existing theory. One issue with this analysis was the large uncertainties on moments and observables calculated from small datasets. While bootstrapping is a valid way to obtain these values, larger yields are necessary to minimise these margins.

The work presented in this thesis appears to agree that there is no preferential transverse polarisation of the Λ_b^0 baryon at these luminosities and centre-of-mass energies. Despite this, the distribution of the 5 angles describing the decay of a polarised Λ_b^0 baryon were measured, and used to calculate the 34 moments associated to the decay. Yields were improved on those published from the previous angular analysis [28], and moments for the $\Lambda_b^0 \rightarrow J/\psi \Lambda$ were found to be consistent with previous measurements. Our measured observables are in good agreement with those published by the previous analysis, and some tension with theoretical values in f_L is relieved. The decision to determine the distribution of moments also allowed for unmeasured observables to be measured and compared to theory, which can then no doubt be used as a cross-check to results from future analyses.

Appendices

A

Definition of Variables used for Selection and Training

Tables 4 and 7 list the variables selected in our stripping requirements, and for neural network training. For easy reference, these variables are defined as;

- Dira Own PV - The angle between a particle's reconstructed momentum vector, and a vector connecting it's origin and decay vertices.
- Impact Parameter - The perpendicular distance between a particle's momentum vector and the interaction point. See Figure 24 for diagram detailing this.
- IP χ^2 - The change in χ^2 of the PV with and without the considered particle.
- Flight Distance - The distance travelled through the detector from its primary vertex to its end vertex.
- FD χ^2 Own PV - The measure of consistency of a particle's primary and decay vertices to be at the same point.
- Vertex χ^2 - The goodness-of-fit of a particle decay vertex.
- τ - The measured lifetime of the particle. How long it exists in the detector before decaying.
- $\tau \chi^2$ - The goodness of the measurement of the decay time.
- Ghost Track Probability - The measurement of how likely a track is to be real, rather than a combination of hits not belonging to a genuine one.
- Muon Particle ID - The likelihood of the particle actually being a muon, using information from the muon systems, calorimeters and RICH detectors to distinguish between these and other similar charged bodies, like kaons or pions.
- Decay Vertex - The vertex at which the particle decays. Measured in x, y and z co-ordinates.
- Decay Vertex χ^2 - The goodness-of-fit of the decay vertex of the particle.
- p_T - The transverse momentum of the particle. The component of the particle momentum perpendicular to the beamline.
- Cone p_T asymmetry - The fraction of the sum of all the transverse momenta of tracks in a cone around the Λ_b^0 baryon that accounts for the tracks used to reconstruct the Λ_b^0 baryon.

B

Comparison of Signal and Background Events for Neural Network Training Variables

The following plots compare distribution of variables used in neural network training. The green histograms are from the 5550 to 5700 MeV/c^2 $\Lambda\mu^+\mu^-$ invariant mass signal region in the simulation files used in training, while the red histograms are from the sideband, with $M(\Lambda\mu^+\mu^-) > 5900 \text{ MeV}/c^2$ from the real data used to teach the network background. Both datasets have had trigger, stripping and preselection requirements applied before creation of the plots. The overlap in variable distributions between signal and background events are indicative of how efficient background rejection cannot be performed with simple 1-dimensional cutting. Events in the overlapping regions cannot be easily labelled as signal or background, and as such it is necessary to analyse other variables also.

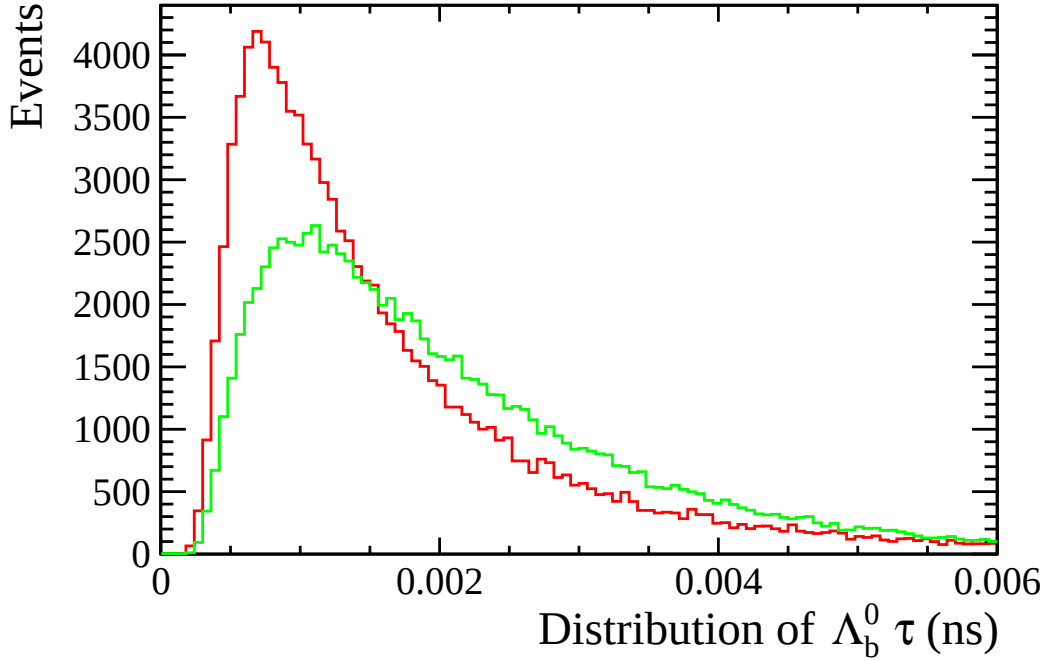


Figure 59: Lifetime of the Λ_b^0 baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

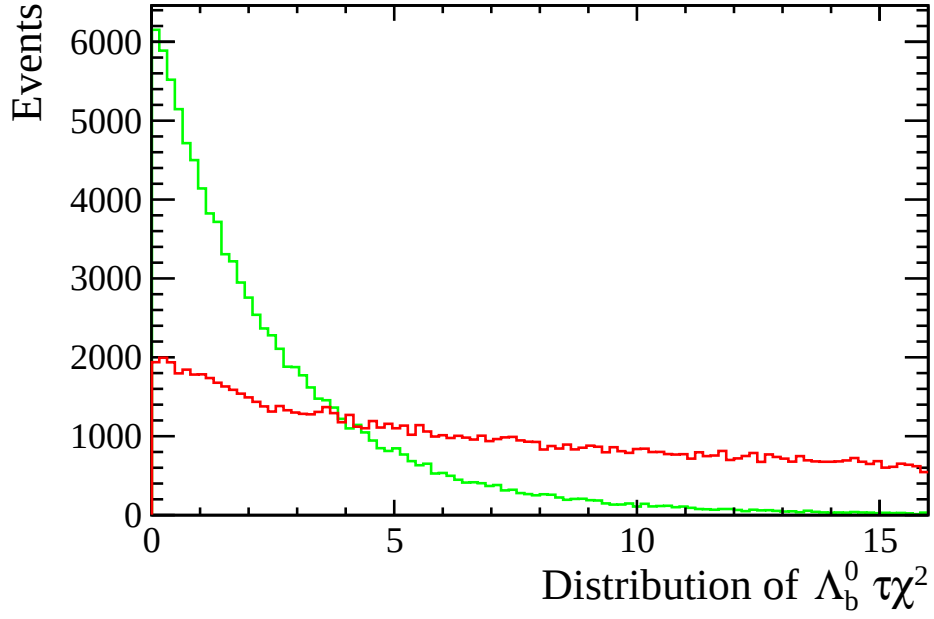


Figure 60: Goodness-of-fit of measured lifetime of the Λ_b^0 baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

