

Search for the rare fully leptonic decay

$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ at LHCb

Slavomira Stefkova

High Energy Physics
Blackett Laboratory
Imperial College London

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

Contents

Declaration of originality	6
Copyright declaration	7
List of Figures	8
List of Tables	18
List of abbreviations and definitions	22
1 Introduction	26
2 Theory	28
2.1 Review of the Standard Model	28
2.2 The principle of Standard Model building	30
2.3 Quantum Chromodynamics	32
2.4 Electroweak unification	33
2.5 The Higgs Mechanism	36
2.6 Fermion mass generation	37
2.7 The Quark Mixing Matrix	39
2.8 Fully Leptonic $P^+ \rightarrow l^+ \nu_l$ decays	44
2.9 Fully Leptonic $B^+ \rightarrow l^+ \nu_l \gamma$ decays	47
2.10 Fully Leptonic $B^+ \rightarrow l^+ l^- l^+ \nu_l$ decays	48
2.11 The $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay model	50

3 The LHCb detector	52
3.1 LHCb Layout	53
3.2 VErtex LOCator	58
3.3 Tracking System	60
3.3.1 Tracking Algorithms	61
3.4 Ring Imaging Detectors	63
3.5 RICH Reconstruction and Performance	65
3.5.1 RICH performance	65
3.6 Calorimetry	67
3.7 Muon Stations	68
3.7.1 Muon Identification	69
3.7.2 Muon Performance	71
3.8 Trigger	72
3.9 Simulation	76
3.9.1 Differences in Simulation And Data ULRIK - not fully finished	77
4 Handling of trimuon correlations at LHCb	79
4.1 Muon PID variables	79
4.1.1 Binary Muon PID variables	80
4.1.2 Muon PID variables based on sharing hits	80
4.1.3 Muon PID variables based on regression techniques	83
4.2 Clones	84
4.3 Probability of $K/\pi \rightarrow \mu$ misidentification at LHCb	88
4.3.1 Specific control sample for $K/\pi \rightarrow \mu$ misid rates	89
4.3.2 Selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^*$	89
4.3.3 Fitting Strategy for $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^*$ decay	89
4.3.4 Results of $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^*$ control sample for $K/\pi \rightarrow \mu$ misid rates	94

5 Discovering (Setting Limit for) $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ at LHCb	99
5.1 Topology of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay at LHCb [ULRIK]	99
5.2 Sources of Backgrounds [ULRIK]	101
5.3 Analysis strategy	102
5.4 Samples [ULRIK]	104
5.4.1 Data Samples [ULRIK]	104
5.4.2 Simulation Samples [ULRIK]	104
5.5 Preselection for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ [ULRIK]	106
5.6 Trigger Selection [ULRIK]	108
5.7 q^2 Selection [ULRIK]	110
5.8 Further Selection [ULRIK]	111
5.8.1 General Features of Multivariate Selections [ULRIK]	112
5.8.2 The Isolation Boosted Decision Tree [ULRIK]	113
5.8.3 The Combinatorial Boosted Decision Tree [ULRIK]	115
5.8.4 The Misid Boosted Decision Tree [ULRIK]	117
5.8.5 Fitting Region Selection	119
5.8.6 Further PID Selection [ULRIK]	120
5.9 Normalisation channel [ULRIK]	124
5.10 Fractional Corrected Mass Error (FCME) Window Split [ULRIK]	125
6 Background Studies	127
6.1 Combinatorial Background	127
6.2 MisID type background	129
6.3 Partially Reconstructed Background	135
6.3.1 Partially Reconstructed Backgrounds, where $D^0 \rightarrow \eta/\eta' X$, where $\eta/\eta' \rightarrow \mu\mu\gamma$	137
6.3.2 Partially Reconstructed $B \rightarrow \eta(') V$ Backgrounds	138
6.4 Rare and resonant $B^+ \rightarrow \pi^+/K^+ \mu^- \mu^+$ backgrounds	139
6.5 Summary	140

7 Mass fits and efficiencies	141
7.1 Efficiency Ratio	141
7.2 Summary of Efficiencies	143
7.2.1 Detector Acceptance Efficiency (GEN)	145
7.2.2 Reconstruction Efficiency (REC)	146
7.2.3 Trigger Efficiency (TRG)	146
7.2.4 Offline Selection (OFF)	147
7.2.5 Combinatorial BDT and Misid BDT efficiency	149
7.2.6 Fitting Range Efficiency (FR)	152
7.2.7 PID Efficiencies (PID)	153
7.3 Mass Fits	155
7.3.1 Normalisation Channel Parametrisation	155
7.3.2 Normalisation Fit	157
7.3.3 Signal Channel Parametrisation	160
7.3.4 Signal Data Fits	171
7.3.5 Signal Fit Systematics	172
7.3.6 Blinded Data Fits	177
7.3.7 Expected Sensitivity	180
7.3.8 Unblinded Data Fits	181
8 Results	186
Bibliography	187
Appendices	194
A Boosted Decision Trees	194
B The <i>sPlot</i> technique	197

C Fitting functions	199
C.1 Crystal Ball function	199
C.2 Double-sided Ipatia function	200
C.3 Rookeys function from <code>ROOTFIT</code> package	200

Declaration of originality

The work presented in this thesis has been done between October 2014 and March 2018. It is a result of my own studies together with the support of Imperial College HEP group and LHCb collaboration. All the analysis work (chapters 4–8) presented in this thesis was performed by myself. All results and plots presented in this thesis that were not the product of my own work are appropriately referenced.

This thesis has not been submitted for any other qualification.

Slavomira Stefkova, July 25, 2018

Copyright Declaration

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

List of Figures

2.1	Unitarity triangle in a complex plane.	41
2.2	Different experimental measurements that constrain the CKM matrix elements together with the fit results from two collaborations (a) UTfit and (b) CKMFitter as of summer 2016. These figures are taken from [4] and [5]. There is a good agreement between the two different results. . .	43
2.3	(a) Annihilation diagram where the initial u -quark state radiates off a virtual photon which decays into a pair of muons and the W^+ decays into a muon and muon neutrino. Most of the contribution to the rate comes from hadronic contribution to the photon. (b) Photon emission from b -quark and (c) final emission from the final state muon.	50
2.4	Distributions for signal simulation. (a) $K^*(\mu^+, \nu_\mu)$ (b) $q(\mu^+, \mu^-)$ distributions under different K^* mass hypotheses. The most flat distribution in $K^*(\mu^+, \nu_\mu)$ is plotted in yellow.	51
3.1	Accelerator complex at CERN. The image is taken from [22].	53
3.2	Schematic slice of LHCb detector in y, z plane where z is defined to be the direction parallel to beamline, and x, y define the plane perpendicular to the beamline. θ , the opening angle in y - z plane with $\theta = 0$ along z -axis. Figure from [23].	54

3.3 (Left) Angular production and acceptance of the b (x-axis) \bar{b} (y-axis) pair produced from pp collision at the LHC. The acceptance of the LHCb detector is the red box and the acceptance of the General Purpose Detector is shown in the yellow box. LHCb covers the region with highest production cross-section at 8 TeV. These plots were produced using PYTHIA8 [24] simulation. Figure from [25].(Right) Probability of interaction per bunch crossing as a function of instantaneous luminosity. Figure from [26].	55
3.4 Integrated luminosity collected in different years of data-taking. This plot is taken from [29] (left). Development of the instantaneous luminosity for ATLAS, CMS and LHCb during LHC fill 2651. After ramping to the desired value of $4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ for LHCb, the luminosity is kept stable in a range of 5% for about 15 hours by adjusting the transversal beam overlap. The difference in luminosity towards the end of the fill between ATLAS, CMS and LHCb is due to the difference in the final focusing at the collision points, commonly referred to as the beta function, β^* . This plot was obtained from [30] (right).	58
3.5 Schematic plot of VELO detector configuration along the beam pipe showing the layout as well as positions while in stable beams (discs have slight overlap) and injection. Figure taken from [31].	59
3.6 Two key variables which quantify performance of the VELO detector. IP resolution which is worse for low momentum tracks (left) and PV resolution dependent on the number of tracks forming the primary vertex N (right). Figures taken from [32].	60

3.7	Visualisation of use of different technology with silicon technology in violet and straw-tube technology in cyan. The Figure was obtained in [33](left). Track types visualisation depending on which track stations provided hits. For the study of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decays only long tracks are considered as muons will travel to the end of the detector leaving the hits all along. Figure is taken from [34] (right).	61
3.8	Momentum resolution of long tracks measured using <i>tag and probe</i> method at LHCb. The decay channel $J/\psi^+ \rightarrow \mu^+ \mu^-$ is analysed.	63
3.9	Separation power for different species of particles in momentum-Cerenkov angle plane (left). Schematic diagram of RICH1 layout (right). Both figures are taken from [31].	64
3.10	Invariant mass distribution of D^0 data sample (in black) overlaid with fit to both background and signal (in blue) (left). An example of kaon ID (red) and MisID (black) efficiency as a function of momentum under two PID hypotheses, $DLLK > 0$ (empty) and $DLLK > 5$ (filled) (right). Both figures are taken from [37].	66
3.11	$\pi^+ \pi^-$ invariant mass distributions obtained using kinematic constraints only (left) and also using PID constraints (right) in order to isolate $B^0 \rightarrow \pi^+ \pi^-$ peak. This figure is taken from [38].	67
3.12	Granularity of ECAL (left) and HCAL (right) detectors. The figure was taken from [31].	68
3.13	(a) Layout of the muon detector x-z plane and (b) x-y plane. This figure is taken from [39].	70
3.14	(a) Probability of correctly identifying muons as a function of momentum p in the bins of p_T for $J/\psi \rightarrow \mu^+ \mu^-$ with IsMuon constraint. (c) Probability of incorrectly identifying pion (b) proton and (d) kaon as muon with IsMuon. This figure is taken from [40].	72

3.15 TOS efficiency as a function of p_T for muon-based decisions (left). TOS efficiency for different decays using L0 hadron trigger lines. This figure is taken from [39].	73
3.16 HLT1 efficiencies of the corresponding triggers using the same proxy as in Figure 3.15. This figure is taken from [41].	74
3.17 Trigger scheme differences between Run I and Run II. Figures obtained from [42]	76
3.18 IP resolution in x-direction comparing the data and simulation output for 2012 data-taking period (left). IP resolution in x-direction comparing the data and simulation output for 2011 data-taking period as a function of angle, ϕ (right). This figures are taken from [32].	78
4.1 Schematic view of the muon station slicing into x-y pads. This is the left quadrant of the M1 station, showing decreasing granularity of the muon stations away from the beam pipe. This figure has been taken from [39]. M1R1 is the innermost region and M1R4 is the outermost region of the M1 station.	81
4.2 (a) nShared variable distribution for the positive muon in $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ decays in (a) simulation and (b) data. Different stripping versions corresponding to 2012 (<i>Stripping 21</i>), 2011 (<i>Stripping 21r1</i>), 2016 (<i>Stripping 26</i>) data-taking are shown. The distributions are normalised to have the same area. There is shift of distribution in <i>Stripping 26</i> towards less isolated tracks. The proportion of muon tracks that share no other hits with other tracks is smaller, whereas the proportion of the tracks sharing hits with other muon track is increasing.	83

4.3 ID and misID probabilities from standard calibration datasets from 2012 (<i>Stripping 21</i>) and 2016 (<i>Stripping 26</i>), binned using the default 2-dimensional binning scheme in momentum p and pseudorapidity η . In this plot, ID and misID rates in the central bin of η , $2.375 < \eta < 3.25$, and the first and second bin in p are compared. This demonstrates that for the same pion ID efficiency, the misID rate is significantly higher in 2016 data.	84
4.4 (a) Visible and (b) corrected mass of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ candidates in 2012 data where all the muons have the same charge. Clear fake peaks, arising from the correlation of several effects in the detector can be seen.	85
4.5 (a) Clone and (b) no clones flight distance properties. It can be seen that <i>clone</i> tracks have their decay vertex placed at the end of the detector, whereas regular good tracks will decay within the VELO	86
4.6 The difference in the OT detector between (a) clones and (b) real tracks in the OT at the distance 9450 mm along the LHCb . <i>Clones</i> are concentrated along the inner edge of the OT . Good muon tracks will cover most of OT evenly.	86
4.7 (a) Peaking clone distribution is visible as all of <i>clone</i> tracks are collinear compared to (b) smooth no clone distribution for $\cos(\theta_B)$	87
4.8 (a) Clone and (b) no clone distribution for sum of all muon nShared. Since in this case the clones are of each other, for the clones there is clear peak at three.	88
4.9 Fit to constrained $J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow \pi^+ K^-)$ mass with all the components for (a)(b) 2011, (c)(d) 2012, (e)(f) 2016. On the left, fit to data with pion ID (giving kaon misid probabilities), on right data with kaon ID (pion misid rates).	93
4.10 (a) $\pi \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ for 2012 data. (b) This is compared to the standard PIDCalib $D^{*+}(\rightarrow D^0(\rightarrow K^+ \pi^-)\pi^+)$ sample.	96

4.11 (a) $K \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ for 2012 data. (b) This is compared to the standard PIDCalib $D^{*+}(\rightarrow D^0(\rightarrow K^+ \pi^-)\pi^+)$ sample.	97
4.12 (a) $\pi \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ for 2016 data. (b) This is compared to the standard PIDCalib $D^{*+}(\rightarrow D^0(\rightarrow K^+ \pi^-)\pi^+)$ sample.	97
4.13 (a) $K \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ for 2016 data. (b) This is compared to the standard PIDCalib $D^{*+}(\rightarrow D^0(\rightarrow K^+ \pi^-)\pi^+)$ sample.	98
5.1 Schematic view of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay. All charged tracks (in filled-blue) are combined into four-vector representing the visible part of the decay (semi filled-blue). Information about invisible neutrino (semi filled-red) are deduced from the conservation of momentum with respect to the direction of the flight of the B^\pm meson.	100
5.2 Distributions for signal MC in using Pythia 6.4 [44] conditions. (a) $K^*(\mu^+, \nu_\mu)$ (b) $q(\mu^+, \mu^-)$ distributions under different K^* mass hypotheses. The most flat distribution in $K^*(\mu^+, \nu_\mu)$ is plotted in yellow.	106
5.3 (a) Signal simulation sample distribution in $minq$ and $maxq$ variables. Values below 980 MeV/ c^2 (red line) are accepted. (b) Combinatorial data sample after <i>stripping</i> selection with no other cuts shows clearly the J/ψ (green) and $\Psi(2S)$ (blue) resonances which are vetoed and the measurement region (red).	110
5.4 An example of decay topology for (a) background and (b) signal.	113
5.5 Isolation score for signal and backgrounds using (a) Run I (b) Run II samples. If isolation fails to find any other track in the event, by default it gives value -2.	114
5.6 Punzi FOM have the optimum working point at 0.47 for Run I and 0.54 for Run II as seen in both figures with a violet line for $\sigma = 3$ and $\sigma = 5$	115

5.7 Comparison of separate and combined training samples and performance on different datasets. Two vertical violet lines represent optimal points in the signal efficiency, for Run I (0.47) and for 2016 (0.34) where the working point of the two BDTs are chosen. Separate training provides greater rejection power in 2016. In Run I training on both datasets provides comparable performance for given optimal signal efficiency. Taking into the account the fact that offline selection slightly differs for 2016, it is advantageous to keep training separately.	117
5.8 The variable with the most discriminative power for both Run I and 2016.	118
5.9 Study of linear correlation between BDT output and $M_{B_{corr}}$ and BDT value for each bin of $M_{B_{corr}}$ in 2016 shows that Combinatorial BDT is relatively flat as a function $M_{B_{corr}}$	119
5.10 Comparison of separate and combined training samples and performance on different datasets. Optimal working point is chosen, see Figure 5.11 and its corresponding signal efficiency in Run I is 0.44 and for 2016 0.37 denoted with a violet line. As the performance is better for 2016 when the training is performed separately, the training is kept separately also to be consistent with previous methodology.	120
5.11 Punzi FOM have the optimum working point at 0.21 for Run I and 0.27 for Run II as seen in both figures with a violet line for $\sigma = 3$ and $\sigma = 5$	121
5.12 (a) Run1 (b) 2016 binned χ^2 fit to misID sample yielding estimates for the number of background events.	121
5.13 (a) The efficiency of applying 2016 combinatorial BDT at optimal working point on the 2016 SS misID sample. It can be seen that (b) combinatorial component of the misID sample has been significantly reduced, where red curve is the distribution after applying the cut.	122
5.14 The variables with the most discriminative power for both misid Run I and II, mostly kinematic properties of different muons.	123

5.15 (a) The resolution of signal simulation with stripping 21 in bins of corrected mass error δ . (b) The resolution of signal simulation with stripping 26 in bins of corrected mass error δ .	126
6.1 (a) Fit to upper mass side band just before application of MVA selection. (b) Fit to $\mu^+\mu^+\mu^+\nu$ same sign sample. (c) Fit to $b\bar{b}$ sample with exponential function. In (b)(c), exponential description is not correct below $4000 \text{ MeV}/c^2$. All plots contain exponential constants.	129
6.2 Diagram of the iterative procedure to establish contamination from decays where pion and kaon are misidentified for muon.	131
6.3 Examples of distributions where misID procedure is applied to obtain yields and shapes for Run I before misID BDT was applied. On the left, unweighted misID distributions (black), weighted with no cross-feed misID distributions (blue) and weighted misID distributions with cross-feed (red) for (a) kaon SS (c) pion SS . On the right, only weighted misID distributions for Run I (b) kaon SS (d) pion SS are shown together with the yield estimates. These shapes are obtained after combinatorial BDT was applied, but before misid BDT was applied. Total yields need to be multiplied by 100 to counteract the prescale that was applied on this data.	134
6.4 Examples of distributions where misID procedure is applied to obtain yields and shapes for Run I before misID BDT was applied. On the left, unweighted misID distributions (black), weighted with no cross-feed misID distributions (blue) and weighted misID distributions with cross-feed (red) for (a) kaon OS (c) pion OS. On the right, only weighted misID distributions for Run I (b) kaon OS (d) pion OS are shown together with the yield estimates. These shapes are obtained after combinatorial BDT was applied, but before misid BDT was applied. Total yields need to be multiplied by 100 to counteract the prescale that was applied on this data.	135

6.5	(a) Signal and partially reconstructed background distributions scaled to their expected ratio after full MVA selection assuming following branching fractions: $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) = 1 \times 10^{-8}$ and $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu) = (4.10 \pm 0.50) \times 10^{-7}$. (b) Full selection efficiency as a function of invariant mass of the proxy pions is constant.	137
7.1	Summary of ratio of efficiencies between 2012 simulation and 2016 simulation with no FCME split. Efficiency values for 2016 are TCK-weighted averaged efficiencies.	145
7.2	(a) Combinatorial BDT response for signal MC and upper mass sideband as well as for normalisation channel MC for Stripping 21 and Stripping 26. The most discriminative variables are (b) p_T of B , (c) muon/kaon p_T and (d) muon/kaon p	151
7.3	(a) Misid BDT response for signal MC and upper mass sideband as well as for normalisation channel MC for Stripping 21 and Stripping 26. The most discriminative variables are (b) p_T of B , (c) muon IP χ^2 and (d) muon/kaon p . Misid BDT responses are plotted with combinatorial BDT already applied.	152
7.4	(a) Visible mass of normalisation and signal simulation. It can be seen that normalisation previous preselection has a sharp cut around B visible mass leading to much higher fitting region efficiency. (b) The corresponding logarithmic version of plot (a).	153
7.5	Fit to 2012 (a) $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ simulation and (c) $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)\pi^+$ simulation under kaon mass hypothesis. On right, the same plots but with logarithmic scale instead.	156
7.6	Fit results in logarithmic scale to (a) Run I (b) Run II $\mu^+\mu^-K^+$ mass spectrum with no fractional corrected mass split, (c)(d) low FCME bin, (e)(f)high FCME bin.	159

7.7	Fit to weighted combined signal <i>cocktail</i> for (a) NO FCME (b) Low FCME and (c) High FCME split.	163
7.8	Fit to weighted combined partially reconstructed background simulation proxy for (a) NO FCME (b) Low FCME and (c) High FCME split.	167
7.9	Binned χ^2 to misid template with no FCME split (b) Low FCME (c) High FCME. In high FCME bin, the distribution of misid is pollutes signal window more than in the low FCME bin. Both full weight error σ_{tot} and partial weight error σ_{par} can be seen.	170
7.10	(a) Non-simultaneous fit, (b) simultaneous fit pulls.	176
7.11	(a) Unbinned maximum likelihood fit to the blinded data in one bin of FCME. Simultaneous unbinned maximum likelihood fit to blinded data after full selection chain in two bins of FCME, with (b) fit to σ_{low} bin, (c) σ_{high} bin.	178
7.12	Simultaneous unbinned extended maximum likelihood fit to unblinded data after full selection chain in two bins of FCME and with $\mathcal{B} = 0$, with (b) fit to σ_{low} bin, (c) σ_{high} bin.	181
7.13	Minimized -log likelihood value at different \mathcal{B}	182
7.14	Simultaneous unbinned maximum likelihood fit to unblinded data after full selection chain in two bins of FCME, with (b) fit to σ_{low} bin, (c) σ_{high} bin.	183
A.1	An example decision tree. The S and B stand for ‘Signal-like’ and ‘Background-like’. The β_i variables refer to the cut values chosen by the machine learning algorithm after the tree has been trained on signal and background samples. The blue ovals represent final nodes called leafs, which each leaf having an associated purity, i.e. the fraction of the weight of a leaf due to signal events.	195

List of Tables

2.1	Quarks and their properties such as flavour, charge and mass. Flavour is a property which distinguishes different species of quarks. One such property is current quark mass which is the mass of the quark itself. The masses are summarized in [1].	29
2.2	Experimental summary of searches for $B^+ \rightarrow l^+\nu$. Tag Hadronic/Semileptonic/Untagged refers to different way data is selected in Belle and BaBar factories.	47
3.1	Running conditions of LHC and LHCb in different years of data-taking. The statistics of LHCb's instantaneous luminosity, \mathcal{L} is extracted using run database information.	57
3.2	Momentum-dependent definition IsMuon variable.	71
4.1	Offline selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ decay.	90
4.2	Summary of the fit parameters and individual component constraints for $B^0 \rightarrow J/\psi K^*$ fit.	91
4.3	It is necessary as this is unrelated to following figures <code>K_InMuonAcc==1.0</code> shows the interpolation of K tracks into muon chambers. It can be seen that both samples agree with each other very well, meaning that measured misid rate is done for the same fraction of considered tracks. This measurement is done in a pseudorapidity region $1.5 < \eta < 5.0$	95

5.1	Summary of signal simulation samples used in this analysis with different decay models. In all cases the daughters of B^\pm are required to be within LHCb acceptance. All of this samples are mixture under magnetic polarity up and magnetic polarity down conditions.	104
5.2	Selection of events based on muon and the B^+ candidate requirements. <i>Stripping selection</i> for the signal decay $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ is the same for both Run1 and 2016 data.	108
5.3	Trigger selection applied on both signal and normalisation samples. . .	109
5.4	Vetoos for J/ψ and $\Psi(2S)$ resonances. As $\min q < 980 \text{ MeV}/c^2$, these vetoes apply to $\max q$ combination only.	111
5.5	Offline selection performed after stripping. Differences can be seen between Run I and Run II datasets.	111
5.6	BDT variables used in both combinatorial and misID Run I and Run II BDTS	113
5.7	Original preselection of events for normalisation channel for $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ for Run I and Run II.	124
6.1	Species region definitions.	131
6.2	Contribution to total $D^0 \rightarrow (\eta/\eta' \rightarrow \mu\mu\gamma) X$ rate made from all the considered decays above. In total, this cumulative contribution is approximately three times smaller than $D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-$. All the branching fractions are obtained from [1].	138
6.3	Different and total contribution to $B^0 \rightarrow \eta(')\rho(\omega)$. All the branching fractions are obtained from [1].	139
7.1	Method of obtaining efficiencies. Most of these efficiencies are evaluated using simulation, however, TRG and PID efficiencies are evaluated using data and/or simulation techniques.	142

7.2	Summary of individual simulation and/or data efficiencies in % necessary for <i>single event sensitivity</i> for signal and normalisation channel. Efficiency values for 2016 are TCK-weighted averaged efficiencies, which will be explained in subsection 7.2.3. The errors considered are of statistical nature, computed using binomial error.	143
7.3	Summary of 16 different TCKs listing properties of candidates necessary to pass L0 and HLT1 selection in 2016. In the final row, the default configuration for 2016 is shown and it corresponds to 288888335 TCK.	148
7.4	Efficiencies of 2016 trigger emulation on MC. Depending on TCK, the efficiencies vary up 10% for L0 level for signal MC and up to 5% for normalisation TCK. This is important as <i>single event sensitivity</i> is sensitive to the ratio of these two efficiencies. This configuration is describing correctly only 35% data with high p_T threshold.	149
7.5	Signal simulation efficiency using PIDCalib efficiencies.	154
7.6	Normalisation MC efficiency using PIDCalib efficiencies.	154
7.7	Summary of the fit parameters and individual component constraints for the fit to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ decays.	157
7.8	$B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ signal yield obtained from fits to $\mu^+ \mu^- K^+$ mass spectrum shown in Figure 7.6.	160
7.9	Signal simulation weights used to create cocktail of mixed Run I (2012) and Run II (2016) cocktail. The cross-sections listed here are not absolute numbers, but rather relative as only its ratio matters.	162
7.10	Summary of number of events that comes from partially reconstructed backgrounds in different bins of FCME, assuming 2012 efficiencies but extrapolating to all samples.	166
7.11	The final misid template is constructed by summing the contribution from Run I and II kaon, pion and proton-like regions for both SS and OS <i>misID</i> contributions. The last column adds cummatively the contributions with respect to the previous row.	169

7.12 Summary of systematic uncertainties.	173
7.13 Signal bias estimate from 10 000 pseudo-experiments for both simultaneous and not simultaneous fit.	176
7.14 For all constrained variables the range is set to be within $\pm 5\sigma$	179
7.15 Fit results for all floating variables in the unblinded data fit. The $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = -1.78 \times 10^{-8}$. Variables R_K^S are the efficiency ratios obtained by normalising the decays to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ decays where S stands for stripping and K for the FCME. $N_{misID_K}^{scaled}$ is the number of misID events, N_{combi_K} is the number of combinatorial events, β_K is the exponential constant, μ_{misID_K} and σ_{misID_K} are the mean and the σ of the CB function.	184
7.16 Resulting yields for different components from the corrected mass fit with $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = -1.78 \times 10^{-8}$	185

List of abbreviations and definitions

$IP\chi^2$ The $IP\chi^2$ is the difference in the χ^2 of the fit to the primary vertex, when the track whose $IP\chi^2$ is being measured is added and then removed.

ALICE A Large Ion Collider Experiment.

ATLAS A Toroidal LHC ApparatuS.

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

CB Crystal Ball function.

CKM Cabibbo-Kobayashi-Maskawa mixing matrix.

CMS Compact Muon Solenoid.

$\text{Cos}(\theta_B)$ The cosine of the angle between the momentum vector of the B^+ meson and the direction of the flight of the B^+ meson from its primary vertex to its secondary vertex .

ECAL Electromagnetic calorimeter.

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

$FD\chi^2$ The **FD** χ^2 is defined as the increase in χ^2 when the primary and secondary vertex are fitted separately as compared to single vertex fit.

FOI Field of Interest.

HCAL hadronic calorimeter.

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

HLT1 First stage of high level trigger.

HLT2 Second stage of high level trigger.

HPD Photomultiplier tubes that collect ĀÑerenkov light.

ID Probability of correctly identifying particle, given PID requirement.

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

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

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

LHC Large Hadron Collider.

LHCb The Large Hadron Collider beauty experiment.

long track Long track is track category which classifies tracks that have hits in the VELO and the T stations. Hits in the TT stations are optional.

LS1 Long Shutdown 1.

M1-5 The five muon stations.

MC Monte Carlo Simulation.

Min IP χ^2 The minimum impact parameter χ^2 is the minimal difference in fit χ^2 (quality of the fit) to the primary vertex between fit with this track added and removed.

misID Probability of incorrectly identifying particle given PID requirement.

MWPCs multi-wire proportional chambers.

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

P_{ghost} Ghost Probability is probability of misreconstruction of the track, where for each track 0 is most signal-like and 1 is most ghost-like. A charged particle is not considered to be a ghost if 70% of the hits match between the reconstructed and simulated true tracks. Similarly, neutral particles are ghosts if simulated particle contributes less than 50% of the reconstructed cluster energy from calorimeter.

PID Particle IDentification.

PRS pre-shower.

PS Proton Synchotron.

PSB Proton Synchotron Booster.

PV Primary Vertex, the pp interaction vertex.

QCD Quantum Chromodynamics.

QED Quantum Electrodynamics.

QFT Quantum Field Theory.

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

RICH1 Ring Imaging Cherenkov providing low momentum **PID** by using Čerenkov radiation.

RICH2 Ring Imaging Cherenkov providing high momentum **PID** by using ĀNerenzov radiation.

SM Standard Model.

SPD Scintillator Pad Detectors.

SPS Super Proton Synchotron.

SV Secondary Vertex.

T1, T2 and T3 Trackers downstream of the magnet composed of silicon micro-strips strips in the inner section and straw tubes in the outer section..

TIS Events which are Triggered Independent of Signal.

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

TOS Events which are Triggered On Signal.

Track χ^2/ndof The track χ^2 pre degree of freedom is the minimal difference in fit χ^2 (quality of the fit) to the primary vertex between fit with this track added and removed.

TT The tracking station upstream of the magnet composed of silicon micro-strips..

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

Vertex χ^2/ndof The vertex χ^2 per degree of freedom in a vertex fit.

Chapter 1

Introduction

The field of particle physics aims is to describe the universe we see today by decomposing everything into fundamental building blocks which then follow certain behaviour according to a given set of rules. The foundation bricks of theoretical formulation of the current best established theory that describe the universe around us, the Standard Model ([SM](#)), were laid last century. Some achievements of the [SM](#) do really leave us breathless with agreement between theoretical and experimental results of ten parts in a billion.

This theory is, however, incomplete as it fails to address several issues. The theory does not include any explanation for the nature of dark matter and it doesn't make any attempt to describe gravity in a quantum field theory framework. Furthermore, fine-tuning of some parameters in the [SM](#) such as the Higgs mass, where parameters get exactly the right value to produce required behaviour, beg questions if there is some symmetry in the model building that is missing. Lastly, as with any model, SM operates with many free parameters that need to be plugged in so that predictions can be made. So why are there exactly so many?

This thesis describes a search for a decay which can help to shed light on some of these parameters with following organization.

In [chapter 2](#) the [SM](#) of particle physics is discussed together with the theoretical and experimental motivation for fully leptonic decays, especially for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$

decay. In [chapter 3](#) the tool to search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decays, the LHCb detector, is detailed. This is then followed by a discussion about how does trimuon signature behave in the detector which is covered in [chapter 4](#). The analysis of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$, the central theme for the thesis is then described in [chapter 5](#) and [chapter 7](#).

Chapter 2

Theory

The Standard Model is without question the most powerful and tested theory of particle physics. It describes and predicts many phenomena very well even though as discussed in previous chapter it fails to address few known issues. In this chapter, the theoretical basis of the Standard Model is first laid out which is then followed by experimental and theoretical consideration of fully leptonic decays.

2.1 Review of the Standard Model

The Standard Model ([SM](#)) of particle physics is currently the most accurate model describing the building blocks of matter, particles, and their interactions via forces. In particular the [SM](#) describes all the fundamental forces but gravity. It is a quantum field theory ([QFT](#)) whereby the dynamics of the system is captured by the most general renormalisable Lagrangian density that is invariant under gauge symmetry. [QFT](#) considers particles to be excited states of an underlying field, also known as quanta. In the [SM](#), particles and forces are the results of interactions between scalar, vector and spinor fields. In general there are two sets of particles. The first set are force-carrying particles also known as bosons, which have integer spin and are quanta of the scalar and vector fields. More specifically, there is the Higgs boson, the only elementary scalar boson in the [SM](#), and vector bosons: gluons, W^\pm , Z and γ . Secondly, there are the

Generation	Flavour	Charge	Constituent Mass
1st	up u	+2/3	$2.2^{+0.6}_{-0.4}$ MeV
1st	down d	-1/3	$4.7^{+0.5}_{-0.4}$ MeV
2nd	charm c	+2/3	1.28 ± 0.03 GeV
2nd	strange s	-1/3	96^{+8}_{-4} MeV
3rd	top t	+2/3	173.1 ± 0.6 GeV
3rd	bottom b	-1/3	$4.18^{+0.4}_{-0.3}$ GeV

Table 2.1: Quarks and their properties such as flavour, charge and mass. Flavour is a property which distinguishes different species of quarks. One such property is current quark mass which is the mass of the quark itself. The masses are summarized in [1].

non-force carrying particles, which are fermions, quanta of spinor fields. Unlike bosons they carry half-integer spin. These can be further classified into two elementary families of particles: quarks, which cannot be observed alone and leptons which can be detected on their own. Out of all of these fundamental particles, those that have mass acquire it by the Higgs mechanism.

Quarks are affected by all three fundamental forces. They come in six different *flavours* and they carry fractional charge as seen in Table 2.1.

There are also 12 leptons in total. Unlike quarks they are not affected by the strong force but also come along in three generations with increasing mass: electrons, muons and taus. They all have their antiparticles and corresponding neutrinos. Much of this thesis is dedicated to the study of the muons or antimuons and their neutrinos.

In the rest of the chapter, the SM formulation is introduced starting with the principle of local gauge invariance explained in section 2.2. The strong and electroweak sectors are described in section 2.3 and section 2.4 and the necessary process of mass generation in the SM, the Higgs mechanism, is covered in section 2.5. The effect of the Higgs mechanism on the electroweak sector is then described in section 2.6 resulting

in the quark mixing matrix detailed in [section 2.7](#). Following sections then discuss the theoretical and experimental status of fully leptonic decays, which are sensitive to elements of the quark mixing matrix. Finally a discussion about decay model used for the search of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ is covered in [section 2.10](#).

2.2 The principle of Standard Model building

In more mathematical terminology, the [SM](#) is a theory that respects $SU(3) \otimes SU(2) \otimes U(1)$ symmetries. In this section the form of the Lagrangian density of the [SM](#) is motivated. Throughout it is assumed that $\hbar = 1$, $c = 1$. The Dirac Lagrangian for a spin- $\frac{1}{2}$ non-interacting or free field ψ (spinor field) for a particle with mass m can be written as

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

where γ^μ are 4×4 Dirac matrices and $\mu \in \{0, 1, 2, 3\}$. By using the Euler-Lagrange equation in the relativistic theory

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\psi_i)}\right) = \frac{\partial\mathcal{L}}{\partial\psi_i} \quad (2.2)$$

for $\bar{\psi}$ in [Equation 2.2](#) the equation

$$i\gamma^\mu\partial_\mu\psi - m\psi = 0 \quad (2.3)$$

can be retrieved. This is the Dirac equation of motion.

The Dirac Lagrangian in [Equation 2.2](#) stays the same under a global phase transformation: $\psi \rightarrow e^{i\phi}\psi$ and $\bar{\psi} \rightarrow e^{-i\phi}\bar{\psi}$. However, under local phase transformation, where ϕ is a function of x^μ , this is not the case any more. In this case

$$\mathcal{L} \rightarrow \mathcal{L} - (\partial_\mu\phi)\bar{\psi}\gamma^\mu\psi. \quad (2.4)$$

By requiring local gauge invariance for the Lagrangian, it is necessary to add a term to counteract the left-over term in [Equation 2.4](#). Let $\lambda = -\frac{\phi(x)}{q}$ and let A_μ be some new (vector) field which transforms as $A_\mu \rightarrow A_\mu + \partial_\mu\lambda$ then the following Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu\psi A_\mu \quad (2.5)$$

stays invariant under a local phase transformation. That is good, however, there is a penalty for introducing a new vector field A_μ which interacts with the spinor field ψ as can be seen in the last part of [Equation 2.5](#). It is now necessary to also introduce a non-interacting term for A_μ .

The Lagrangian for the non-interacting vector field for a particle with mass m_A and field strength $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is

$$\mathcal{L} = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{8\pi}m_A^2 A^\mu A_\mu. \quad (2.6)$$

In order not to spoil the local gauge invariance, it is required that $m_A = 0$. Hence the full Dirac Lagrangian with local phase invariance introduces a massless vector field A^μ and is of the form

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu}, \quad (2.7)$$

which can be recognized as the Lagrangian for quantum electrodynamics ([QED](#)), whereby the electrons and positrons (quanta of spinor field) are interacting with photons (quanta of vector field). In other words, A_μ is the electromagnetic potential and $q = e$, the current density is hence $J^\mu = e\bar{\psi}\gamma^\mu\psi$. This represents the $U(1)_{EM}$ part of the SM.

Upgrading from global invariance of the non-interacting Lagrangian in [Equation 2.2](#) to local invariance in one step can be achieved by defining *covariant derivative*

$$\mathcal{D}_\mu = \partial_\mu + iqA_\mu, \quad (2.8)$$

where the secret ingredient is to transform the partial derivative in the same way as the field itself under local gauge transformation.

2.3 Quantum Chromodynamics

To require gauge invariance under a local transformation is a powerful tool and it is used throughout the SM building. In this section the development of Lagrangian for Quantum chromodynamics (QCD) is explained. QCD describes strong interactions or nuclear binding forces and makes use of quarks (q) whereby most of the time they are observed to be bound either in pairs - mesons ($q\bar{q}$) - or triplets - baryons (qqq). The interactions between quarks and gluons are described by the $SU(3)_C$ gauge group. The conserving charge associated with the strong force is known as color, hence the subscript C. It was experimentally established that there are 3 colors and borrowing from color theory used by painters these colors are red, blue and green. The quark carries color and antiquark anticolor making mesons and baryons colorless. There are 8 quanta of strong interactions known under the name gluons.

With these constraints, and by requiring the free Lagrangian to be invariant under local $SU(3)$ transformation similarly to the QED case, *covariant derivative*

$$\mathcal{D}_\mu = \partial_\mu - ig_s \frac{\lambda^a}{2} G_\mu^a \quad (2.9)$$

that respects $SU(3)$ symmetry is obtained. λ^a are Gell-Mann matrices, $a \in \{1..8\}$ (8 possible gluons) and g_s is the strong coupling. The field strength for the gluon field is defined as $G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu + g_s f^{abc} G_b^\mu G_c^\nu$, where f^{abc} are so-called structure constants which satisfy the following commutation relation:

$$\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2} \right] = if^{abc} \frac{\lambda^c}{2}. \quad (2.10)$$

As compared to the QED field strength, there is an additional term involving gluon fields themselves, causing cubic and quartic gluon interactions, which were not present before.