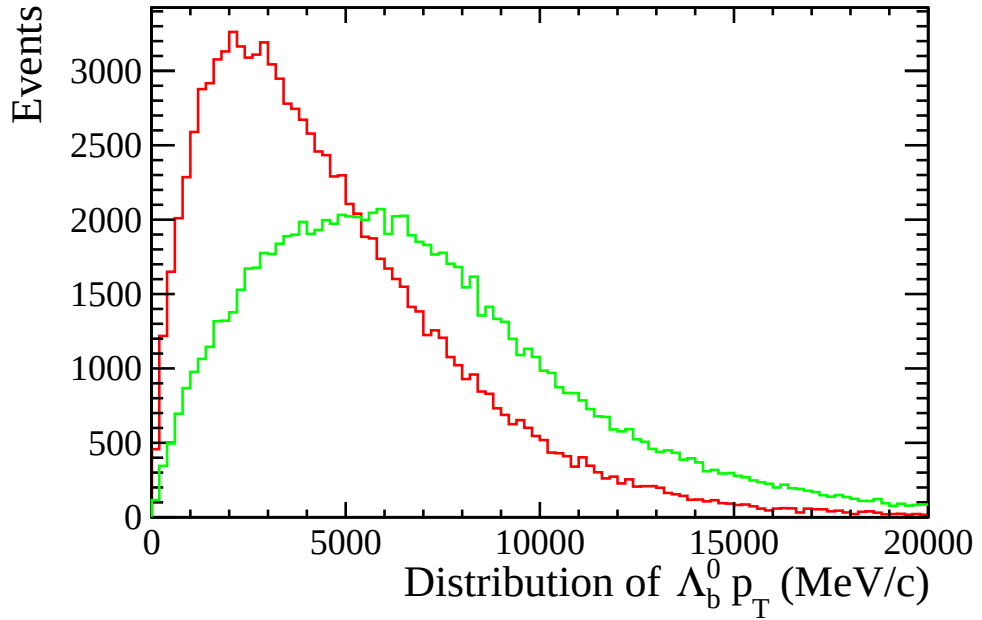


Figure 61: Transverse momentum of the Λ_b^0 baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

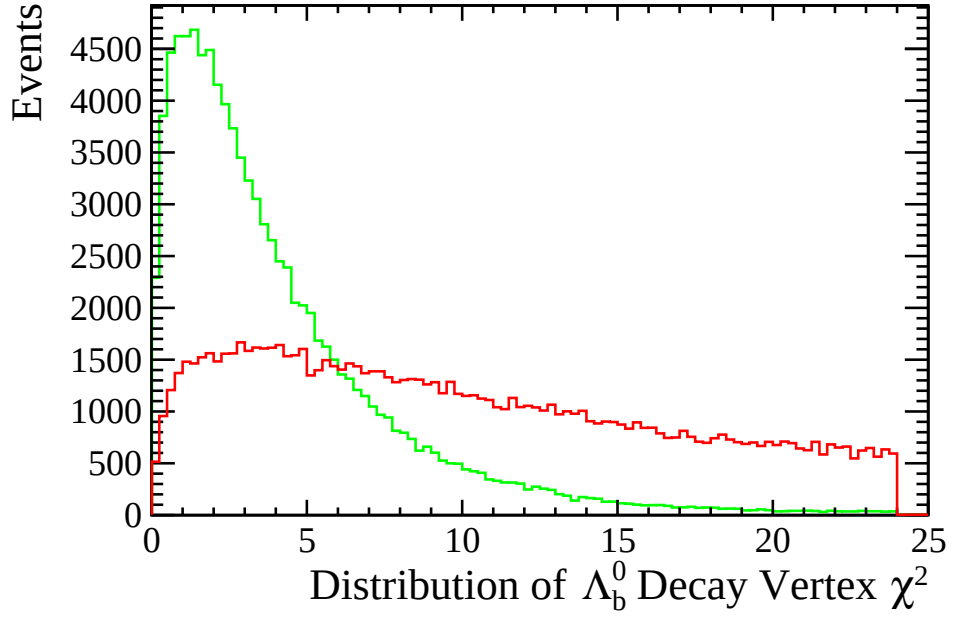


Figure 62: Goodness-of-fit of the decay vertex of the Λ_b^0 baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

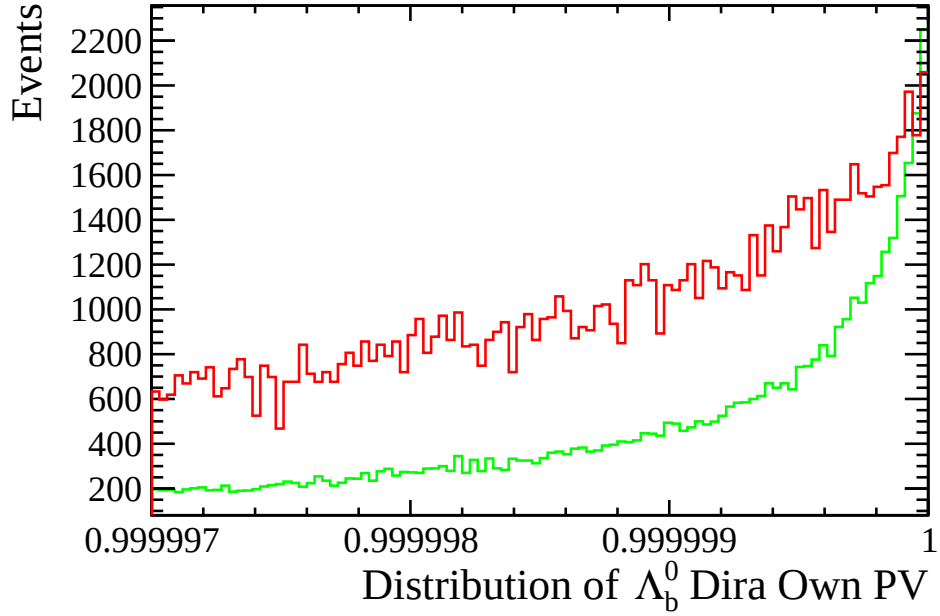


Figure 63: Cosine of the direction angle of the Λ_b^0 baryon momentum vector from its creation vertex. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

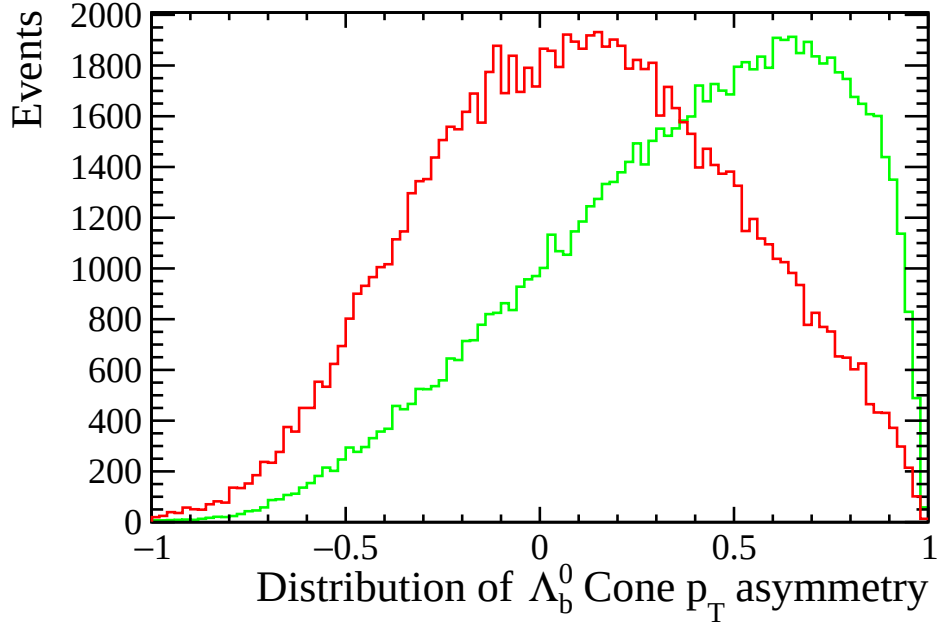


Figure 64: Measured asymmetry of event to background transverse momenta of all tracks in a cone about the Λ_b^0 . The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