Another interesting behaviour of strong interaction is that the quarks are not observed alone. This is due *confinement* which can be understood within the framework of QFT theory by observing evolution of the coupling strength g as a function of energy

scale, also known as β function. The β -function for a coupling constant g in the SM to the two loop contribution takes the following form:

$$\beta_g = \mu \frac{dg}{d\mu} = \frac{1}{16\pi^2} \beta_g^{(1)} + \frac{1}{(16\pi^2)^2} \beta_g^{(2)}, \quad (2.11)$$

where $\beta_g^{(1)}$, $\beta_g^{(2)}$ denote the one-loop and two-loop contributions respectively, and μ is the energy scale. For the strong interaction, unlike electromagnetic and weak interaction, β_g is negative. For low energies as $\mu \rightarrow 0$ the coupling is very high and hence quarks cannot be observed on their own, *confinement*. On the other hand as $\mu \rightarrow \infty$, or at high energies, the coupling gets small, particles get decoupled, which is known as *asymptotic freedom*.

The full Lagrangian density for the strong interaction reads as:

$$\mathcal{L}_{QCD} = i\bar{\psi}\gamma^\mu \mathcal{D}_\mu \psi - m\bar{\psi}\psi - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi + g_s \bar{\psi}\gamma^\mu \frac{\lambda^a}{2} \psi G_\mu^a - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a, \quad (2.12)$$

where the interaction between quarks and gluons is encoded in the third term.

2.4 Electroweak unification

The idea behind unification of weak and electromagnetic interactions is very powerful, as it has to accommodate for forces that act with very different strength with force-carrying particles that are both massive (W^\pm, Z) and massless (γ). Furthermore W^\pm bosons only couple to left-handed particles, whereas the Z^0 boson couple to both left and right-handed particles. To aid with the situation, the spinor field can be decomposed into left-handed and right-handed (chiral) spinor components

$$\psi = \psi_L + \psi_R = P_L \psi + P_R \psi, \quad (2.13)$$

where $P_L = \frac{1-\gamma^5}{2}$ and $P_R = \frac{1+\gamma^5}{2}$ are known to be projection operators. By calling these operators left-handed and right-handed, there is a misconception that ψ_L is a helicity

eigenstate, but this is only true given the particle in question is massless. These spinors are known to have chirality - known as left or right-handedness. Helicity is rather a projection of the spin on the direction of the momentum.

The spinor field decomposition then have influence on electromagnetic and weak currents as well as fermion mass term. Firstly the fermion mass term mixes both left handed and right handed spinors as

$$m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R). \quad (2.14)$$

On the other hand, the electromagnetic current is not mixing the left and right handed components, since $\bar{\psi}\gamma^\mu\psi = \bar{\psi}_R\gamma^\mu\psi_R + \bar{\psi}_L\gamma^\mu\psi_R$. Finally weak charged current only acts on left-handed fermions as $\frac{1}{2}\bar{\psi}\gamma^\mu(1 - \gamma_5)\psi = \bar{\psi}_L\gamma^\mu\gamma_L$.

Another observation is that the charged weak interaction only couples leptons within the three generations. This motivates left-handed isospin doublets where for the first generation of fermions:

$$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad L_L = \begin{pmatrix} e_L \\ \nu_L \end{pmatrix}, \quad (2.15)$$

and right-handed isospin singlets for up-type quarks, down-type quarks and charged leptons:

$$u_R = (u_R, c_R, t_R), \quad d_R = (d_R, s_R, b_R), \quad l_R = (e_R, \mu_R, \tau_R). \quad (2.16)$$

The simplest group with doublet representation is $SU(2)$ and in combination with the electromagnetic interaction forms $SU(2)_L \otimes U(1)_Y$. The conserving charges are inter-related

$$Q = I^3 + \frac{1}{2}Y, \quad (2.17)$$

where I refers to weak isospin, Y refers to weak hypercharge, and Q is electric charge.

Again by assuming gauge invariance under local transformation the *covariant derivative* of $SU(2)_L \otimes U(1)_Y$ is

$$D_\mu = \partial_\mu + i\frac{g}{2}W_\mu^i \frac{\sigma^i}{2} - i\frac{g'}{2}B_\mu. \quad (2.18)$$

Here σ^i are Pauli matrices, g, g' are the electroweak couplings and W_μ^i where $i \in \{1, 2, 3\}$ and B_μ are the vector fields that should be corresponding to W^\pm, Z^0, γ . The field strengths are defined as $B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu$ and $W_{\mu\nu}^i = \partial^\mu W_\nu^i - \partial^\nu W_\mu^i + g\epsilon^{ijk}W_\mu^j W_\nu^k$.

The real charged bosons corresponding to W^\pm arise as linear combinations of W_μ^i , for $i \in (1, 2)$ as

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \\ W_\mu^- &\equiv W_\mu^-, \\ W_\mu^+ &\equiv W_\mu^+. \end{aligned} \quad (2.19)$$

The neutral bosons are obtained using W_μ^3 and B_μ in a similar fashion in the following way:

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \quad (2.20)$$

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W, \quad (2.21)$$

where θ_W angle is known as weak mixing angle and can be determined experimentally from the masses of the Z and W^\pm bosons by the relation $\cos\theta_W = \frac{M_W}{M_Z}$. So far, however, there was no consideration of how bosons or fermions for that matter become massive which will be covered in the next section.

The full Lagrangian of electroweak theory then consists of the kinetic part

$$\mathcal{L}_{kin} = -\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W_i^{\mu\nu}W_{\mu\nu}^i \quad (2.22)$$

where for $W_i^{\mu\nu}$ like in QCD there is cubic and quartic self interaction amongst the gauge fields. Then there are interactions between the quark-lepton fields and the gauge bosons where it is conventional to split this into two categories according to the charge of gauge bosons. This is what gives rise to charged and neutral currents for the electroweak interactions. So employing the physical gauge boson representation, charged current

Lagrangian \mathcal{L}_{CC} and neutral current Lagrangian \mathcal{L}_{NC} for one family of fermions reads as

$$\mathcal{L}_{CC} = -\frac{g}{2\sqrt{2}} \left[W_\mu^\dagger [\bar{v}\gamma^\mu(1-\gamma_5)l + \bar{u}\gamma^\mu(1-\gamma_5)d] + h.c. \right], \quad (2.23)$$

$$\mathcal{L}_{NC} = -g\sin\theta_W(\bar{l}\gamma^\mu l)A_\mu - \frac{g}{2\cos\theta_W} \sum_{\psi=\nu,l} \bar{\psi}_i \gamma^\mu(g_V^i - g_A^i \gamma_5) \psi_i Z_\mu \quad (2.24)$$

where the famous (V-A) structure of the weak charged current can be seen. The first part of \mathcal{L}_{NC} represents known electromagnetic interaction $e = g\sin\theta_W$.

If the field is considered to be under U(1) charge then it was shown that this gauge field was invariant in the QED case. However under SU(2), only left-handed fields transform and hence for the fermionic mass term, which mixes right-handed and left-handed terms as shown in [Equation 2.14](#) breaks the gauge invariance. For this very reason and also to give mass to the gauge bosons the Higgs mechanism is introduced.

2.5 The Higgs Mechanism

The Higgs mechanism introduce a new scalar field with potential V into the model. Through the process known as spontaneous symmetry breaking it allows fermions and gauge bosons to have a mass term in their Lagrangians while retaining gauge invariance. Let ϕ be a doublet of complex scalar fields where

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (2.25)$$

where $\phi^+ = \frac{\phi_1 + i\phi_2}{\sqrt{2}}$ and $\phi^0 = \frac{\phi_3 + i\phi_4}{\sqrt{2}}$ so that $\phi^\dagger\phi = \frac{\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2}{2}$. The Lagrangian for this field is then

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) + V = (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2, \quad (2.26)$$

where D_μ is given in [Equation 2.18](#), and V is the famous Mexican hat potential where the x-axis is ϕ_1 and the y-axis is ϕ_2 . It is required that $\lambda > 0$ to have a ground state.

By finding the ground state - or the stable minimum - of this potential with $\mu^2 < 0$, one gets infinite number of this minima such that

$$\phi\phi^\dagger = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2}. \quad (2.27)$$

This is the same as saying that the minimum is independent of direction as it lies on the circle of minima. As the minimum is usually known as vacuum, v is called vacuum expectation value. By choosing a particular minimum one fixes the direction and the symmetry of $SU(2) \otimes U(1)$ is spontaneously broken, meaning that the overall theory is symmetrical but the ground state exhibits antisymmetry. By convention, the direction $\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$ is chosen. Detailing both real and imaginary part of the fields the direction can be translated so that $\phi_3 = \frac{v}{2}, \phi_1 = \phi_2 = \phi_4 = 0$. This allows for generation of three massive bosons W^\pm and Z^0 and massless γ of the electroweak theory. The Higgs boson itself arises as an excited quanta around the minimum

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \quad (2.28)$$

2.6 Fermion mass generation

Moreover, introducing an additional scalar doublet into the model fixes the broken gauge symmetry for fermionic mass mentioned in [Equation 2.14](#) as it is possible to construct the fermion-scalar interaction Lagrangian that is gauge invariant, usually denoted as Yukawa Lagrangian \mathcal{L}_Y . It is made up of the leptonic part and the quark part:

$$\mathcal{L}_Y = \mathcal{L}_L + \mathcal{L}_Q. \quad (2.29)$$

The leptonic term for one family of leptons using definitions in [Equation 2.15](#) and [Equation 2.16](#) is

$$\mathcal{L}_L = g_l (\bar{L}_L \phi l_R + \bar{l}_R \phi^\dagger L_L), \quad (2.30)$$

With $\phi_c = \begin{pmatrix} \phi^0 \\ \phi^- \end{pmatrix}$ the full three generation quark term is

$$\mathcal{L}_Q = y_{ij}^u \bar{Q}_L^i \phi u_R^j + y_{ij}^d \bar{Q}_L^i \phi_c d_R^j + h.c., \quad (2.31)$$

where h.c stands for Hermitian conjugate, i, j are the generations, y^q are 3×3 matrices defining strengths between generations with each y^q . After spontaneous symmetry breaking (Equation 2.28), the leptonic interaction term becomes

$$\mathcal{L}_L = \frac{g_l v}{\sqrt{2}} (\bar{l}_L l_R + \bar{l}_R l_L) + \frac{g_l}{\sqrt{2}} (\bar{l}_L l_R + \bar{l}_R l_L) H = m_l (\bar{l}_L l_R + \bar{l}_R l_L) (1 + \frac{H}{v}), \quad (2.32)$$

where the mass term is then defined as $m_l = \frac{g_l v}{\sqrt{2}}$. In a similar way for quarks,

$$\mathcal{L}_Q = \frac{v}{\sqrt{2}} (y_{ij}^d \bar{u}_L^i u_R^j + y_{ij}^d \bar{d}_L^i d_R^j + h.c.) (1 + \frac{H}{v}). \quad (2.33)$$

where the quark masses are grouped into 3×3 complex matrices of up-type quark (down-type quark) $M_{ij}^u = \frac{v}{\sqrt{2}} y_{ij}^u$ ($M_{ij}^d = \frac{v}{\sqrt{2}} y_{ij}^d$). So in conclusion before the spontaneous breakdown of the electroweak symmetry, all quarks and leptons are massless. Once the Higgs scalar field acquires a vacuum expectation value implying a broken symmetry, quarks and leptons acquire mass. It is possible then to decompose a complex matrix into two distinguishable unitary and one diagonal matrix. The mass matrices can be diagonalised by unitary transformations U_{uL} and U_{dL} in the following way:

$$\begin{aligned} \mathcal{M}_u &= U_{uL}^\dagger M^u U_{uR} = \text{Diag}\{m_u, m_c, m_t\}, \\ \mathcal{M}_d &= U_{dL}^\dagger M^d U_{dR} = \text{Diag}\{m_d, m_s, m_b\}. \end{aligned} \quad (2.34)$$

This way of diagonalising mass matrices is the most general case of a weak basis transformation which transforms a system to different basis without altering the physics. Such transformation is equivalent to changing quark fields from the basis of flavour eigenstates to that of mass eigenstates. In particular for $q \in (u, d)$

$$\begin{aligned} q'_L &\equiv U_{qL} q_L, \\ q'_R &\equiv U_{qR} q_R. \end{aligned} \quad (2.35)$$

This change into the mass eigenstate basis does not affect most of the Lagrangian. More specifically, there will be no change to the L_{NC} in [Equation 2.24](#) when expressed in mass eigenstates (hence at tree-level there are no flavour changing neutral-currents in the [SM](#)), however, charged current L_{CC} in [Equation 2.36](#) is affected. Due to the diagonalisation of the mass matrices L_{CC} now includes non diagonal couplings for the current as seen in the L_{CC} for all three fermion generations:

$$\mathcal{L}_{CC} = -\frac{g}{2\sqrt{2}} \left[W_\mu^\dagger \left[\sum_l \bar{v} \gamma^\mu (1 - \gamma_5) l + \sum_{ij} \bar{u}_i \gamma^\mu (1 - \gamma_5) V_{ij} d_j \right] + h.c. \right], \quad (2.36)$$

In this equation there is a new term $V_{ij} = V_{CKM} = U_{uL} U_{dL}^\dagger$ which is the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix. From [Equation 2.34](#) it follows that $V_{CKM}^\dagger V_{CKM} = 1$, or that CKM mixing matrix is unitary by assuming that only the charged current via W will lead to a transition from the up-quark to down-type quark sector. Therefore the CKM matrix elements provide the probabilities of how a W^\pm bosons decay.

2.7 The Quark Mixing Matrix

As mentioned above, from the transformation of the mass matrix using two unitary matrices one obtains the CKM matrix which exhibits a strong hierarchy in the size of the matrix elements. From the previous discussion the quark mixing matrix is a 3×3 complex unitary matrix yielding 18 parameters to start with. Unitarity of the CKM matrix implies that matrix elements are orthonormal, reducing the count of free parameters to 9. Further, 5 out of 6 quark phases can be absorbed into the redefinition of the quark field, cutting the number of parameters down to 4 parameters, three quark mixing angles and one CP violating phase. There are many different parametrisations which are all mathematically equivalent to the CKM matrix, but the standard parametrisation of the CKM matrix for flavour mixing is the following:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \quad (2.37)$$

$$= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\exp(-i\delta) \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\exp(i\delta) & c_{12}c_{23} - s_{12}s_{23}s_{13}\exp(i\delta) & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\exp(i\delta) & -c_{12}s_{23} - s_{12}c_{23}s_{13}\exp(i\delta) & c_{23}c_{13} \end{pmatrix}, \quad (2.38)$$

where $s_{ij} = \sin(\theta_{ij})$ and $c_{ij} = \cos(\theta_{ij})$, θ_{12} , θ_{23} , θ_{13} are Euler angles and θ_{12} is also known as the Cabibbo angle.

A parametrisation reflecting the hierarchical nature in flavour mixing, which is an expansion in terms of the small parameter λ , was introduced by Wolfenstein [2]. The four Wolfenstein parameters are related to the standard parametrization via the following expressions:

$$\begin{aligned} \lambda &= s_{12}, \\ A\lambda^2 &= s_{23}, \\ A\lambda^3(\rho - i\eta) &= s_{13}\exp(-i\delta), \end{aligned} \quad (2.39)$$

$$V_{\text{CKM}_{\text{Wolfenstein}}} = \left(\begin{array}{ccc} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{array} \right) + \mathcal{O}(\lambda^4). \quad (2.40)$$

A geometrical interpretation of CP violation is offered by the concept of unitarity triangles. Unitarity of the CKM matrix can be summarized by two sets of orthogonality

relations: $\sum_k |V_{ik}|^2 = \sum_i |V_{ik}|^2 = 1$ for all i generations and $\sum_k V_{ik} V_{jk}^* = 0$ for all $i \neq j$. One of the unitary constraints of the CKM matrix explicitly states:

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0. \quad (2.41)$$

Dividing this constraint by $V_{cd} V_{cb}^*$ and using relation of $\bar{\rho}$ and $\bar{\eta}$ to ρ and η

$$\rho + i\eta = \frac{\sqrt{A^4 \lambda^4} (\bar{\rho} + i\bar{\eta})}{\sqrt{1 - \lambda^2} [A^4 \lambda^4 (\bar{\rho} + i\bar{\eta})]} \quad (2.42)$$

where $\bar{\rho}$ and $\bar{\eta}$ are defined

$$\bar{\rho} \approx \rho - \frac{\rho \lambda^2}{2}, \quad (2.43)$$

$$\bar{\eta} \approx \eta - \frac{\eta \lambda^2}{2}, \quad (2.44)$$

the constraint can be pictorially represented in the $\bar{\rho}$ and $\bar{\eta}$ plane as in Figure 2.1. The

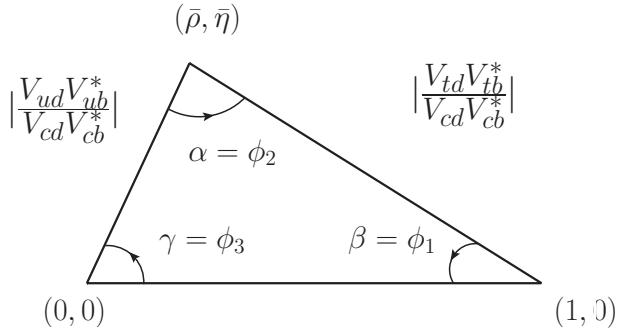


Figure 2.1: Unitarity triangle in a complex plane.

area of the six different triangles are half of the Jarlskog invariant J , a quantifier of CP violation, which is defined as $\text{Im}[V_{ij} V_{kl} V_{il}^* V_{kj}^*]$ [3]. It is interesting to notice that the SM with its parameters may or may not violate CP. Only after measuring the parameter it is possible to determine the CP non-conservation. J vanishes only if the mixing angle $\theta_{ij} = \{0, \pi/2\}$; $\delta = \{0, \pi\}$. So measurements of J allows to verify that the CKM matrix is complex and hence different mixing for quarks and anti-quarks is obtained, providing theoretical grounding for CP violation.

The CKM matrix elements which comprises of magnitudes and phases can be determined in different ways but the most precise option employs a global fit to all

available measurements (as seen in [Figure 2.2](#)) while imposing the [SM](#) constraints such as the unitarity of [CKM](#) matrix. Hence, the most precise measurement of the [CKM](#) matrix magnitudes [export 2018 citation PDG](#) to-date is

$$|V_{\text{CKM}}| = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix}, \quad (2.45)$$

with non-zero Jarlskog invariant $J = (3.18 \pm 0.15) \times 10^{-5}$. Highlighted is the result for magnitude of the V_{ub} matrix element $|V_{ub}|$ which is the element with highest fractional uncertainty on its value. Therefore precise measurement of this element is very important and was the original motivation for the analysis of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$. Moreover, as displayed in [Figure 2.2\(a\)\(b\)](#), the measurement of $|V_{ub}|$ (orange circle)(green circle) together with $\sin(2\beta)$ measurement (green band)(blue band) constrain the apex of the triangle. This means that these two measurements together with other measurements test the unitarity of the [CKM](#) matrix, one of the fundamental assumptions of the [SM](#).

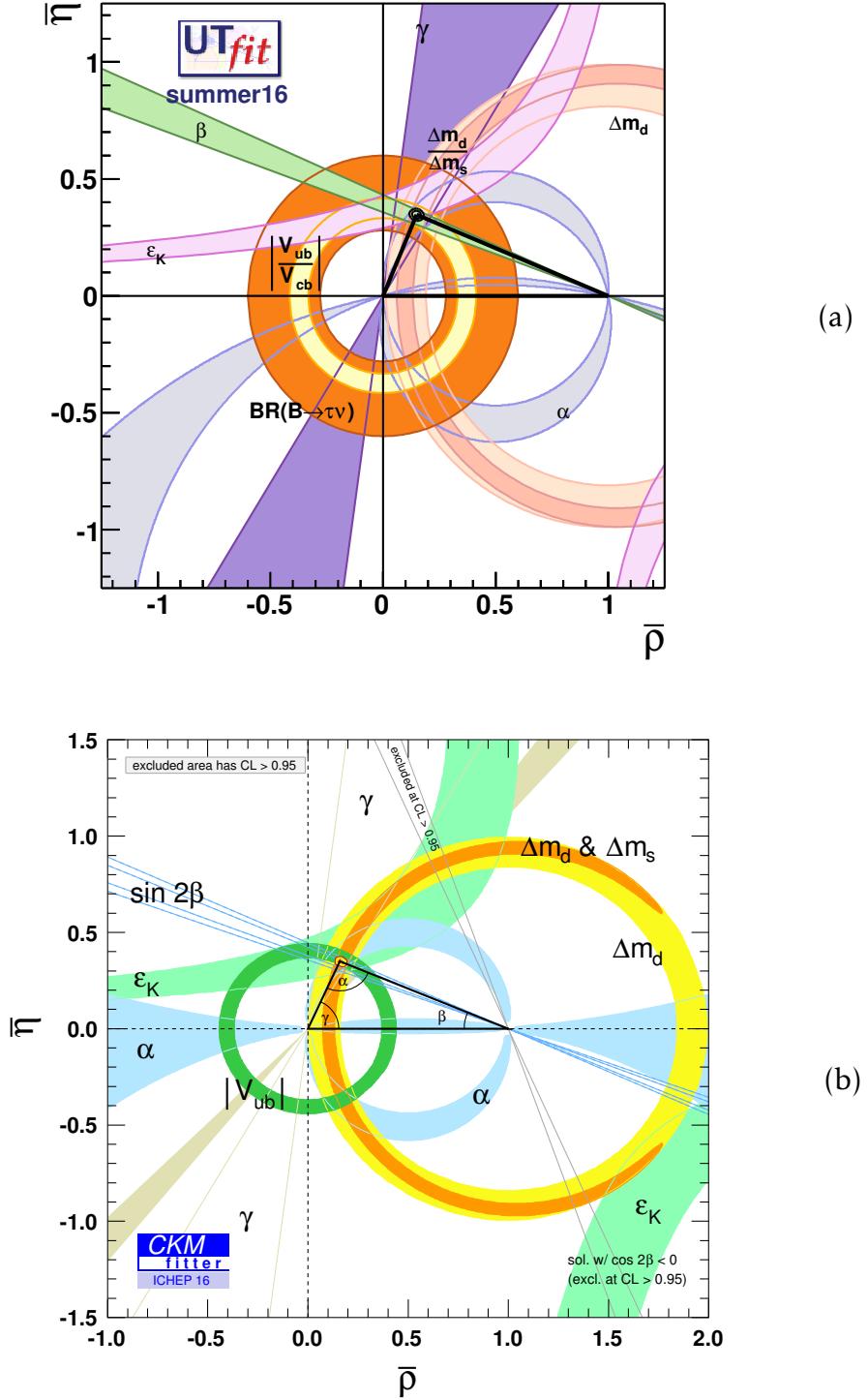


Figure 2.2: Different experimental measurements that constrain the CKM matrix elements together with the fit results from two collaborations (a) UTFit and (b) CKMFitter as of summer 2016. These figures are taken from [4] and [5]. There is a good agreement between the two different results.

2.8 Fully Leptonic $P^+ \rightarrow l^+ \nu_l$ decays

Purely leptonic decays that proceed via annihilation-type diagrams of pseudoscalar mesons (P) are of great interest for flavour physicists because they allow to make:

- either measurements of the CKM matrix elements,
- or measurements of leptonic decay constants,
- or measurements of new physics effects.

The first two types of measurements are possible because the decay rates of $P^+ \rightarrow l^+ \nu_l$ decays are sensitive to the product of the appropriate CKM matrix element ($V_{q_1 q_2}$ where q_1 and q_2 are constituent quarks of the pseudoscalar meson) and decay constant f_P , related parameter arising from the strong interaction. In more detail, the decay width of a fully leptonic decay of a pseudoscalar meson in the SM to the lowest order can be expressed as

$$\Gamma(P^+ \rightarrow l^+ \nu_l) = \frac{G_F^2 m_{P^+} m_{l^+}^2}{8\pi} \left[1 - \frac{m_{l^+}^2}{m_{P^+}^2} \right]^2 f_P^2 |V_{q_1 q_2}|^2, \quad (2.46)$$

where G_F is the Fermi constant, m_{P^+} and m_{l^+} are the pseudoscalar meson and lepton masses, respectively. This decay width can be compared to that of $\tau \rightarrow l \nu \bar{\nu}$ [6]

$$\Gamma(\tau \rightarrow l \nu \bar{\nu}) = \frac{G_{l\tau}^2 m_\tau^5}{192\pi^3} \left[1 - f\left(\frac{m_l^2}{m_\tau^2}\right) \right], \quad (2.47)$$

where $G_{l\tau}$ is relevant Fermi constant. In this case $f(x) = 18x - 8x^3 + x^4 + 12x^2 \log(x)$ represents correction due to finite lepton mass in final state. Corrections arising from the W propagator effects are negligible for this decay and are not considered here and nor are radiative corrections so that only the lowest order contributions are considered. As compared to Equation 2.46 the decay width is significantly higher.

So in order to measure the CKM matrix amplitude, knowledge of f_P must be inferred. f_P can be calculated using the lattice QCD techniques and together with experimental

determination of the decay rates provides provide a way to determine relevant amplitude squared of the relevant CKM matrix element assuming there is no contribution from new physics (new physics). More conventionally, CKM magnitudes are determined from semileptonic decays, but in this case the sensitivity to a different type of current is given [this is in pdg i dont understand why is it wrong]. In purely leptonic decays axial-vector flavour-changing currents ($q_1\gamma_\mu\gamma_5 q_2$) are probed as opposed to vector current ($q_1\gamma_\mu q_2$) in semileptonic case.

Vice versa, assuming unitarity of CKM triangle and experimental determination of relevant $V_{q_1 q_2}$ one can obtain experimental determination of the decay constants and compare it with theoretical prediction.

Last, but not least, is of course the measurement of presence of new physics in these decays. Especially appealing is the presence of new particles which would manifest themselves in the decay rates of heavier pseudoscalars ($D_{(s)}$ or B). Example of such new particles include charged Higgs bosons, H^\pm , coming from so-called Type II of two-Higgs-doublet models [7] [8] [9] or leptoquarks [9]. In this case, considering $B^+ \rightarrow l^+\nu_l$ decay, four-fermion interaction between W^\pm and H^\pm would modify the SM decay width [Equation 2.46](#) to

$$\Gamma(B^+ \rightarrow l^+\nu_l) = \frac{G_F^2 m_{B^+} m_{l^+}^2}{8\pi} \left[1 - \frac{m_{l^+}^2}{m_{B^+}^2} \right]^2 f_P^2 |V_{ub}|^2 \times r_H, \quad (2.48)$$

where

$$r_H = [1 - \tan^2 \beta(m_{B^+}^2/m_{H^\pm}^2)]^2. \quad (2.49)$$

Here $\tan \beta = \frac{v_2}{v_1}$, where v_i are the vacuum expectation values for the Higgs doublets. In order to have enhancing effect for the rate of $B^+ \rightarrow l^+\nu_l$ decay (to have $r_H > 1$), $\tan \beta/m_{H^\pm} > 0.27 \text{ GeV}^{-1}$. The experimental limit already present strong lower bound on $M_{H^\pm} > 600 \text{ GeV}$ [10] making most of the parameter space in $\tan \beta$ and m_{H^\pm} satisfying the enhancing condition to be above 0.27 GeV^{-1} .

Given current tensions arising in flavour physics searches, especially concerning lepton non-universality, ratio of rates between $P \rightarrow \tau\nu$, $P \rightarrow \mu\nu$ and $P \rightarrow e\nu$ should be

measured. In the ratios the decay constant f_P cancels out making such measurements a good tool for lepton universality tests.

As seen in [Equation 2.46](#), a purely leptonic final state going through $P \rightarrow W^* \rightarrow l\nu$ is suppressed by m_l^2 , also known as helicity suppression. This suppression occurs as a result of angular momentum conservation. In case of $B^+ \rightarrow l^+\nu$, the B^+ is a spin-0 particle and hence its decay products should have spin 0 combined, or in other words, be anti-aligned. Neutrinos in the [SM](#) are always produced left-handed. As the spin of the antilepton and the neutrino should be anti-aligned, the antilepton also needs to be left-handed (to have negative helicity). However, the weak current only couples to right-handed antiparticles. Therefore, the antilepton has to be boosted in order to have different helicity. For massless particles such a helicity flip is not possible making this decay impossible. The lighter the lepton the larger the velocity and hence higher boost is necessary, making decays to lighter leptons rarer even though they have bigger kinematic phase space available.

Concentraing on decays where pseudoscalar of interest is B^\pm , the latest experimental measurements for rates of $B^+ \rightarrow l\nu$ decays have been performed by B factories, finding evidence for $B^+ \rightarrow \tau^+\nu$ and first sign of $B^+ \rightarrow \mu^+\nu$ as seen in [Table 2.2](#). These results are to be compared with the [SM](#) prediction $\mathcal{B}(B^+ \rightarrow \tau^+\nu) = (0.82 + 0.03 - 0.02) \times 10^{-4}$ [5] which is obtained by using the unitarity-constrained V_{ub} value aggregated from other measurements and lattice calculations of f_B .

With helicity suppressed rates and very limited signature in the detector (one charged track for muons and electron, more charged tracks for taus, but also more missing energy depending on the reconstruction channel) searching for such decays is still very challenging. In order to make measurements of the same kind (CKM precision measurements, decay constants measurements, new physics searches), fully leptonic decays with photons can be considered.

Process	Experiment	Tag	\mathcal{B}	Significance (σ)
$B^+ \rightarrow \tau^+ \nu$	Belle [11]	Hadronic	$(0.72^{+0.27}_{-0.25} \pm 0.11) \times 10^{-4}$	3.0
$B^+ \rightarrow \tau^+ \nu$	Belle [12]	Semileptonic	$(1.25 \pm 0.28 \pm 0.27) \times 10^{-4}$	3.8
$B^+ \rightarrow \tau^+ \nu$	Belle [12]	Average	$(0.91 \pm 0.22) \times 10^{-4}$	4.6
$B^+ \rightarrow \tau^+ \nu$	BaBar [13]	Hadronic	$(1.83^{+0.53}_{-0.49} \pm 0.24) \times 10^{-4}$	3.8
$B^+ \rightarrow \tau^+ \nu$	BaBar [14]	Semileptonic	$(1.7 \pm 0.8 \pm 0.2) \times 10^{-4}$	2.3
$B^+ \rightarrow \tau^+ \nu$	BaBar [13]	Average	$(1.79 \pm 0.48) \times 10^{-4}$	-
$B^+ \rightarrow \mu^+ \nu$	Belle [15]	Untagged	$(6.46 \pm 2.22 \pm 1.60) \times 10^{-7}$	2.4

Table 2.2: Experimental summary of searches for $B^+ \rightarrow l^+ \nu_l$. Tag Hadronic/Semileptonic/Untagged refers to different way data is selected in Belle and BaBar factories.

2.9 Fully Leptonic $B^+ \rightarrow l^+ \nu_l \gamma$ decays

The helicity suppression can be lifted by considering the decay with an additional photon radiated from the B^+ meson, at the cost of the electromagnetic suppression with coupling constant α_{em} . Consequently, the branching fraction for radiative decays can be comparable or even larger than the corresponding fraction for purely leptonic decays. It has been shown that $R_B^\mu = \frac{\Gamma(B \rightarrow \mu \nu \gamma)}{\Gamma(B \rightarrow \mu \nu)} \approx (1 - 20)$ making $\mathcal{B}(B \rightarrow \mu \nu \gamma) \approx (10^{-7} - 10^{-6})$ [16].

As compared to photonless decays, the amplitude of the decay will have a contribution from both the axial-vector weak current as well as the vector current. The differential decay width with $\frac{1}{m_b}$ and radiative corrections at next-to-leading logarithmic order calculated in [17] is given by

$$\frac{d\Gamma}{dE_\gamma} = \frac{\alpha_{em} G_F^2 |V_{ub}|^2}{48\pi^2} m_B^4 (1 - x_\gamma) x_\gamma^3 [F_A^2 + F_V^2], \quad (2.50)$$

where $x_\gamma = 2E_\gamma/m_B$, F_A is the axial form factor and F_V is the vector form factor defined

as

$$F_V(E_\gamma) = \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + [\xi(E_\gamma) + \frac{Q_u m_B f_B}{(2E_\gamma)^2} + \frac{Q_b m_B f_B}{2E_\gamma m_b}], \quad (2.51)$$

$$F_A(E_\gamma) = \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + [\xi(E_\gamma) - \frac{Q_u m_B f_B}{(2E_\gamma)^2} - \frac{Q_b m_B f_B}{2E_\gamma m_b} + \frac{Q_l f_B}{E_\gamma}]. \quad (2.52)$$

Here Q_l, Q_u, Q_b are the charges of the lepton, up quark, and bottom quark, respectively, and $R(E_\gamma, \mu)$ is a radiative correction calculated at the energy scale μ and m_b is the mass of the b quark.

The first term in [Equation 2.51](#) and [Equation 2.52](#) represents the leading-power contribution in the heavy-quark expansion. Note that this term is the same for the vector and axial form factor. The second terms are $\frac{1}{m_b}$ power corrections relative to the leading term. Further corrections have been discussed in [\[18\]](#).

Recent measurement of the radiative $B^+ \rightarrow l^+ \nu_l \gamma$, where l^+ is either e^+ or μ^+ was performed by Belle using hadronic tagging on their full data sample [\[19\]](#). The search yielded $\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) < 3.4 \times 10^{-6}$ and $\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) < 6.1 \times 10^{-6}$.

2.10 Fully Leptonic $B^+ \rightarrow l^+ l^- l^+ \nu_l$ decays

In LHCb, the most optimal approach due to the detector capabilities is to measure this kind of decay by decaying the photon into pair of muons, see [Figure 2.3\(a\)](#). If the naive expectation of only taking into account photon conversion into two muons is adopted, then the expected branching fraction for this analysis is $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu) \approx 10^{-8}$. However, such estimate is not correct because there are other contributions to the total decay rate as shown in the first theoretical prediction for $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu)$ in [\[20\]](#) based on Vector Meson Dominance (VMD) model. This theoretical prediction yields $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu) \approx 1 \times 10^{-7}$ and the rest of this section is a short summary of this publication.

The VMD model was formulated to describe the interaction between photons and hadrons before quantum chromodynamics was formulated. It is an approximative

model where the photon is treated to be made of both purely electromagnetic component and vector meson component. This idea originates in the fact that both photon and vector mesons have the same quantum numbers $J^{PC} = 1^{- -}$ and if two particles have the same quantum numbers then they mix. Therefore, there is mixing between photons and vector mesons.

As mentioned previously, there are different contributions to the amplitude of the $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$. Using the VMD model, it is not surprising that the biggest contribution arises from the photon emission from the valence u -quark of the B meson. In this case, the contribution from the $\rho(770)$ and $\omega(782)$ resonances are included in the calculation. Secondly, the contribution of photon emission from the b -quark is studied, effectively creating excited B^+ , B^{*+} intermediate resonance state. Thirdly, the photon can be emitted from the final-state lepton, a process known as Bremsstrahlung. All these different contributions to the decay amplitude are shown in [Figure 2.3](#). To obtain the total amplitude, the sum of the matrix elements of the three contributions is calculated in the limit where m_l is set to zero.

In this publication the amplitude of $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$ is estimated by calculating $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- e^+ \nu)$ amplitude first and then adding a negative interference term that arises due to the identical fermions in the final state doubling the number of possible diagrams. The numerical calculation yields $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- e^+ \nu) \approx 1.3 \times 10^{-7}$ and $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) \approx 1.0 \times 10^{-7}$.

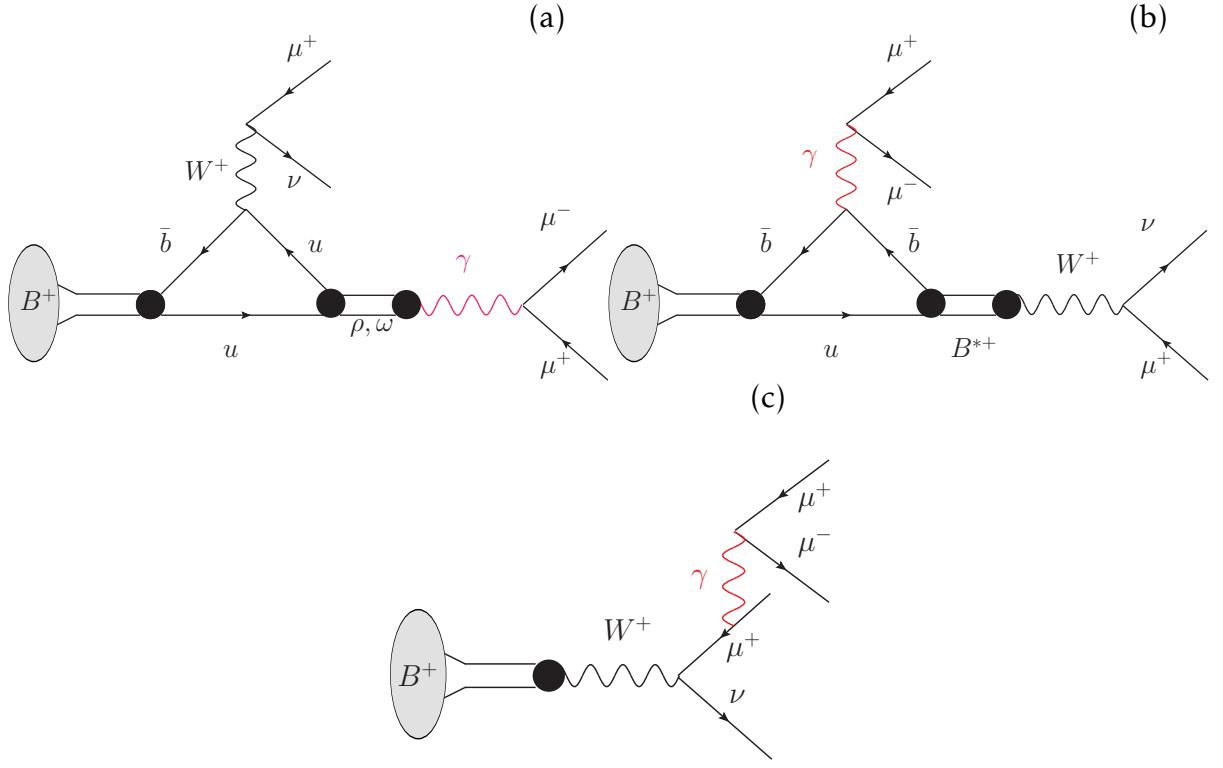


Figure 2.3: (a) Annihilation diagram where the initial u -quark state radiates off a virtual photon which decays into a pair of muons and the W^+ decays into a muon and muon neutrino. Most of the contribution to the rate comes from hadronic contribution to the photon. (b) Photon emission from b -quark and (c) finally emission from the final state muon.

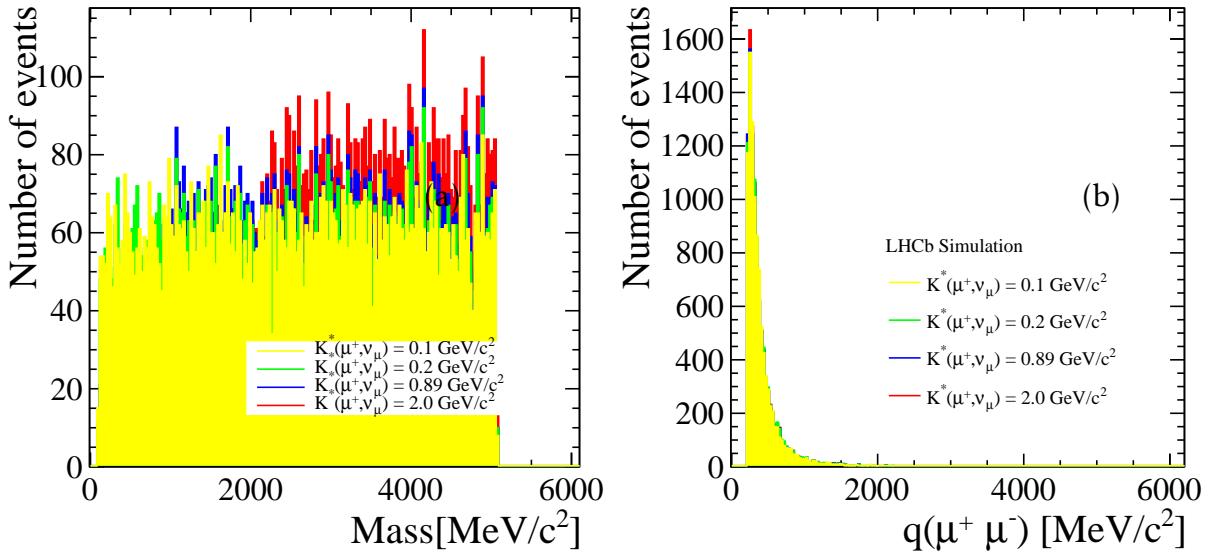
2.11 The $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay model

As the search for the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay is the first of its kind, a simulation that describes this type of decay was not available.

For any decay, it is possible to use phase space model, *PHSP*, which only takes into account the kinematic constraints of the decay without taking into account any input from theoretical considerations as the matrix element is constant. This is not satisfactory for decays where there are intermediate virtual photons or vector meson resonances.

The following decay model is developed to take into account a more correct behaviour of the decay as shown in [Figure 2.3](#).

The decay proceeds through a virtual W decaying to $\mu^+\nu$ and a virtual photon decaying to a muon pair. This has similar structure to $B^+ \rightarrow (K^{*+})\mu^+\mu^-$ decay, where the K^{*+} can take the role of the virtual W decay. By using the *BTOSLLBALL* model [21], traditionally used for $B^+ \rightarrow (K^{*+})l^+l^-$ decays, but modifying properties of the K^{*+} to those of virtual W (having mass of $0.1 \text{ GeV}/c^2$ and width 50 GeV), it is possible to obtain a good approximation to the correct features of the decay. This is visible in [Figure 5.2](#), where there is a characteristic photon pole for low $q(\mu^+, \mu^-)$, invariant mass of the opposite muon pair, and flat distribution for $K^*(\mu^+, \nu_\mu)$, invariant mass of the muon and neutrino pair. This decay model will be further referred to as SALLY's ([can I do this?](#)) model.



[Figure 2.4](#): Distributions for signal simulation. (a) $K^*(\mu^+, \nu_\mu)$ (b) $q(\mu^+, \mu^-)$ distributions under different K^* mass hypotheses. The most flat distribution in $K^*(\mu^+, \nu_\mu)$ is plotted in yellow.

Chapter 3

The LHCb detector

In this section, an overview of the accelerator complex at CERN as well as the physics motivation behind the LHCb detector and its design will be described.

CERN has built one of the most exciting laboratories to study elementary particle interactions. Its complex set of particle accelerators and detectors is shown in Figure 3.1. The process of accelerating protons starts with the source of protons. Protons are obtained from a hydrogen gas bottle by applying an electric field separating hydrogen into protons and electrons. The first proton accelerator in the chain, Linac 2, accelerates the protons to the energy of 50 MeV. Linac 2 is a tank composed of several chambers where the resonant cavities are tuned to a specific frequency creating potential differences in them, which then make the protons accelerate. The protons are then injected into the Proton Synchrotron Booster (PSB), where they are accelerated further to 1.4 GeV. The next in line is the Proton Synchrotron (PS) reaching energy of 25 GeV. Before either entering the Large Hadron Collider (LHC) or North Area (mainly used as testing facility for experiment upgrades) the Super Proton Synchrotron (SPS) is the last accelerator in the chain. Here proton acceleration to 450 GeV is achieved.

LHC is a complex machine which accelerates beams of protons in opposite directions in $\sim 27\text{km}$ long circular tunnel. It is located 50-157 m below ground on the border of Switzerland and France. Once the desired energy is achieved proton-proton (pp) or ion collisions happen at four distinct points, where different detectors with different

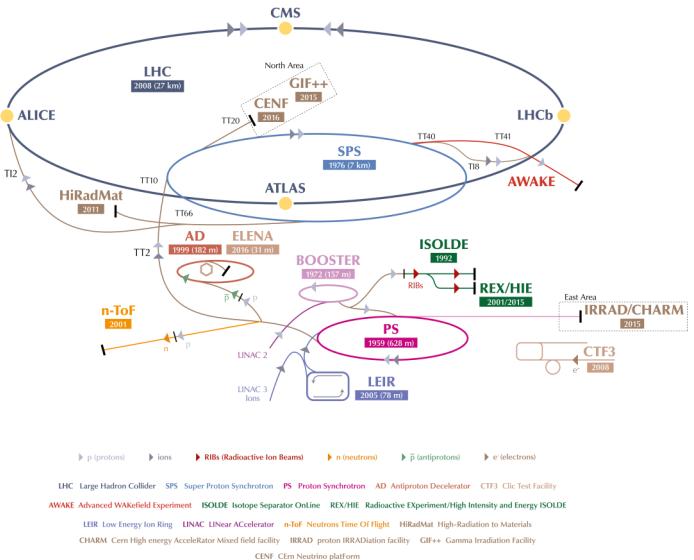


Figure 3.1: Accelerator complex at CERN. The image is taken from [22].

physics focus are located. These are [ATLAS](#), [CMS](#), [ALICE](#) and [LHCb](#). The search for the decay $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ was performed using data obtained at [LHCb](#).

3.1 LHCb Layout

[LHCb](#), seen in [Figure 3.2](#), differs from the other general purpose detectors on the [LHC](#) ring as its main aim is to study properties of heavy particles containing b or c quarks. This is possible as this experiments was designed to have the geometrical acceptance and unique vertex resolution as well as excellent particle identification (PID) suitable for beautiful and charming physics.

Studies of B mesons can happen either at positron-electron colliders or at hadron colliders. The advantage of positron-electron collider is that the information about all the event is known, as the collision point is in the centre of the detector that surrounds it. This gives an overall constraint on collision information, unlike in hadron collider B factory [LHCb](#). Contrary to the two general purpose detectors in hadron collider,

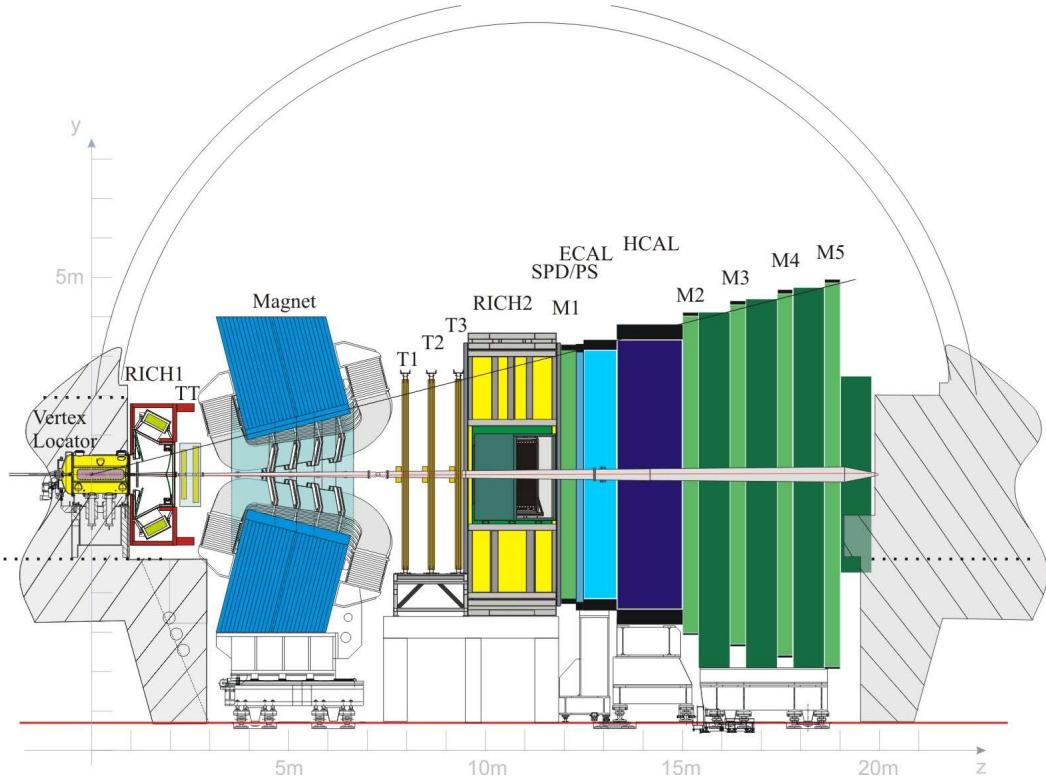


Figure 3.2: Schematic slice of [LHCb](#) detector in y,z plane where z is defined to be the direction parallel to beamline, and x,y define the plane perpendicular to the beamline. θ , the opening angle in y - z plane with $\theta = 0$ along z -axis. Figure from [23].

where the collisions are occurring in the centre of the detector, [LHCb](#)'s collision point is located at one end of the detector, hence its description as a forward single-arm spectrometer.

This disadvantage is, however, compensated by production mechanism of $b\bar{b}$ and $c\bar{c}$ in pp interactions, which occurs predominantly via gluon-gluon fusion. In this process, each gluon will carry part of proton's momentum. If the two gluons from two protons carry significantly different momentum, the $b\bar{b}$ system will be boosted with respect to the pp rest frame, either in forward or backward cone closely to the beamline, as can be seen in [Figure 3.3](#) (Left).

The angular coverage of [LHCb](#) is formally defined using pseudorapidity η ,

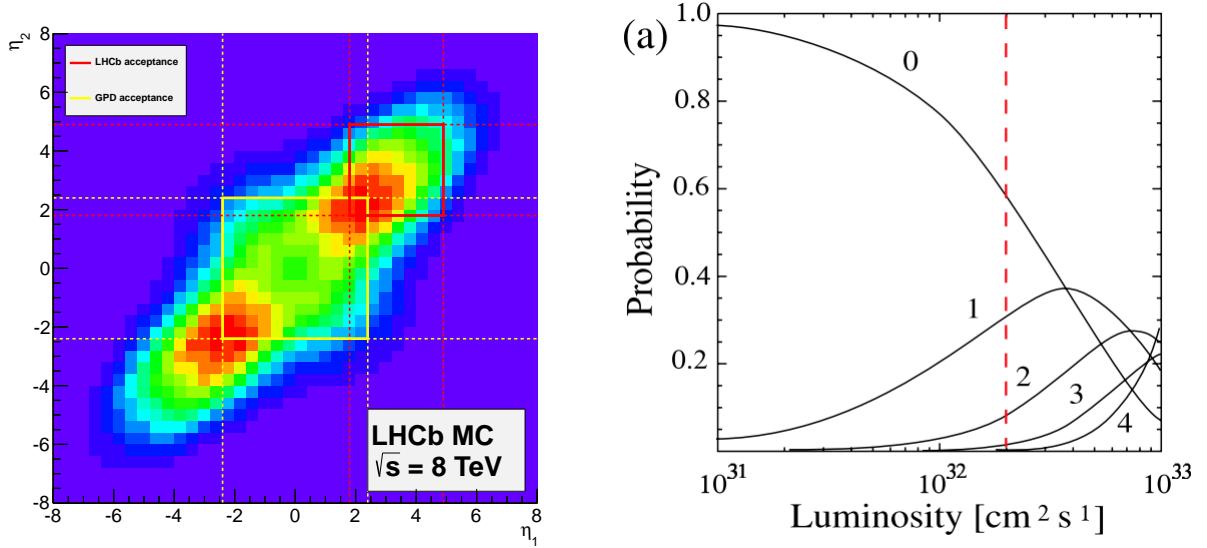


Figure 3.3: (Left) Angular production and acceptance of the b (x-axis) \bar{b} (y-axis) pair produced from pp collision at the LHC. The acceptance of the LHCb detector is the red box and the acceptance of the General Purpose Detector is shown in the yellow box. LHCb covers the region with highest production cross-section at 8 TeV. These plots were produced using PYTHIA8 [24] simulation. Figure from [25].(Right) Probability of interaction per bunch crossing as a function of instantaneous luminosity. Figure from [26].

$$\eta = -\ln\left(\tan \frac{\theta}{2}\right) \quad (3.1)$$

where θ is the polar angle measured from the beam axis. The LHCb detector was built to cover the region $2 < \eta < 5$. The production cross-section of the fundamental process of $pp \rightarrow b\bar{b}X$ was measured in this region yielding, $\sigma(pp \rightarrow b\bar{b}X) = 75.3 \pm 5.4 \pm 13.0 \text{ } \mu\text{b}$ at 7 TeV [27] and $144 \pm 1 \pm 21 \text{ } \mu\text{b}$ at 13 TeV [28], which shows that the production cross-sections scales roughly linearly with the centre-of-mass energy. Assuming design conditions of LHCb, listed in Table 3.1, 2 fb^{-1} of data (equivalent to 2012 dataset) would correspond to 10^{12} of $b\bar{b}$ pairs being produced in a full 4π region. 27% of these $b\bar{b}$ pairs are produced in the LHCb's acceptance.

Despite the impressive statistics of $b\bar{b}$ pairs available to [LHCb](#), the bottleneck arises from the much more copious inelastic background. It mostly originates from soft QCD processes which are related to the amount of pile-up, the visible number of pp interaction in the visible events. By looking at the probability of the number of pp interaction per bunch crossing as a function of luminosity, shown in [Figure 3.3](#), it can be noted that the maximum probability for only one pp interaction (and hence minimizing the background) is found to be at $\sim 2 \times 10^{32} cm^{-2}s^{-1}$. This was the reason behind the [LHCb](#)'s design luminosity.[just trying to explain why we wanted to run at the lumi of 2, is it better to discuss why we run at 4?](#) In addition, to keeping the occupancy of the detector reasonable for physics analyses, global event cuts, GECs, on the occupancy variables are put in the place. Only events with 600 (in 7,8 TeV) and 450 (in 13 TeV) hits and less, corresponding to the track density in the particular part of the detector, are allowed to be processed. As the majority of the branching fractions measurements at [LHCb](#) are measured with respect other branching fraction, so there is no bias being introduced by the GECs.

As [LHCb](#) requires much lower luminosity compared to other [LHC](#) detectors, there is an [LHCb](#)-specific control of luminosity known as *luminosity levelling*. This procedure achieves stable instantaneous luminosity by controlling that the two beams do not collide straight head-on at collision point, but are moved with respect to each other. It limits the effects of luminosity decay, which can lead to trigger alterations during specific data taking run, resulting in systematic uncertainties.

So far, the detector has been running since 2010 collecting data corresponding to integrated luminosity as summarized in [Table 3.1](#). As compared to [ATLAS](#) and [CMS](#) the integrated luminosity is much lower, due to allowed luminosity conditions. In 2017, most of the data was taken at $\sqrt{s} = 13$ TeV, with small luminosity collected at $\sqrt{s} = 5$ TeV mainly to fuel studies involving productions comparisons at different energies. The Run I data-taking (2010-2012) was paused by Long Shutdown 1 ([LS1](#)) and followed with the Run II data-taking period (2015-2018).

In the following sections, a brief discussion of the different subdetectors, shown

year	\sqrt{s} [TeV]	$\mathcal{L} [\times 10^{32} \text{cm}^{-2}\text{s}^{-1}]$	Integrated Recorded Luminosity [fb $^{-1}$]
Design	Up to 14	2	-
2011	7	$\sim 3.0\text{-}3.5$	1.1
2012	8	~ 4.0	2.1
2015	13	$\sim 0.5\text{-}4.5$	0.3
2016	13	~ 4.0	1.7
2017	13	$\sim 4.0\text{-}6.0$	1.7

Table 3.1: Running conditions of [LHC](#) and [LHCb](#) in different years of data-taking. The statistics of [LHCb](#)'s instantaneous luminosity, \mathcal{L} is extracted using run database information.

in [Figure 3.2](#), is presented. The vertexing at [LHCb](#) is performed with vertex locator system, also known as VELO, is described in [section 3.2](#). The tracking system at [LHCb](#) consisting of trackers before magnet (TT), and three tracking stations behind the magnet (T1, T2, T3) are highlighted in [section 3.3](#). The particle identification is provided by two Ring Imaging Cherenkov counters (RICH1 and RICH2), which are detailed in [section 3.4](#). No particle physics experiment is complete without calorimeter system, discussed in which consists of a Scintillator Pad Detector and Preshower (SPD/PS), an electromagnetic calorimeter (ECAL) and finally a hadronic calorimeter (HCAL). The muon system positioned at the end of the detector, consisting from five muon chambers are described in [section 3.6](#). The trigger chain as well as the simulation chain are discussed in [section 3.8](#),[section 3.9](#). Particular emphasis is given to the muon detectors and the simulation of [LHCb](#).

Sally corrected until now .

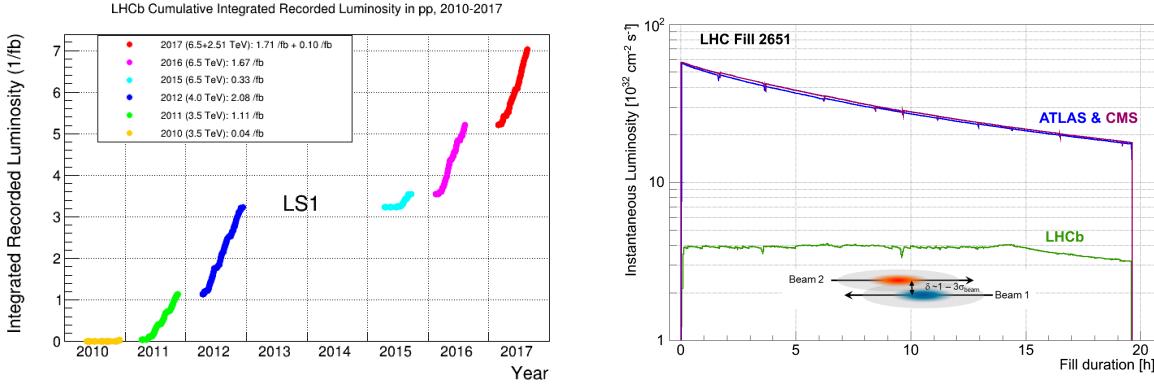


Figure 3.4: Integrated luminosity collected in different years of data-taking. This plot is taken from [29] (left). Development of the instantaneous luminosity for **ATLAS**, **CMS** and **LHCb** during LHC fill 2651. After ramping to the desired value of $4 \times 10^{32} \text{cm}^{-2} \text{s}^{-1}$ for LHCb, the luminosity is kept stable in a range of 5% for about 15 hours by adjusting the transversal beam overlap. The difference in luminosity towards the end of the fill between **ATLAS**, **CMS** and **LHCb** is due to the difference in the final focusing at the collision points, commonly referred to as the beta function, β^* . This plot was obtained from [30] (right).

3.2 VErtex LOcator

The closest detector around the collision point is VErtex LOcator (**VELO**). This silicon-strip based detector, that extends 1 m along the beam axis, is primarily used to distinguish signal-like events from prompt background. The typical differing property of a B hadron decay includes large impact parameter (**IP**), the minimal distance between the track and primary vertex, in addition to significantly higher transverse momentum, p_T . Therefore, the main tasks of this subdetector include finding:

- primary vertices positions
- secondary vertices of short-lived particles (heavy quark hadrons)
- tracks that did NOT originate from primary vertex
- .

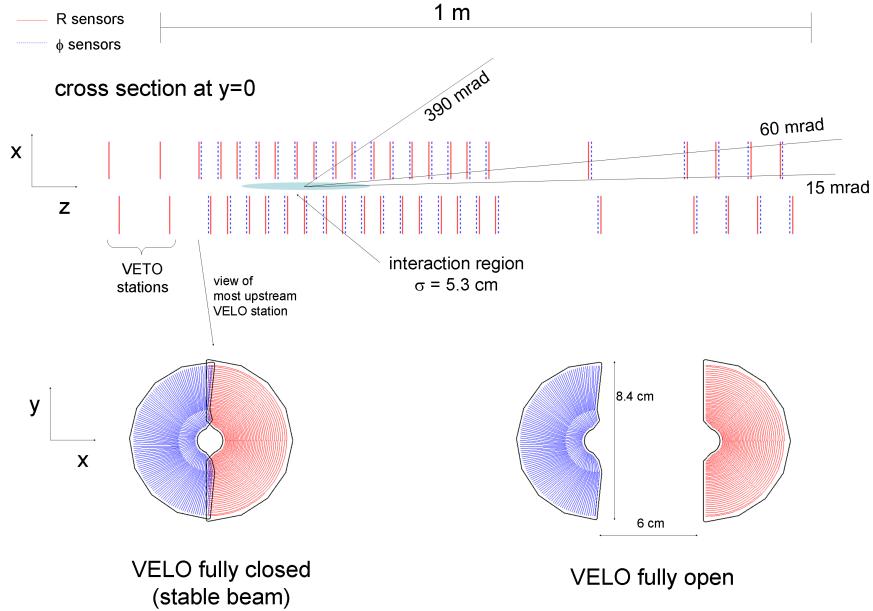


Figure 3.5: Schematic plot of **VELO** detector configuration along the beam pipe showing the layout as well as positions while in stable beams (discs have slight overlap) and injection. Figure taken from [31].

The detector consists of two sets of 21 silicon modules positioned around the beam pipe, where each module has 2 types of half-moon-shaped discs as seen in Figure 3.5. In the first type, the strips are arranged to provide radial information (R), whereas the second type provides azimuthal (ϕ) information. As pp interaction point brings high radiation dose for this detector, the first sensitive strip starts only at a distance of 8 mm once stable beams are declared. Throughout the beam injection, when the beam radius may be larger, the two sets are moved away 3 cm perpendicularly from the interaction point. For the R sensor, the individual module's strip pitch, distance between two strips, varies from 38 μm to 102 μm away from the beam pipe, so that the hit occupancy is roughly even as a function of distance away from the beam pipe. Each **VELO** half is kept within an aluminium welded box causing material overlap once stable beams are declared. These boxes then create their own vacuum which is different to the nominal

LHC vacuum in order to protect the detector from any electromagnetic interference with the beam.

This setup brings outstanding hit resolution ($4\text{-}40\,\mu\text{m}$), which in turn allows for very high IP and very good primary vertex (PV) resolution, as seen in Figure 3.6. This is indispensable not only in order to perform the precise measurements of B and D lifetimes, but also to resolve oscillations caused by $B_s^0 - \bar{B}_s^0$ mixing occurring at 3 trillion Hz rate.

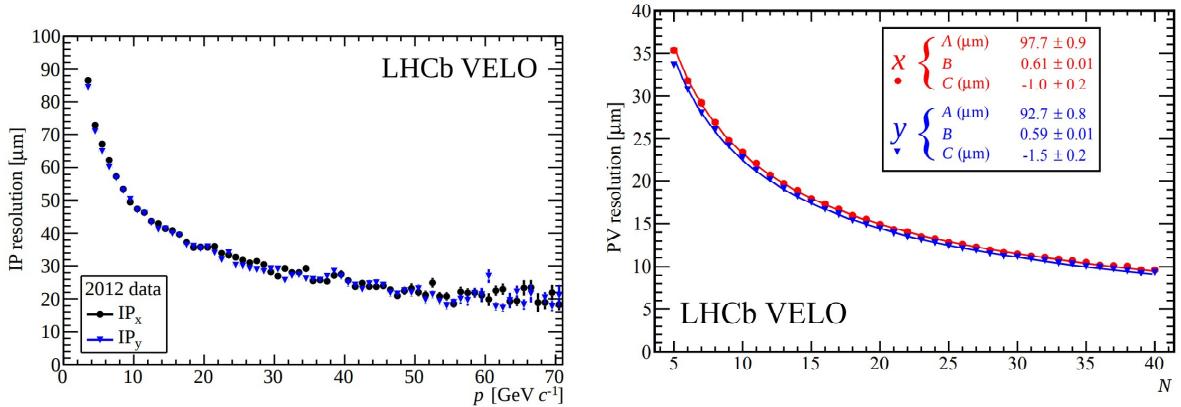


Figure 3.6: Two key variables which quantify performance of the VELO detector. IP resolution which is worse for low momentum tracks (left) and PV resolution dependent on the number of tracks forming the primary vertex N (right). Figures taken from [32].

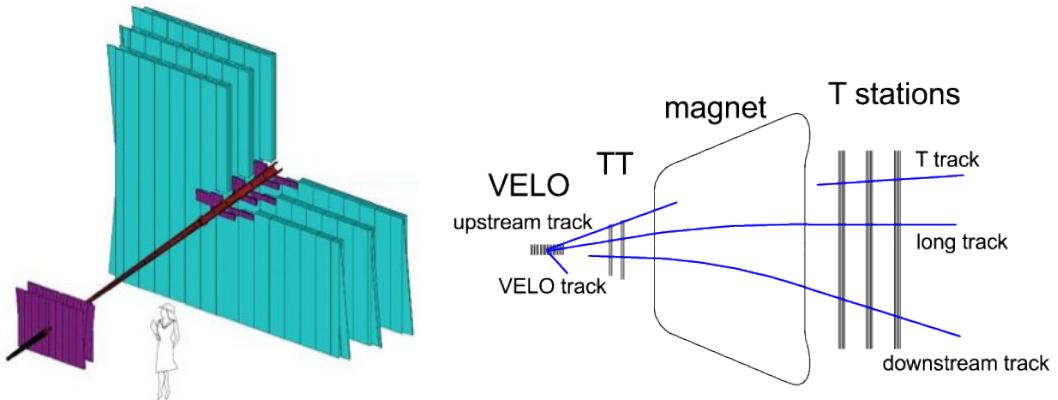
3.3 Tracking System

In addition to tracking information provided by VELO, the trajectories of charged particles are monitored by series of tracking subdetectors. The main task of these tracking subdetectors is to provide efficient reconstruction and precise measurement of particle's momentum. There are four tracking stations apart from VELO: Tracker Turicensis (TT), positioned upstream from magnet, and T1, T2 and T3 tracking stations on the other side from the magnet. The 10 m dipole magnet with ≈ 4 Tm integrated field provides enough strength to bend charged particles with p of $200\,\text{GeV}/c^2$.

Two different detection technologies are used in these trackers reflecting the nature of track occupancy as function of distance from beam pipe. The tracker's part close to the beam pipe, **TT** station together with central region of **T1**, **T2** and **T3**, also known as Inner Tracker (**IT**), expects higher occupancy and makes use of the silicon microstrip detection mechanism. The outer part of **T1**, **T2** and **T3** stations, also known as Outer Tracker (**OT**), is made of a straw-tube detectors. Straw-tube detector measures the hit position by measuring the drift-time of ionized electrons. Use of the two technologies are seen in [Figure 3.7](#).

3.3.1 Tracking Algorithms

Different particles will leave different footprint in the detector. Charged particles will form tracks. Depending on presence of hits in individual subdetectors, they are grouped into several categories, visualized in [Figure 3.7](#).



[Figure 3.7: Visualisation of use of different technology with silicon technology in violet and straw-tube technology in cyan. The Figure was obtained in \[33\]\(left\). Track types visualisation depending on which track stations provided hits. For the study of \$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu\$ decays only long tracks are considered as muons will travel to the end of the detector leaving the hits all along. Figure is taken from \[34\] \(right\).](#)

Most of the physics analyses use **long tracks**, tracks leaving hits in **VELO** and **T1**, **T2** and **T3**, as they give most precise momenta estimates. **VELO** tracks leave hits only

in R and Φ sensor, but not in any other tracking stations. VELO tracks are formed by particles which must have left [LHCb](#) acceptance or they come from particles produced backwards and hence are useful for [PV](#) reconstruction. Upstream tracks are formed by tracks leaving hits in [VELO](#) and [TT](#) only. These are usually low momentum particles, which are bent out [LHCb](#) acceptance while traversing the magnet. Long-lived particles such as Λ or K_s^0 will only decay outside of the [VELO](#) acceptance and hence will produce no hits until [TT](#) and [T1, T2 and T3](#) forming downstream tracks. T-track is track type that only have hits in [T1, T2 and T3](#). Again this could be due to presence of long-lived particles or due to secondary interactions in the detector.

In general, the track reconstruction software starts with *pattern recognition*, where several hits in one part of a tracking subdetector are identified and form *track seeds*, which are then extrapolated and combined with hits in other tracking subdetector provided this subdetector sits in low magnetic field. The long track candidates are formed and fitted with a Kalman filter [35], where because of the material present in the detector, corrections for energy losses as well as multiple scattering are incorporated.

Sometimes *pattern recognition* may combine random hits into a track, *ghost track*, or several tracks could be made out of same hits, *clone track*. Presence of these tracks are heavily suppressed with different techniques - such as establishing ghost probability (P_{ghost}) - variable based on the output of neural network combining track χ^2 , quality of the track, and missing hits in the subdetectors.

Uncertainty on mass is one of the crucial parameters to minimize as it provides opportunity for high precision measurements by better separations from backgrounds. It strongly correlates with momentum resolution that is obtained using tracking. Resulting relative momentum uncertainty (0.5-1.1%) on [long tracks](#) using $J/\psi \rightarrow \mu^+ \mu^-$ using *tag and probe* can be seen in [Figure 3.8](#). It varies logarithmically with increasing momentum.

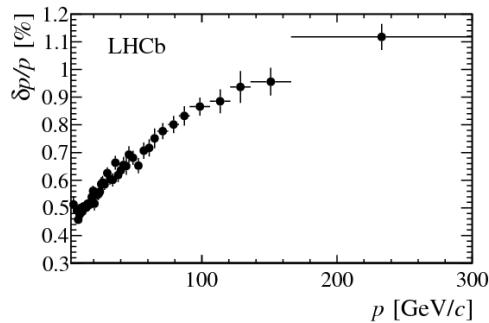


Figure 3.8: Momentum resolution of long tracks measured using *tag and probe* method at [LHCb](#). The decay channel $J/\psi^+ \rightarrow \mu^+\mu^-$ is analysed.

3.4 Ring Imaging Detectors

Particle identification, [PID](#), at [LHCb](#) relies heavily on two dedicated Ring Imaging subdetectors, [RICH](#). These detectors take advantage of the phenomena, emission of Cerenkov light, which happens when a charged particle travels through a medium at a speed faster than the phase velocity of light in that medium. This cone of light is emitted at an angle θ with respect to the charged particle's trajectory. Using the knowledge of refractive index of the medium, n , and momentum p that is measured using tracking, mass m of the particle can be obtained through:

$$\cos \theta_c = \frac{\sqrt{m^2 + p^2}}{pn}. \quad (3.2)$$

As the momentum and mass are intrinsic properties of passing particle, the momentum identification range is limited by the choice of medium, also known as radiator. For very low-momentum particle, as $\cos \theta_c \rightarrow 1$, the particle is not producing any Cerenkov light cone. At the very high momentum, as $\cos \theta_c \rightarrow 1/n$, there is saturation point as all species of particles will emit the light at the same Cherenkov angle, hence all the

discriminating power will be lost.

Low momentum (2-60 GeV) particles are identified in the upstream RICH1 detector and high momentum particles (15-100) GeV are analyzed downstream in RICH2. RICH1 covers $\pm 25\text{-}300$ mrad in x-z plane, ± 250 mrad in the x-y plane, using either gaseous Aerogel ($n = 1.03$) and C_4F_{10} ($n = 1.0014$) as radiators. RICH2 has more limited acceptance of $\pm 15\text{-}120$ mrad in x-y plane and ± 100 mrad in x-z plane and uses CF_4 as radiator, with lower $n = 1.0005$. The discrimination power between different particles can be seen Figure 3.9.

Both RICH1 and RICH2 use set of spherical primary mirrors to guide the photons onto the flat secondary mirrors which are then further focused into Cerenkov rings onto the surface of Hybrid Photon Multipliers, (HPD). The schematic view of a particle passing through RICH1 can be also be seen in Figure 3.9.

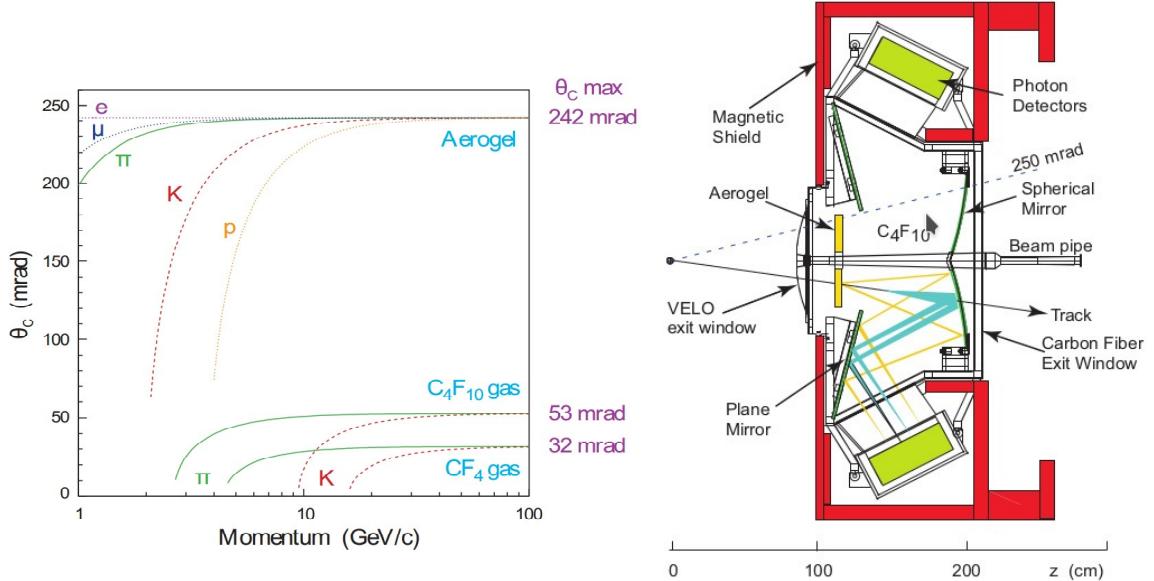


Figure 3.9: Separation power for different species of particles in momentum-Cerenkov angle plane (left). Schematic diagram of RICH1 layout (right). Both figures are taken from [31].

3.5 RICH Reconstruction and Performance

In order to establish species of particles for each track, the Cerenkov angle is combined with the track momentum measured by tracking. In practice, however, as **RICH** detectors operate in high track density environment, many Cerenkov rings will be overlapping and hence a complex pattern recognition algorithm is deployed [36].

For each event, the **RICH** computes full event likelihood that is consistent with assigning pion mass hypothesis for all tracks given the observed hit distribution read out by **HPD**. The algorithm then iterates through all other possible particle species, (e, μ, π, K , proton, deuteron), assigning new full event likelihood for a given track, having all other hypotheses fixed. The mass hypothesis with the highest full event likelihood is assigned to the track and this process is repeated for all the tracks in the event, until no improvement is found.

Results of this algorithm provide likelihood variables, $DLLx$, that quantify the strength of the chosen species hypothesis against pion hypothesis,

$$DLLx = \log(\mathcal{L})_x - \log(\mathcal{L})_\pi \quad x \in e, \mu, K, \text{proton}, \text{deuteron}. \quad (3.3)$$

By calculating $DLLx_1 - DLLx_2$, one can obtain discriminative strength between any two species.

3.5.1 RICH performance

In order to measure the performance of the **PID** computed by **RICH**, populous calibration samples with very little background contamination are required. In order not to bias results, these samples have no **PID** constraints themselves and are reconstructed solely using kinematic information. For studies of pion/kaon efficiencies, $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$ background-subtracted samples are used, whereby the daughter tracks of D^0 become proxies for evaluation. The probability of correctly identifying kaon given certain constraint on $DLLK$, identification efficiency (**ID**), and probability of mistakenly swapping pion identification, **MisID** efficiency, are summarized in

Figure 3.10. Identification probabilities of $\approx 85\%$ with misIDentification rate of $\approx 3\%$ provide invaluable discriminating separation between kaon and pion.

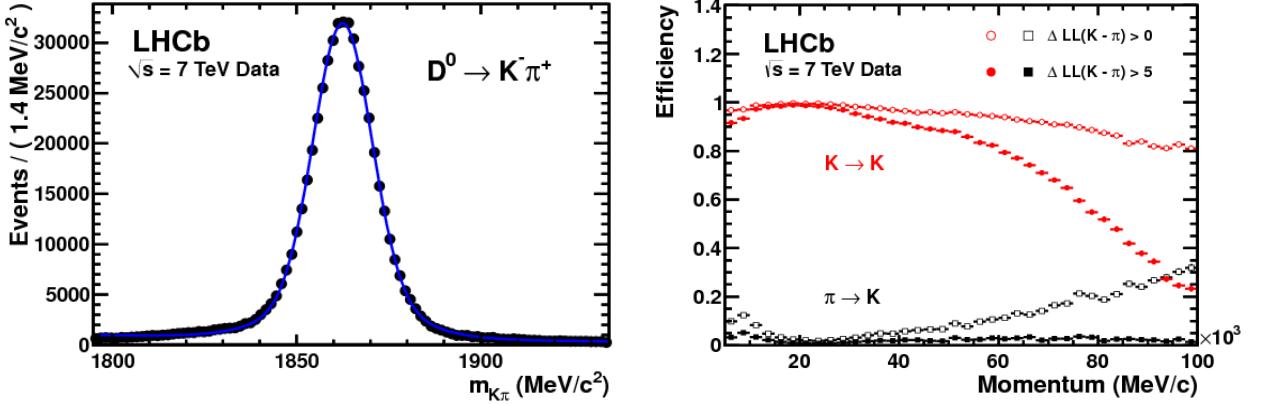


Figure 3.10: Invariant mass distribution of D^0 data sample (in black) overlaid with fit to both background and signal (in blue) (left). An example of kaon ID (red) and MisID (black) efficiency as a function of momentum under two PID hypotheses, $DLLK > 0$ (empty) and $DLLK > 5$ (filled) (right). Both figures are taken from [37].

In search for B^0 and B_s^0 decaying to h^+h^- , where $h \in K, \pi, \pi^+\pi^-$ invariant mass spectra with and without PID $DLLx$ requirements can be seen in Figure 3.11. These plots clearly demonstrate increase in sensitivity searching for $B^0 \rightarrow \pi^+\pi^-$ signal amongst other components.