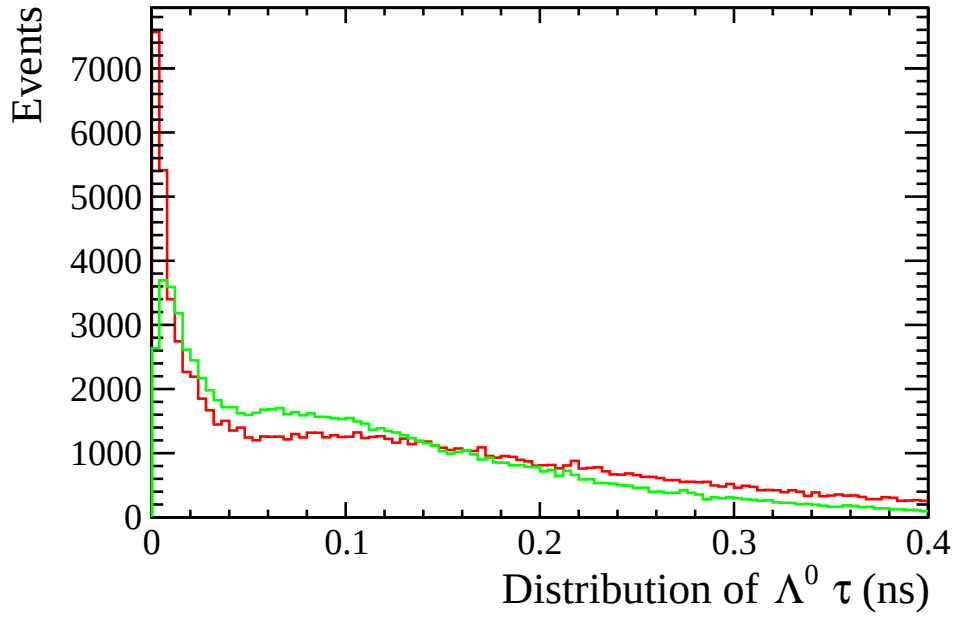


Figure 65: Lifetime of the Λ baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

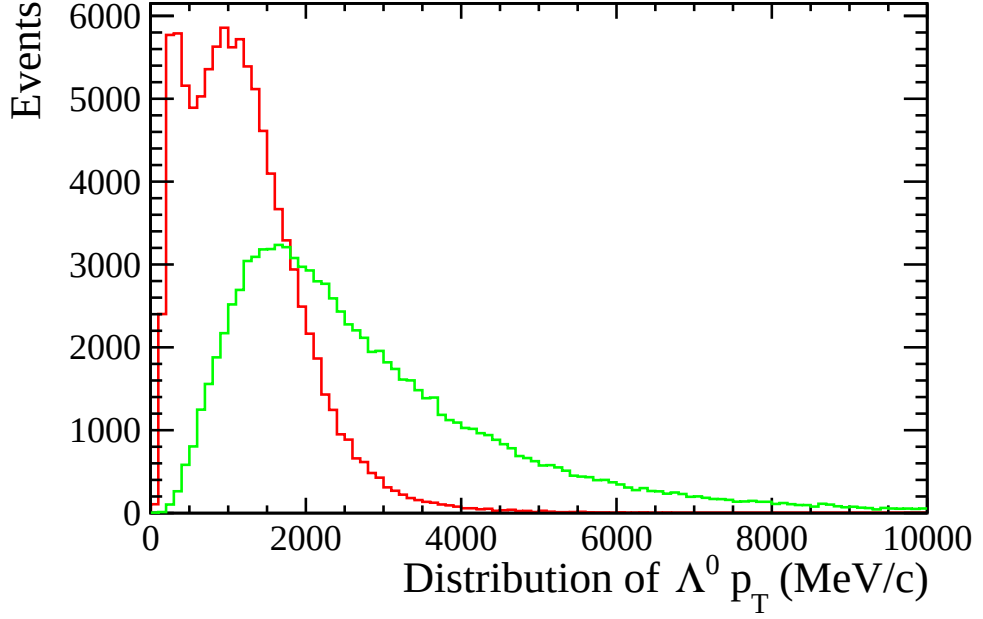


Figure 66: Transverse momentum of the Λ baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background. The structure just above 500 MeV/c arises due to a combination of the tight Λ baryon mass restriction and the trigger requirements we apply, which kills these low p_T events.

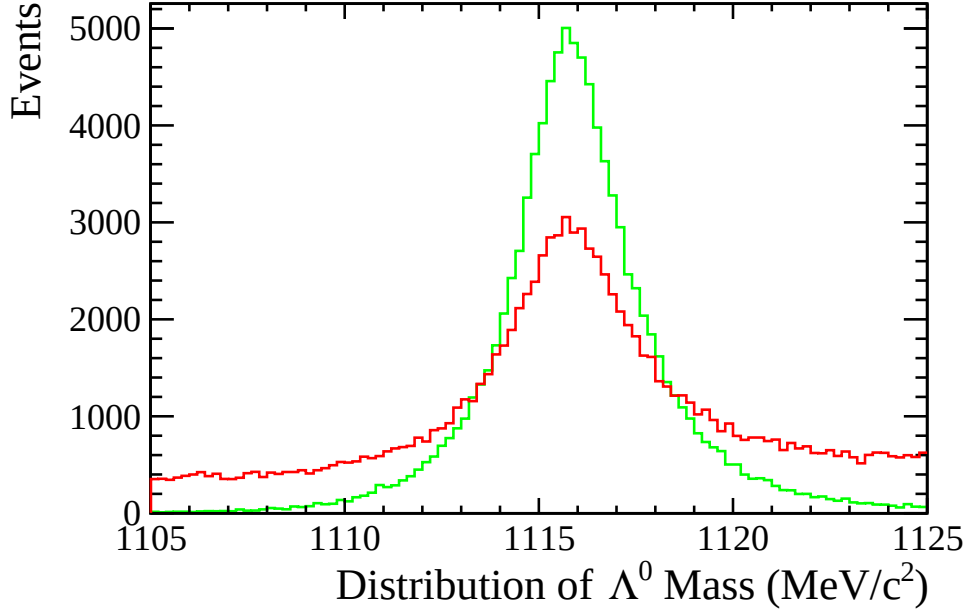


Figure 67: Invariant $p\pi^-$ mass distribution used to reconstruct the Λ baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

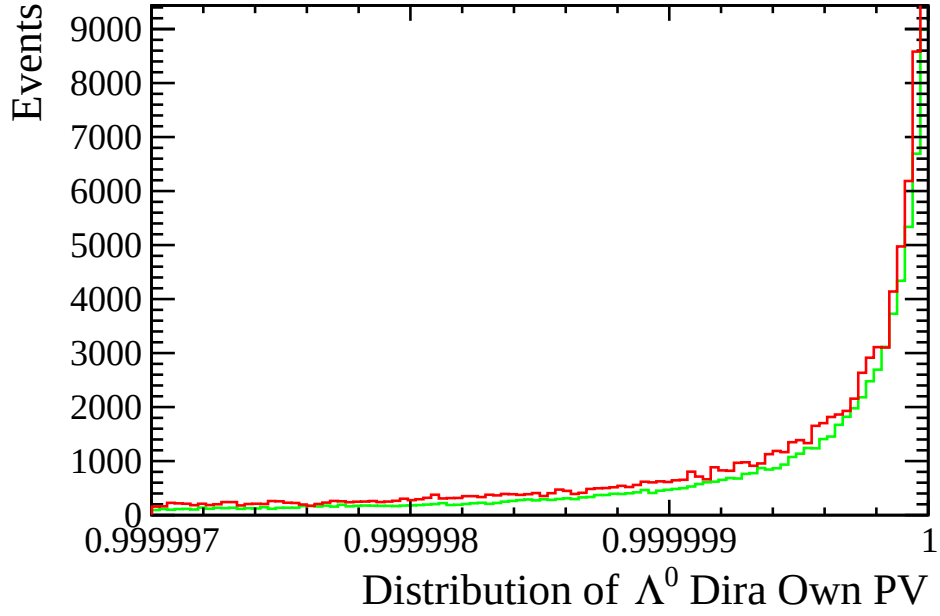


Figure 68: Cosine of the direction angle of the Λ baryon momentum vector from its creation vertex. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

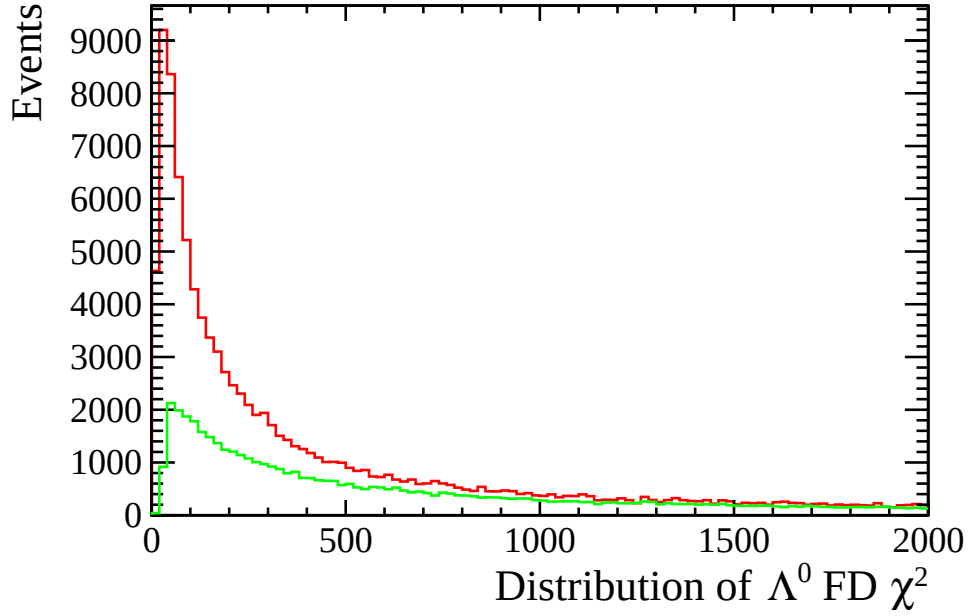


Figure 69: Flight distance fit goodness of the Λ baryon. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

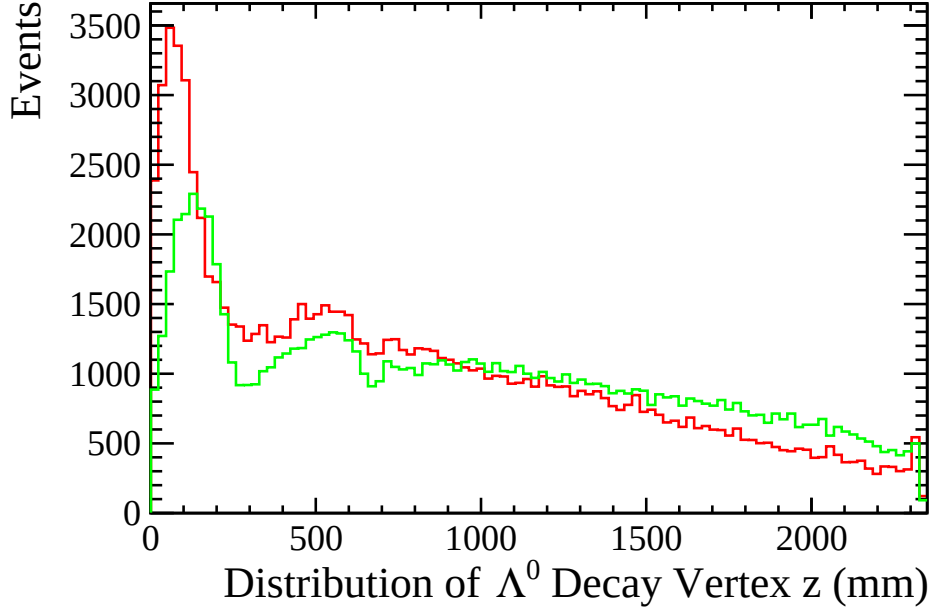


Figure 70: Z-component of the Λ baryon decay vertex. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background. The features at 250 and 650 mm correspond to sources of background events caused by material boundaries in the detector [79]. The minima occur as these events are removed via preselection and trigger requirements.

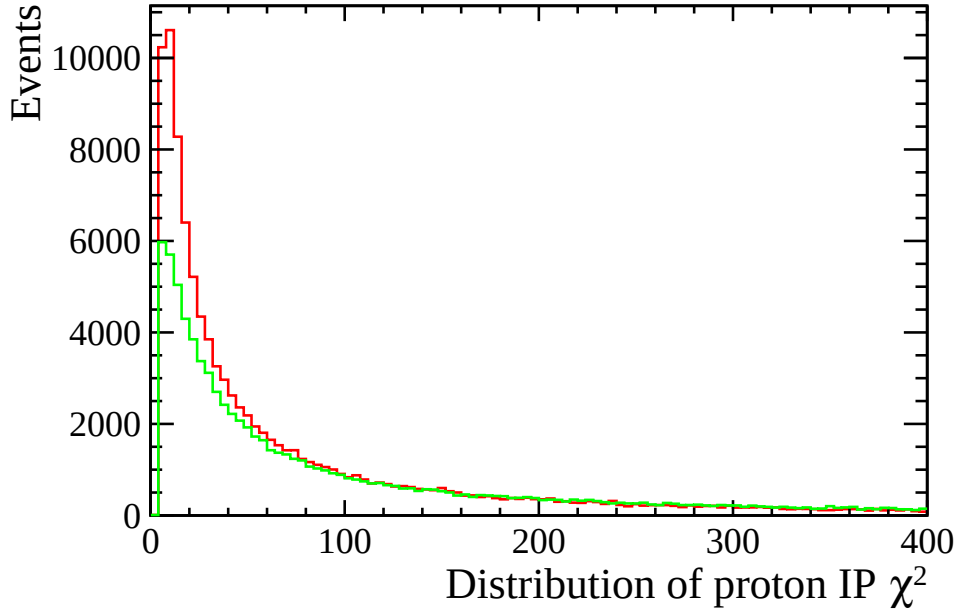


Figure 71: $\text{IP}\chi^2$ of the proton. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

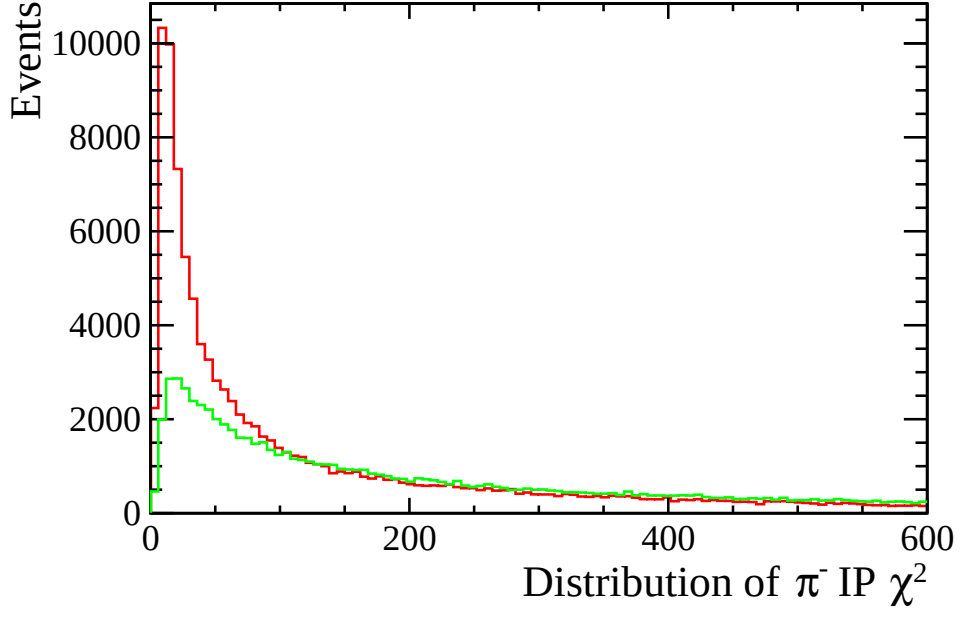


Figure 72: $\text{IP}\chi^2$ of the pion. The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