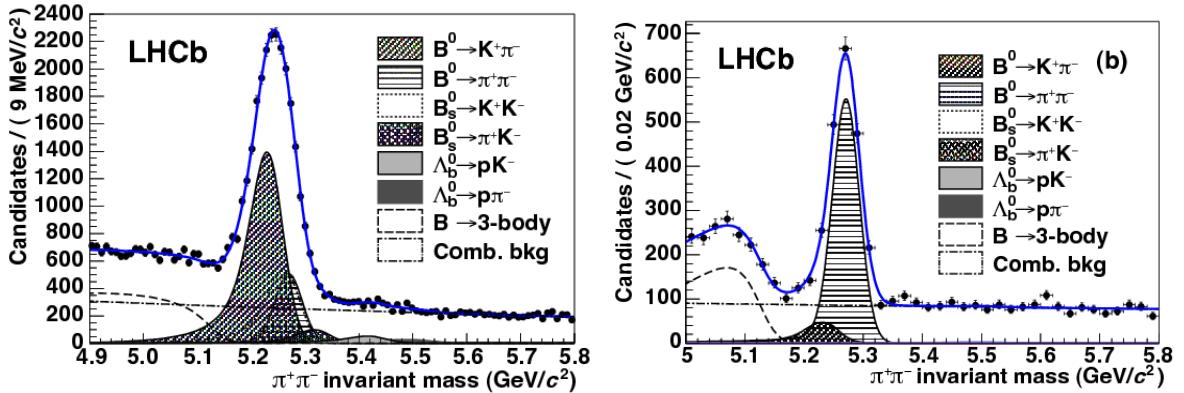


Figure 3.11: $\pi^+\pi^-$ invariant mass distributions obtained using kinematic constraints only (left) and also using PID constraints (right) in order to isolate $B^0 \rightarrow \pi^+\pi^-$ peak. This figure is taken from [38].

3.6 Calorimetry

As many other particle physics detectors, LHCb is equipped with series of subdetectors providing separation and PID tool for electrons, pions and photons. This separation is achieved because different particles interact at different distances, producing differently shaped showers of light. This part of detector is not only integral to the way LHCb trigger system works but it also provides precise measurement of energies of these objects. All the subcomponents discussed here operate on the same principle. The light from the scintillating material is guided to photomultiplier tubes by wavelength shifting fibres.

Electrons, pions and photons firstly encounter two planes of scintillating tiles: Scintillating Pad Detector (SPD), Preshower Detector (PRS) intersected by a wall of lead. The SPD senses the passage of charged particles whereas neutral particles will not be affected, distinguishing electrons from photons. The wall of lead initiates the electromagnetic shower, where photons are converted into electron-positron pairs, depositing sizable energy in the PRS allowing electron/pion separation.

The following Electromagnetic Calorimeter (ECAL) is based on sampling shashlik-

type technology, where scintillating tiles are alternated by lead plates measuring the energy deposit of electromagnetic showers. As the best energy resolution requires full energy deposit of energetic photons along the **ECAL**, the thickness is equivalent to 25 radiation lengths. The resulting resolution of **ECAL** is $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%$, where E is in GeV.

On the other hand, **HCAL** sandwiches iron instead of lead as the absorber with thickness of 5.6 interaction length only, achieving resolution of $\frac{\sigma_E}{E} = \frac{70\%}{\sqrt{E}} \oplus 10\%$ in beam tests. This poorer resolution however fulfils the requirements necessary for the main purpose of this detector, hadron trigger. Away from beampipe the granularity of cells is coarser to mirror the track occupancy as seen in [Figure 3.12](#).

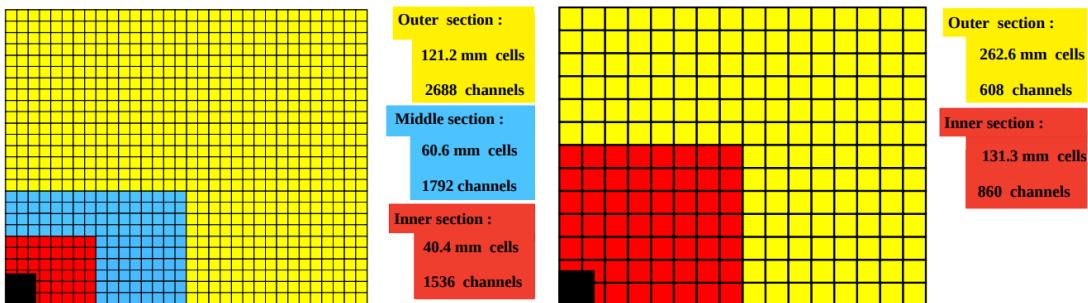


Figure 3.12: Granularity of **ECAL** (left) and **HCAL** (right) detectors. The figure was taken from [31].

3.7 Muon Stations

Muons are considered to be of fundamental importance to many flagship analyses by **LHCb**, such as the search for the rare $B_s^0 \rightarrow \mu^+ \mu^-$ decay . Analysis of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ of course relies heavily on good performance of this part of detector. Muon stations are positioned at the end of the detector, taking advantage of the low probability of interaction of muon previously in the detector.

LHCb's five rectangular muon stations M1-5 are positioned before and after calorime-

try system, with first station M1 upstream of the [SPD](#), and four stations (M2-M5) downstream of [HCAL](#) as shown in [Figure 3.13](#). The M1 station consists of 12 sets of three gas electron multiplier foils (triple-GEMs) in the region closest to the beam pipe, resisting the highest dose of radiation due to the highest particle flux. Its main use lies in improving the p_T resolution by $\approx 10\%$. M2-M5 station each consist of 276 multi-wire proportional chambers ([MWPCs](#)) filled with Ar – CO₂ – CF₄ gas mixture. They are interlayered with 0.8 m iron walls, to provide stopping target to all particles, other than muons with momentum higher than 6 GeV/c. In order to ease the accessibility, like in [VELO](#), all the stations are split into two independent mechanical sides, also known A and C side.

Each station is then further segmented into four increasingly larger regions away from the beam, R1 to R4. All the regions were constructed to cover the same acceptance, keeping the track occupancy constant across the station. The granularity of the readout is higher in the horizontal plane to take advantage of magnet's horizontal bending plane.

Both GEM and [MWPCs](#) operate on a same principle. In each station, position in the $x - y$ plane is determined by ionizing electrons that come from muons passing through the detector, which are then attracted either to the closest anode mesh or wire mesh. The trigger is fired if the corresponding rectangular region in each station registered positive binary decision. This means the efficiency of each station must be $\geq 99\%$ to give overall 95% trigger efficiency. Geometrical layout covers $\approx 20\%$ muons originating in semileptonic b decays.

3.7.1 Muon Identification

Apart from triggering events with high enough p_T muons, muon stations provide necessary PID information for muon analyses. Offline variables mostly used for muon ID by analysts are

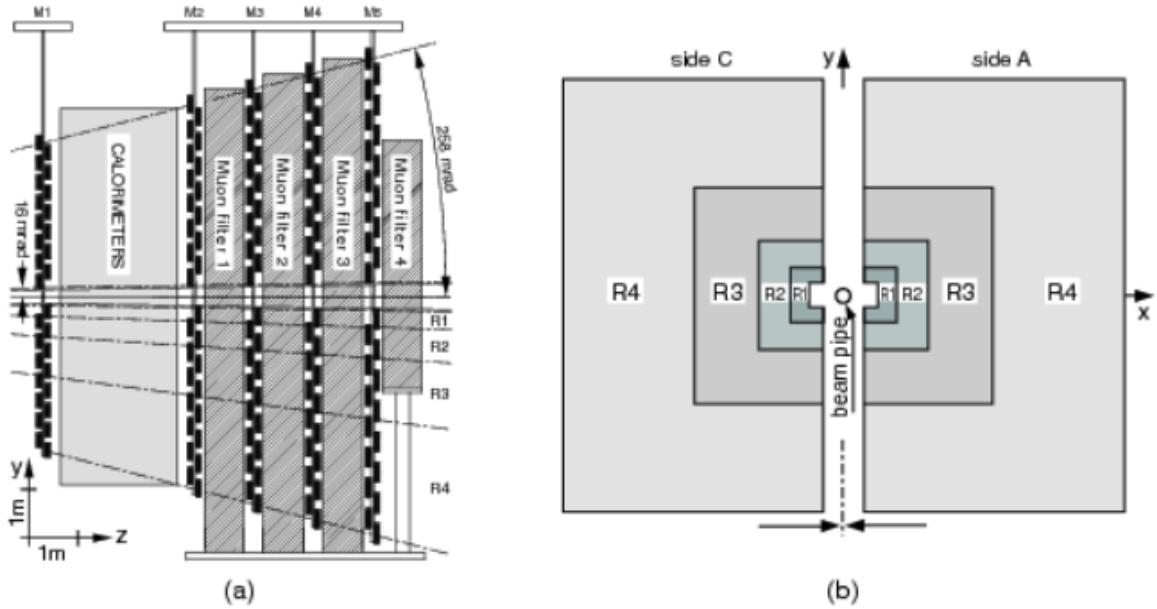


Figure 3.13: (a) Layout of the muon detector x-z plane and (b) x-y plane. This figure is taken from [39].

- **IsMuon:** Boolean decision of muon candidates with momentum-dependent categorisation. Long tracks with $p > 3 \text{ GeV}/c$ are extrapolated to muon stations yielding $x - y$ coordinates in $M2 - M5$, considering only tracks within acceptance. For each station, search for the hit information within elliptical area defined by momentum, field of interest (FOI), is performed. The hit requirements are summarized in [Table 3.2](#).
- **muDLL:** Difference in log likelihoods computed using muon and non-muon hypothesis. These hypotheses are based on the proximity/distance D^2 of the track extrapolation into the muon stations and corresponding closest sensed hits in those stations. Muon-like particle will tend to have sharper distribution in D^2 as compared to other species. Protons were chosen to be the other species for the calibration purposes. They give broader distribution as they originate either as punch-through protons (protons coming from showers not fully contained in

[HCAL](#)), protons having coincident hit position to true muon, and random hits.

- **DLLmu:** For each track global likelihood is produced, by combination of muon and non-muon likelihood from [muDLL](#), with the [RICH](#) different mass hypothesis likelihoods, and calorimetry likelihood exploiting the energy deposits information. Like in [RICH](#) likelihoods, the default hypothesis corresponds to separation between muon and pion hypothesis.

Particle Momentum p	Hits in Muon Stations
$3 \text{ GeV}/c < p < 6 \text{ GeV}/c$	M1 & M2
$6 \text{ GeV}/c < p < 10 \text{ GeV}/c$	M1 & M2 & (M3 M4)
$10 \text{ GeV}/c < p$	M1, M2, M3 and M4

Table 3.2: Momentum-dependent definition [IsMuon](#) variable.

3.7.2 Muon Performance

As in hadron performance measurements, muon ID is determined using high statistics decay channel $J/\psi \rightarrow \mu^+ \mu^-$ using *tag and probe* method. MisID rates of kaon and pion are computed using the same decay channels, which were used for identification of hadrons, $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$. The summary of [IsMuon](#) [ID](#) and misID rates are presented in [Figure 3.14](#). Very high ID rate (above 90%) for relatively low misID probability (below 10%) is key to analyses with muons in a final state. But the least performing are the low p_T muons where the identification suffers because these muons can end up outside of the [LHCb](#) acceptance and misID rates for kaon and pions are significantly higher in low momenta region as the dominant process causing this are prompt muons from decay-in-flight.

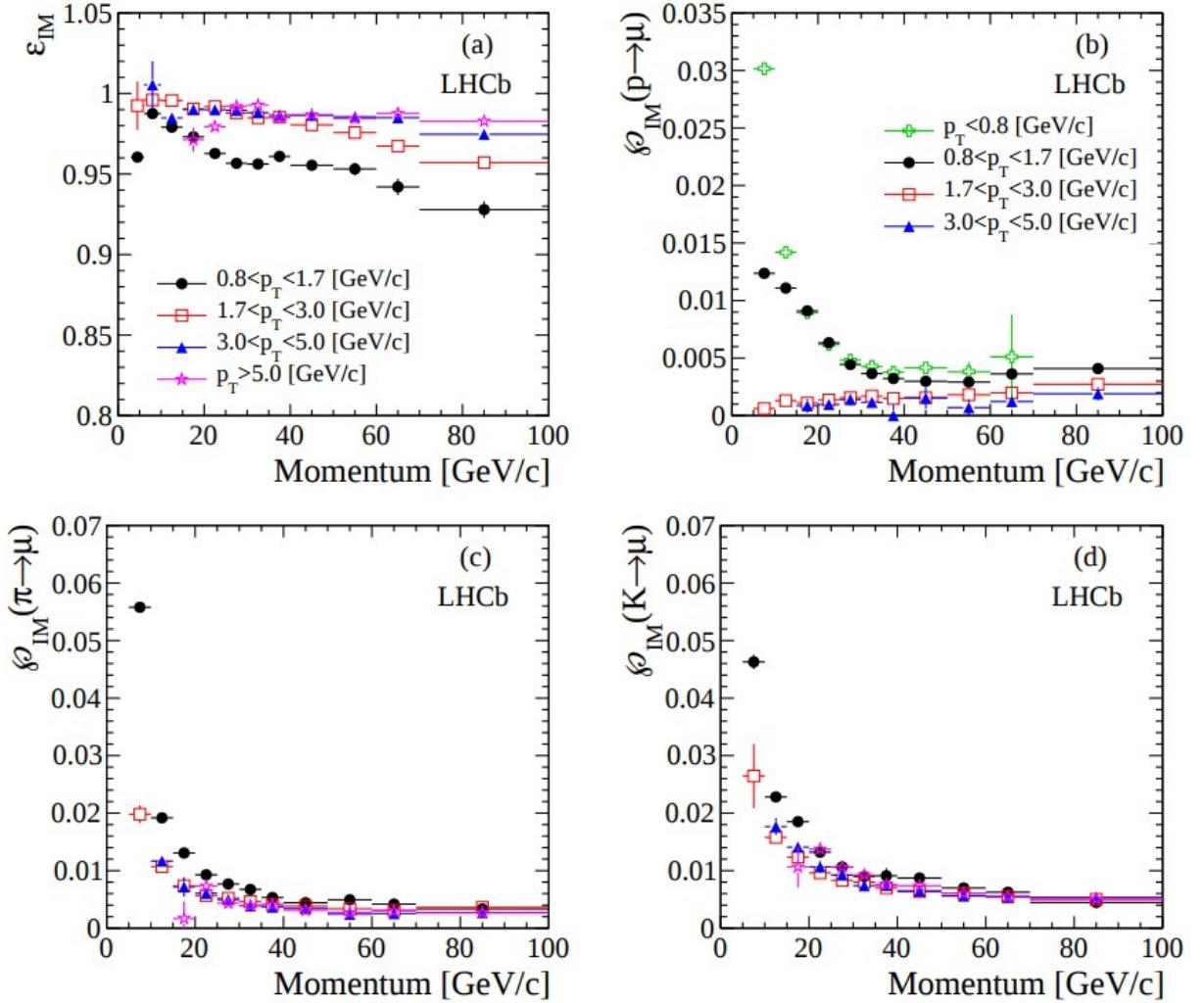


Figure 3.14: (a) Probability of correctly identifying muons as a function of momentum p in the bins of p_T for $J/\psi \rightarrow \mu^+ \mu^-$ with IsMuon constraint. (c) Probability of incorrectly identifying pion (b) proton and (d) kaon as muon with IsMuon. This figure is taken from [40].

3.8 Trigger

Nowadays, big-data physics experiments have to make decisions on what kind of data they want to keep. The choice of interesting events is performed by series of decisions, which is cumulatively known as trigger. **LHCb** trigger system was build

around constraints posed by the run conditions, read-out capabilities and available disk space. In Run I and Run II LHCb has at its disposal the multistage trigger consisting of hardware-based level 0 trigger (**L0**) and software-based high level trigger (**HLT**).

In the end, selected events have their trigger decisions categorized by different type. An event where signal candidate caused the trigger to fire is known to be Trigger on Signal (**TOS**). An event where it is non-signal like particle causing the trigger decision to occur, Trigger Independent of Signal (**TIS**) is used. Finally, if only by combination of signal particle together with other particle's properties in the event produce affirmative decision, then these events are categorized as **TIS & TOS = TISTOS**.

L0 reduces the rate of data from 40 MHz to 1 MHz by employing five trigger decisions, also known as lines. First three lines make decision using calorimeter information about the transverse energy, E_T , whether it is photon, electron or hadron causing the shower energy deposit. Two other lines are reading out information from the muon system by looking for transverse momentum, p_T , of muon and dimuon (two muon tracks) objects. Efficiencies of L0 muon triggers are evaluated using $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ decays. Hadron trigger efficiency in different decay channels can be seen in [Figure 3.15](#).

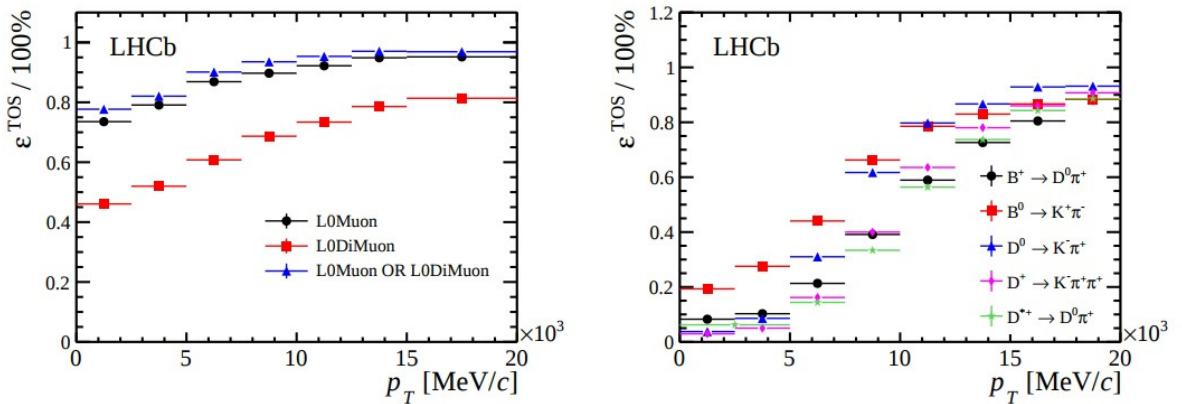


Figure 3.15: **TOS** efficiency as a function of p_T for muon-based decisions (left). **TOS** efficiency for different decays using L0 hadron trigger lines. This figure is taken from [\[39\]](#).

Software-based **HLT** then further reduces the rate from 1 MHz down to 40 – 80 kHz which can be safely stored to disk. The first stage of the **HLT**, (**HLT1**), performs limited track reconstruction and hence makes decision based on the presence of charged tracks in the event. **HLT1** uses **VELO** hits to reconstructs **PVs** and **VELO** tracks by using 3D pattern recognition. As **LHCb**'s primary mission is to study decays of hadrons containing b and c quark, **HLT1** will make decision based on the track segments being displaced (having high **IP**) with respect to the **PV**. For events selected by the **L0Muon**, an attempt is made to match the **VELO** tracks to hits observed in the vertical plane in the muon chambers due to magnet bending plane. By computing the track χ^2 , the potential muon track candidates are selected. Finally, the **VELO** tracks and muon tracks are extrapolated into the **OT** or **IT** trackers, allowing for so called *forward tracking*, whereby p and p_T requirements are imposed to reduce processing time. Each track is then fitted with fast Kalman filter providing the χ^2 of the fit. The corresponding performance of **HLT1** trigger lines are shown in Figure 3.16.

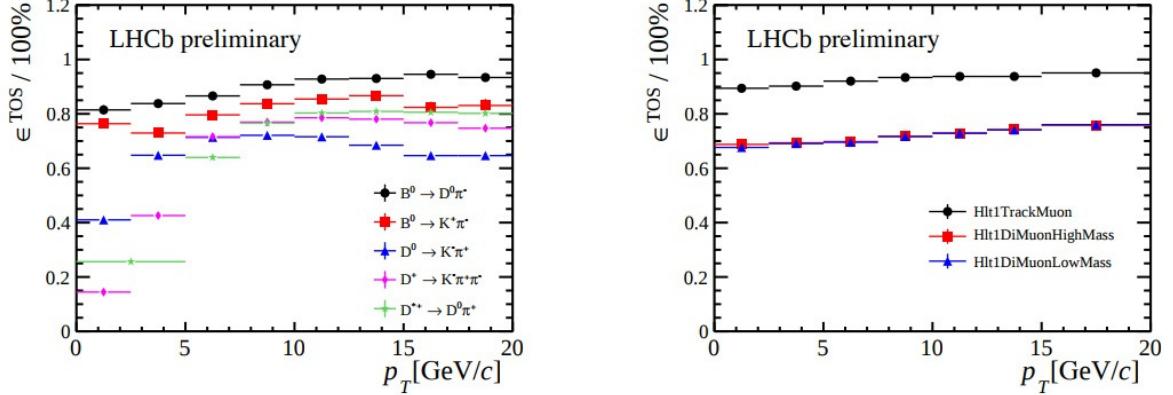


Figure 3.16: **HLT1** efficiencies of the corresponding triggers using the same proxy as in Figure 3.15. This figure is taken from [41].

The second stage **HLT2** reduces the rate to 5 kHz that can be safely written to disk. **HLT2** consists of a series of decisions based on a full reconstruction of either groups of decays or specific decay modes. *Topological triggers* exploit the vertex and track

information (topology) of b -hadron decays. By employing multivariate techniques 2,3 or 4-body decays away from PV are reconstructed. To account for decays where final state particle is not fully reconstructed, corrected mass serves as an input variable in the the BDT. Dedicated lines are also written to reconstruct muon and dimuon channels allowing for both prompt J/ψ and $B \rightarrow J/\psi X$ studies. Finally *Exclusive triggers* concentrating on selecting events with $c\bar{c}$ do selection very similar to the offline selection but without PID cuts and with *prescales*, only allowed in a certain fraction of events, is applied.

Between the Run I and Run II period there has been a change in how the software trigger operates, which can be seen in Figure 3.17. As more timing budget was introduced for both HLT1 and HLT2, LHCb took advantage in upgrading the trigger system by introducing update of calibration and alignment constants of the relevant subdetectors before the data is sent to permanent disk. *Online reconstruction*, defined as being produced at trigger farm, became the same as the *offline reconstruction*, defined as reconstruction made when data reached the permanent disk. Hence, there is enhancement of available information, such as the PID in the HLT, which can be then used at the trigger level. (Shall I Mention Turbo?).

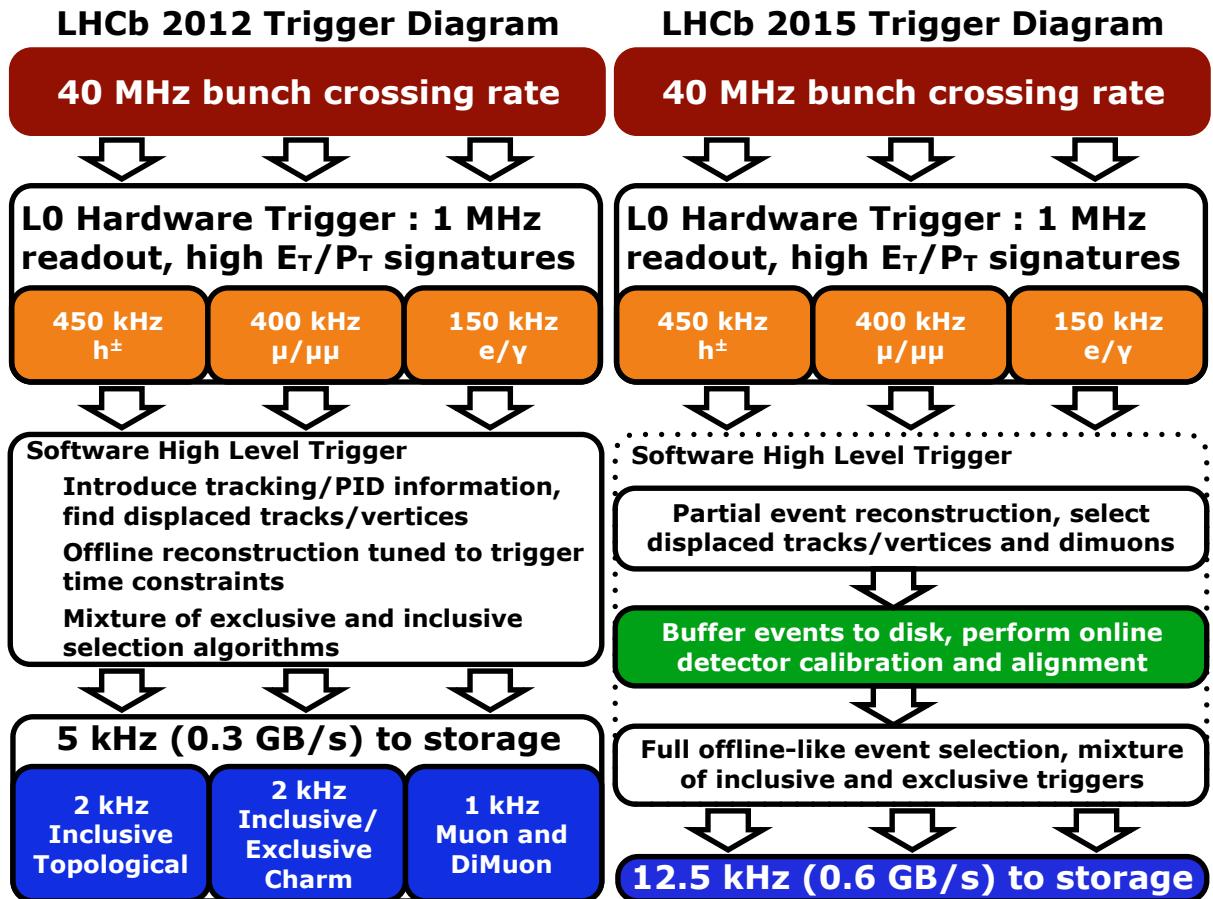


Figure 3.17: Trigger scheme differences between Run I and Run II. Figures obtained from [42]

3.9 Simulation

In order to optimise the event selections, extract efficiencies and model the backgrounds, a full Monte Carlo Simulation MC can be produced starting from simulation of the pp collision to detector readout of the decay of interest produced. The pp collisions within LHCb configuration [43] are simulated with Pythia 6.4 [44] and Pythia 8.1 [24]. LHCb specific settings are mostly related to running conditions: luminosity, number of collisions per bunch crossing as well as contamination from other bunches, *spill-over*.

In the pp collision, the b and c production mechanisms are simulated and then the following $b\bar{b}$ or $c\bar{c}$ pair is hadronized into hadrons of interest. In this thesis and the

analysis presented, B^+ is the hadron of interest. Hadrons are then further decayed using EVTGEN [45] into the chosen decay products. In this stage, different physics models or inputs from theory can be configured. At the same time some initial CPU-friendly selection is established, usually requiring the hadrons to be contained within the forward detector's acceptance. In order to account for the effects of QED radiative corrections, PHOTOS [46] algorithm can be used. All of this combined establishes *generator-level simulation* of LHCb.

In the next phase, *detector simulation*, the interactions of all the particles with the detector, transport, as well as detector's response are simulated using the C++ GEANT4 toolkit [47], [48]. LHCb's interface to GEANT4 is detailed in [49].

3.9.1 Differences in Simulation And Data

ULRIK - not fully finished

Despite the complexity and best intention of the LHCb simulation, there are several shortcomings that require correction treatment. The most affected variables necessary for physics analyses that one needs to consider are IP resolution, track reconstruction efficiencies, PID variables and track occupancy.

The IP resolution shows better trend in the simulation than in the data due to the mismodelling of material description of Velo simulation. As shown in Figure 3.18 IP resolution does greatly differ depending on the variation of material density of Velo. Around $\phi = \pm\pi/2$, where the two Velo parts overlap, the material difference causes the discrepancy. It can be corrected either by reweighting to data or by smearing the resolution with Gaussian distribution.

Track reconstruction efficiency is also not reproduced very well in certain kinematical bins, again due to modelling of scattering interactions.

The most critical problem that needs to be addressed in the presented analysis are the inaccuracies of PID variables, which are mismodelled in the simulation. The origin of this problem arises as a consequence of much lower estimate of low momentum

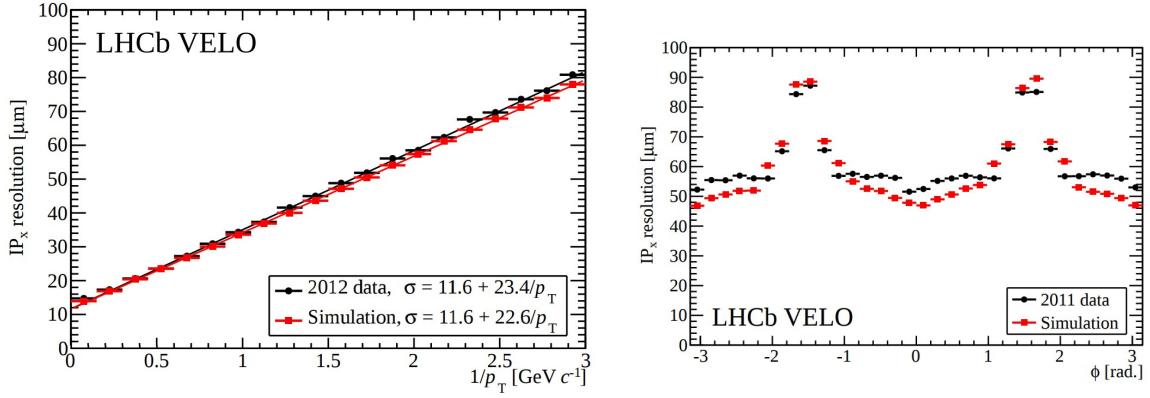


Figure 3.18: IP resolution in x-direction comparing the data and simulation output for 2012 data-taking period (left). IP resolution in x-direction comparing the data and simulation output for 2011 data-taking period as a function of angle, ϕ (right). This figures are taken from [32].

tracks in the detector making the photoelectron background underestimated. This results in better performance of separation power in simulation and is corrected using real data calibration.

Usually therefore PID efficiency is obtained from data. More specifically, by using high-yield and relatively background-free calibration channels, where the species of the particle can be deduced from kinematics of the decay. Standard set of these channels are "housed" in a PIDCalib package [50]. In this package, PID efficiency can be computed in a given kinematic region of interest.

Chapter 4

Handling of trimuon correlations at LHCb

asked for more formal wording, is this ok? This chapter discusses issues associated with three muons passing through the detector. Two collimated muons may traverse through the same parts of the detector if they bear the same charge, causing problems in resolving their individual tracks. Therefore, ghosts and clones are much more likely to occur. In LHCb, plethora of muon PID variables are used to suppress these types of spurious tracks. However, the usage of PID variables in an analysis in LHCb brings its own challenges. As the simulation is not able to estimate PID efficiencies correctly, most of the PID efficiencies are taken from control samples. New control samples for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ are considered as the PID efficiencies depend strongly on number of muons in the detector and in standard misid control samples there are no other muons in the event.

4.1 Muon PID variables

In addition to the muon identification variables mentioned in subsection 3.7.1, there is a further set of criteria for selecting muons. In this section a summary of the variables used in the analysis of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ is discussed.

4.1.1 Binary Muon PID variables

Similar to `isMuon` shown in [Table 3.2](#), there are more binary variables, such as `isMuonTight`, that can help with classification of muons. As its name suggests, `isMuonTight` has stronger conditions to satisfy as compared to `isMuon`.

In each muon station (M1-5) a field of interest, FOI is defined as

$$FOI_{x,y} = \rho_0{}_{x,y} + \rho_1{}_{x,y} \cdot \exp\left(\frac{\rho_2{}_{x,y} \cdot p}{\text{GeV}/c}\right), \quad (4.1)$$

where x, y are the dimensions perpendicular to the direction of the beam, p is momentum of the muon, $\rho_i{}_{x,y}$ are three dimension-dependent parameters tuned to give the best performance, by maximizing efficiency to misid rate.

When a muon passes through the detector, it leaves hits ($h_{x,y}$ coordinate) in pad with size $pad_{x,y}$ of each muon station. From the tracks formed in the tracking part of the detector, extrapolation coordinates $E_{x,y}$ are obtained by extrapolating the tracks into the muon stations. The hits are considered to be within the FOI if they satisfy the condition that $\|h_d - E_d\| < FOI_d \cdot pad_d$ for both $d=x,y$.

The detector information is read out in the x and y direction separately. The pad slicing according to this read-out scheme is known as *physical* slicing of pads. However, as seen in [Figure 4.1](#), the overlapping x and y *physical* pads of can be grouped into *logical* pads, which give information about x and y simultaneously. This leads to two groups of hits according to pad type: uncrossed hits - registered within *physical* pads only, and crossed hits - given by *logical* pads. Whereas `isMuon` only requires positive decision from uncrossed hits, `isMuonTight` requires positive decision based on crossed hits.

4.1.2 Muon PID variables based on sharing hits

Another way of identifying muon tracks is based on the variable, `nShared`, which identifies the number of tracks with shared hits in the muon stations. For each hit within the FOI of an extrapolated track, the `nShared` algorithm will check whether

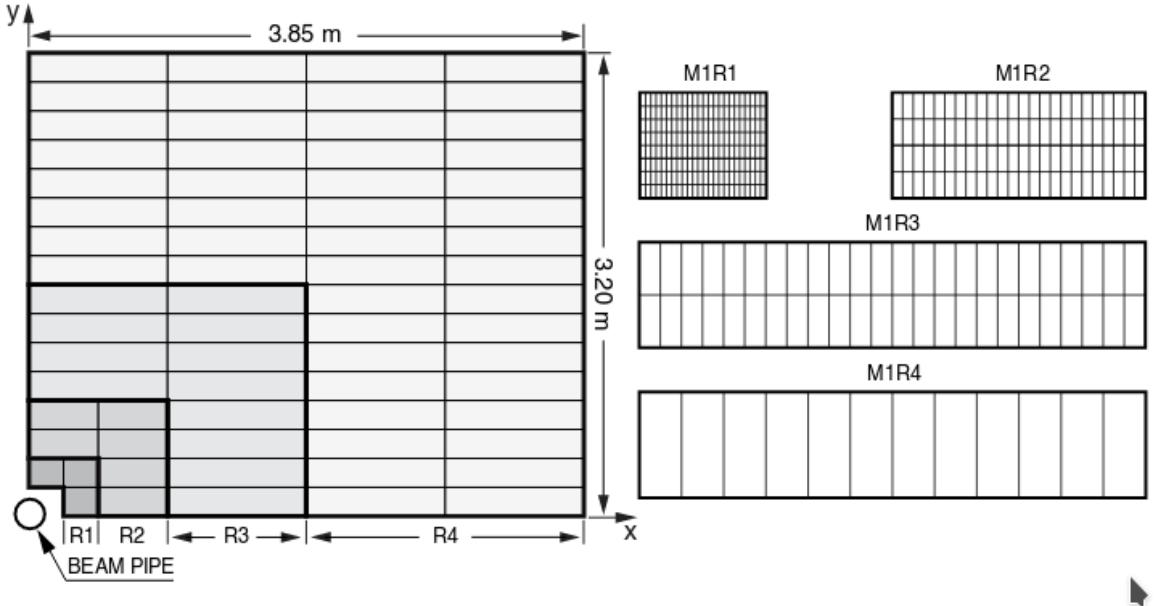


Figure 4.1: Schematic view of the muon station slicing into x-y pads. This is the left quadrant of the M1 station, showing decreasing granularity of the muon stations away from the beam pipe. This figure has been taken from [39]. M1R1 is the innermost region and M1R4 is the outermost region of the M1 station.

any other track was built using the given hit. In this case, the `nShared` variable of the muon track which has the bigger distance between the extrapolation coordinates and the hit coordinates is increased by 1. Hence this integer `PID` variable helps suppressing *ghost* tracks and *clones* if no tracks have hits in common with the owner of the track (`nShared=0`).

The muon identification software algorithms evolved significantly between the processing of Run I and Run II data. This included bug fixes, improvements and the introduction of new bugs. In the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ analysis, this has to be taken into account.

The first feature that is different between Run I and Run II arises from the calculation of the distance between the extrapolation and the hit in `nShared` algorithm. In *Stripping 21* used for 2012 and *21r1* used for 2011 data, it was discovered that the distance

between an extrapolated track and a hit was wrongly calculated. This mistake was corrected before *Stripping 23*, used for analysing 2015 data.

Secondly, information from M1 station was used to calculate distances, even though M1 information is not usually used for Muon ID algorithms. For analysts, this feature was present across all reconstruction software, meaning that simulation and data is affected in the same way.

In *Stripping 23*, the Muon ID algorithm was rewritten to adapt for parallelisation that needs to be done in order to meet the criteria for the upgrade of **LHCb**. There were two mistakes introduced prior to 2015 data taking. Firstly, an array was defined with 4-elements [0, 3] to store information about x and y coordinates of the hits. However, an iteration occurred by filling elements 1 to 4 of the array (M2-M5 station) resulting in a 5-element array where the 0-th element was not filled. Despite this, it turns out to be well-behaved and has no impact on physics.

Further in the process, however, this information is used to calculate the sum and average of distances per station between the hits and extrapolations. This algorithm again iterates over [0, 3] arrays, meaning that no information is used from the M5 muon station. This obviously has an effect, but again it is consistent across the versions of the reconstruction software used for the processing of Run II data.

The interplay between all these features for $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ decays, can be seen in Figure 4.2, which see shift in distribution of nShared for 2016 data taking, making the muons less isolated.

Using the same calibration channels as in subsection 3.7.2, misID and ID rates can be seen in Figure 4.3. As the tracks tend to be less isolated in *Stripping 26*, typical of non-signal like events, the misID rate is expected to be higher for the same working point (ID efficiency). While the issues highlighted here can be fixed with a reprocessing of the data, this is not expected to happen before 2019 or 2020.

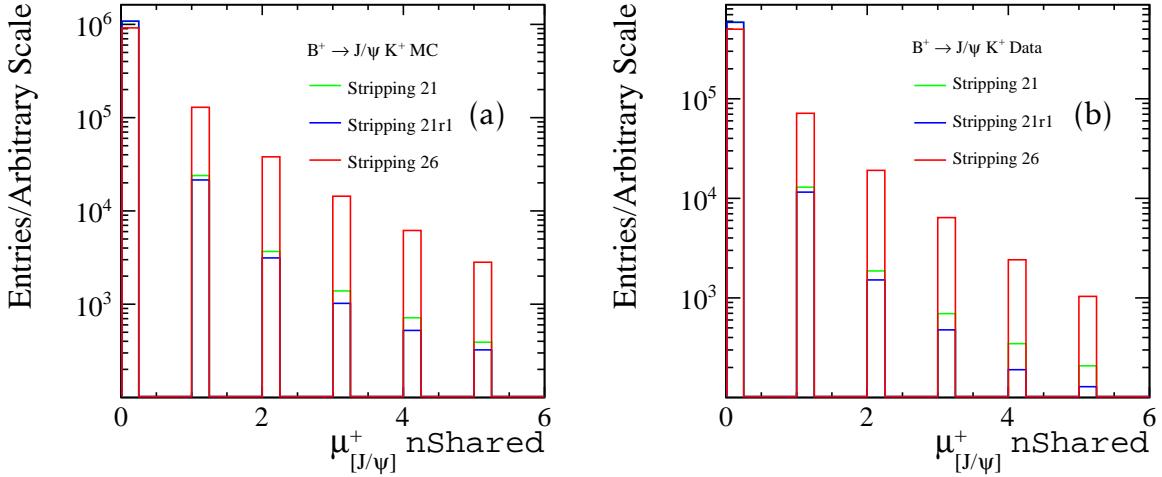


Figure 4.2: (a) n_{Shared} variable distribution for the positive muon in $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ decays in (a) simulation and (b) data. Different stripping versions corresponding to 2012 (*Stripping 21*), 2011 (*Stripping 21r1*), 2016 (*Stripping 26*) data-taking are shown. The distributions are normalised to have the same area. There is shift of distribution in *Stripping 26* towards less isolated tracks. The proportion of muon tracks that share no other hits with other tracks is smaller, whereas the proportion of the tracks sharing hits with other muon track is increasing.

4.1.3 Muon PID variables based on regression techniques

Similar to the DLL_{μ} variable in subsection 3.7.1, which combines all the information from the detector into a global likelihood, it is possible to feed all the different variables to a neural network, which can then produce an output corresponding to the probability of a particle to be of a certain species. Probn_{nx} , where x is the species of interest, is calculated and can be used also for muon identification. Compared to DLL_x variables, Probn_{nx} variables tend to have smaller correlation with the kinematics of the particle, and hence are more useful with decays where particles are soft, such as $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$. As with any machine learning algorithm, the selection of both the training sample and the input variables are important. In Run I, there were two tunings (trainings) introduced V2 and V3, with more input variables in V2. Depending on the species of

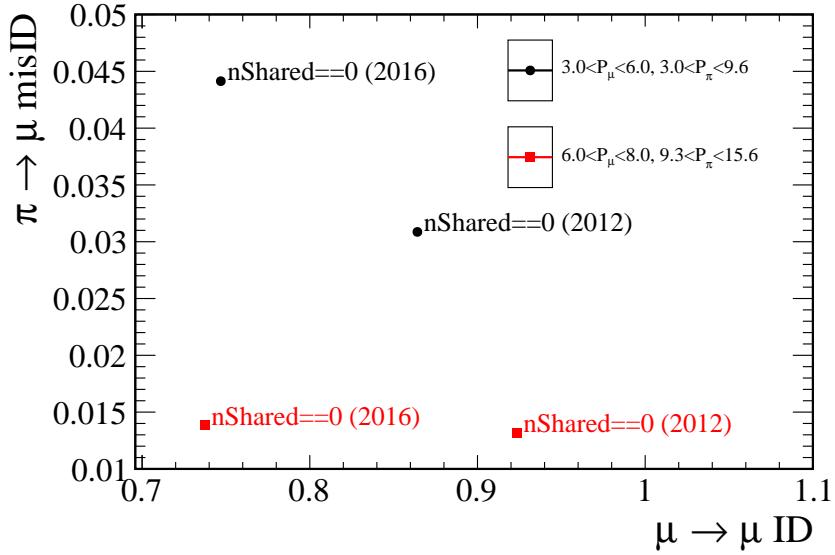


Figure 4.3: ID and misID probabilities from standard calibration datasets from 2012 (*Stripping 21*) and 2016 (*Stripping 26*), binned using the default 2-dimensional binning scheme in momentum p and pseudorapidity η . In this plot, ID and misID rates in the central bin of η , $2.375 < \eta < 3.25$, and the first and second bin in p are compared. This demonstrates that for the same pion ID efficiency, the misID rate is significantly higher in 2016 data.

particles, V2 or V3 performed better. In the analysis of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$, Probnnx_V2 is used.

4.2 Clones

When analysing decays with two muons of opposite charge, the LHCb magnet bends these two muons in two opposite directions. With two muons of the same sign, the muons will instead bend in the same direction and can stay close together in both the tracking system and the muon detectors. This causes trouble for the tracking algorithm as it distinguishes these two tracks less well. It is even possible that these two same sign muon tracks are not genuine tracks, but rather subtracks or a copy of another track,

clone tracks. Two tracks are clones if they share at least 70% of the hits in the **VELO** and at least 70% of the hits in the other T-stations. Of course, once it is established that two tracks share this percentage of hits, it has to be established which track is the clone track. This decision is based on the total number of hits and the **track χ^2/ndof** comparison of the two tracks.

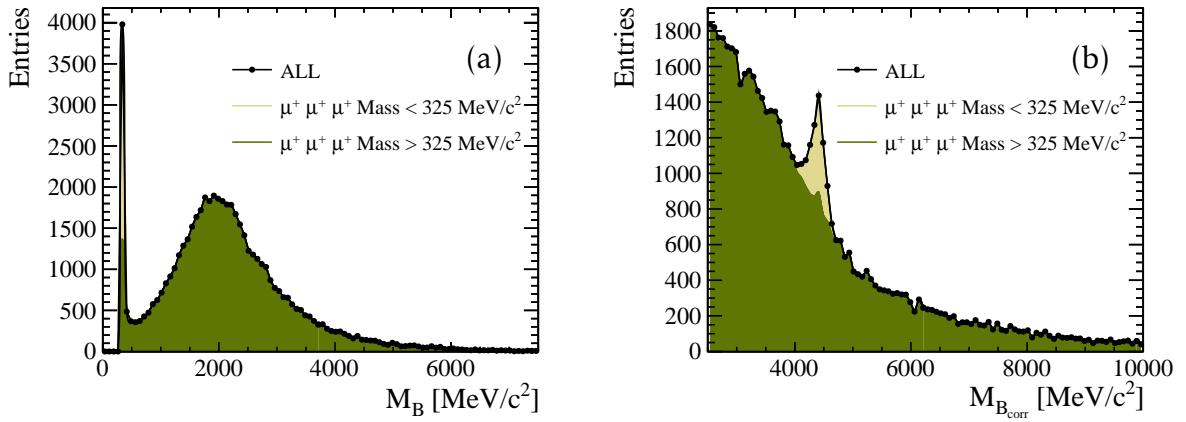


Figure 4.4: (a) Visible and (b) corrected mass of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ candidates in 2012 data where all the muons have the same charge. Clear fake peaks, arising from the correlation of several effects in the detector can be seen.

In search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$, two muons have the same charge, and hence are affected by the *clones*, which needs to be understood. In a control sample from data corresponding to 2012 data-taking period, which have three muon candidates of the same charge, the effect is even more prominent and can create potentially *fake peaks* in visible mass spectrum. *Clones* peak at well defined visible mass

$$M_B = \sqrt{(3 \times M_\mu)^2} \approx 318 \text{ MeV}/c^2 \quad (4.2)$$

Once translated into corrected mass, these *fake peaks* are smeared and look like genuine resonances with resolution as seen in Figure 4.4.

The shape shape emulating a genuine resonance arises as a collective effect from vertexing, tracking and trigger selection. As there are three parallel tracks, the vertex of the system is not well defined. However, the vertex fitting of the **PV** and **SV** is functional

and vertex χ^2/ndof is good as these tracks are subtracks of each other. However, *clones* can be differentiated by the position of the decay vertex of B , Figure 4.5 as well as the transverse position of the track in the tracking, **OT** as seen in Figure 4.6.

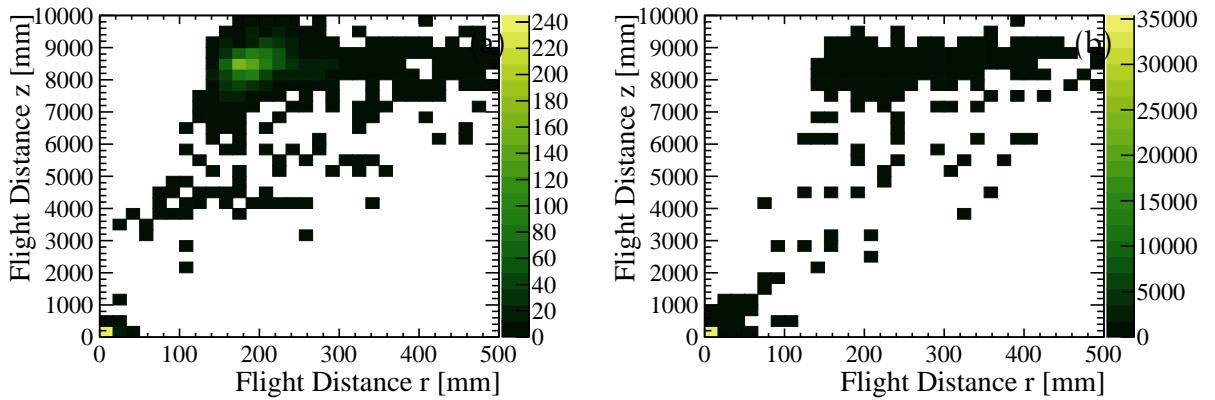


Figure 4.5: (a) Clone and (b) no clones flight distance properties. It can be seen that *clone* tracks have their decay vertex placed at the end of the detector, whereas regular good tracks will decay within the **VELO**.

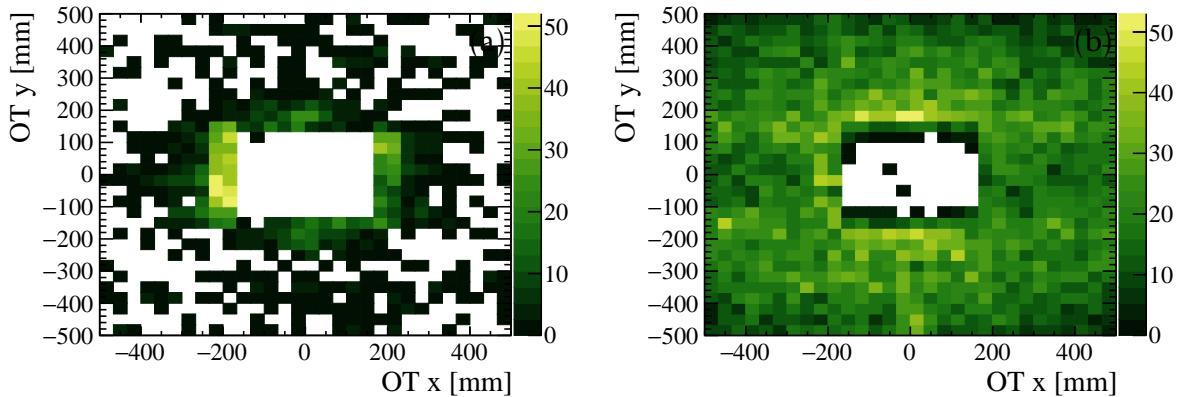


Figure 4.6: The difference in the **OT** detector between (a) clones and (b) real tracks in the **OT** at the distance 9450 mm along the **LHCb**. *Clones* are concentrated along the inner edge of the **OT**. Good muon tracks will cover most of **OT** evenly.

With this typical path for the clones there is a fixed angle of the clones through the

detector (the angle between the muon momentum and the z-axis), which is calculated using information from [OT](#) as

$$\arctan(\theta) = \arctan\left(\frac{\text{FD radius}}{\text{FD distance along } z}\right) = \arctan\left(\frac{200 \text{ mm} \text{ (Figure 4.6)}}{8500 \text{ mm}}\right) = 0.023 \text{ rad.} \quad (4.3)$$

With the L0Muon p_T threshold of 1.76 GeV/c for 2012 [41], the typical momentum from about 75 to 120 GeV/c is yielded because

$$p = 1.76 \text{ GeV/c} / \sin\left(\arctan\left(\frac{200 \text{ mm}}{8500 \text{ mm}}\right)\right). \quad (4.4)$$

The angle between B flight and trimuon momentum vector, $\cos(\theta_B)$, will also be fixed and have typical value of 0.7 mrad as seen in [Figure 4.7](#).

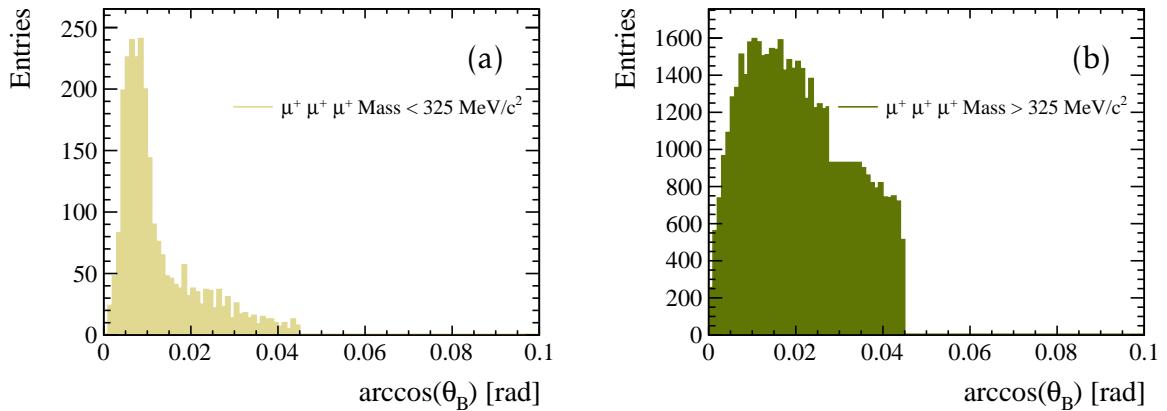


Figure 4.7: (a) Peaking clone distribution is visible as all of *clone* tracks are collinear compared to (b) smooth no clone distribution for $\cos(\theta_B)$.

Hence, missing p_T in the direction of the flight can be calculated using $\cos(\theta_B)$ and typical p ,

$$p_T = 100 \text{ GeV/c} \times \sin(0.0007) = 0.7 \text{ GeV/c}, \quad (4.5)$$

corrected mass $M_{corr} = \sqrt{M^2 + |p_T^2|} + |p_T| = 4.2 \text{ GeV}/c^2$, using missing p_T from [Equation 4.5](#) and visible mass of *clones* from [Equation 4.2](#).

In order to suppress these tracks in analysing $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$, where two muons have the same sign, any distinguishing features mentioned could be used. But the most powerful PID-wise is requiring nShared=0 in I, as this requirement removes all of the clones, as seen in Figure 4.8. For Run II, due to the introduced bugs, such strong requirement would harm signal efficiency too much so combination of nShared<2 and isMuonTight=1 is applied. *this should remove them as well as nShared is increased for the non-owner track only- i assume that owner track will be that of nshared=0. I.e below nShared==3 for the clones, so for two muons nShared==2 so if there are two tracks, is ok*

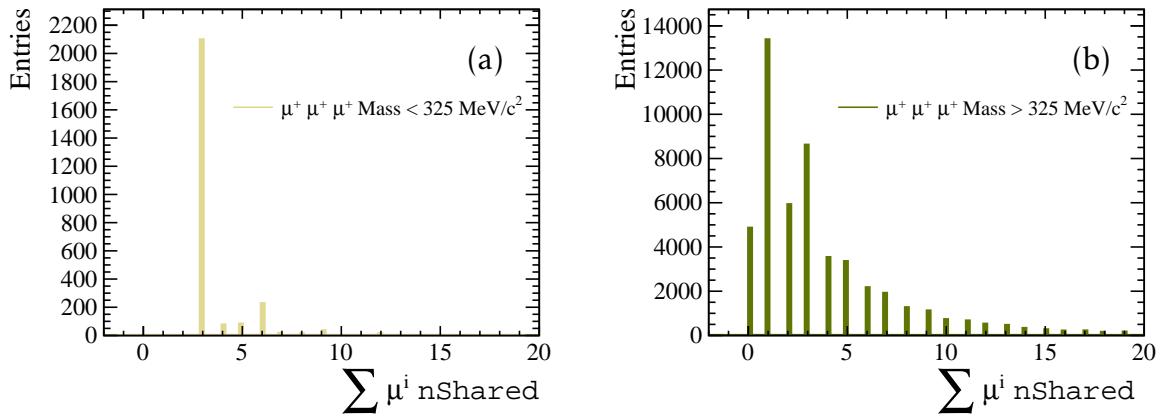


Figure 4.8: (a) Clone and (b) no clone distribution for sum of all muon nShared. Since in this case the clones are of each other, for the clones there is clear peak at three.

4.3 Probability of $K/\pi \rightarrow \mu$ misidentification at LHCb

Usually, in order to estimate background coming from misidentification of particles as muons in the detector, data samples with particles of known (non-muon) type are identified from the kinematics of decay chains. From these samples, probabilities of mis-identification are derived as discussed in subsection 3.5.1. However, the three muon signature will induce problems for PID variables that are correlated with the number of muons in the detector and specific data samples that incorporates this correlation have to be used for measuring the mis-identification probability.

4.3.1 Specific control sample for $K/\pi \rightarrow \mu$ misid rates

A platform that LHCb analysts usually use to extract the misid and id efficiencies, as described in subsection 3.5.1, is known as PIDCalib package [50]. It contains samples where the identity of the particle is known purely from kinematics. In this PIDCalib package, such a control sample for K/π is obtained from $D^{*+}(\rightarrow D^0(\rightarrow K^+\pi^-)\pi^+)$. These statistically populated background-free *sWeighted* samples, for which it is possible to extract misID and ID rates as a function of kinematics given certain PID criteria, do not have other muons in the final state.

More specifically, the topology of the mis-ID background component, which is two real muon tracks with an additional *fake* muon track is very different to PIDCalib sample $D^{*+}(\rightarrow D^0(\rightarrow K^+\pi^-)\pi^+)$, where there are no muons in the final state.

For this reason, $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*(\rightarrow K^+\pi^-)$ is used instead. While not as common as $D^{*+}(\rightarrow D^0(\rightarrow K^+\pi^-)\pi^+)$ decay, it still has high statistics and can be isolated with little background. It mimics the two real muon plus fake muon correctly and will be used to obtain pion and kaon misID probabilities.

4.3.2 Selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$

Data samples for each year of data taking were obtained from *stripping line* dedicated to look for this type of decay. The sample can be used for mis-ID studies of the hadrons as no particle identification is applied on them. Some initial selection was applied together with the more stringent $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ selection. The trigger criteria were applied on the J/ψ candidate rather than on the B candidate. (but on J/ψ candidate rather than B). The full additional selection is summarized in Table 4.1 is used.

4.3.3 Fitting Strategy for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ decay

After all the selection, the residual background needs to be modelled.

The signal component, $B^0 \rightarrow J/\psi K^*$, is obtained by fixing the shape from simulation apart from the mean μ and the width σ . It is fitted with a double-sided Hypatia

Idea	Cut
ID K^*	$ m(K\pi) - m_{PDG}(K_0) < 100 \text{ MeV}/c^2$
Compatible with PIDCalib kin comparison with pidcalib	for $K, \pi, p_T > 250 \text{ MeV}/c$
Compatible with PIDCalib	for $\mu, p_T > 800 \text{ MeV}/c$
Muon swap veto	$ m((h \rightarrow \mu)\mu) - m_{PDG}(J/\psi) > 60 \text{ MeV}/c^2$
Veto $B^+ \rightarrow K^+ \mu^+ \mu^-$	$\max(m(K^+ \mu^+ \mu^-)), m((\pi^+ \rightarrow K^+) \mu^+ \mu^-) < 1040 \text{ MeV}/c^2$
Veto $B_s^0 \rightarrow \phi \mu^+ \mu^-$	$m(K(\pi \rightarrow K)) > 1040 \text{ MeV}/c^2$
ID muons	mu1_ProbNNmu>0.5 and mu2_ProbNNmu>0.5
For kaon misID rates:	
ID pion	DLLK < 0 DLLp < 0 and IsMuon==0
For pion misID rates:	
ID kaon	DLLK > 0 and DLLK-DLLp > 0 and IsPion==0

Table 4.1: Offline selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$ decay.

function [51] (more in [section C.2](#)).

Background that peaks in the upper mass sideband, coming from heavier $B_s^0, \bar{B}_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ is also modelled using simulation, using the same function as signal but with μ offset by the difference between known B_s^0 and B^0 .

It is also possible that kaons and pions are swapped between themselves. Background coming from $K \leftrightarrow \pi$ swaps is modelled from simulation where mass hypotheses were swapped. Its distribution is fitted with a double sided Crystal Ball function [52] (more in [section C.1](#)).

Possibility of misidentified background comes from decay of $\Lambda_b \rightarrow K^- p \mu^+ \mu^-$ where the proton is misidentified as a pion. This background is modelled from simulation and fitted with a RooKeys p.d.f (more in [section C.3](#)).

Finally a combinatorial component is modelled by the exponential function.

The mass of the J/ψ was *constrained* to its nominal mass, procedure also known as a

mass constraint. It yields new estimates for track parameters of the final state particles, from which a new kinematic refit is done.

In order to obtain K/π misid probabilities an unbinned maximum likelihood fit to $\mu^+\mu^-\pi^+K^-$ mass between 5150 - 5450 MeV/ c^2 was performed. This fit with parameters listed in Table 4.2 give the yield of all the components.

Fit Parameter	Status
Yields	
$N_{B^0 \rightarrow J/\psi K^*}$ (Signal)	free
$N_{K\pi swaps}$	free
$N_{\Lambda_b \rightarrow J/\psi K^- p}$	free
$N_{B_s \rightarrow J/\psi K^*}$	free
$N_{Combinatorial}$	free
Signal Shape Parameters	
$\mu_{B^0 \rightarrow J/\psi K^*}$	constrained from signal MC
$\sigma_{B^0 \rightarrow J/\psi K^*}$	constrained from signal MC
Others	fixed from MC
$K\pi swaps$ Shape Parameters	
$\Lambda_b \rightarrow J/\psi K^- p$ Shape Parameters	
$B_s \rightarrow J/\psi K^*$ Shape Parameters	
$\mu_{B_s \rightarrow J/\psi K^*}$	Offset by $\mu_{B^0 \rightarrow J/\psi K^*}$
Others	fixed from signal MC
Combinatorial Shape Parameters	
exponential par.	free

Table 4.2: Summary of the fit parameters and individual component constraints for $B^0 \rightarrow J/\psi K^*$ fit.

The actual determination of the misid rate was obtained using a statistical subtraction of background, known as the *sPlot* technique [53], as the samples are not fully

background-free. The same method is also used in the `PIDCalib` package. In the *sPlot* method, the invariant mass distribution is fitted with no PID applied and each event is assigned *sWeights*, probabilities that a given event is a signal-like or a background-like. Then, through the *sPlot* technique, background is subtracted. Signal component can be then calculated by summing all the *sWeights* for all the candidates. The misid probabilities are finally obtained by dividing this signal component sum of *sWeights* with PID applied and with no PID applied. This misid probabilities are then considered within some kinematic partitioning, bins of p, η .

The misid rate was also cross-checked with another method, the *fit twice method*. This is because *sPlot* technique relies on the fact that there is no correlation between the control variables (p, η) and the discriminating variable (invariant mass) for both signal and background. This assumption may not be true, especially for background, and it can introduce biases.

The *fit twice method* consists of fitting $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$ before and after the PID requirement in a given kinematic (p, η) bin separately. Misid probabilities are then obtained as the ratio of signal yields arising from these two fits.

It was shown that these two methods yields very similar results, hence, for purposes of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ analysis *sWeight* values will be used. Fits to Run I and Run II data for both kaon and pion misid studies can be seen in [Figure 4.9](#).

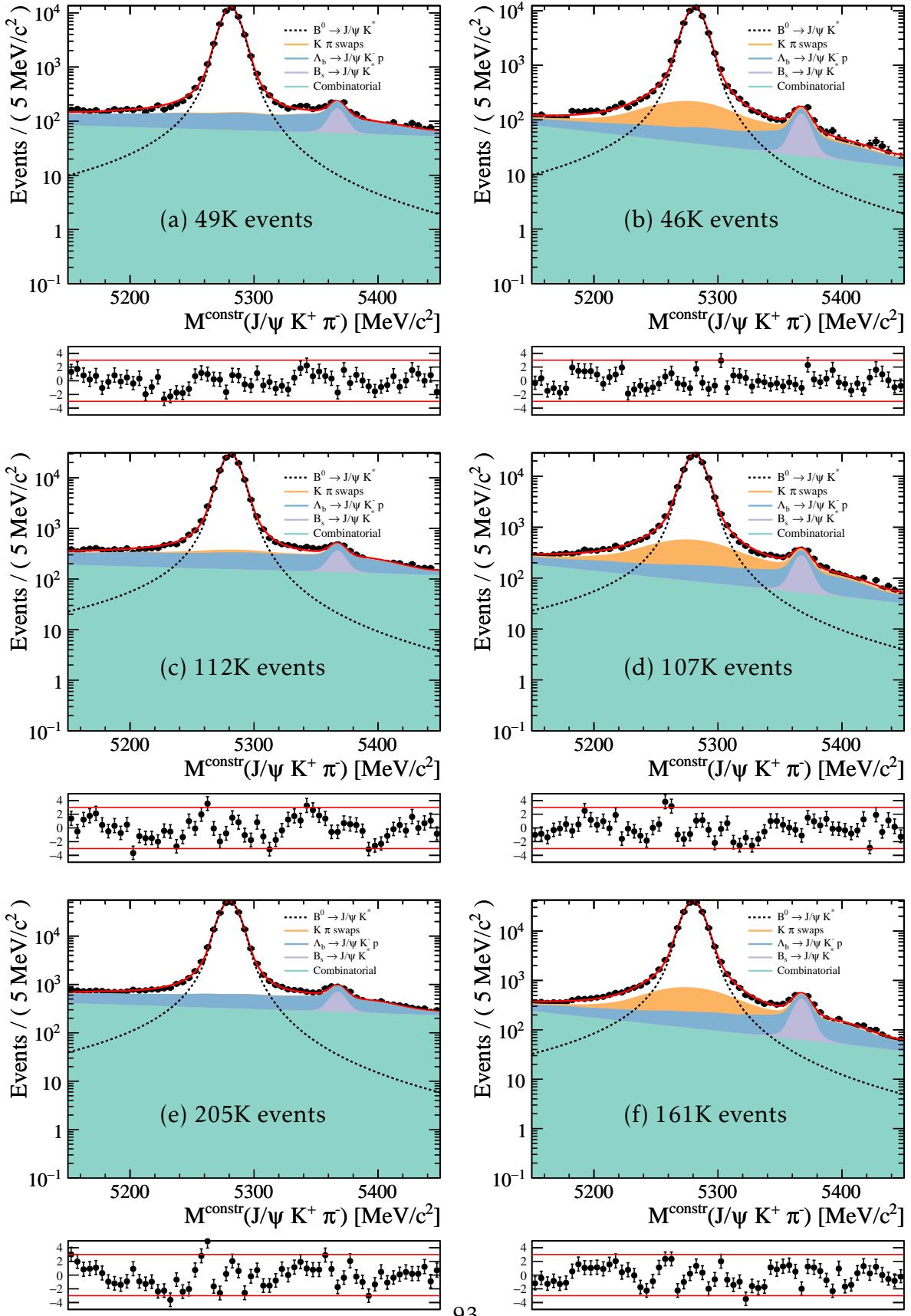


Figure 4.9: Fit to constrained $J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow \pi^+ K^-)$ mass with all the components for (a)(b) 2011, (c)(d) 2012, (e)(f) 2016. On the left, fit to data with pion ID (giving kaon misid probabilities), on right data with kaon ID (pion misid rates).

4.3.4 Results of $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$ control sample for $K/\pi \rightarrow \mu$ misid rates

Using the *sWeight* method, misID rates for kaons and pions can be obtained. In order to rule out that any disagreement of PID performance is caused by the fraction of the real kaon and pions tracks within the muon fiducial area, following check is performed. For both control samples, kaon sample from the $D^{*+}(\rightarrow D^0(\rightarrow K^+\pi^-)\pi^+)$ events and the kaon sample from the $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$ events, the probability of correctly identifying kaon is computed, given that only extrapolated tracks that fall within muon acceptance are considered. This is achieved by requiring that given kaon track has `InMuonAcceptance==1.0`. It can be seen in [Table 4.3](#) that the ID performance is the same across nearly all of the momentum range for kaons, showing that the fraction of the kaons tracks within muon acceptance is very similar. The same check was performed for pions tracks as well.

Hence using `InMuonAcc==1.0` tracks for both pion and kaon allows to perform study of the misID probabilities within the two calibration samples. In [Figure 4.10](#), the $\pi \rightarrow \mu$ misID probability for different PID hypotheses are studied. As it can be noticed, the more stringent the muon selection on the pion track, the lower the probability of misidentification.

In general the agreement is good in the low momentum regions between the two samples as shown in [Figure 4.10](#). These pions are softer and hence they will spread out more in the magnetic field, causing less interference with other two real muons in decay. However, in high momentum region, the pion will follow a path through the muon system that is more similar to the path of the muon of the same charge in the $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$ decay. The influence of other two real muons in high momenta region will lead to bigger disagreement as these two real muons leave hits in the muon chambers close to the collimated pion track, making the rate of `IsMuon==1.0` (pink) is higher.

This disagreement is decreased by requiring `nShared==0.0` (blue), as having two

p range [MeV/c]	$D^{*+}(\rightarrow D^0(\rightarrow K^+\pi^-)\pi^+)$	$B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*(\rightarrow K^+\pi^-)$	Ratio
3000 - 6000	0.77±0.0016	0.83±0.0047	1.1±0.0065
6000 - 9300	0.93±0.00030	0.95±0.0019	1.0±0.0020
9300 - 10000	0.96±0.00037	0.97±0.0031	1.0±0.0033
10000 - 12600	0.97±0.00014	0.97±0.0017	1.0±0.0017
12600 - 15600	0.98±0.00011	0.97±0.0017	0.99±0.0018
15600 - 17500	0.98±0.00013	0.96±0.0024	0.98±0.0025
17500 - 21500	0.98±8.9e-05	0.96±0.0018	0.98±0.0018
21500 - 27000	0.98±7.8e-05	0.96±0.0018	0.98±0.0019
27000 - 32000	0.98±8.8e-05	0.96±0.0024	0.98±0.0025
32000 - 40000	0.98±8.0e-05	0.96±0.0022	0.98±0.0022
40000 - 60000	0.97±7.5e-05	0.95±0.0021	1.0±0.0022
60000 - 70000	0.96±0.00016	0.96±0.0043	1.0±0.0046
70000 - 100000	0.95±0.00013	0.94±0.0044	0.99±0.0046

Table 4.3: **It is necessary as this is unrelated to following figures** `K_InMuonAcc==1.0` shows the interpolation of K tracks into muon chambers. It can be seen that both samples agree with each other very well, meaning that measured misid rate is done for the same fraction of considered tracks. This measurement is done in a pseudorapidity region $1.5 < \eta < 5.0$.

other collimated muons to share hits will be more likely. The effect of other `PID` variables can also be seen, but it is harder to interpret as these depend on several variables.

Even though this disagreement is decreased, it can still be noted that for the high momenta region $\pi \rightarrow \mu$ Figure 4.10 and $K \rightarrow \mu$ Figure 4.11 rate is 2 to 3 times higher with additional two real muon tracks, which is significant, as this means that if these mmissid rates are used to be parametrically applied on samples to estimate the background, the standard control sample would underestimate the misid component by the

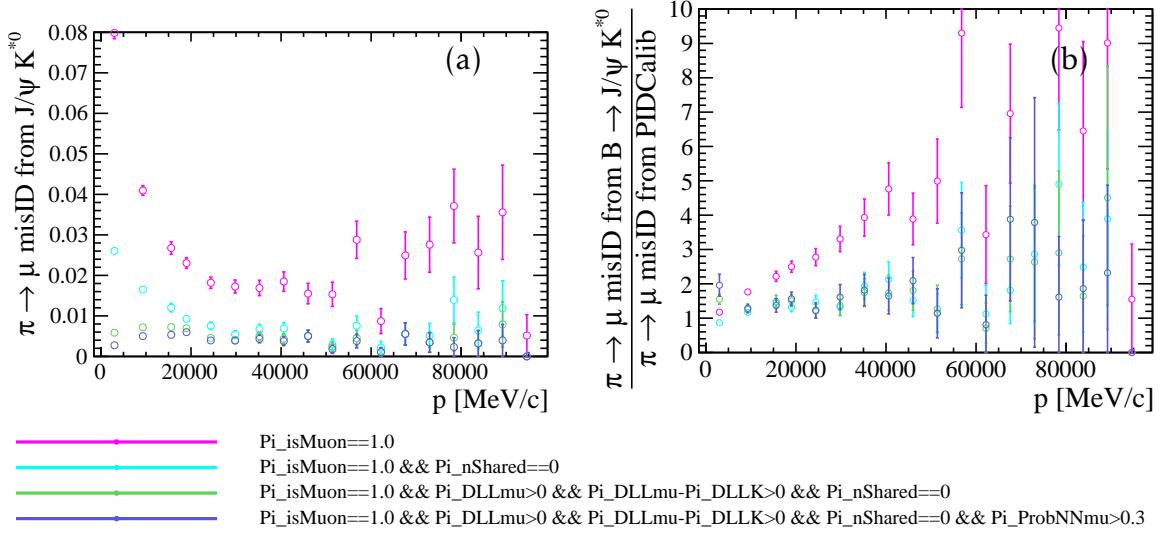


Figure 4.10: (a) $\pi \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ for 2012 data. (b) This is compared to the standard PIDCalib $D^{*+}(\rightarrow D^0(\rightarrow K^+ \pi^-)\pi^+)$ sample.

same factor.

In conclusion, it was shown that the standard misID samples are not good proxy for estimating the misid probabilities as there is interference from two other muons in the event. Instead, the misid probabilities that are used in calculations for the misid background for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ are obtained from *Sweighted* $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$. Remaining effects of taking this sample for calibration are considered as a systematic uncertainty. forward reference to misid templating.

Due to the different PID definitions of nShared between Run I and II, different PID requirement are tested. Results for $\pi \rightarrow \mu$ and $K \rightarrow \mu$ are summarized in [Figure 4.12](#) and [Figure 4.13](#). The misID probabilities in 2016 for also show the same momentum dependent trend as in 2012.

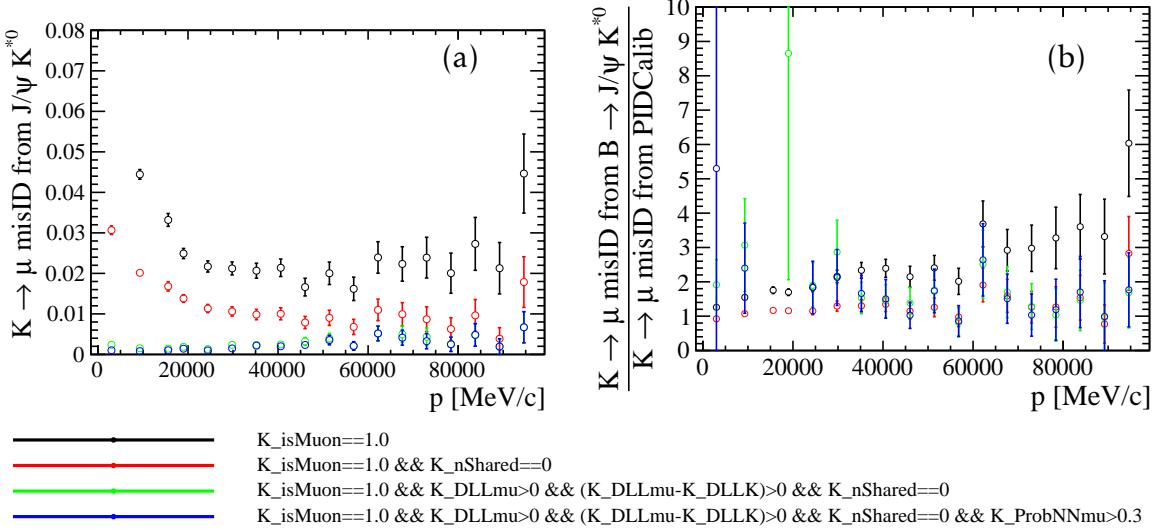


Figure 4.11: (a) $K \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ for 2012 data. (b) This is compared to the standard PIDCalib $D^{*+}(\rightarrow D^0(\rightarrow K^+ \pi^-)\pi^+)$ sample.

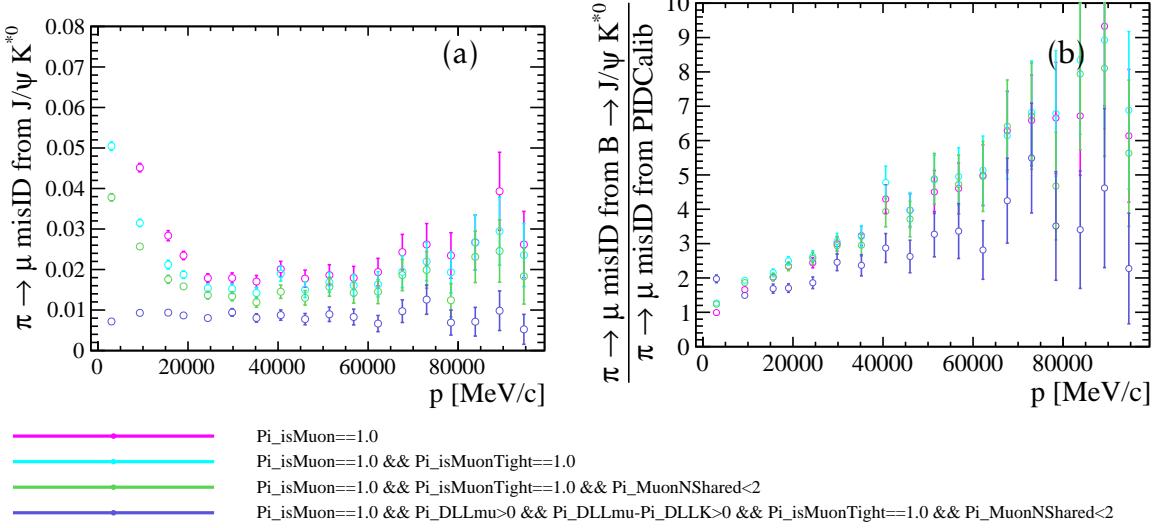


Figure 4.12: (a) $\pi \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ for 2016 data. (b) This is compared to the standard PIDCalib $D^{*+}(\rightarrow D^0(\rightarrow K^+ \pi^-)\pi^+)$ sample.