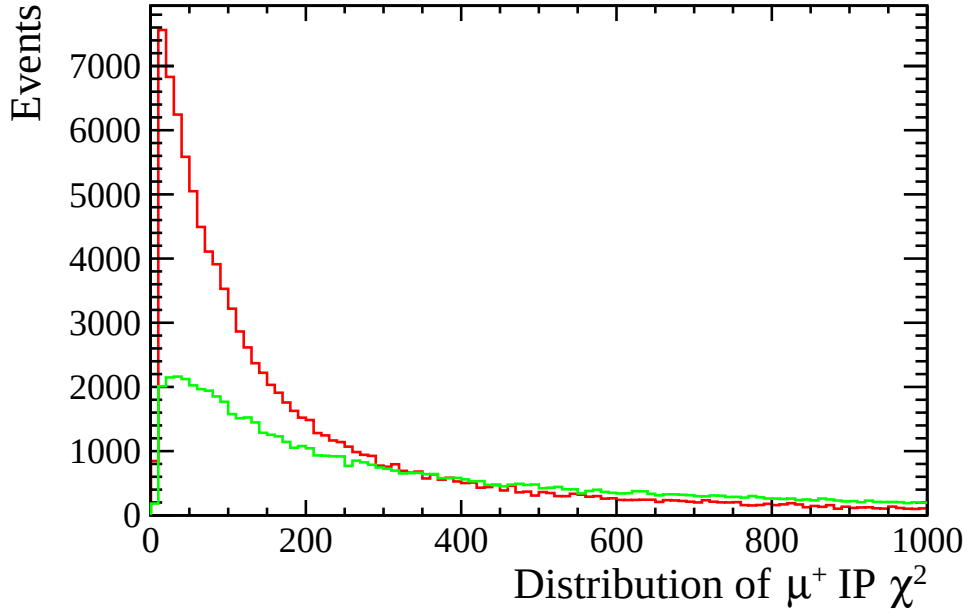


Figure 73: $\text{IP}\chi^2$ of μ^+ . The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

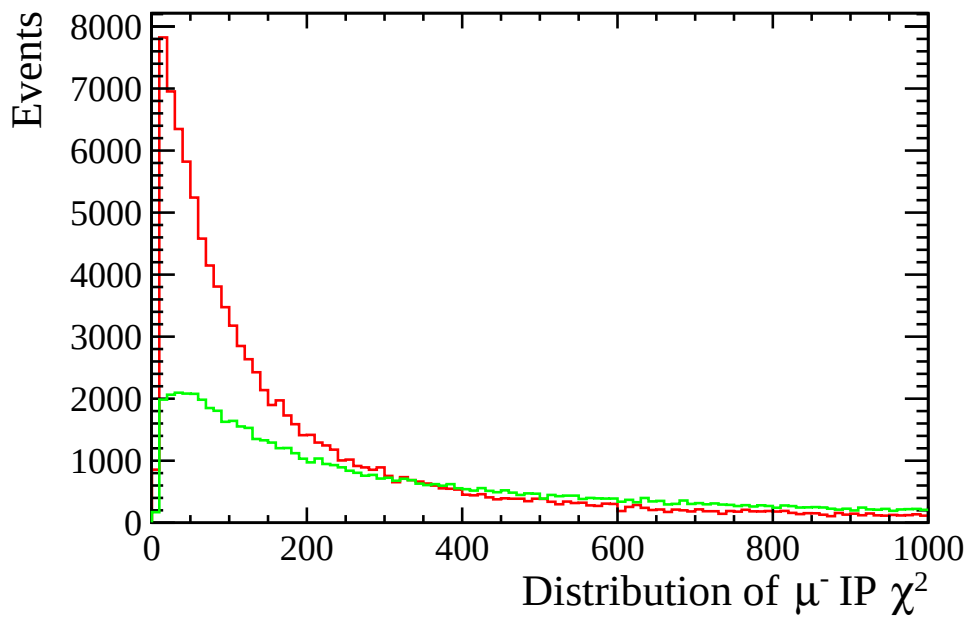


Figure 74: $\text{IP}\chi^2$ of μ^- . The green distribution is signal from simulated data, the red is the distribution in the real data upper side band, representing background.

C

Distributions of Pull Plots used in Moment Validation

Distributions are required to be Gaussian, with a mean of 0 and width of 1 in order to show unbiased moment generation. Pull is calculated as $(M_i^{\text{Extracted}} - M_i^{\text{True}})/\sigma(M_i)$.

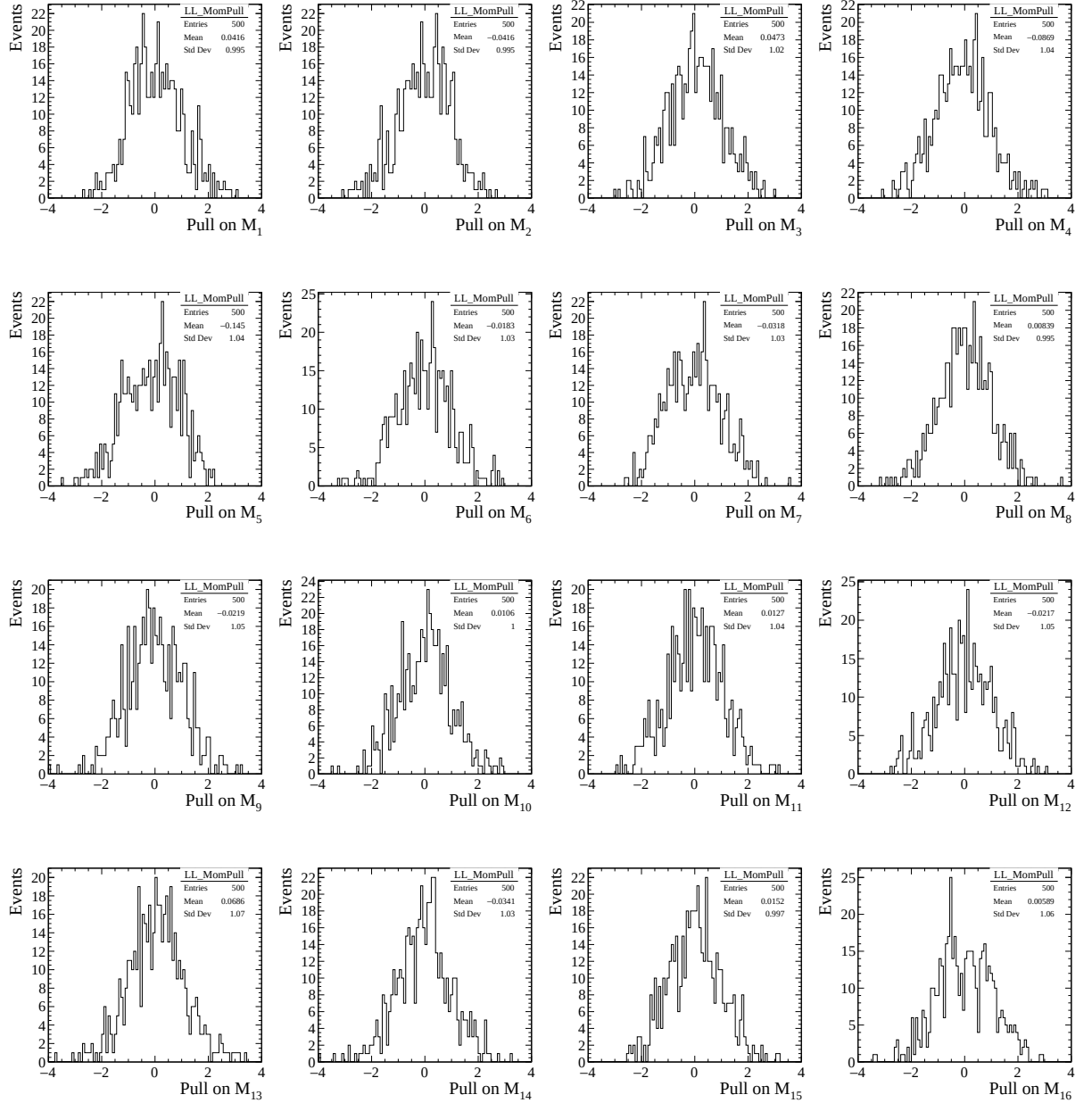


Figure 75: Pull distributions of Moments 1 - 16 produced from 500 simulation studies, each with 108 LL signal events, and the relevant signal efficiency model.