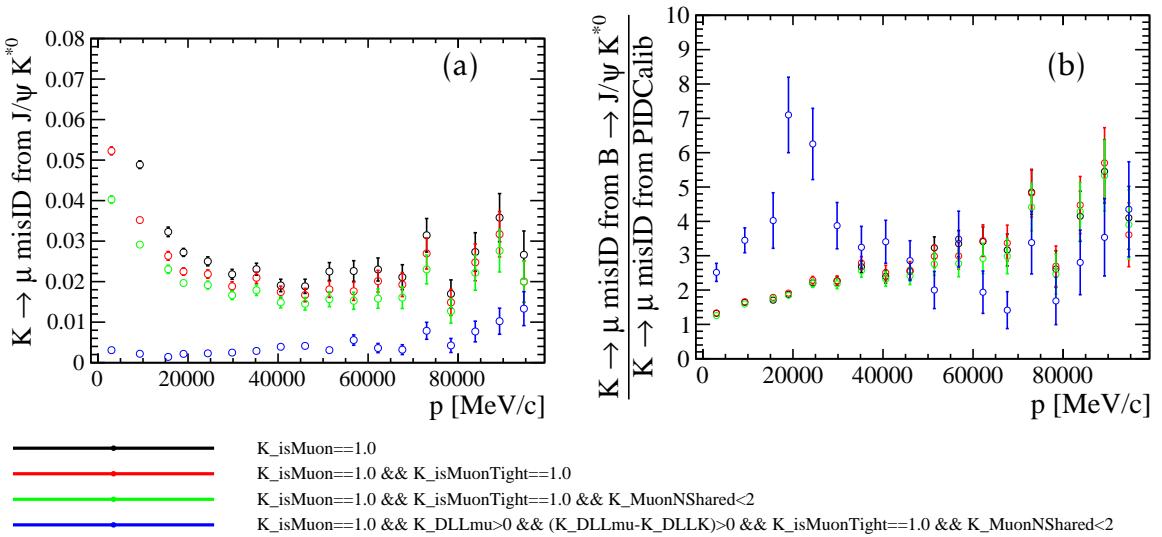


Figure 4.13: (a) $K \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow K^+ \pi^-)$ for 2016 data. (b) This is compared to the standard PIDCalib $D^{*+}(\rightarrow D^0(\rightarrow K^+ \pi^-)\pi^+)$ sample.

Chapter 5

Discovering (Setting Limit for)

$$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu \text{ at LHCb}$$

LHCb's flagship analyses contain several muons in the final state coming from differently flavoured B mesons. Despite being in this category, search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ is limited by the rareness of its occurrence as well as different backgrounds that can mimic its signature in the detector. Moreover, presence of invisible neutrino does induce uncertainties into reconstruction. The following chapter 5 will concentrate on characterisation of backgrounds as well as selection that is performed in order to reduce these backgrounds.

5.1 Topology of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay at LHCb **ULRIK**

Upon hadronisation from $b\bar{b}$ pair a B^\pm particle will travel less than a centimetre in the laboratory frame of reference before it decays. This allows reconstruction of a primary vertex **PV** and its decay vertex, *secondary vertex* **SV**. By joining these vertices, the direction as well as flight distance (**FD**), can be established. In order to infer information about the kinematic properties of B^\pm , the decay products are studied. All three muons are used to reconstruct the visible four-momentum. By conservation of momentum with respects to the direction of the flight of B^\pm , the neutrino is assigned all missing momentum transverse to the direction of the flight of the B^\pm meson. A

schematic diagram of the decay topology can be seen in Figure 5.1.

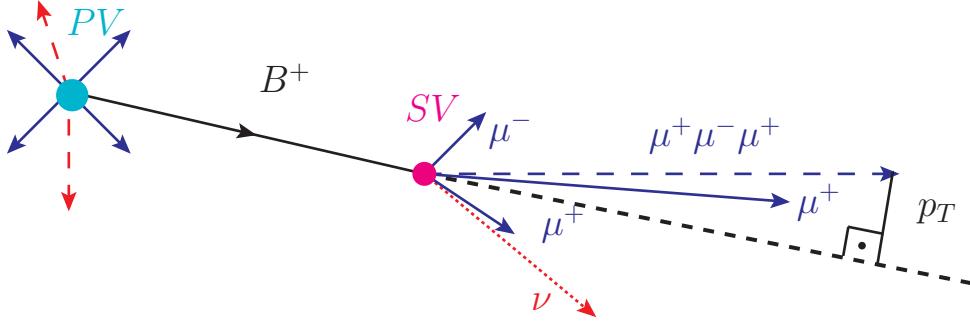


Figure 5.1: Schematic view of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay. All charged tracks (in filled-blue) are combined into four-vector representing the visible part of the decay (semi filled-blue). Information about invisible neutrino (semi filled-red) are deduced from the conservation of momentum with respect to the direction of the flight of the B^\pm meson.

Combining all information allows for reconstruction of the *corrected mass* that plays similar role to invariant mass in fully reconstructed decays. Invariant mass is usually used in LHCb as the distribution from which the yield of a signal decay is determined through a fit. This particular quantity is used as it distinguishes well signal and background shapes with minimal modelling assumptions.

Corrected mass is defined as

$$M_{corr} = \sqrt{M^2 + |p_T^2| + |p_T|}, \quad (5.1)$$

where the M^2 is the invariant visible mass squared and p_T^2 is the missing momentum squared transverse to the direction of the B^+ meson flight. The corrected mass of the B^\pm meson will be denoted as $M_{B_{corr}}$.

M_{corr} can be thought of as the minimal correction to the visible mass to account for the missing neutrino information. The resolution on the *corrected mass* hence becomes a critical quantity that needs to be understood. As the method of reconstruction of corrected mass relies heavily on the knowledge of the B^\pm meson flight direction, the resolution of PV position and SV vertex is crucial. Let $\vec{x}_{PV} = \{x_{PV}, y_{PV}, z_{PV}\}$, $\vec{x}_{SV} =$

$\{x_{SV}, y_{SV}, z_{SV}\}$ be PV and SV vertex position and $\vec{p} = \{p_x, p_y, p_z\}$ be the visible trimuon momentum. Then the missing transverse momentum to the direction of the flight p_T (momentum of the neutrino) as shown in [54] is

$$p_T^2 = \left| \vec{p} - (\vec{x}_{SV} - \vec{x}_{PV}) \frac{\vec{p} \cdot (\vec{x}_{SV} - \vec{x}_{PV})}{|(\vec{x}_{SV} - \vec{x}_{PV})|^2} \right|^2. \quad (5.2)$$

In general in order to propagate error on $f(x, y, z)$, where x, y, z are independent variables, the variance of $f(x, y, z)$ is given as

$$\langle f^2 - \langle f \rangle^2 \rangle = \langle f(x + \delta x, y + \delta y, z + \delta z)^2 - f(\langle x \rangle, \langle y \rangle, \langle z \rangle)^2 \rangle \quad (5.3)$$

Using first order Taylor expansion of variance and rewriting into the matrix form:

$$\begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix} \begin{bmatrix} \delta x^2 & \delta x \delta y & \delta x \delta z \\ \delta y \delta x & \delta y^2 & \delta y \delta z \\ \delta z \delta x & \delta z \delta y & \delta z^2 \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix} \quad (5.4)$$

In this formalism let f is the *corrected mass* and x, y, z are variables on which *corrected mass* depends. Using Equation 5.1, these independent variables are visible mass four-vector, $p_{3\mu} = \{E, p_x, p_y, p_z\}$, and missing p_T (defined in Equation 5.2), which in turn depends on \vec{p} , \vec{x}_{PV} and \vec{x}_{SV} .

In this case, let $x = \vec{x}_{PV}$, $y = \vec{x}_{SV}$, $z = p_{3\mu}$ and COV being the covariance matrix, the error (square root of variance) on *corrected mass*, $\delta_{corr m}$ is calculated as

$$\delta_{corr m} = \sqrt{\langle f^2 - \langle f \rangle^2 \rangle} = \sqrt{\nabla_{x_{PV}}^T \text{COV}_{x_{PV}} \nabla_{x_{PV}} + \nabla_{x_{SV}}^T \text{COV}_{x_{SV}} \nabla_{x_{SV}} + \nabla_{p_{3\mu}}^T \text{COV}_{p_{3\mu}} \nabla_{p_{3\mu}}}. \quad (5.5)$$

It was shown in [54] that the $\delta_{corr m}$ is mostly dominated by vertex position terms.

5.2 Sources of Backgrounds ULRIK

The largest background that looks similar to signal comes from *cascade decays*, where the semileptonic $b \rightarrow c \rightarrow s$ or $\bar{b} \rightarrow \bar{c} \rightarrow \bar{s}$ transitions occur. A typical example of this

type of background in hadronic terms is $B^+ \rightarrow (\bar{D}^0 \rightarrow (K^+ \rightarrow \mu^+ \nu) \mu^+ \nu)$, where the K^+ meson is subsequently misidentified as muon. Because the K^+ meson is misidentified as a muon, this type of background is denoted as misID background.

All background sources that contain at least one misidentified particle are categorized as misID. If the sign of the misidentified particle agrees with the sign of the mother B^\pm , it belongs to the same sign misID background (*SS misID*) background. In the event where opposite sign particle to the mother B^\pm is misidentified, this background will be referred to as (*OS misID*) background. *OS misID* background is expected to have smaller rate as the misidentified particle would have to proceed via decays with additional particles.

As the hadronisation of a $b\bar{b}$ pair leads to the creation of two b hadrons, each with their own decay chain, it is possible to mix up the decay products of the two to create a single fake signal candidate. This type of background is denoted combinatorial background.

Then presence of a neutrino in a final state allows for certain uncertainty regarding the information of the fourth decay product. If some of the tracks of the decays are not reconstructed, either because they are neutral, or they are charged but with too low momentum to be found by the tracking algorithm, it means that the missing information may be attributed to the neutrino. *Missing tracks* will hence create partially reconstructed background. Some of the most dangerous are $B^+ \rightarrow D\mu^+\nu$ type partially reconstructed backgrounds where $B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu\nu$, where $\mathcal{B}(D^0 \rightarrow K^-\pi^+\mu^+\mu^-) \approx 4.17 \times 10^{-6}$ and $B^+ \rightarrow D^0\mu\nu \approx 10\%$. This predicts $\mathcal{B}(B^+ \rightarrow K^-\pi^+\mu^+\mu^-) \approx 1 \times 10^{-7}$.

Decays that proceed via hadronic resonances such as $B^+ \rightarrow \rho/\omega \mu^+\nu$, followed by $\rho/\omega \rightarrow \mu^+\mu^-$ are part of signal as mentioned in **Sally-add reference to theory chapter, also maybe write about other backgrounds**.

5.3 Analysis strategy

Sally-add references to chapters

The analysis of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ decay is divided into several different parts; signal selection, optimisation, normalisation, fitting and limit setting. Throughout this document, charge conjugates of the decays are assumed unless stated otherwise. Results presented are based on the analysis of the full 3 fb^{-1} Run I dataset as well $\approx 1.7 \text{ fb}^{-1}$ Run II data from 2016. Data from 2015 is not used due to the very high pT threshold for the muon triggers used during that year, resulting in a very low signal efficiency. Additionally the search will be conducted in a particular $\min q^2 = \min(q^2(\mu_1^+, \mu^-), q^2(\mu^-, \mu_2^+))$ region.

To perform the search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$, a specific preselection was applied to form potential signal candidates. To reconstruct the mass of the B^+ with missing information about the neutrino [Equation 5.1](#) is used. A simulation sample that mimics the decay of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ passing through preselection was used to develop further discriminating selection. To get the selection efficiency for different types of backgrounds, different proxy samples are used. For more details about samples used see [section 5.4](#).

Throughout the analysis there was a blinding procedure put in place for this search in order not to bias the result. The signal region $4500 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 5500 \text{ MeV}/c^2$ was blinded until full strategy and sensitivity was evaluated. The signal dataset corresponding to this selection is known as **blinded signal dataset**. After unblinding, data with full mass spectrum was obtained and is known as **full signal dataset**.

Combinatorial background, which arises as random combinations of tracks passing the preselection, is taken from the upper corrected $\mu^+ \mu^- \mu^+$ mass side band, $M_{B_{\text{corr}}} > 5.5 \text{ GeV}$, where very few signal candidates are expected.

5.4 Samples ULRIK

5.4.1 Data Samples ULRIK

Results presented in this thesis are based on the analysis of the full 3 fb^{-1} Run I dataset at $\sqrt{s} = 7, 8 \text{ TeV}$ and 1.7 fb^{-1} Run II data at $\sqrt{s} = 13 \text{ TeV}$.

5.4.2 Simulation Samples ULRIK

For signal simulation, three different decay models were exploited and are summarized in [Table 5.1](#).

Channel	Year	Pythia	EVTGEN	Size	Stage
Simulation used for fitting mass shapes					
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 6.4 [44]	PHSP	0.5M	<i>generator-level+detector</i>
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 8.1 [24]	PHSP	0.5M	<i>generator-level+detector</i>
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 6.4 [44]	INSP	0.5M	<i>generator-level+detector</i>
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 8.1 [24]	INSP	0.5M	<i>generator-level+detector</i>
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2016	Pythia 8.1 [24]	INSP	1.0M	<i>generator-level+detector</i>
Simulation used for evaluating <i>generator-level</i> efficiencies					
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 6.4 [44]	PHSP	25000	<i>generator-level</i>
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 6.4 [44]	INSP	25000	<i>generator-level</i>
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 8.1 [24]	INSP	25000	<i>generator-level</i>
Simulation used for cross checking of $\text{min}q^2$ selection					
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 6.4 [44]	NIKI	25000	<i>generator-level</i>

Table 5.1: Summary of signal simulation samples used in this analysis with different decay models. In all cases the daughters of B^\pm are required to be within [LHCb](#) acceptance. All of this samples are mixture under magnetic polarity up and magnetic polarity down conditions.

The full phase space model, *PHSP*, only takes into account the kinematic constraints of the decay without taking into account any input from theoretical considerations as the matrix element is constant.

Sally: move it to theory and check ulrik's comment In order to produce simulation with a decay model which is more representative of the spin structure involved, the following strategy is adapted. In this simulation approach, the decay proceeds as follows: B^\pm decays into W^\pm and a pair of opposite sign muons and then W^\pm is decayed to $\mu^\pm \nu$. *BTOSLLBALL* model [21], traditionally used for $B \rightarrow (K, K^*) l^+ l^-$ decay, with the form factor calculations can be used to simulate $B^\pm \rightarrow W^\pm l^+ l^-$ decay. After that, W^\pm is decayed to $\mu^\pm \nu$ using *PHSP*. For semileptonic $b \rightarrow s l^+ l^-$ transitions, there is a characteristic photon pole for low $q(\mu^+, \mu^-)$, invariant mass of the opposite muon pair, and flat distribution for $K^*(\mu^+, \nu_\mu)$, invariant mass of the muon and neutrino pair. In order to achieve this, a new pseudo-particle is introduced to *EVTGEN* with specific properties, $K^*(\mu^+, \nu_\mu)$, and the best output can be seen to be for a particle $K^*(\mu^+, \nu_\mu)$ with mass to be set to $0.1 \text{ GeV}/c^2$, and width, corresponding to $\tau = 1.3 \times 10^{-17}$ nanoseconds as can be seen in [Figure 5.2](#). This procedure was also applied for the charge conjugate case. This model is denoted as *INSP* and is used as default in mass fits and efficiency calculations.

Finally, exclusively for this decay, a new decay model *B2MuMuMuNu* was added to *EVTGEN*, based on work performed by theorist Nikola Nikitin (write more once theory chapter is done and refer to it.). This model denoted as *NIKI*, is used mainly for validation purposes.

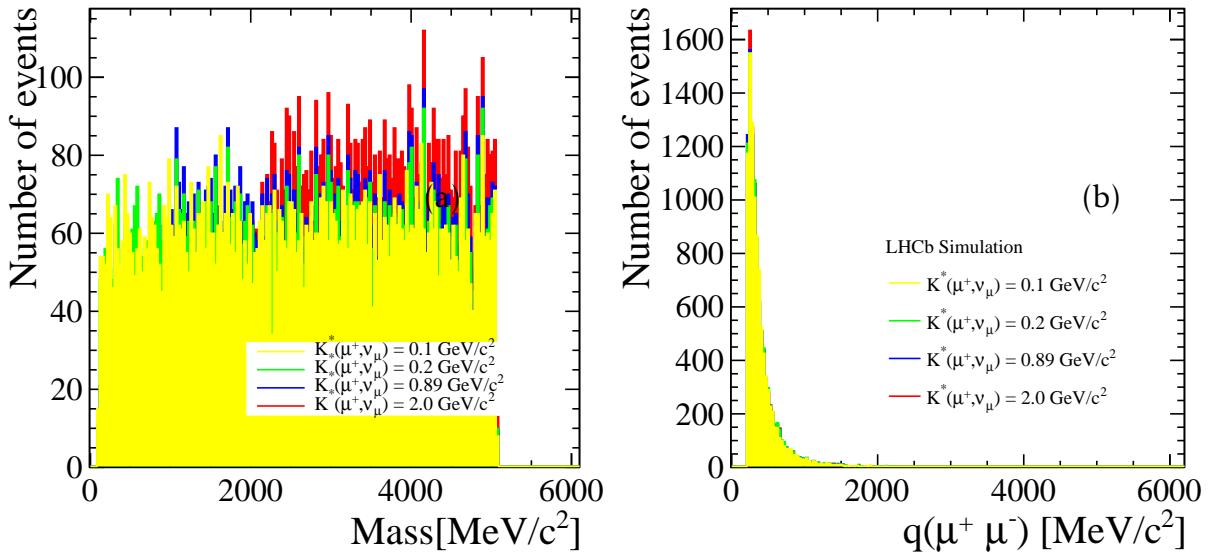


Figure 5.2: Distributions for signal MC in using Pythia 6.4 [44] conditions. (a) $K^*(\mu^+, \nu_\mu)$ distributions under different K^* mass hypotheses. (b) $q(\mu^+, \mu^-)$ distributions under different K^* mass hypotheses. The most flat distribution in $K^*(\mu^+, \nu_\mu)$ is plotted in yellow.

5.5 Preselection for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ **ULRIK**

In order to fit within the LHCb computing model, an initial set of selection criteria is applied during the data processing known as *stripping*. Each of the criteria are discussed below and a summary can be found in Table 5.2.

Firstly, all three muon tracks are required to have a significant IP with respect to the primary vertex. Minimum Impact Parameter χ^2 , ($\text{min IP}\chi^2$), gives the minimum significance of a particle's trajectory to the primary vertex. Hence by requiring $\text{min IP}\chi^2 > 9$ for muons is consistent with the hypothesis that the muon is 3σ away from the primary vertex and hence can be well differentiated. In addition, the change in the χ^2 if PV and SV vertices are fitted separately as opposed to common vertex fit, FD χ^2 , suppresses prompt backgrounds.

Each muon track is required to have good track χ^2 per number of degrees of freedom of the fit (ndof), ($\text{track } \chi^2/\text{ndof}$), as well as low P_{ghost} . This removes spurious tracks as

well as tracks with low quality.

Each muon candidate is also identified with initial basic [PID](#) variables. Firstly muons are chosen due to their signature in the muons stations with the binary `isMuon` decision. Secondly, muons candidates are chosen such that it is more likely that the candidate is a muon than a pion or kaon using global DLLmu variables defined in [subsection 3.7.1](#). This reduces the background from misidentified muons.

In order to only select events which are compatible with the three muons originating from the same point in the space, ([vertex \$\chi^2/\text{ndof}\$](#)), the χ^2 of the trimuon vertex per degree of freedom fit is required to be small. This decreases the contamination from *cascade decays* where the particle with the c quark content from $b \rightarrow c \rightarrow s$, such as D , would have non-negligible lifetime leading to higher [vertex \$\chi^2/\text{ndof}\$](#) .

Requiring that B^+ direction points in the same direction as the line from [PV](#) to [SV](#), ($\cos(\theta_B)$ - which measures the angle between these two vectors), is close to unity translates into a well reconstructed event, which minimizes combinatorial background, where random track makes this pointing worse. Putting bounds on the mass window, whether it is *visible* or *corrected* mass, also suppresses combinatorial events.

Candidate	Stripping Selection
muon	$\min \text{IP} \chi^2 > 9$
muon	$p_T > 0$
muon	$\text{track } \chi^2/\text{ndof} < 3$
muon	$\text{DLLmu} > 0$
muon	$\text{DLLmu-DLLK} > 0$
muon	isMuon==true
combination	$\cos(\theta_B) > 0.999$
	$p_T > 2000 \text{ MeV}$
	$\text{FD } \chi^2 > 50$
	$\text{vertex } \chi^2/\text{ndof} < 4$
	$0 \text{ MeV}/c^2 < M_B < 7500 \text{ MeV}/c^2$
	$2500 \text{ MeV}/c^2 < M_{B_{corr}} < 10000 \text{ MeV}/c^2$

Table 5.2: Selection of events based on muon and the B^+ candidate requirements. *Stripping selection* for the signal decay $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ is the same for both Run1 and 2016 data.

5.6 Trigger Selection ULRIK

In order to obtain triggered data, $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ candidates are required to pass a certain set of trigger decisions at [L0](#), [HLT1](#) and [HLT2](#) levels summarized in [Table 5.3](#). It can be noted that the decision is applied at the mother B^\pm level. In particular, `Bplus_L0MuonDecision_T0S` decision, means that one of the muons from B^\pm in an event has triggered and made a positive decision.

As discussed in [section 3.8](#) `L0MuonDecision` decides on whether an event is accepted depending on the p_T of a muon and the number of hits in the [SPD](#). Run I can be split into 2011 and 2012 conditions where, in 2011 the most used threshold for positive decision is $1.48 \text{ GeV}/c$ [55] and $1.76 \text{ GeV}/c$ [41]. Run I [SPD](#) rate only accepts events

Trigger Selection
Bplus_L0MuonDecision_TOS
Bplus_H1t1TrackMuonDecision_TOS
Bplus_H1t2TopoMu2BodyBBDTDecision_TOS
Bplus_H1t2TopoMu3BodyBBDTDecision_TOS
Bplus_H1t2DiMuonDetachedDecision_TOS
Bplus_H1t2DiMuonDetachedHeavyDecision_TOS

Table 5.3: Trigger selection applied on both signal and normalisation samples.

below 600. In Run II, the trigger thresholds varied more but the most representative acceptance for muon p_T was above 1.85 GeV/c with **SPD** multiplicity below 450.

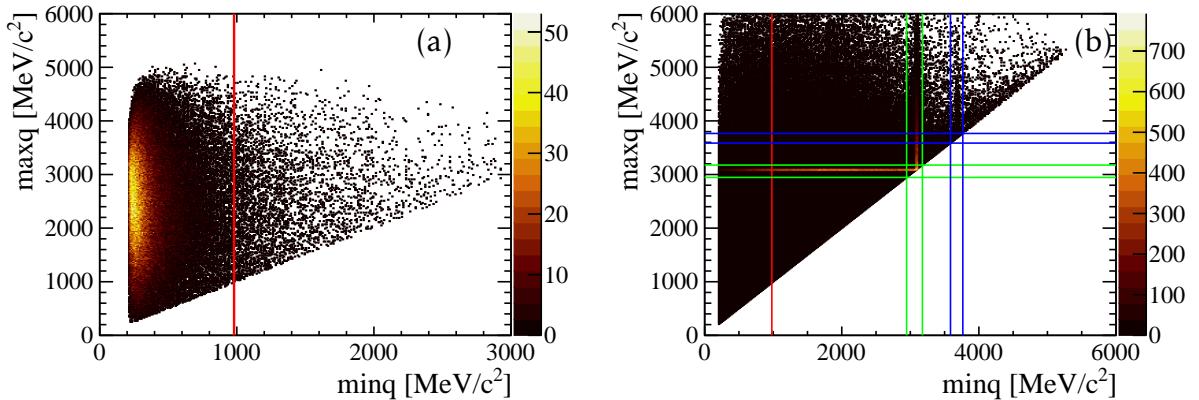
H1t1TrackMuonDecision accepts events where at least one identified muon muon has to pass thresholds on $IP\chi^2$, p_T and p . This favours muons arising from b - and c -hadron decays. There has to be at least one muon (`isMuon==true`) in its final state with certain kinematic thresholds on p and p_T . For example, in 2011 the identified muons that triggered positive decision had to have p above 8 GeV/c [55].

At **HLT2** level, the candidates are required to pass through at least one of the four decisions. **H1t2TopoMu[2,3]BodyBBDTDecision** belong to the *topological triggers* category with an extra requirement of a particle in a candidate being identified by `isMuon` decision. **H1t2DiMuonDetachedDecision** and **H1t2DiMuonDetachedHeavyDecision** reconstruct decays with two muons in a final state. The two lines differ in that they are optimised for heavy and light dimuon pairs respectively. For example, **H1t2DiMuonDetachedDecision** accepts events with dimuon p_T above 1.5 GeV/c and with mass above 1 GeV/ c^2 , whereas **H1t2DiMuonDetachedHeavyDecision** accepts dimuon pairs with any p_T but above 2.95 GeV/ c^2 in mass. The reason why these lines are called detached are because individual muons are required to have high $IP\chi^2$.

5.7 q^2 Selection **ULRIK**

In the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay, two pairs of opposite sign muons can be formed, namely $q^2(\mu_1, \mu_2)$ and $q^2(\mu_2, \mu_3)$ where $\mu_1 = \mu^+, \mu_2 = \mu^-, \mu_3 = \mu^+$. From the two invariant mass squared pairs one can define, $\min q^2 = \min[q^2(\mu_1, \mu_2), q^2(\mu_2, \mu_3)]$ and $\max q^2 = \max[q^2(\mu_1, \mu_2), q^2(\mu_2, \mu_3)]$. This measurement is made in region where $\sqrt{\min q^2} = \min q < 980$ MeV/c² for two reasons: most of the contributions to the amplitude of the decay is below this value and combinatorial background is greatly reduced if $\min q^2 < 1$ (GeV/c²)², see [Figure 5.3](#).

In order to remove backgrounds that proceed via resonant J/ψ and $\Psi(2S)$ contributions, vetoes in invariant mass are placed in the corresponding regions, see [Table 5.4](#) for more details.



[Figure 5.3](#): (a) Signal simulation sample distribution in $\min q$ and $\max q$ variables. Values below 980 MeV/c² (red line) are accepted. (b) Combinatorial data sample after *stripping* selection with no other cuts shows clearly the J/ψ (green) and $\Psi(2S)$ (blue) resonances which are vetoed and the measurement region (red).

Veto	q [MeV/ c^2]
J/ψ	$!(2946.0 < q < 3176.0)$
$\Psi(2S)$	$!(3586.0 < q < 3766.0)$

Table 5.4: Veto for J/ψ and $\Psi(2S)$ resonances. As $\min q < 980 \text{ MeV}/c^2$, these vetoes apply to $\max q$ combination only.

5.8 Further Selection ULRIK

Further selection was performed as seen in [Table 5.5](#). This selection further suppresses backgrounds but is different to what is described above as it requires a different treatment in Run I and Run II due to the different definitions of variables as seen in [subsection 4.1.2](#).

Idea	Object	Run I Selection	Run II Selection
Clean	Muon	-	<code>IsMuonTight==1.0</code>
Clone and ghost	Muon	<code>Nshared==0</code>	<code>Nshared<2</code>
		discussed in subsection 4.1.2	
Fit Region	B	$4000 < M_{B_{corr}} < 7000 \text{ MeV}/c^2$	$4000 < M_{B_{corr}} < 7000 \text{ MeV}/c^2$
		discussed in subsection 5.8.5	
Bkg Removal	event	Combinatorial BDT selection	Combinatorial BDT selection
		discussed in subsection 5.8.3	
Bkg Removal	event	Misid BDT selection	Misid BDT selection
		discussed in subsection 5.8.4	
Optimize FOM	Muon	<code>Probnnmu>0.35</code>	<code>Probnnmu>0.35</code>
		discussed in subsection 5.8.6	
		(FOM defined Equation 5.6)	

Table 5.5: Offline selection performed after stripping. Differences can be seen between Run I and Run II datasets.

The sections below comment on the more exact features of this further selection.

5.8.1 General Features of Multivariate Selections ULRIK

All the multivariate classifiers in the search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay use TMVA’s [56] implementation of Boosted Decision Tree (BDT) with the AdaBoost algorithm. The multivariate selections used in the search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay are the isolation BDT detailed in [subsection 5.8.2](#), combinatorial BDT detailed in [subsection 5.8.3](#) and misid BDT detailed in [subsection 5.8.4](#).

The rationale for these selections is the following. As the background study of inclusive $b\bar{b}$ simulation sample shows that there will be two dominating backgrounds, combinatorial background and misID background. In order to reduce these backgrounds, two consecutive multivariate classifiers are used. The first multivariate classifier is developed to remove efficiently combinatorial background and a second multivariate classifier will help to control the contamination from misID decays. As one of the key variables that provide the greatest separation power in these two multivariate classifiers is another BDT output, the isolation BDT.

Cross-validation is one of the useful methods used within MVAs which improves the chance of good performance of the predictive model on an independent dataset. In this way, biases due to simple sample split into training and testing subsample, could be overcome. In general, it helps also with overfitting when the model of the classifier is sensitive to fluctuations. The cross-validation method used in both combinatorial BDT and misid BDT is known as the *k-folding* technique [57].

In particular, both background and signal samples are randomly split into k similar size subsamples. Then the BDT is trained on the $k - 1$ signal/background subsamples, which are subsequently tested on the remaining last subsamples. This process is repeated k -times for all possible combinations, hence the name of cross-validation. In the last step, the k results produced from k folds are averaged yielding final estimate. In the combinatorial and misid BDT, the number of folds used is $k=10$. Both of the BDT

classifiers the same set of variables listed in [Table 5.6](#).

$B^+ p$	$\min \text{IP} \chi^2$ of all three muons	$\cos(\theta_B)$
$B^+ p_T$	p_T of all three muons	B^+ FD χ^2
B^+ vertex χ^2/ndof	$\min \text{IP} \chi^2$ of all three muons	Isolation variable
B^+ lifetime		(subsection 5.8.2)

Table 5.6: BDT variables used in both combinatorial and misID Run I and Run II BDTs

5.8.2 The Isolation Boosted Decision Tree ULRIK

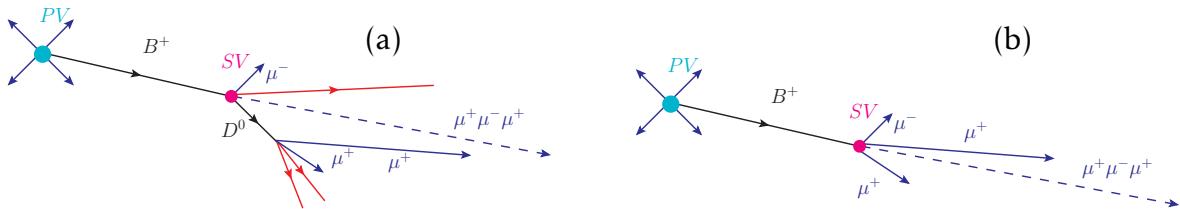


Figure 5.4: An example of decay topology for (a) background and (b) signal.

The vast majority of the backgrounds that share the possibility of contaminating $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ signal have one property in common: they have more tracks associated with the decay. It is hence possible to use multivariate analysis (MVA) techniques to establish how *isolated* the signal trimuon vertex is as compared to background trimuon vertex as seen in [Figure 5.4](#).

The isolation quality of the vertex is determined with BDT. This regression algorithm classifies the event to be more signal-like or background-like according to different track and vertex properties, the *isolation variables*.

The signal proxy for the isolation BDT was trained and tested with $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}$ simulation sample, where all tracks apart from the $p \mu^-$ signal tracks are taken into the account. The background sample was formed with $\Lambda_b^0 \rightarrow (\Lambda_c \rightarrow p) \mu^- \bar{\nu}$ tracks, also disregarding the $p \mu^-$ tracks. The isolation BDT is based on the weights obtained from

these samples, which are computed in [58], and are a result of other's people work. These weights are however then applied parametrically on $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ signal and background proxies, as they share similar topology with respect to isolation properties.

The *isolation variables* include track p_T , the opening angle between track's momentum and momentum of the combined signal/background visible system, the [track \$\chi^2/\text{ndof}\$](#) , the ghost probability of the track P_{ghost} , $IP\chi^2$ of the track with respect to signal/background [SV](#) and [PV](#).

The Isolation BDT response peaks between -1 and 0 for isolated tracks (signal-like) and between 0 and 1 for non-isolated tracks (background-like). The output of this BDT can be seen in [Figure 5.5](#) for both types. Backgrounds shown include combinatorial background and misID type background. In the analysis, there is no explicit selection on this variable, but it is used as one of the input variables for the combinatorial and misid BDTs.

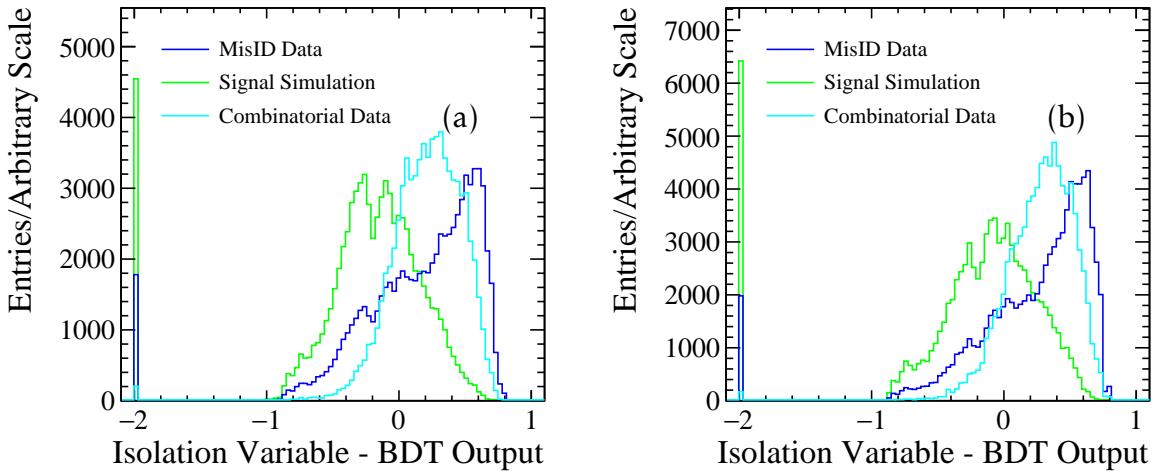


Figure 5.5: Isolation score for signal and backgrounds using (a) Run I (b) Run II samples. If isolation fails to find any other track in the event, by default it gives value -2.

5.8.3 The Combinatorial Boosted Decision Tree ULRIK

One of the most prominent backgrounds is combinatorial background and to reduce its contamination while keeping the signal as high as possible combinatorial BDT is trained. To obtain combinatorial BDT discriminant, a simulated sample for signal and upper mass sideband data sample ($M_{B_{corr}} > 5.5$ GeV/ c^2) for background are used as training and testing samples. These samples passed through preselection, trigger, q^2 selection stage and are using input variables mentioned in [Table 5.6](#).

As the branching fraction and hence the number of signal events is unknown, the metric known as the Punzi figure of merit (FOM) [59], is used to find an optimal working point. It is defined as

$$FOM = \frac{\varepsilon_S}{\sqrt{B + \sigma/2}}, \quad (5.6)$$

where ε_S is the signal efficiency of the selection and B refers to the number of background candidates, σ is the significance. In this case, the significance 3σ is used, but it was checked that there is no change to optimal working point if it is varied to 5σ , as seen in [Figure 5.6](#).

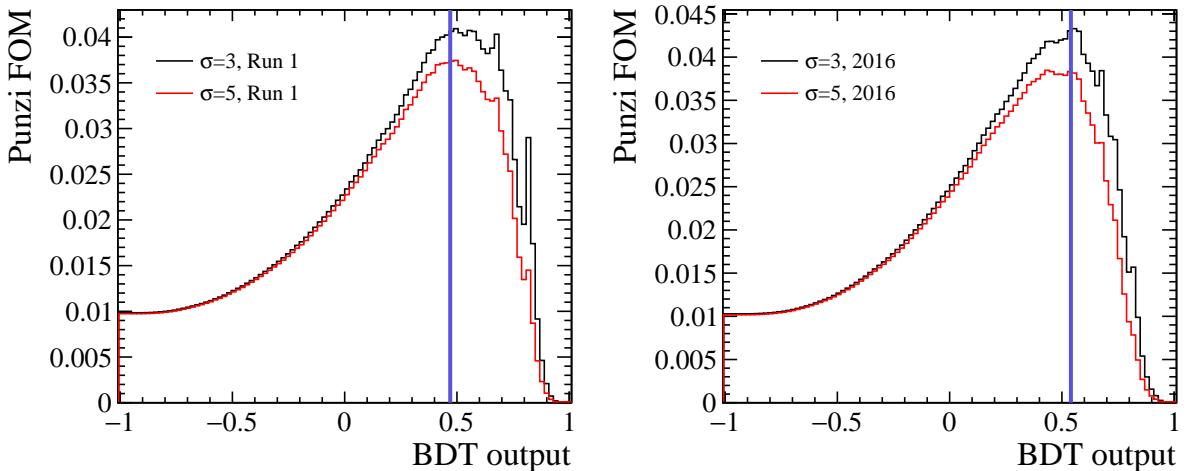


Figure 5.6: Punzi FOM have the optimum working point at 0.47 for Run I and 0.54 for Run II as seen in both figures with a violet line for $\sigma = 3$ and $\sigma = 5$.

The FOM is computed in the mass region, $4.5 \text{ GeV}/c^2 < M_{B_{corr}} < 5.5 \text{ GeV}/c^2$, which is also known as a blinded region. To estimate the number of background candidates in the blinded region, the final fit strategy described in [add reference to fit] is used to fit the data, yielding around 10000 in Run I, and 9000 combinatorial candidates in Run II. The yields are extracted from fits to data integrating the combinatorial part of the total background P.D.F in the blinded region.

In order to accommodate different offline selections between Run I and Run II, separate BDTs are trained for different Runs. Combined training of all of the datasets was also performed but it does not lead to any improvement. Results of the comparison between separate and combined training can be seen in [Figure 5.7](#). Different intrinsic properties (such as the number of trees used) and variables (such as two-particle vertices) have been explored to see whether improvement in discrimination of the BDT can be achieved but the configuration here proves to be the most optimal.

In both Combinatorial BDTs, the most discriminating variables are the isolation variable [subsection 5.8.2](#), B^+ vertex χ^2/ndof , min IP χ^2 of the muons and p_T of B^+ . Combinatorial muon comes more from somewhere else in the event and hence its min IP χ^2 is worse as compared to the signal, making B^+ vertex χ^2/ndof worse. Moreover, as this combinatorial muon comes from somewhere else, other tracks may accompany it making the isolation variable a good discriminant. The combinatorial muon also tends to have higher momentum and hence p_T of B^+ is higher. Distributions for these different variables can be seen in [Figure 5.8](#).

It is also important that there is no skewing of the mass distribution for the background as this could lead later to modelling issues with these different background components. This was checked by looking at the behaviour of BDT output in different bins of $M_{B_{corr}}$. If the BDT value stays flat then the background will not be skewed, which is the case for Run II as seen [Figure 5.9](#). This is also the case in Run I.

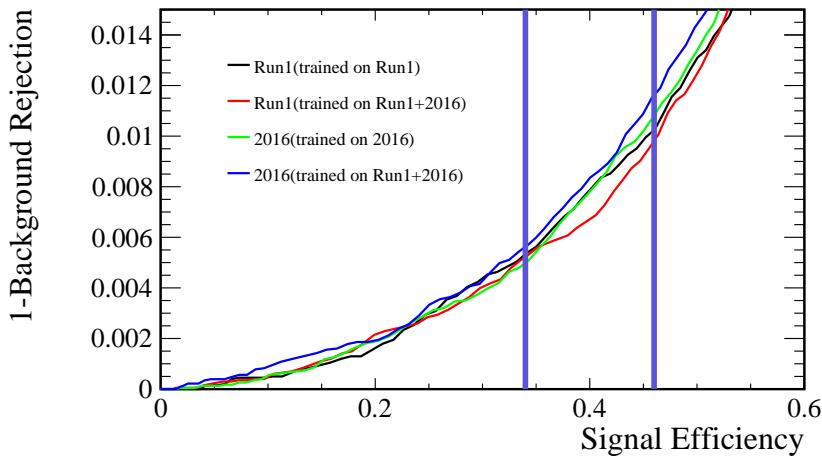


Figure 5.7: Comparison of separate and combined training samples and performance on different datasets. Two vertical violet lines represent optimal points in the signal efficiency, for Run I (0.47) and for 2016 (0.34) where the working point of the two BDTs are chosen. Separate training provides greater rejection power in 2016. In Run I training on both datasets provides comparable performance for given optimal signal efficiency. Taking into the account the fact that offline selection slightly differs for 2016, it is advantageous to keep training separately.

5.8.4 The Misid Boosted Decision Tree ULRIK

In the same way, the classifier that distinguishes well between signal and misID background was developed. The misID sample, that is used for training and testing, was obtained the same way as the signal but with one of the muons not identified as the muon. Rather, this third particle will be identified either as a proton, pion or kaon. talk about the misid parametrisation. As before, Run I and Run II are trained and used separately on the relevant datasets, as shown in [Figure 5.10](#).

Optimisation metric for this classifier was again was Punzi FOM in a blinded region. The Punzi FOM for Run I and Run II as a function of BDT cut can be seen in [Figure 5.11](#) for both significances of $\sigma = \{3, 5\}$.

To obtain the number of background events, the default fitting strategy for misID

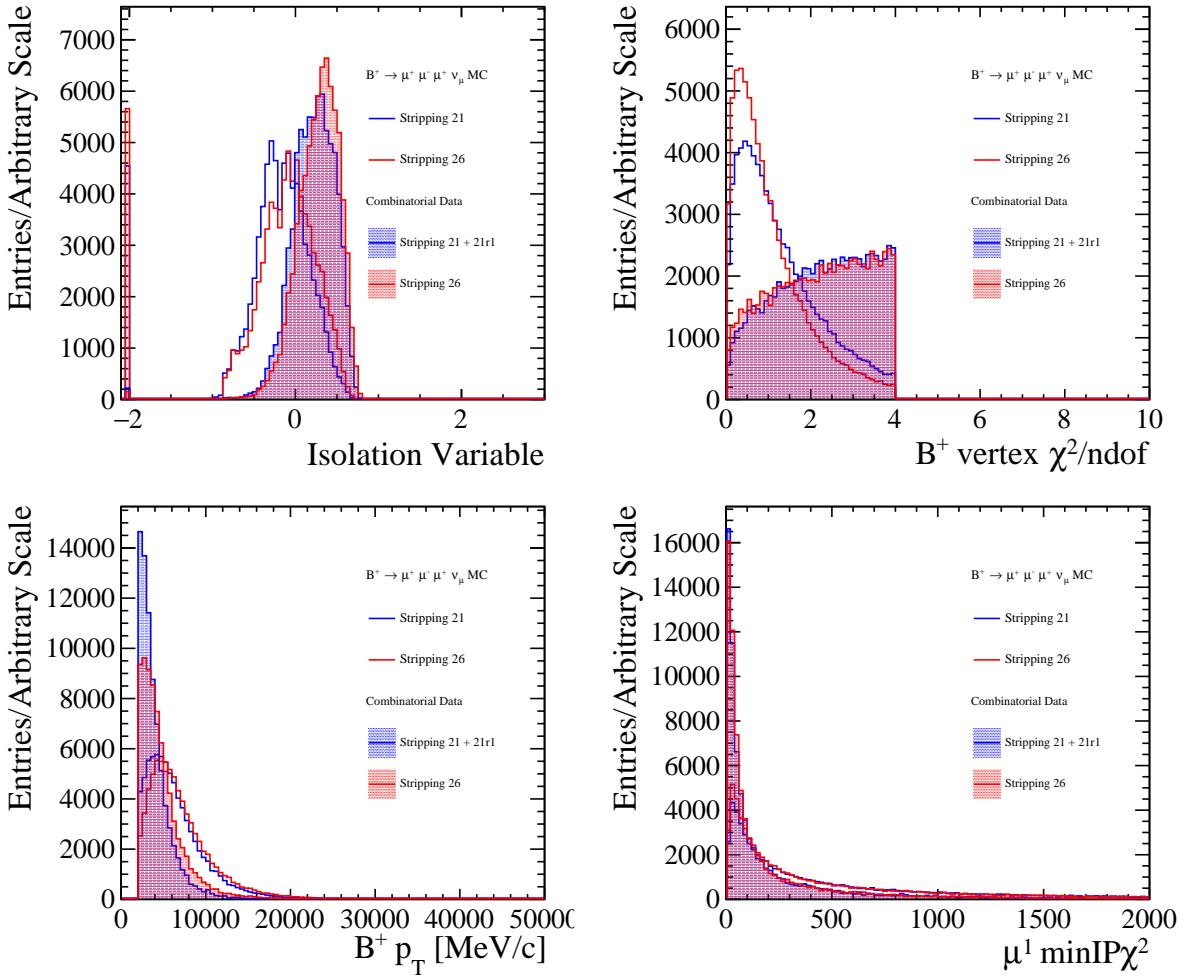


Figure 5.8: The variable with the most discriminative power for both Run I and 2016.

is used [talk about the misid fit], where the total yield need to be multiplied by 100 in order to counter balance the prescale used at pre-selection stage. To obtain the yield, the binned χ^2 fit [forw ref misid]. The binned χ^2 fits to the misID templates are shown in Figure 5.12 yielding 2400 unparametrized misID candidates in Run I polluting the signal window in the prescaled sample, and 2200 Run in II.

The misID background can proceed also through combination with random muon and hence by applying the combinatorial BDTs on the misID samples, this "combinatorial" component in the misID samples should be reduced and misID samples that are left should consist of true cascade decays. This can be seen in Figure 5.13.

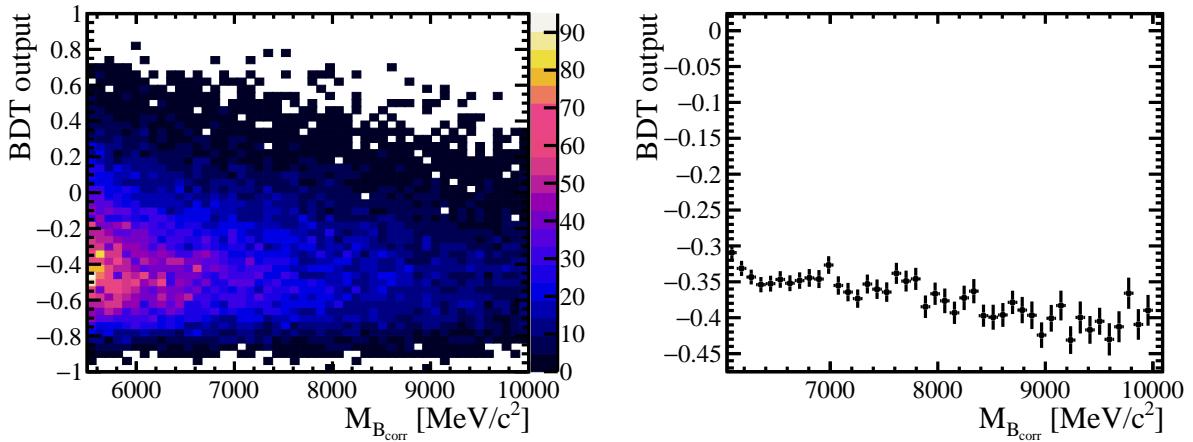


Figure 5.9: Study of linear correlation between BDT output and $M_{B_{corr}}$ and BDT value for each bin of $M_{B_{corr}}$ in 2016 shows that Combinatorial BDT is relatively flat as a function $M_{B_{corr}}$.

The most powerful variables that distinguish the signal from misID background are the kinematic properties of the misidentified muon, namely p_T , p and $\text{min IP} \chi^2$. Misidentified muons tend to be softer than in the signal case as they come from cascades via D^0 and its excited states. The $\text{min IP} \chi^2$ distribution will be different as the misid muon can proceed D^0 , rather than directly from the B . The kinematic distributions are also different for the two real muons in signal and background misid. The real muon that has the same charge as B tends to be softer for the signal case whereas the real muon that has opposite charge proceeding via D will be harder, as seen Figure 5.14.

5.8.5 Fitting Region Selection

Because the signal mass distribution is expected to be in a more narrow window around corrected B^+ mass peak and the exponential description of combinatorial background is not correct below 4000 MeV/ c^2 , as shown in [forwardref], the fitting region selection in which the measurement will be made is $4000 \text{ MeV}/c^2 < M_{B_{corr}} < 7000 \text{ MeV}/c^2$.

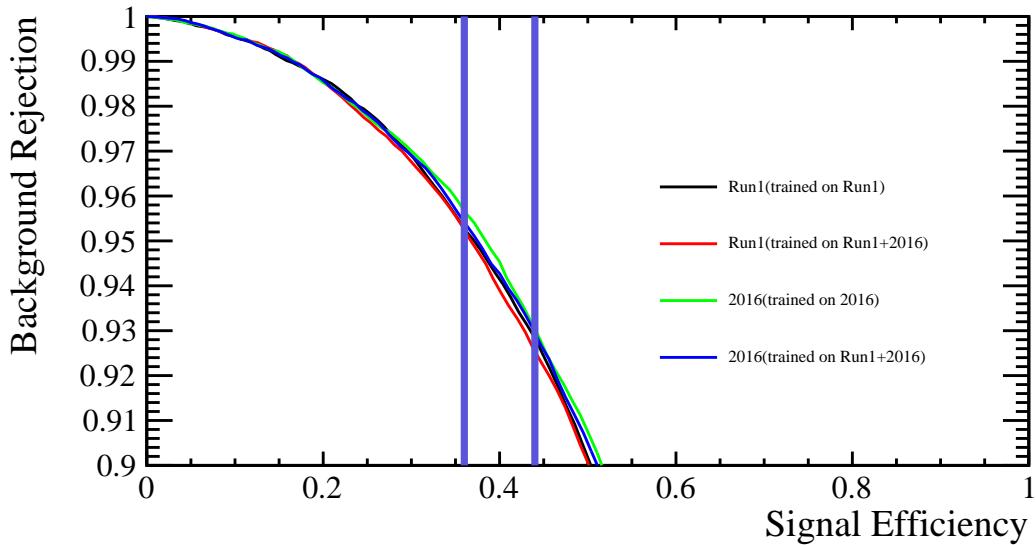


Figure 5.10: Comparison of separate and combined training samples and performance on different datasets. Optimal working point is chosen, see Figure 5.11 and its corresponding signal efficiency in Run I is 0.44 and for 2016 0.37 denoted with a violet line. As the performance is better for 2016 when the training is performed separately, the training is kept separately also to be consistent with previous methodology.

5.8.6 Further PID Selection **ULRIK**

After classifiers to reduce combinatorial and misID backgrounds are trained and applied and the fitting region is defined, further PID selection is performed. This can be done as the preselection had relatively loose DLL requirements and hence it is possible to improve the performance by using cuts on additional PID variables. In the optimisation procedure, different hypotheses were tested, such as cuts on Probnnmu, Probnnpi, and ProbnnK variables and their combinations. The optimisation was performed in a such a way as to optimize Punzi FOM with $\sigma = 3$ (Equation 5.6) in a blinded signal region, by performing full blinded data fit [make reference to this once it is written], but with fits to Run I and Run II data separately.

For each PID hypothesis, Punzi FOM was calculated. In both cases, in I and II $\text{Probnnmu} > 0.35$ yielded highest Punzi FOM.

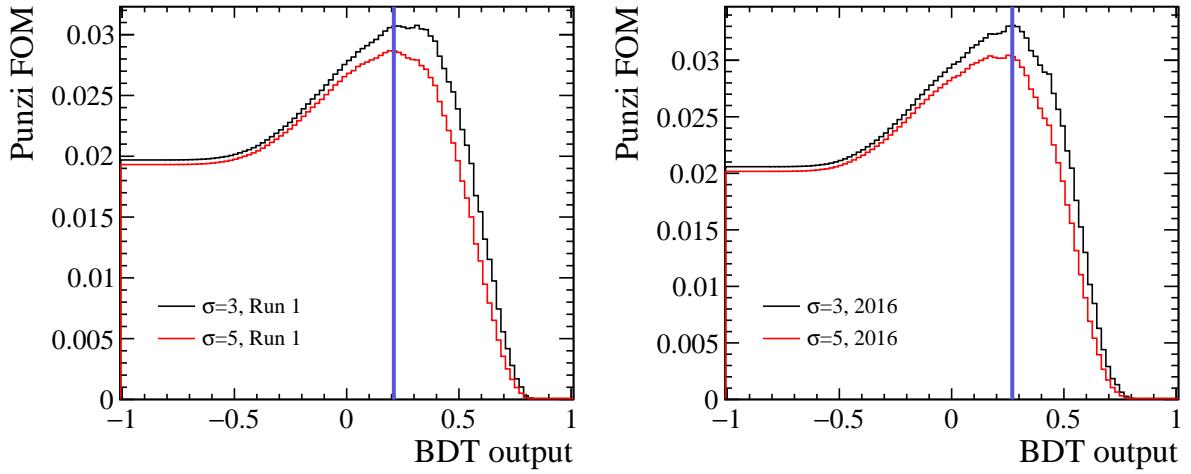


Figure 5.11: Punzi FOM have the optimum working point at 0.21 for Run I and 0.27 for Run II as seen in both figures with a violet line for $\sigma = 3$ and $\sigma = 5$.

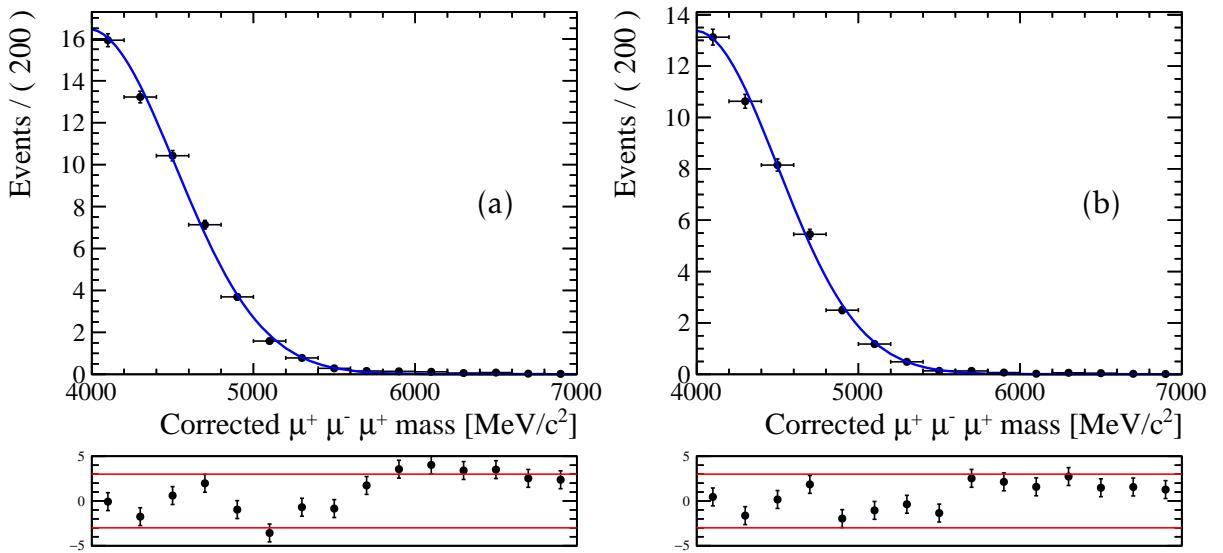


Figure 5.12: (a) Run1 (b) 2016 binned χ^2 fit to misID sample yielding estimates for the number of background events.

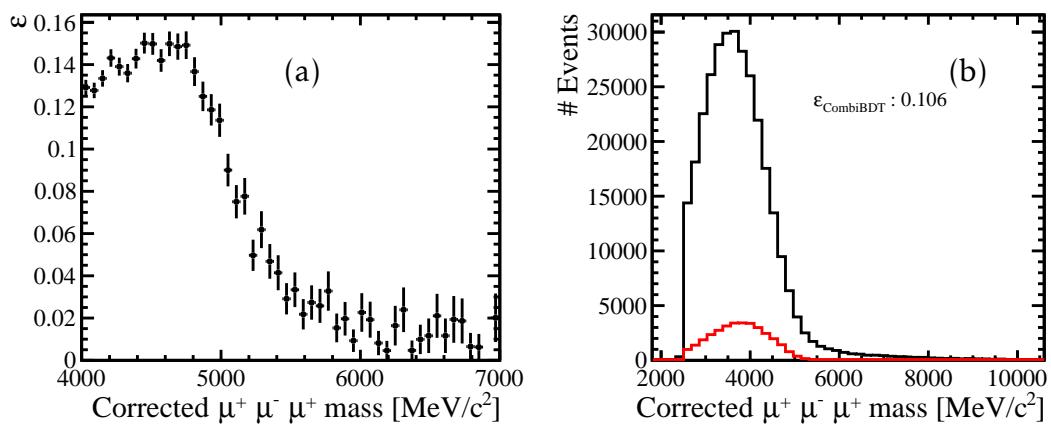


Figure 5.13: (a) The efficiency of applying 2016 combinatorial BDT at optimal working point on the 2016 SS misID sample. It can be seen that (b) combinatorial component of the misID sample has been significantly reduced, where red curve is the distribution after applying the cut.

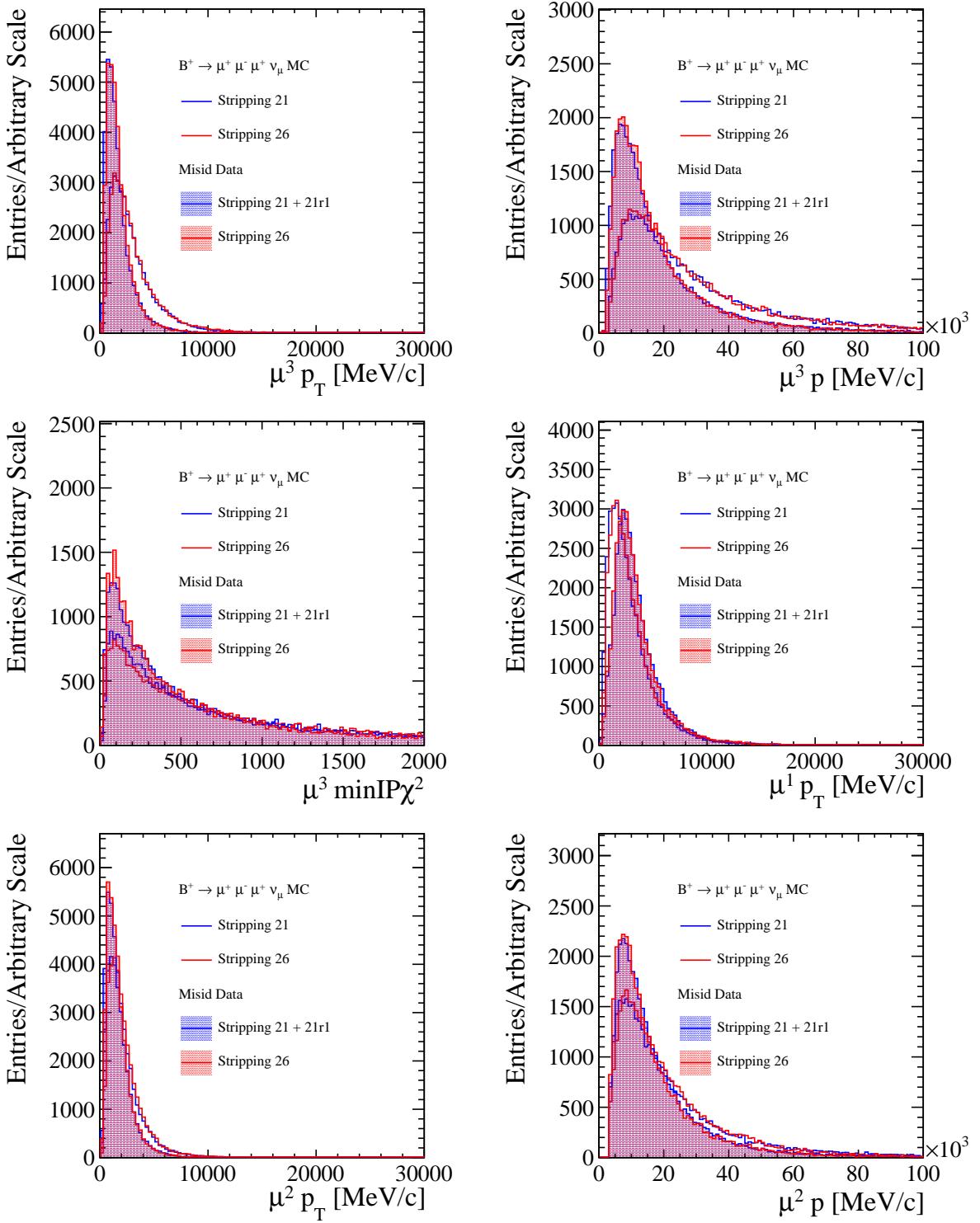


Figure 5.14: The variables with the most discriminative power for both misid Run I and II, mostly kinematic properties of different muons.

5.9 Normalisation channel **ULRIK**

The normalisation channel used in this analysis is $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ as it is clean, well understood, and well-populated channel that is similar to signal. This means that many systematic uncertainties will cancel. Normalising the signal decay to this decay also means that absolute efficiencies, luminosity, the b-quark cross-section or fragmentation fractions will be cancelled. With the same number of tracks, it will also give a small uncertainty in the tracking efficiency. There are, however, few differences in the selection that need to be underlined.

Firstly, the preselection stream from which this sample is taken has different requirements seen in [Table 5.7](#). As compared to preselection of the signal channel shown in [Table 5.2](#), this preselection is less tight. To unify and impose same kind of preselection so that the tracks chosen are of a good quality and away from [PV](#), the preselection for signal channel is applied on the top of the original preselection.

Candidate	Stripping Selection
muon	$p_T > 500$ MeV
muon	$DLL_{\text{mu}} > 0$
kaon	$PT > 500$ MeV
kaon	$\text{track } \chi^2/\text{ndof} < 5$
kaon	$DLL_K > 0$
dimuon	$\text{vertex } \chi^2/\text{ndof} < 16$
dimuon	$ M(\mu^+, \mu^-) - M_{PDG}(J/\psi) < 80$ MeV/ c^2
combination	$\text{vertex } \chi^2/\text{ndof} < 10$
combination	$5150 \text{ MeV}/c^2 < M_B < 5450 \text{ MeV}/c^2$
combination	B lifetime > 0.2 ps

Table 5.7: Original preselection of events for normalisation channel for $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ for Run I and Run II.

Secondly, this decay proceeds via J/ψ resonance and hence the q^2 veto for J/ψ and $\Psi(2S)$, listed in [Table 5.4](#), is not applied but rather reversed as seen in [Table 5.7](#). As the third particle is kaon rather than muon, there is no explicit choice of $minq$ region.

And finally the kaon candidates are required to have [PID](#) criteria consistent with being a kaon. In addition to preselection already imposing $DLLK > 0$, $DLLp - DLLK < 5$ is required to make a distinction with protons. To assure that the kaon is not confused with the muon, $IsMuon == 0.0$ is imposed. But only kaon tracks, which have the properties that they could be within geometrical muon acceptance, $InMuonAcc == 1.0$, are considered.

5.10 Fractional Corrected Mass Error (FCME) Window

Split **ULRIK**

Information about a particular fitting procedure put in place is discussed in this section. In order to increase sensitivity, but not to decrease statistics as all the previous selection leads to a low-statistics regime, it was decided that the fitting procedure for the final fit will be in two bins of fractional corrected mass error (FCME), defined as

$$\sigma_{\{\text{lowFCME}, \text{highFCME}\}} = \frac{\delta}{M_{B_{corr}}}, \quad (5.7)$$

where δ is the corrected mass error. Because the corrected mass error has clear dependence on resolution (see [Figure 5.15](#)), this split will split the data into two bins of resolution increasing the sensitivity for observation.

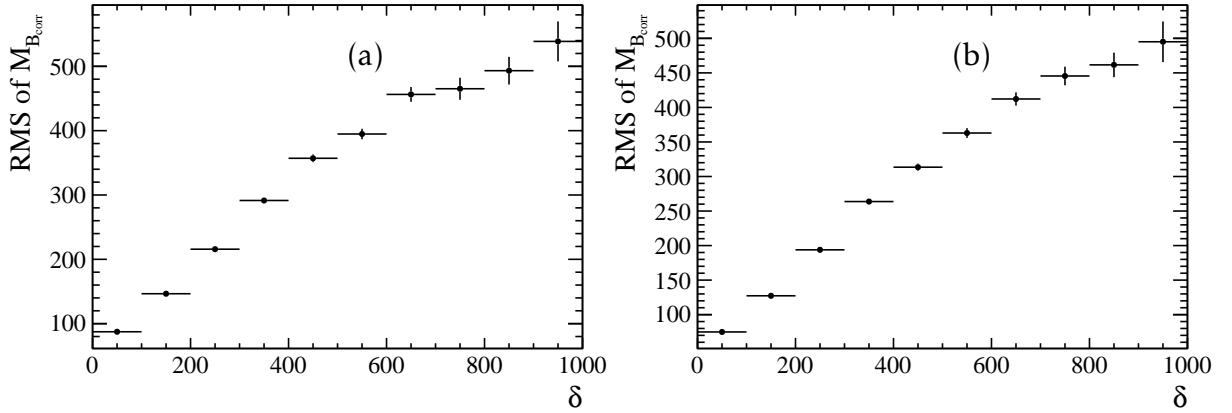


Figure 5.15: (a) The resolution of signal simulation with stripping 21 in bins of corrected mass error δ . (b) The resolution of signal simulation with stripping 26 in bins of corrected mass error δ .

The split boundary was chosen in such a way as to keep $\sim 50\%$ of signal in σ_{lowFCME} and $\sim 50\%$ signal in high σ_{highFCME} . Numerically this corresponds to

$$\sigma_{\{\text{lowFCME}, \text{highFCME}\}} = \begin{cases} \sigma_{\text{lowFCME}} & \text{if } \frac{\delta}{M_{B_{corr}}} < 0.0225, \\ \sigma_{\text{highFCME}} & \text{if } \frac{\delta}{M_{B_{corr}}} > 0.0225. \end{cases} \quad (5.8)$$

However, in order to look at the consistency of this two bin strategy also one bin of fractional corrected mass error strategy is performed and will be denoted as σ_{NOFCME} .

Chapter 6

Background Studies

The decay $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ is a fully leptonic decay with good potential for eliminating many types of the backgrounds. In this chapter parametrisation and estimations for all considered backgrounds are sketched. A quick summary of all backgrounds that passed the stringent selection were provided in section 5.2.

In this chapter summary of considered backgrounds is provided where combinatorial background is described in section 6.1, misidentified background in section 6.2, different classes of partially reconstructed background are considered in section 6.3 and finally rare and resonant backgrounds considered in section 6.4.

6.1 Combinatorial Background

Combinatorial background comes around as a random combination of tracks from different b -decay chains. The usual method at LHCb of estimating the amount and the shape of this background include extrapolation from the upper mass data sideband to the signal region. In this case the upper mass sideband is defined as $M_{B_{\text{corr}}} > 5500 \text{ MeV}/c^2$ and the signal region is defined to be $4500 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 5500 \text{ MeV}/c^2$. The characteristic shape for this background can be described by exponential function up to certain point, where this point is the primary discussion of this section. Since tight selection resulted in low-statistics data samples the extrapolation from upper mass

sideband introduces a big uncertainty on the exponential constant and cannot be used to estimate the correct shape and yield of this background. What can be done, however, is to assume the exponential shape for combinatorial component and let the exponential constant be a floating parameter in the data fit. This method for estimation of the combinatorial component will be mentioned in signal data mass fits ??.

In the rest of this section exponential parametrisation of this background between $4000 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 7000 \text{ MeV}/c^2$ is motivated. This is important as the final fitting region was chosen in such a way as to make sure that combinatorial background is exponential in this entire fitting region.

Apart from the nominal upper mass data sideband sample, two other samples are analysed as proxies for this type of background. Despite the fact that these samples are also scarcely populated they are studied altogether to determine in which mass regions the combinatorial background can be considered exponential. Firstly, same sign data sample was studies (the same sample as in [section 4.2](#)), where this sample consists of $\mu^+ \mu^+ \mu^+ \nu$ events passing all selection up to MVA selection to have sufficient statistics. Secondly an inclusive $b\bar{b}$ simulation sample consisting of events where two muons with $p > 3 \text{ GeV}/c$ are required to be present alongside with a third muon. On the top these events have to satisfy all the stripping selection outlined in [Table 5.2](#).

As seen in [Figure 6.1\(b\)\(c\)](#), the exponential component stops rising at $M_{B_{\text{corr}}} = 4000 \text{ MeV}/c^2$. Hence the choice of fitting region $4000 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 7000 \text{ MeV}/c^2$.

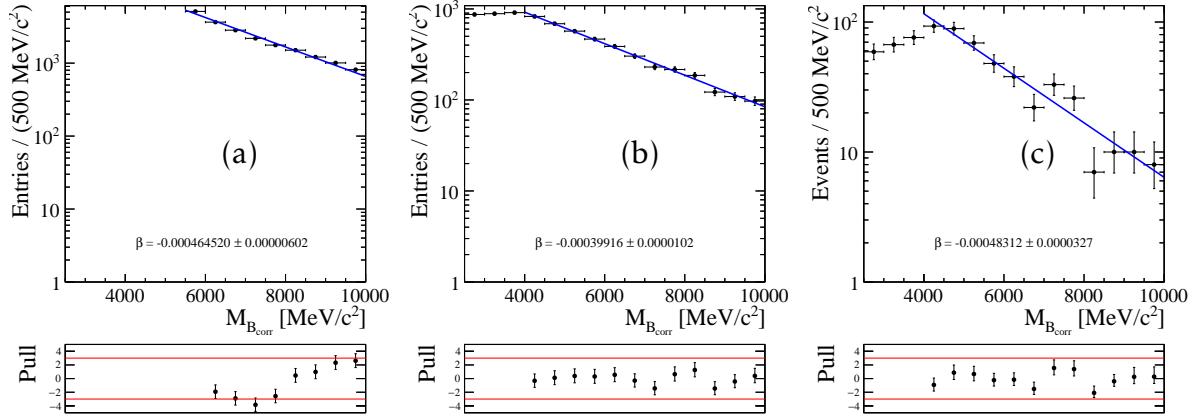


Figure 6.1: (a) Fit to upper mass side band just before application of MVA selection. (b) Fit to $\mu^+ \mu^+ \mu^+ \nu$ same sign sample. (c) Fit to $b\bar{b}$ sample with exponential function. In (b)(c), exponential description is not correct below $4000 \text{ MeV}/c^2$. All plots contain exponential constants.

This background is heavily suppressed with dedicated MVA selection described in subsection 5.8.3.

6.2 MisID type background

MisID background is one of the most prominent backgrounds that is expected to be present. This type of background proceeds mostly via cascade decays, where $B^+ \rightarrow (\bar{D}^0 \rightarrow hX\mu^-\nu)\mu^+\nu$ and then $h \in [K^+, \pi^+]$ are misidentified as muons. The contributions from decays where two muons are correctly identified as muons and third track is consistent with proton passing all the selection is also considered, however this contribution is very limited.

As discussed in section 5.2 there are two possibilities for the charge for the misidentified background. In one case the sign of the misidentified particle agrees with the sign of the mother B , *SS misID* background. The opposite case is denoted as *OS misID* decays, which arises less often as it requires decays with more additional particles. These two types of backgrounds are studied using data-driven method described below. Finally,

also double misID employing same data-driven methods was studied, where there are two hadrons misidentified as muons, however, the double misID contribution proved to be insignificant.

To determine the amount and the shape of the misID background, data sample with the same selection as for the signal sample is obtained with one marginal difference - **no PID cut** on one muon, either positive or negative. As the muon misID rate is different for pions and kaons [40], the species of the hadron, h must be determined at first. The strategy for this purpose is to isolate the hadron into separate hadron **PID** regions, and to determine the cross-feed of one region into the other. For this, an iterative procedure as shown in [Figure 6.2](#) is applied ignoring insignificant proton cross-feed. This iterative procedure hence splits the misidentified data sample into **PID** regions, where the hadron candidate is consistent with the kaon, pion and proton hypotheses. For this procedure, probabilities of identifying a given species with given **PID** requirement are taken from dedicated control samples in **PIDCalib** package [50] discussed in [subsection 4.3.1](#). The **PID** performance is highly dependant on kinematic properties of the misidentified particle and hence the estimation is performed in bins of momentum p and pseudorapidity η . At the beginning of the procedure, the number of misidentified events of given species is assumed to be zero, and the cross-feed between regions is calculated assuming that the pion, kaon and protons regions are pure pions, kaons and protons. This procedure then corrects the distributions by taking into the account this initial cross-contamination. This procedure is repeated until the number of total misidentified particles does not change significantly from one iteration to another.

Once the cross-feed between the different hadron species has been taken into account, the final step is to calculate the probability for a specific hadron to pass the stringent muon **PID** requirements applied in the analysis. The presence of the two real muons in the $\mu^+ \mu^- hX$ background increases probability to misidentify the hadron as a muon, mainly due to sharing of hits in the muon stations. Therefore the hadron misID probability is obtained from a dedicated control sample designed to emulate the topology of the mis-identified background, $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^*(\rightarrow K^+ \pi^-)$, as shown

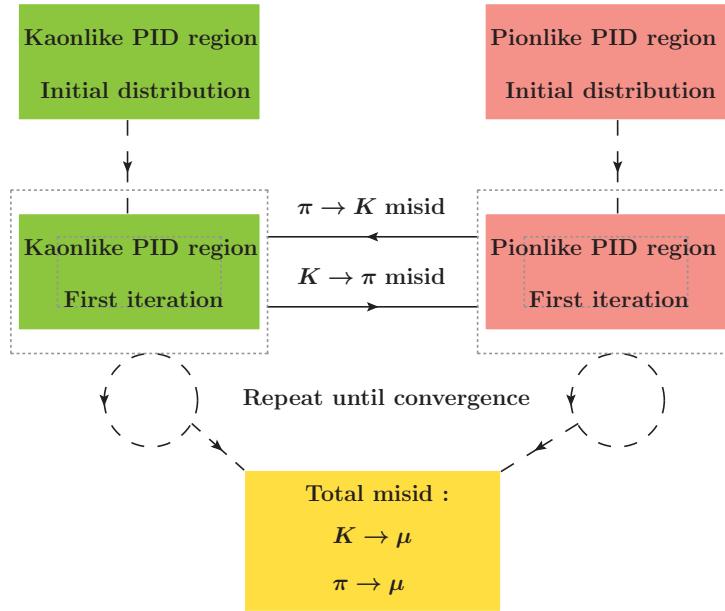


Figure 6.2: Diagram of the iterative procedure to establish contamination from decays where pion and kaon are misidentified for muon.

in subsection 4.3.4.

This process can be summarized mathematically in a following way:

- The proton-, pion- and kaon-like regions are defined in Table 6.1.

Region	PID cuts
Proton-like	$DLL_p > 5, DLL_p - DLL_K > 5$
Kaon-like	$DLL_K > 0, DLL_p - DLL_K < 5$
Pion-like	$DLL_K < 0, DLL_p < 5$

Table 6.1: Species region definitions.

- ID efficiencies are obtained from PIDCalib in bins of p, η for all three regions.
- MisID efficiencies are obtained from the specific calibration sample compensating having two other muons in the sample in subsection 4.3.4 in the bins of p, η .

- In order to account for cross-contamination between the kaon and pions species the following procedure is applied:
 - The data in each region is binned to obtain two dimensional $N(p, \eta)$ distributions, where p is momentum and η is pseudorapidity. The true kinematical distributions for kaons and pions are given by

$$n(p, \eta)_{\pi/K}^0 = \frac{N(p, \eta)_{\pi/K}}{\epsilon(p, \eta)_{\pi/K}}. \quad (6.1)$$

where $\epsilon(p, \eta)_{\pi/K}$ are efficiencies obtained from PIDCalib tables.

- To correct for the cross-feed between pion and kaon regions, following algorithm which corrects the original distribution is applied:

$$n(p, \eta)_\pi^{i+1} = \frac{N(p, \eta)_\pi - M(p, \eta)_{K \rightarrow \pi} n(p, \eta)_K^i}{\epsilon(p, \eta)_\pi}, \quad (6.2)$$

$$n(p, \eta)_K^{i+1} = \frac{N(p, \eta)_K - M(p, \eta)_{\pi \rightarrow K} n(p, \eta)_\pi^i}{\epsilon(p, \eta)_K}. \quad (6.3)$$

Here, $n(p, \eta)_\pi^i$ $n(p, \eta)_K^i$ together with the misID binned efficiencies $M(p, \eta)_{K \rightarrow \pi}$ and $M(p, \eta)_{\pi \rightarrow K}$ are estimating the-cross contamination between two regions.

- At each iteration, the total number of misID particles of the type $\pi \rightarrow \mu$ and $K \rightarrow \mu$ events are given by

$$\sum_{p, \eta} n(p, \eta)_\pi^i M(p, \eta)_{\pi \rightarrow \mu} \quad (6.4)$$

$$\sum_{p, \eta} n(p, \eta)_K^i M(p, \eta)_{K \rightarrow \mu} \quad (6.5)$$

- This procedure is repeated until the change in total misID between iterations is less than 0.1%. Typical number of iterations depends on the size of the sample. For big samples the convergence is achieved after two or three iterations. For small samples this is achieved after six iterations on average.

- For each event in both kaon-like and pion-like sample, $w_{cross-feed}$ = probability of being misidentified particle including the cross-contamination correction is calculated

$$w_{cross-feed} = \frac{n(p, \eta)_\pi^{final} \times M(p, \eta)_{\pi \rightarrow \mu}}{N(p, \eta)_\pi^0}, \quad (6.6)$$

$$w_{cross-feed} = \frac{n(p, \eta)_K^{final} \times M(p, \eta)_{K \rightarrow \mu}}{N(p, \eta)_K^0}. \quad (6.7)$$

- The number of misidentified events and the shape are obtained by reweighting the pion-like and kaon-like datasets by the $w_{cross-feed}$.

Examples of misID distributions with unweighted, weighted by probability with no cross-feed correction, and weighted with cross-feed correction can be seen in [Figure 6.3](#) for the *SS misID* and [Figure 6.4](#) for the *OS misID*. These are the misID distributions before misID BDTs are applied, which minimize the contamination of this background as discussed in [subsection 5.8.4](#).