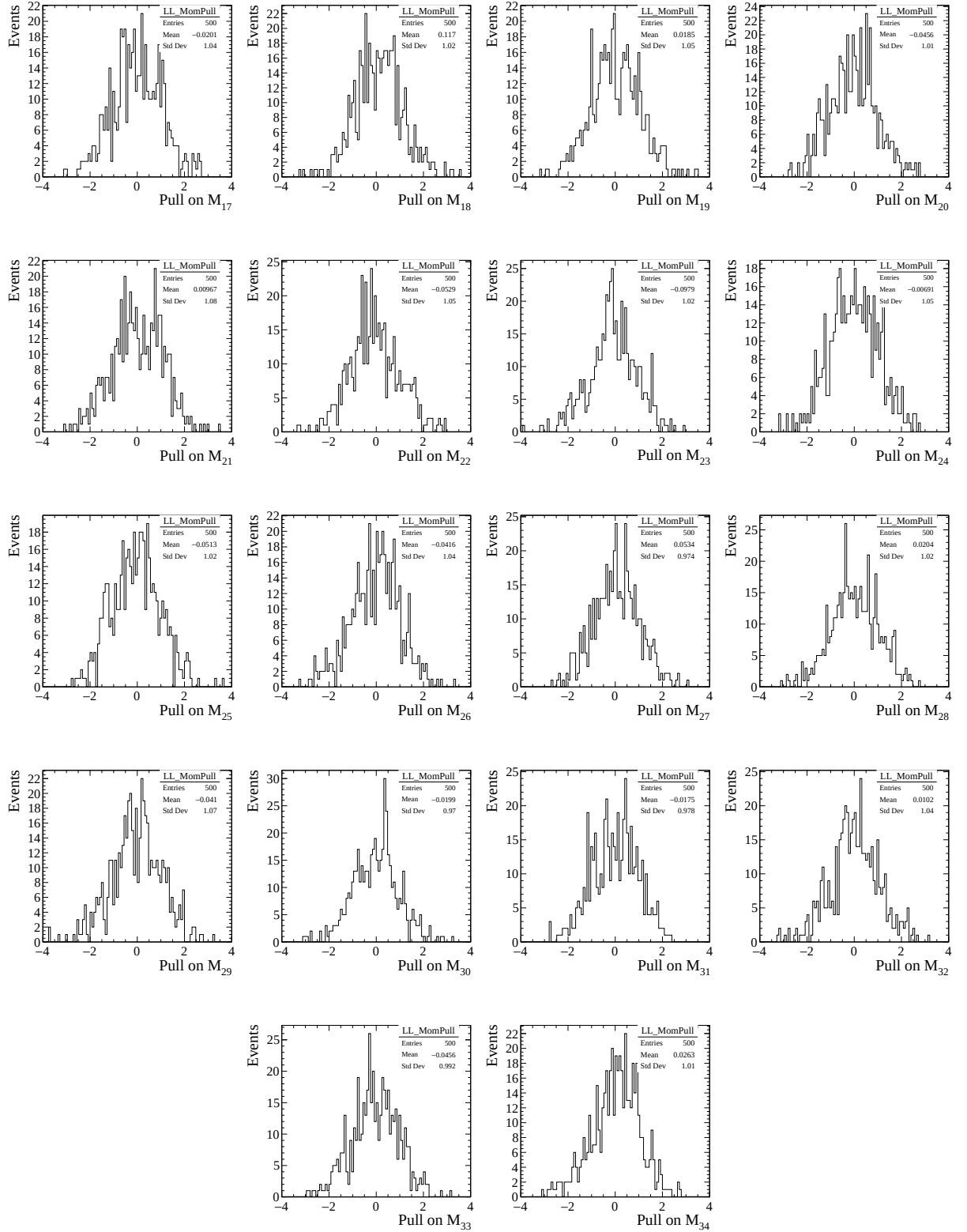


Figure 76: Pull distributions of Moments 17 - 34 produced from 500 simulation studies, each with 108 LL signal events, and the relevant signal efficiency model.

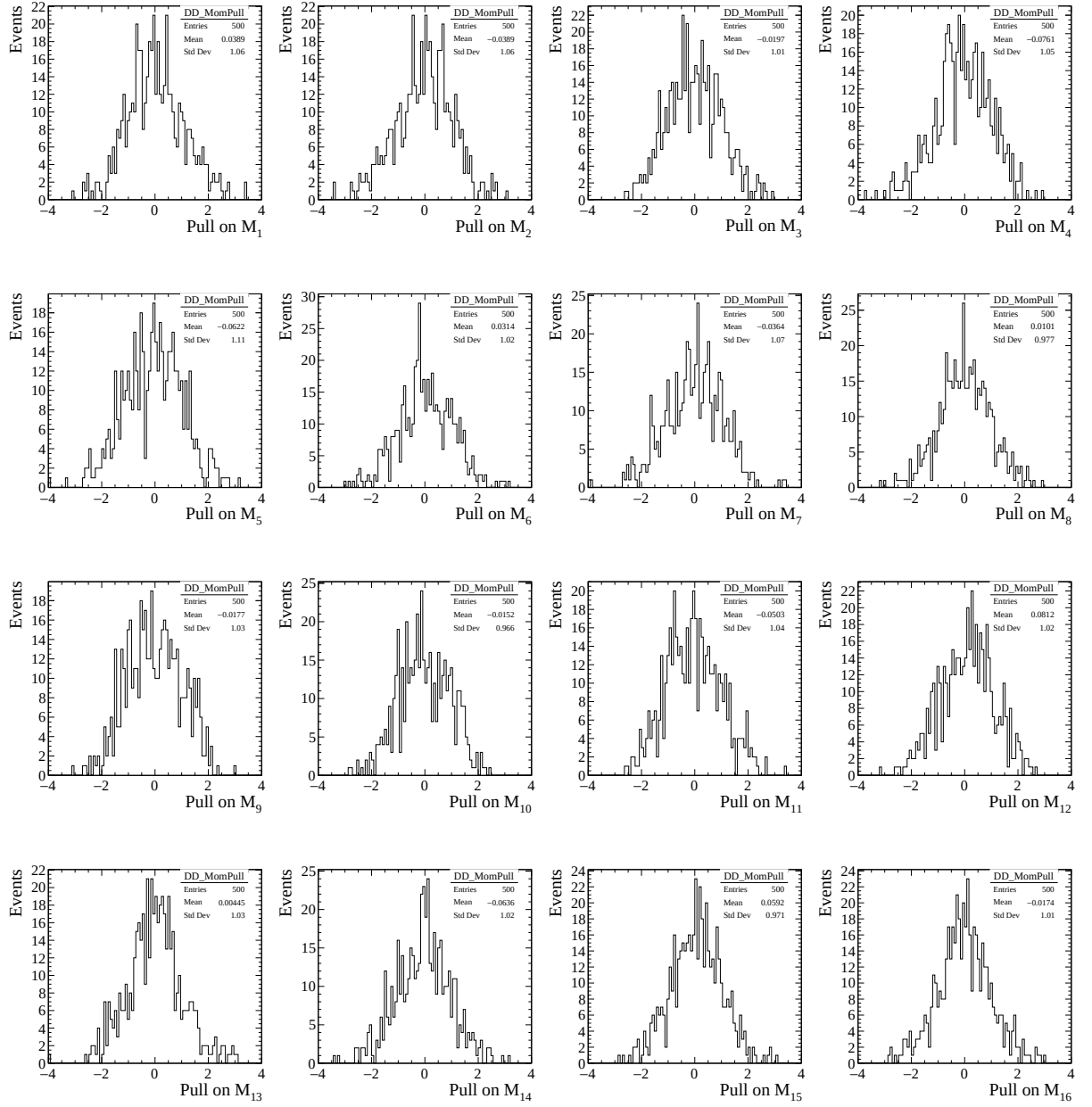


Figure 77: Pull distributions of Moments 1 - 16 produced from 500 simulation studies, each with 162 DD signal events, and the relevant signal efficiency model.

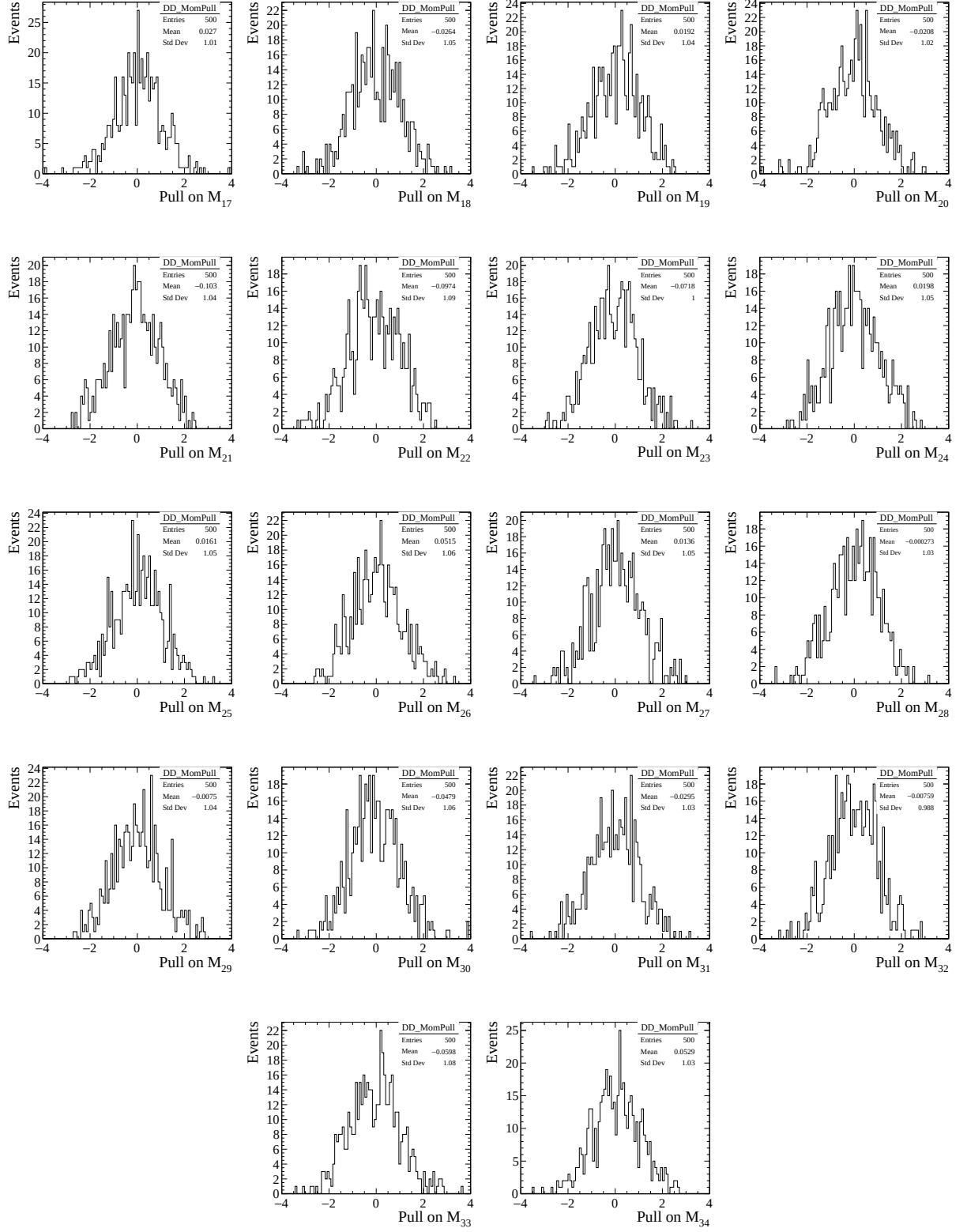


Figure 78: Pull distributions of Moments 17 - 34 produced from 500 simulation studies, each with 162 DD signal events, and the relevant signal efficiency model.

9 Figure References

- Figure 5 is taken from Reference [6].
- Figure 6 is taken from information in References [28, 31].
- Figure 2 is taken from https://map.gsfc.nasa.gov/universe/uni_matter.html, NASA/WMAP Science Team
- Figure 9 is taken from the CERN Document Server. By Esma Anais Mobs, 18/10/2016 - <https://cds.cern.ch/record/2225847?ln=en>. Figure was edited to increase the size of the key.
- Figures 10, 14, 15, 16, 17, 18, 19, 12, 13, 20, 21 and 22 are all taken from Reference [43].
- Figure 23 is taken from Wikimedia Commons.
- Figure 58 is taken from Reference [31].

10 Literature References

References

- [1] K.A. Olive et al., (*Particle Data Group*), Chin. Phys. **C38**, 090001 (2014) and 2015 update.
- [2] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22**, 4 (1961).
- [3] P. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 16 (1964).
- [4] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett **19**, 21 (1967).
- [5] B. R. Martin and G. Shaw, *Particle Physics, Third Edition, 2008*, Page 36. Published by Wiley.
- [6] T. Blake and T. Gershon and G. Hiller, *Rare b hadron decays at the LHC*, Annu. Rev. Nucl. Part. Sci. **65** (2015) 113, **arXiv:1501.03309**.
- [7] LHCb collaboration and A. Augusto Alves, Jr et al., *The LHCb Detector at the LHC*, JINST **3** S08005 (2008).
- [8] M. Kobayashi, T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics, **49**, 2 (1973).
- [9] R. R. Horgan, Z. Liu, S. Meinel, M. Wingate, *Rare B decays using lattice QCD form factors*, Proceedings of Science (LATTICE2014) **372**, (2015) **arXiv:1501.00367v2**.

- [10] W. Altmannshofer and D. M. Straub, *New physics in $b \rightarrow s$ transitions after LHC run 1*, Eur. Phys. J. C (2015) 75:382 **arXiv:1411.3161v4**.
- [11] LHCb collaboration and R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay $B_s^0 \rightarrow \phi \mu^+ \mu^-$* , JHEP **09** (2015) 179 **arXiv:1506.08777v3**.
- [12] LHCb collaboration and R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \rightarrow K^{(*)} \mu^+ \mu^-$* JHEP **06** (2014) 133 **arXiv:1403.8044v3**.
- [13] S. L. Glashow, J. Iliopoulos, L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry* Phys. Rev. D **2**, 1285 (1970).
- [14] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10**, 531 (1963).
- [15] J. H. Christenson, J. W. Cronin, V. L. Finch, R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett, **13**, 4 (1964).
- [16] C. L. Bennett *et al.*, *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results*, **arXiv:1212.5225v3**.
- [17] G. Bertone, *Particle Dark Matter, 2010*, Cambridge University Press.
- [18] S. A. Malik *et al.*, *Interplay and Characterization of Dark Matter Searches at Colliders and in Direct Detection Experiments*, Physics of the Dark Universe, Vol 9-10, 51-58 (2015) **arXiv:1409.4075v3**, CERN-PH-TH/2014-180.
- [19] A. D. Sakharov, *Violation of CP invariance, C asymmetry, and the baryon asymmetry of the universe*, Journal of Experimental and Theoretical Physics. **5**: 24–27 (1967).
- [20] F. Jegerlehner, *The hierarchy problem and the cosmological constant problem in the Standard Model*,
- [21] J. Bulava, K. Jansen, A. Nagy, *Constraining a fourth generation of quarks: non-perturbative Higgs boson mass bounds*, Phys. Lett. **B723** 1-3 (2013), **arXiv:1301.3416v1**.
- [22] M. A. Schmidt, A. Y. Smirnov, *Neutrino Masses, and a Fourth Generation of Fermions*, Nucl. Phys. **B857**, 1, (2012) , **arXiv:1110.0874v3**.
- [23] D. Atwood, S. K. Gupta, A. Soni, *Detecting Fourth Generation Quarks at Hadron Colliders*, JHEP **1206**, 105, (2012).
- [24] S. P. Martin, *A Supersymmetry Primer*, **arXiv:hep-ph/9709356v7**.
- [25] M. Dine, *Supersymmetry and String Theory: Beyond the Standard Model, 2007*, p. 169. Published by Cambridge University Press.
- [26] G. Jungman, M. Kamionkowski, K. Griest, *Supersymmetric Dark Matter*, Phys. Rept. **267** 195-373 (1996) **arXiv:hep-ph:9506380v1**.

- [27] L. Roszkowski, E. M. Sessolo, A. J. Williams, *What next for the CMSSM and the NUHM: Improved prospects for superpartner and dark matter detection*, JHEP **1408** (2014).
- [28] LHCb collaboration and R. Aaij *et al.*, *Differential branching fraction and angular analysis of $\Lambda_b^0 \rightarrow \mu^+ \mu^-$ decays*, JHEP **06** (2015) 115, **arXiv:1503.07138v2**.
- [29] P. Böer, T. Feldmann and D. van Dyk, *Angular Analysis of the Decay $\Lambda_b^0 \rightarrow \Lambda(\rightarrow N\pi)\ell^+\ell^-$* , JHEP **01** (2015) 155, **arXiv:1410.2115**.
- [30] Thomas Gutsche and Mikhail A. Ivanov and Jurgen G. Korner and Valery E. Lyubovitskij and Pietro Santorelli, *Rare baryon decays $\Lambda_b^0 \rightarrow \Lambda l^+ l^-$ ($l = e, \mu, \tau$) and $\Lambda_b^0 \rightarrow \Lambda \gamma$: Differential and total rates, lepton- and hadron-side forward-backward asymmetries*, **arXiv:1301.3737**.
- [31] W. Detmold and S. Meinel, *$\Lambda_b^0 \rightarrow \Lambda \ell^+ \ell^-$ form factors, differential branching fraction and angular observables from lattice QCD with relativistic b quarks*, Phys. Rev. **D93**, 074501 (2016), **arXiv:1602.01399v1**.
- [32] CDF Collaboration and T.Aaltonen *et al.*, *Observation of the Baryonic Flavor-Changing Neutral Current Decay $\Lambda_b^0 \rightarrow \Lambda \mu \mu$* , Phys. Rev. Lett. **107**:201802, (2011) **arXiv:1107.3753v2**.
- [33] LHCb collaboration and R. Aaij *et al.*, *Measurement of the differential branching fraction of the decay $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$* , Phys. Lett. **B725**, 25 (2013) **arXiv:1306.2577**.
- [34] M. Shifman, *Quark-Hadron Duality*, **arXiv:hep-ph/0009131v1**
- [35] LHCb collaboration and 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** 27 (2013), **arXiv:1302.5578**.
- [36] R. Mohanta, A. K. Giri, *Forth generation effect on Λ_b decays*, Phys. Rev. **D82**, 094022 (2012), **arXiv:1010.1152v3**.
- [37] F. Zolfagharpour, V. Bashiry, *Double Lepton Polarization in $\Lambda_b \rightarrow \Lambda \ell^+ \ell^-$ Decay in the Standard Model with Fourth Generations Scenario*, Nucl. Phys. **B796**, 294-319 (2008), **arXiv:0707.4337v1**.
- [38] A. J. Buras, *Flavour Expedition to the Zeptoverse*, PoS FWNP, (2015) 003, **arXiv:1505.00618v1**.
- [39] R. E. Caflisch, *Monte Carlo and quasi-Monte Carlo methods*, Acta Numerica Vol. 7, Cambridge University Press, 1998, p. 1-49.
- [40] Lyndon Evans and Philip Bryant, *LHC Machine* JINST **3** (2008) S08001.

- [41] *The accelerator complex*, CERN Document Server, <https://cds.cern.ch/record/1997193>, Jan (2012).
- [42] C. O’Luanaigh, *Experiments*, CERN Document Server, <http://cds.cern.ch/record/1997374>, Aug (2012).
- [43] LHCb collaboration and R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. **A 30**, (2015) 1530022.
- [44] LHCb collaboration and 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. **C 40** (2016) 011001, **arXiv:1509.00292**.
- [45] LHCb collaboration and R. Aaij *et al.*, *Measurement of B meson production cross-sections in proton-proton collisions at $\sqrt{s} = 7$ TeV*, JHEP **08** (2013) 117, **arXiv:1306.3663v2**.
- [46] P. Collins (on behalf of the LHCb collaboration), *The LHCb Upgrade*, **arXiv:1108.1403v1**.
- [47] LHCb collaboration and R. Aaij *et al.*, *Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \rightarrow J/\psi K^- p$ decays*, Phys. Rev. Lett. **115** 072001 (2015), **arXiv:1507.03414v2**.
- [48] LHCb collaboration and R. Aaij *et al.*, *Observation of the rare $B_s^0 \rightarrow \mu^+ \mu^-$ from the combined analysis of CMS and LHCb data*, Nature **522**, 68-72 (2015), **arXiv:1411.4413v2**.
- [49] W. Flegel, M. Losassa, F. Rohner, *LHCb Magnet, Technical Design Report*, CERN/LHCC/2000-007.
- [50] LHCb VELO Group and R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, **arXiv:1405.7808v2**.
- [51] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C 73** (2013) 2431, **arXiv:1211.6759v2**.
- [52] LHCb collaboration and R. Aaij *et al.*, *LHCb Calorimeters Technical Design Report*, CERN/LHCC/2000-036.
- [53] T. Head (on behalf of the LHCb Trigger Group) *The LHCb trigger system*, JINST **9** (2014) C09015.
- [54] M. P. Altarelli, F. Teubert, *B Physics at LHCb*, Int. J. Mod. Phys. **A23**, 5117-5136 (2008) **arXiv:0802.1901v2**.
- [55] J. Albrecht, *The LHCb Trigger System: Present and Future*, J. Phys.: Conf. Ser. **623** 012003, (2015).

- [56] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** P02013 (2013) **arXiv:1210.6861**.
- [57] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. **A559** (2006) 190.
- [58] Information on ROOT can be found at its homepage - <https://root.cern.ch/>.
- [59] W. Verkerke, D. Kirkby, *The RooFit toolkit for data modeling*, **arXiv:physics/0306116**.
- [60] Information on RooFit can be found at roofit.sourceforge.net.
- [61] Information on MINUIT can be found at <https://seal.web.cern.ch/seal/snapshot/work-packages/mathlibs/minuit/>
- [62] Muriel Pivk and Francois R. Le Diberder, *sPlot: a statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555**: 356-369, 2005, **arXiv:physics/0402083**.
- [63] Twiki page - <https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbStripping>.
- [64] Configuration found at <https://twiki.cern.ch/twiki/bin/view/LHCb/Stripping21r1Configuration>.
- [65] Configuration found at <https://twiki.cern.ch/twiki/bin/view/LHCb/Stripping21Configuration>.
- [66] T. Sjöstrand, S. Mrenna, P. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** 852-867 (2008), **arXiv:0710.3820v1**.
- [67] 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**, 1155 (2010).
- [68] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462**, 152 (2001).
- [69] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45**, 97 (2006), **arXiv:hep-ph/0506026**.
- [70] GEANT4 collaboration, J. Allison *et al.*, *Geant4 developments and applications* IEEE Trans. Nucl. Sci. **53**, 270 (2006); GEANT4 collaboration, S. Agostinelli *et al.*, *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506**, 250 (2003).
- [71] M. Clemencic *et al.*, *The LHCb simulation application, GAUSS: Design, evolution and experience*, J. of Phys. : Conf. Ser. **331**, 032023 (2011).
- [72] Members of the LHCb RICH group, *LHCb RICH1 Engineering Design Review Report*, LHCb 2004-121.

- [73] Information on the Background Category may be found at http://lhcb-release-area.web.cern.ch/LHCB-release-area/DOC/davinci/releases/latest/doxygen/df/d88/class_i_background_category.html
- [74] 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.
- [75] LHCb collaboration and R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \rightarrow K^* \mu^+ \mu^-$* , JHEP **06**, 133, (2014), **arXiv:1403.8044**.
- [76] S. Okubo, R. E. Marshak, and E. C. G. Sudarshan, *V – A Theory and the Decay of the Λ Hyperon*, Phys. Rev. **113**, (1959), 944.
- [77] B. Efron, *The Jackknife, the Bootstrap, and other Resampling Plans, 1982*, Department of Statistics, Stanford University.
- [78] E. Gibney, *China plans super collider*, Nature **511**, 394-395 (2014).
- [79] LHCb Collaboration, *LHCb reoptimized detector design and performance : Technical Design Report*, CERN-LHCC-2003-030.