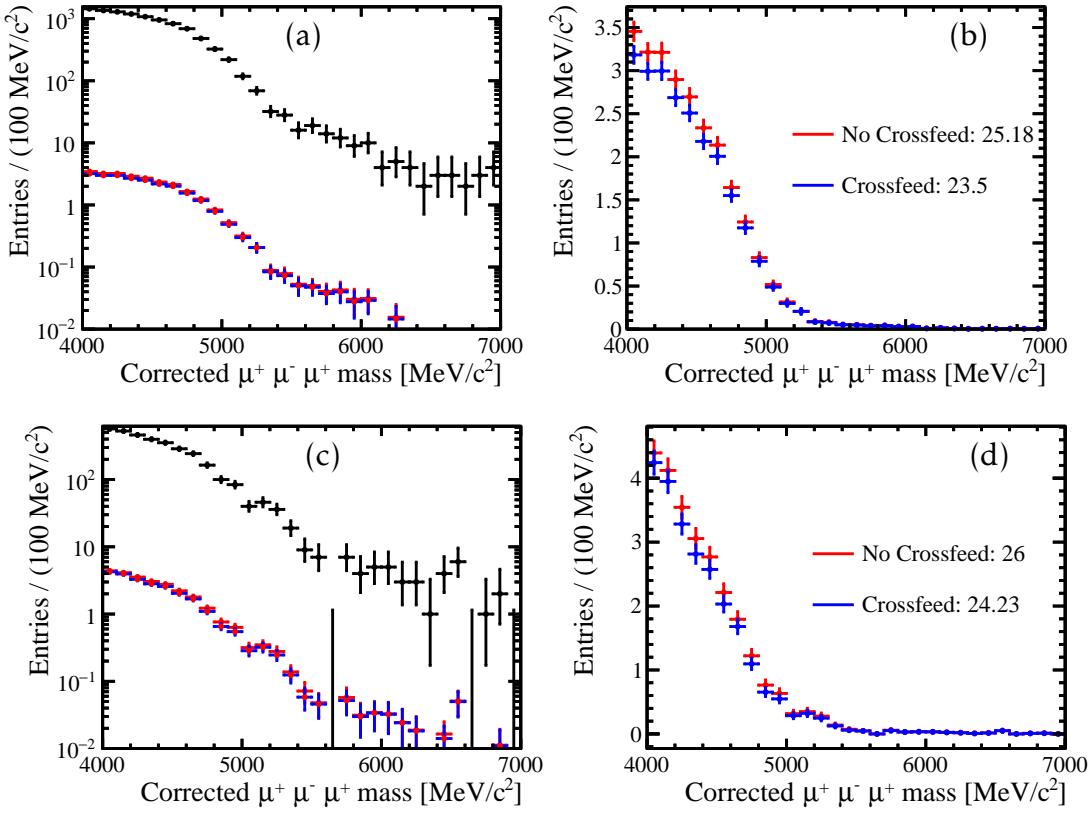


Figure 6.3: Examples of distributions where misID procedure is applied to obtain yields and shapes for Run I before misID BDT was applied. On the left, unweighted misID distributions (black), weighted with no cross-feed misID distributions (blue) and weighted misID distributions with cross-feed (red) for (a) kaon SS (c) pion SS . On the right, only weighted misID distributions for Run I (b) kaon SS (d) pion SS are shown together with the yield estimates. These shapes are obtained after combinatorial BDT was applied, but before misid BDT was applied. Total yields need to be multiplied by 100 to counteract the prescale that was applied on this data.

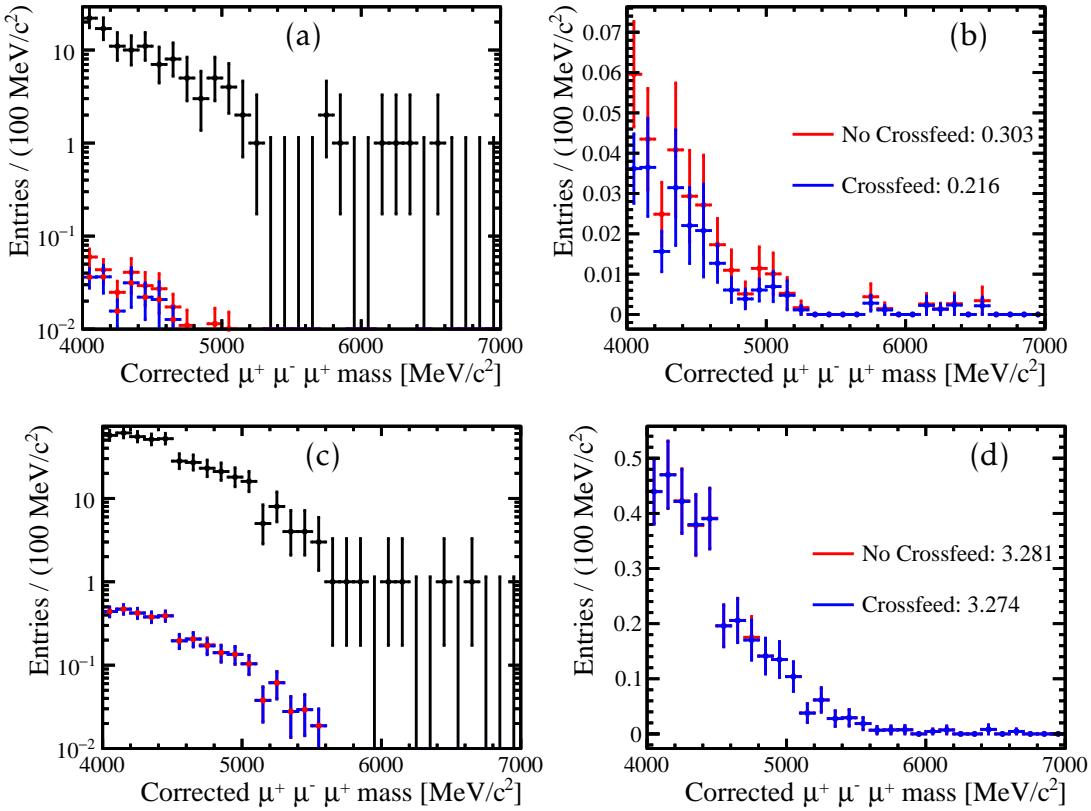


Figure 6.4: Examples of distributions where misID procedure is applied to obtain yields and shapes for Run I before misID BDT was applied. On the left, unweighted misID distributions (black), weighted with no cross-feed misID distributions (blue) and weighted misID distributions with cross-feed (red) for (a) kaon OS (c) pion OS. On the right, only weighted misID distributions for Run I (b) kaon OS (d) pion 0S are shown together with the yield estimates. These shapes are obtained after combinatorial BDT was applied, but before misid BDT was applied. Total yields need to be multiplied by 100 to counteract the prescale that was applied on this data.

6.3 Partially Reconstructed Background

Partially reconstructed background can occur by missing or misidentifying one or more particle tracks in the decay. The common feature for this type of backgrounds is that

the corrected or reconstructed mass of the B will be lower than in the signal case.

In order to estimate both the shape and the size of the partially reconstructed backgrounds, one of the most dangerous example is studied: $B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu\nu$. The expected $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu^+\nu)$ is obtained by amalgamating $\mathcal{B}(D^0 \rightarrow K^-\pi^+\mu^+\mu^-) = (4.17 \pm 0.12 \pm 0.40) \times 10^{-6}$ [60] and $\mathcal{B}(B^+ \rightarrow Dl^+\nu X) = (9.8 \pm 0.7) \times 10^{-2}$ [1] yielding $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu^+\nu) \approx (4.10 \pm 0.50) \times 10^{-7}$.

The shape of this background is investigated with inclusive simulation samples containing also higher excited resonances of D^{*0}, D_2^{*0} , and so on. This simulation has one imperfection: it has two charged pions rather than muons coming from the D^0 decay, which are reconstructed as signal. In this study the effect of missing particles on the corrected mass shape is investigated hence these two pions become good proxies for the muons given the muon and pion have very similar mass. The only problem that could arise is if the selection efficiency was not constant as a function of the dipion mass, $M(\pi^+\pi^-)$, as this would lead to shaping of the background, potentially underestimating the contributions from the resonant ω and ρ region, which are present with the two muons.

For this reason all muon cuts from selection apart from the PID are also applied to pions. Relative efficiency ratios including all the efficiencies after MVA stage are obtained, where for signal the total selection efficiency is $\epsilon_{B^+ \rightarrow \mu^+\mu^-\mu^+\nu}^{total} = (2.63 \pm 0.03) \times 10^{-3}$ and for partially reconstructed background $\epsilon_{B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu^+\nu}^{total} = (6.82 \pm 0.07) \times 10^{-4}$. Assuming the branching fractions for $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) = 1 \times 10^{-8}$ and for $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu^+\nu) = (4.10 \pm 0.50) \times 10^{-7}$, relative contamination between signal and partially reconstructed background results in [Figure 6.5\(a\)](#).

To check the fact that there is no dangerous shaping of the background using this particular proxy simulation sample the full selection efficiency in bins of dipion mass is plotted. The efficiency flatness shown in [Figure 6.5\(b\)](#) means that this selection does not have model dependence and hence it is safe to use for shape estimates for the partially reconstructed backgrounds.

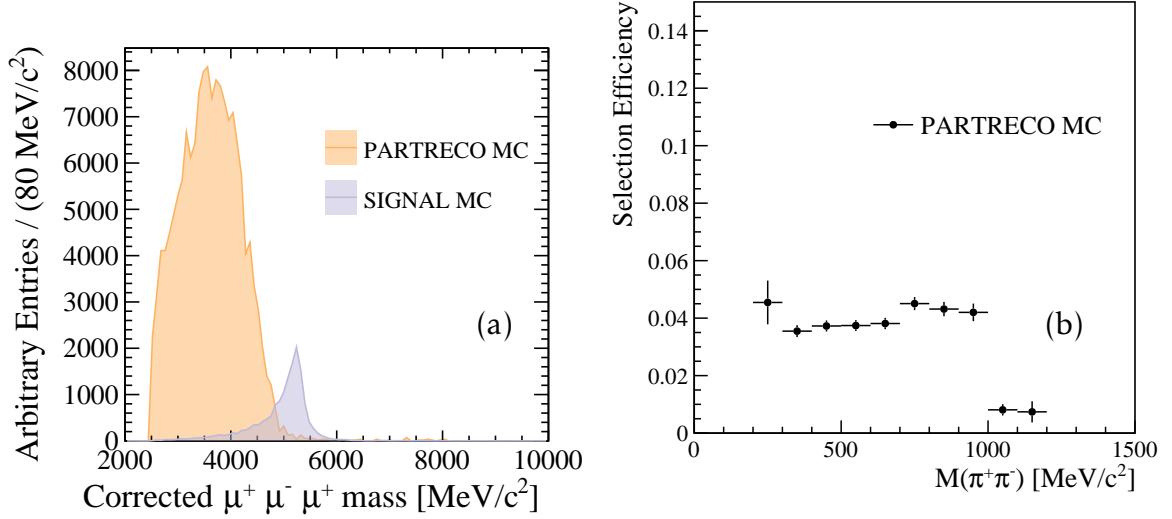


Figure 6.5: (a) Signal and partially reconstructed background distributions scaled to their expected ratio after full MVA selection assuming following branching fractions: $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = 1 \times 10^{-8}$ and $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu) = (4.10 \pm 0.50) \times 10^{-7}$. (b) Full selection efficiency as a function of invariant mass of the proxy pions is constant.

The most powerful part of selection that eliminates this part of the background is isolation as partially reconstructed background decay had usually has more tracks. In order to estimate contamination in the final fit, normalisation with respect to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ is used as shown in section 7.3.3.

6.3.1 Partially Reconstructed Backgrounds, where $D^0 \rightarrow \eta/\eta' X$, where $\eta/\eta' \rightarrow \mu\mu\gamma$

In the previous partially reconstructed sample, the background that proceed via η/η' from D resonance is not considered, as it is not part of the inclusive simulation. The selection efficiency of such decays is expected to have very similar values as in the partially reconstructed sample proxy, because the reconstructed particles are the same.

In this section, the total estimate for the branching fraction of partially reconstructed backgrounds proceeding via η/η' from D resonances is computed. Full inclusive rate

$\mathcal{B}(D^0 \rightarrow \eta/\eta'X)$ is $\sim 10\%$). However, the most relevant decay chains are the ones where the mass of the missed particle is small. This is because if only light particle is missed, the shape of corrected mass of partially reconstructed background comes closest to the signal region. Such decay chains are considered in [Table 6.2](#).

It can be seen that total cumulative contribution is much smaller than the one considered with $D^0 \rightarrow K^+\pi^-\mu^+\mu^-$, where $\mathcal{B}(D^0 \rightarrow K\pi^+\mu^+\mu^-) = (4.17 \pm 0.12(\text{stat}) \pm 0.40(\text{syst})) \times 10^{-6}$ [60]. No further consideration hence is necessary for this type of decay.

Process	\mathcal{B}	Contribution to $\mathcal{B}(D^0 \rightarrow (\eta/\eta' \rightarrow \mu\mu\gamma)X)$
$\mathcal{B}(\eta \rightarrow \mu\mu\gamma)$	$(3.10 \pm 0.40) \times 10^{-4}$	-
$\mathcal{B}(\eta' \rightarrow \mu\mu\gamma)$	$(1.08 \pm 0.27) \times 10^{-4}$	-
$\mathcal{B}(D^0 \rightarrow \eta'\pi^0)$	$(9.10 \pm 1.40) \times 10^{-4}$	$(9.80 \pm 2.90) \times 10^{-8}$
$\mathcal{B}(D^0 \rightarrow \eta'\pi^+\pi^-)$	$(4.50 \pm 1.70) \times 10^{-4}$	$(4.90 \pm 2.20) \times 10^{-8}$
$\mathcal{B}(D^0 \rightarrow 2\eta)$	$(1.70 \pm 0.02) \times 10^{-3}$	$(5.30 \pm 0.70) \times 10^{-7}$
$\mathcal{B}(D^0 \rightarrow 2\eta)$	$(1.70 \pm 0.02) \times 10^{-3}$	$(5.30 \pm 0.70) \times 10^{-7}$
$\mathcal{B}(D^0 \rightarrow \underline{\eta}\eta')$	$(1.06 \pm 0.27) \times 10^{-3}$	$(3.30 \pm 0.90) \times 10^{-7}$
$\mathcal{B}(D^0 \rightarrow \eta\underline{\eta}')$	$(1.06 \pm 0.27) \times 10^{-3}$	$(1.10 \pm 0.40) \times 10^{-7}$
$\mathcal{B}(D^0 \rightarrow \eta\phi)$	$(1.40 \pm 0.50) \times 10^{-4}$	$(4.30 \pm 1.60) \times 10^{-8}$
Total	-	$(1.69 \pm 0.15) \times 10^{-6}$

Table 6.2: Contribution to total $D^0 \rightarrow (\eta/\eta' \rightarrow \mu\mu\gamma)X$ rate made from all the considered decays above. In total, this cumulative contribution is approximately three times smaller than $D^0 \rightarrow K^+\pi^-\mu^+\mu^-$. All the branching fractions are obtained from [1].

6.3.2 Partially Reconstructed $B \rightarrow \eta(')V$ Backgrounds

The backgrounds with $\eta(')$ resonances from partially reconstructed decays that proceed via D were considered in previous [subsection 6.3.1](#). In this section backgrounds with $\eta(')$ along with vector resonances ω/ρ coming directly from B are estimated. The total

branching fractions for these processes are listed in [Table 6.3](#) and since they are very small this type of background is discarded and will not be considered further.

Process	\mathcal{B}
$\mathcal{B}(B^0 \rightarrow \omega\eta')$	$(1.00 \pm 0.50) \times 10^{-6}$
$\mathcal{B}(B^0 \rightarrow \rho\eta')$	$< 5 \times 10^{-7}$
$\mathcal{B}(B^0 \rightarrow \omega\eta)$	$(9.00 \pm 4.00) \times 10^{-7}$
$\mathcal{B}(B^0 \rightarrow \rho\eta)$	$< 5 \times 10^{-7}$
$\mathcal{B}(\eta \rightarrow \mu\mu\gamma)$	$(3.10 \pm 0.40) \times 10^{-4}$
$\mathcal{B}(\eta' \rightarrow \mu\mu\gamma)$	$(1.08 \pm 0.27) \times 10^{-4}$
$\mathcal{B}(\rho \rightarrow \mu\mu)$	$(4.55 \pm 0.28) \times 10^{-5}$
$\mathcal{B}(\omega \rightarrow \mu\mu)$	$(9.00 \pm 3.10) \times 10^{-5}$
Process	Contribution to $B^0 \rightarrow (\eta(') \rightarrow \mu\mu\gamma)(\rho(\omega) \rightarrow \mu\mu)$
$\mathcal{B}(B^0 \rightarrow (\omega \rightarrow \mu\mu)(\eta \rightarrow \mu\mu\gamma))$	$(7.10 \pm 1.00) \times 10^{-15}$
$\mathcal{B}(B^0 \rightarrow (\omega \rightarrow \mu\mu)(\eta' \rightarrow \mu\mu\gamma))$	$(2.50 \pm 0.60) \times 10^{-15}$
$\mathcal{B}(B^0 \rightarrow (\rho \rightarrow \mu\mu)(\eta \rightarrow \mu\mu\gamma))$	$<(2.50 \pm 1.40) \times 10^{-14}$
$\mathcal{B}(B^0 \rightarrow (\rho \rightarrow \mu\mu)(\eta' \rightarrow \mu\mu\gamma))$	$<(1.00 \pm 0.60) \times 10^{-14}$
Total	$<(4.50 \pm 1.60) \times 10^{-14}$

Table 6.3: Different and total contribution to $B^0 \rightarrow \eta(')\rho(\omega)$. All the branching fractions are obtained from [\[1\]](#).

6.4 Rare and resonant $B^+ \rightarrow \pi^+/K^+\mu^-\mu^+$ backgrounds

The resonant backgrounds arising through $B^+ \rightarrow (J/\psi \rightarrow \mu^-\mu^+)X^+$ and $B^+ \rightarrow (\psi(2S) \rightarrow \mu^-\mu^+)X^+$ decay chains are eliminated because of the $c\bar{c}$ veto as discussed in [section 5.7](#).

It is, however, necessary to evaluate the impact of the rare equivalent of this background, namely $B^+ \rightarrow \mu^+\mu^-\pi^+$ decays, where π^+ is misidentified as muon. The

$\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \pi^+) = 1.79 \pm 0.23 \times 10^{-8}$ [1]. The contribution of this background is accounted for in the [section 6.2](#), but it is crosschecked as this particular background peaks just under the corrected mass of B . For the same rare decay but with kaon instead, $B^+ \rightarrow \mu^+ \mu^- K^+$, the mass is expected to be shifted away from this peak because of the higher kaon mass.

In order to establish the severity of this background, $B^+ \rightarrow \mu^+ \mu^- \pi^+$ simulation for Run I and Run II is reconstructed where the muon mass hypothesis is used for the pion track candidate. This means that energy of this candidate is recalculated. After, the same selection as in the signal case is applied. The expected number of $B^+ \rightarrow \mu^+ \mu^- \pi^+$ decays after all selection in a given Run can be calculated by normalising to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ decays. In the end 0.06 (0.03) $B^+ \rightarrow \mu^+ \mu^- \pi^+$ events are expected in Run I (II) which is negligible given that there are ~ 17 expected signal events with $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = 1 \times 10^{-8}$. Hence no further specific action for this background is taken, however, its contribution is directly accounted for in the [section 6.2](#).

6.5 Summary

In conclusion, different backgrounds that can mimic the signal were studied. From all considered backgrounds only combinatorial, misID and partially reconstructed backgrounds have considerable contribution after all the selection and hence need to be modelled. Exact contribution of these three backgrounds is shown in ??.

Chapter 7

Mass fits and efficiencies

To be able to translate observed signal events into branching fraction estimate, the normalisation channel of $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ is used. Both, for signal and normalisation channel the absolute efficiencies, luminosity, the b-quark cross-section or fragmentation fractions will cancel. There are, however, efficiencies that will not cancel and will be necessary for the final limit setting procedure. In this section, methods of obtaining efficiencies of selection for normalisation and signal channel are described as well as efficiencies themselves.

7.1 Efficiency Ratio

As this measurement is performed in a particular $\min q^2$ region, discussed in [section 5.7](#), all signal efficiencies are calculated with the $\min q^2$ selection imposed. Overall selection efficiency for signal, ε^s , and normalisation, ε^n , includes contributions from the detector acceptance efficiency labelled (GEN); the reconstruction selection efficiency (REC); the offline selection efficiency comprising of trigger (TRG), J/ψ and $\Psi(2S)$ veto (OFF), MVA based selection efficiency (CombiBDT and MisidBDT); fitting region selection efficiency (FR); the efficiency of the PID requirement (PID). The summary of method used to extract signal efficiency is shown in Table [Table 7.1](#). For normalisation channel, there is no $\min q^2$ region selection and hence full (*generator-level+detector*) simulation is used everywhere apart from ε_{GEN} , *generator-level*.

Component	Method
$\varepsilon_{GEN}, \varepsilon_{REC}$	I
$\varepsilon_{TRG}, \varepsilon_{OFF}, \varepsilon_{BDTs}, \varepsilon_{FR}$	II
ε_{PID}	III

Table 7.1: Method of obtaining efficiencies. Most of these efficiencies are evaluated using simulation, however, TRG and PID efficiencies are evaluated using data and/or simulation techniques.

The three methods for signal efficiency determination are listed:

- Method I - The first two efficiencies, $\varepsilon_{GEN}, \varepsilon_{REC}$, for signal are obtained using privately generated simulation from [Table 5.1](#) using

$$\varepsilon_{GEN,minq^2} \times \varepsilon_{REC,minq^2} = \frac{N_{in_acc,minq^2}}{N_{generated,minq^2}} \times \frac{N_{REC,minq^2}}{N_{in_acc,minq^2}}, \quad (7.1)$$

$$N_{in_acc,minq^2} = N_{in_acc} \times \varepsilon_{minq^2}. \quad (7.2)$$

In Equation [Equation 7.2](#), ε_{minq^2} is obtained by dividing number of generated events in *generator-level* simulation (mentioned in [Table 5.1](#)) with $minq^2$ condition imposed, $N_{generated,minq^2}$, to total number of generated events, $N_{generated}$. N_{in_acc} is the number of events in *generator-level+detector* simulation before reconstruction, $N_{REC,minq^2}$ is the number of events after reconstruction with $minq^2$ condition.

- Method II - Divide number of events that passed the selection by total number of events prior to this particular selection step.
- Method III - Data-driven approach using `PIDCalib` package explained in [section 3.9](#) of determining PID efficiency is used.

Having all the individual efficiencies the relative efficiency with no FCME split, $R_{NOFCME}^{(21,26)}(\varepsilon)$, can be calculated

$$R_{\text{NOFCME}}^{(21,26)}(\varepsilon) = \frac{\varepsilon^s}{\varepsilon^n} = \frac{\varepsilon_{\text{GEN}}^s}{\varepsilon_{\text{GEN}}^n} \times \frac{\varepsilon_{\text{REC}}^s}{\varepsilon_{\text{REC}}^n} \times \frac{\varepsilon_{\text{TRG}}^s}{\varepsilon_{\text{TRG}}^n} \times \frac{\varepsilon_{\text{OFF}}^s}{\varepsilon_{\text{OFF}}^n} \times \frac{\varepsilon_{\text{CombiBDT}}^s}{\varepsilon_{\text{CombiBDT}}^n} \times \frac{\varepsilon_{\text{MisidBDT}}^s}{\varepsilon_{\text{MisidBDT}}^n} \times \frac{\varepsilon_{\text{FR}}^s}{\varepsilon_{\text{FR}}^n} \times \frac{\varepsilon_{\text{PID}}^s}{\varepsilon_{\text{PID}}^n}, \quad (7.3)$$

where 21, 26 denote the stripping version for Run I and Run II.

7.2 Summary of Efficiencies

The summary of individual efficiencies together with total efficiency, which is calculated for signal using nominator of [Equation 7.3](#) and for normalisation denominator of [Equation 7.3](#), is given in [Table 7.2](#).

Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$		$B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$	
	2012	2016	2012	2016
ε_{GEN}	18.56 ± 0.11	19.59 ± 0.07	16.22 ± 0.02	17.39 ± 0.02
ε_{REC}	10.84 ± 0.03	12.40 ± 0.01	17.74 ± 0.01	20.03 ± 0.00
ε_{TRG}	74.22 ± 0.13	74.83 ± 0.05	77.79 ± 0.03	79.12 ± 0.01
ε_{OFF}	88.20 ± 0.11	88.30 ± 0.05	100.00 ± 0.00	100.00 ± 0.00
$\varepsilon_{\text{CombiBDT}}$	47.25 ± 0.18	34.28 ± 0.07	50.89 ± 0.05	39.73 ± 0.02
$\varepsilon_{\text{MisidBDT}}$	43.58 ± 0.26	36.80 ± 0.12	51.12 ± 0.07	44.64 ± 0.02
ε_{FR}	92.30 ± 0.21	93.77 ± 0.10	99.59 ± 0.01	99.91 ± 0.00
ε_{PID}	63.15 ± 0.50	62.27 ± 0.27	68.53 ± 0.11	65.63 ± 0.04
$\varepsilon_{\text{total}}$	0.1581 ± 0.0020	0.1182 ± 0.0008	0.3974 ± 0.0011	0.3203 ± 0.0005

Table 7.2: Summary of individual simulation and/or data efficiencies in % necessary for *single event sensitivity* for signal and normalisation channel. Efficiency values for 2016 are TCK-weighted averaged efficiencies, which will be explained in [subsection 7.2.3](#). The errors considered are of statistical nature, computed using binomial error.

Hence resulting values for relative no fractional corrected mass split efficiency ratios

defined in [Equation 7.3](#) are

$$\begin{aligned} R_{\text{NOFCME}}^{21}(\varepsilon) &= \frac{(1.58 \pm 0.02) \times 10^{-3}}{(3.97 \pm 0.01) \times 10^{-3}} = (3.98 \pm 0.05) \times 10^{-1}, \\ R_{\text{NOFCME}}^{26}(\varepsilon) &= \frac{(1.18 \pm 0.01) \times 10^{-3}}{(3.20 \pm 0.00) \times 10^{-3}} = (3.69 \pm 0.03) \times 10^{-1}. \end{aligned} \quad (7.4)$$

Including fractional corrected mass split efficiency ratios defined in [Equation 5.8](#) are

$$\begin{aligned} R_{\text{lowFCME}}^{21}(\varepsilon) &= \frac{(7.44 \pm 0.12) \times 10^{-4}}{(2.33 \pm 0.01) \times 10^{-3}} = (3.20 \pm 0.05) \times 10^{-1}, \\ R_{\text{highFCME}}^{21}(\varepsilon) &= \frac{(8.37 \pm 0.13) \times 10^{-4}}{(1.65 \pm 0.01) \times 10^{-3}} = (5.09 \pm 0.08) \times 10^{-1}, \\ R_{\text{lowFCME}}^{26}(\varepsilon) &= \frac{(6.51 \pm 0.05) \times 10^{-4}}{(2.15 \pm 0.00) \times 10^{-3}} = (3.03 \pm 0.02) \times 10^{-1}, \\ R_{\text{highFCME}}^{26}(\varepsilon) &= \frac{(5.33 \pm 0.05) \times 10^{-4}}{(1.05 \pm 0.00) \times 10^{-3}} = (5.06 \pm 0.04) \times 10^{-1}. \end{aligned} \quad (7.5)$$

As it can be noticed, different selections that were optimised for Run I and II yield different overall as well as individual efficiencies. This results in small differences in sensitivity between Run I and Run II. To better understand where does this difference come from, ratio of individual relative efficiencies as a function of stripping version is plotted in the [Figure 7.1](#). The difference can be attributed to different BDTs.

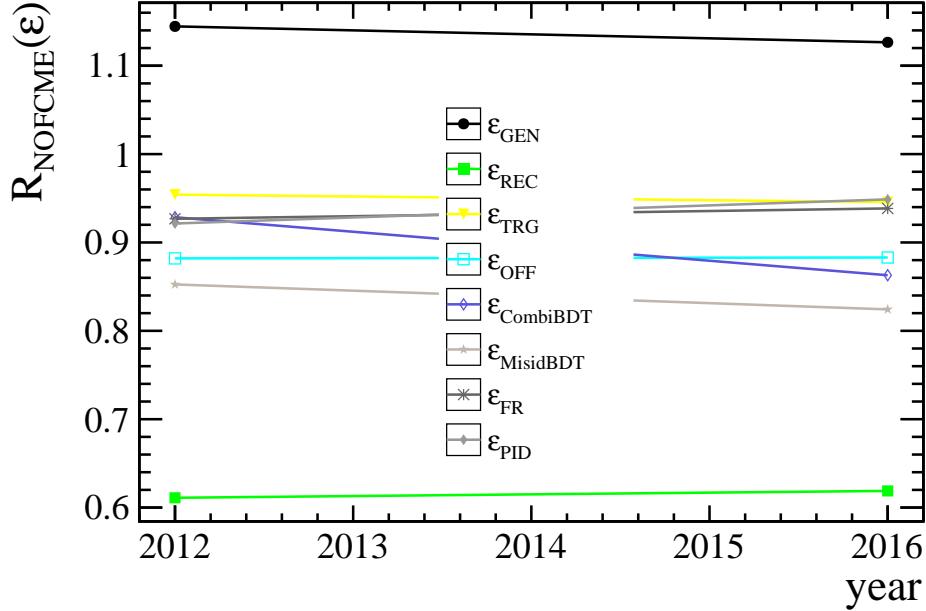


Figure 7.1: Summary of ratio of efficiencies between 2012 simulation and 2016 simulation with no FCME split. Efficiency values for 2016 are TCK-weighted averaged efficiencies.

More detailed discussion on individual efficiencies is covered in following subsections.

7.2.1 Detector Acceptance Efficiency (GEN)

For charged particles detector acceptance efficiency describes the fraction of decays contained in the polar angle region of [10, 400] mrad. For 2012 and 2016 simulation samples, the overall detector acceptance efficiency will be the average for two possible magnetic polarity conditions: down, up. For 2012 this will be also averaged with two different simulation versions: Pythia 6.4 [44] and Pythia 8.1 [24].

The hierarchy of generator level efficiencies $\epsilon_{\text{GEN}}^s > \epsilon_{\text{GEN}}^n$ is expected as the muon is lighter than kaon making kaon more likely to be softer and at larger angle, therefore outside of the acceptance.

7.2.2 Reconstruction Efficiency (REC)

The reconstruction efficiency is calculated on simulated events which have passed the detector acceptance. For signal, this efficiency consists of reconstruction and stripping, detailed in [Table 5.2](#). For normalisation it consists from reconstruction, stripping, **and on the top** signal stripping is applied. This is done so that selections in normalisation and signal channel are kept as similar as possible and the fact that the signal selection has tighter cuts as explained in [section 5.9](#). However, it should be noted that reconstruction efficiency reflects stripping selection **without the PID cuts** for both signal and normalisation. This is because **PID** is badly modelled in simulation and hence will be accounted for separately and at the end of the selection chain.

The hierarchy of reconstruction level efficiencies $\varepsilon_{REC}^s < \varepsilon_{REC}^n$ is also expected as many variables in a signal stripping are based on alignment of the mother B with its daughters. For fully reconstructed normalisation channel this is expected to be the case, whereas for not fully reconstructed decays alignment requirements make selection tighter.

7.2.3 Trigger Efficiency (TRG)

The trigger efficiency is calculated on the top of (GEN) and (REC) efficiency. In order to extract the trigger efficiency, full simulation for both signal and normalisation is used. It should be noted though that at **LHCb**, full simulation is produced based on a certain trigger configuration. Trigger configuration key, TCK, represents unique code for exact conditions the data have been triggered with at L0, HLT1 and HLT2, notably thresholds of certain quantities such as p , p_T .

Therefore if default TCK for simulation is representative for the whole considered dataset then the efficiency can be extracted directly from the simulation produced, which is the case for the Run I data.

However in Run II the trigger thresholds have been changing often resulting in 16 different TCKs with very different p , p_T thresholds, see Table [Table 7.3](#) for

full detail. In the third column, luminosity proportion for 2016 is given. It can be seen that the default simulation in 2016 (corresponding to TCK decimal key 288888335) only represents around 35% of the data. For this reason, the trigger efficiencies for 2016 data have been obtained by emulation of the trigger on simulation for L0 and HLT1 level for each individual TCK, creating 16 TCK-based simulations. This trigger emulation to extract efficiencies was tested with the default trigger configuration (TCK 288888335) to validate the emulation and the correct efficiencies have been recovered. It should be noted that small differences arise from difference between *offline* and HLT1 container for PVs which stores the information about min IP χ^2 as the PV finding-algorithm is different, but these have negligible effect.

mentioning this just in case of reproducibility, but maybe not necessary.

In order to obtain the average efficiency for Run II, the 16 TCK efficiencies are weighted by the proportion of luminosity corresponding to the integrated luminosity for a given TCK over the full 2016 integrated luminosity. The integrated luminosity per TCK was extracted by looking at API version of the LHCb rundatabase.

The full trigger luminosity for Run II is calculated by averaging the luminosity-weighted efficiencies, as seen in [Table 7.4](#). This averaged efficiency is going to be given as a final efficiency for 2016 from now on.

For the HLT2 level, there were no significant changes of thresholds and are hence efficiencies are obtained from full simulation regardless. The systematic effect of this assumption will be listed in the systematic uncertainties chapter [SALLY - ADD SECTION REFERENCE TO SYSTEMATICS](#).

Run I trigger efficiency is determined directly by looking at default TCK as it is representative of the whole dataset.

7.2.4 Offline Selection (OFF)

In this section offline efficiencies are discussed. These include J/ψ and $\Psi(2S)$ veto signal efficiency that were mentioned in [Table 5.4](#), where $2946.0 < |m(\mu^+ \mu^-)| < 3176.0$

TCK dec	TCK hex	% \mathcal{L} %	\mathcal{L} pb^{-1}	HLT1TrackMuon				L0Muon	
				P_{ghost}	p_μ [MeV]	$p_T(\mu)$ [MeV]	$\min \text{IP}_X^2$	SPD_{mult}	$p_T(\mu)$ [MeV]
2016 MD 0.859656fb^{-1}									
287905280	0x11291600	0.769	12.74	–	6.0	0.91	10	450	14
287905283	0x11291603	2.11	35.01	–	6.0	0.91	10	450	23
287905284	0x11291604	1.50	24.78	–	6.0	0.91	10	450	27
287905285	0x11291605	4.73	78.42	–	6.0	0.91	10	450	31
288822793	0x11371609	4.35	72.14	0.2	6.0	1.1	35	450	27
288822798	0x1137160e	1.37	22.756	0.2	6.0	1.1	35	450	27
288888329	0x11381609	0.414	6.86	0.2	6.0	1.1	35	450	31
288888334	0x1138160e	1.912	31.70	0.2	6.0	1.1	35	450	31
288888335	0x1138160f	34.7	575.25	0.2	6.0	1.1	35	450	37
2016 MU 0.798156fb^{-1}									
288495113	0x11321609	6.45	107.00	–	6.0	0.91	10	450	27
288626185	0x11341609	7.12	118.06	–	6.0	0.91	10	450	27
288691721	0x11351609	1.42	23.46	0.2	6.0	1.1	35	450	27
288757257	0x11361609	25.0	414.62	0.2	6.0	1.1	35	450	27
288888337	0x11381611	2.66	44.13	0.2	6.0	1.1	35	450	31
288888338	0x11381612	5.41	89.75	0.2	6.0	1.1	35	450	33
288888339	0x11381613	0.0685	1.136	0.2	6.0	1.1	35	450	27
MC Default									
1362630159	0x5138160f	–	–	0.2	6.0	1.1	35	450	37

Table 7.3: Summary of 16 different TCKs listing properties of candidates necessary to pass L0 and HLT1 selection in 2016. In the final row, the default configuration for 2016 is shown and it corresponds to 288888335 TCK.

for J/ψ veto and $3586.0 < |m(\mu^+ \mu^-)| < 3766.0$ veto for $\Psi(2S)$ veto. For normalisation channel this is non applicable as the normalisation decay proceeds via J/ψ resonance. As trigger efficiencies for Run II are TCK-dependant, luminosity-weighted average is used. Similarly for all 2016 efficiencies from now on that are calibrated from the

TCK	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$			$B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$		
	ϵ_{L0}	ϵ_{HLT1}	ϵ_{HLT2}	ϵ_{L0}	ϵ_{HLT1}	ϵ_{HLT2}
287905280	0.921	0.999	0.831	0.891	0.997	0.943
287905283	0.905	0.999	0.845	0.878	0.998	0.953
287905284	0.894	0.999	0.855	0.867	0.998	0.962
287905285	0.88	0.999	0.868	0.854	0.998	0.973
288495113	0.894	0.999	0.855	0.867	0.998	0.962
288626185	0.894	0.999	0.855	0.867	0.998	0.962
288691721	0.894	0.957	0.873	0.867	0.94	0.965
288757257	0.894	0.957	0.873	0.867	0.94	0.965
288822793	0.894	0.957	0.873	0.867	0.94	0.965
288822798	0.88	0.957	0.886	0.854	0.941	0.976
288888329	0.894	0.957	0.873	0.867	0.94	0.965
288888334	0.88	0.957	0.886	0.854	0.941	0.976
288888335	0.848	0.958	0.911	0.821	0.941	0.999
288888337	0.88	0.957	0.886	0.854	0.941	0.976
288888338	0.871	0.957	0.895	0.844	0.941	0.984
288888339	0.89	0.957	0.877	0.864	0.94	0.968
Weighted efficiency	0.876	0.967	0.884	0.849	0.953	0.978

Table 7.4: Efficiencies of 2016 trigger emulation on MC. Depending on TCK, the efficiencies vary up 10% for L0 level for signal MC and up to 5% for normalisation TCK. This is important as *single event sensitivity* is sensitive to the ratio of these two efficiencies. This configuration is describing correctly only 35% data with high p_T threshold.

simulation are weighted averages, unless stated otherwise.

7.2.5 Combinatorial BDT and Misid BDT efficiency

Efficiencies of MVA selection are also evaluated on simulation samples. These efficiencies are obtained using samples that passed (GEN), (REC), (TRG) and (OFF) cuts.

Specific MVA for combinatorial background suppression (see [subsection 5.8.3](#)) and misID background suppression (see [subsection 5.8.4](#)) are applied to the simulation samples. As the optimisation led to different BDTs depending on the data-taking period, these are then applied parametrically to relevant simulation samples. The results were listed in ??.

For Misid and Combinatorial BDT selection, normalisation $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ channel retains more signal than the $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ channel. This is due to the kaon/muon p and p_T kinematics difference as seen in [Figure 7.2](#) and [Figure 7.3](#), where the kaon track is generally harder than the muon track. Kaon reconstruction efficiency is worse than muon reconstruction efficiency because about 11% of the kaons cannot be reconstructed due to hadronic interactions that occur before the last T station [34], implying that the p_T of B is on average harder for normalisation channel. As these two quantities are high in BDT importance ranking as mentioned in [subsection 5.8.3](#), this makes normalisation MC more efficient. In Misid BDT selection, again the kinematics of B and the $\text{min IP} \chi^2$ of the oppositely charged muon to B is more signal like than signal.

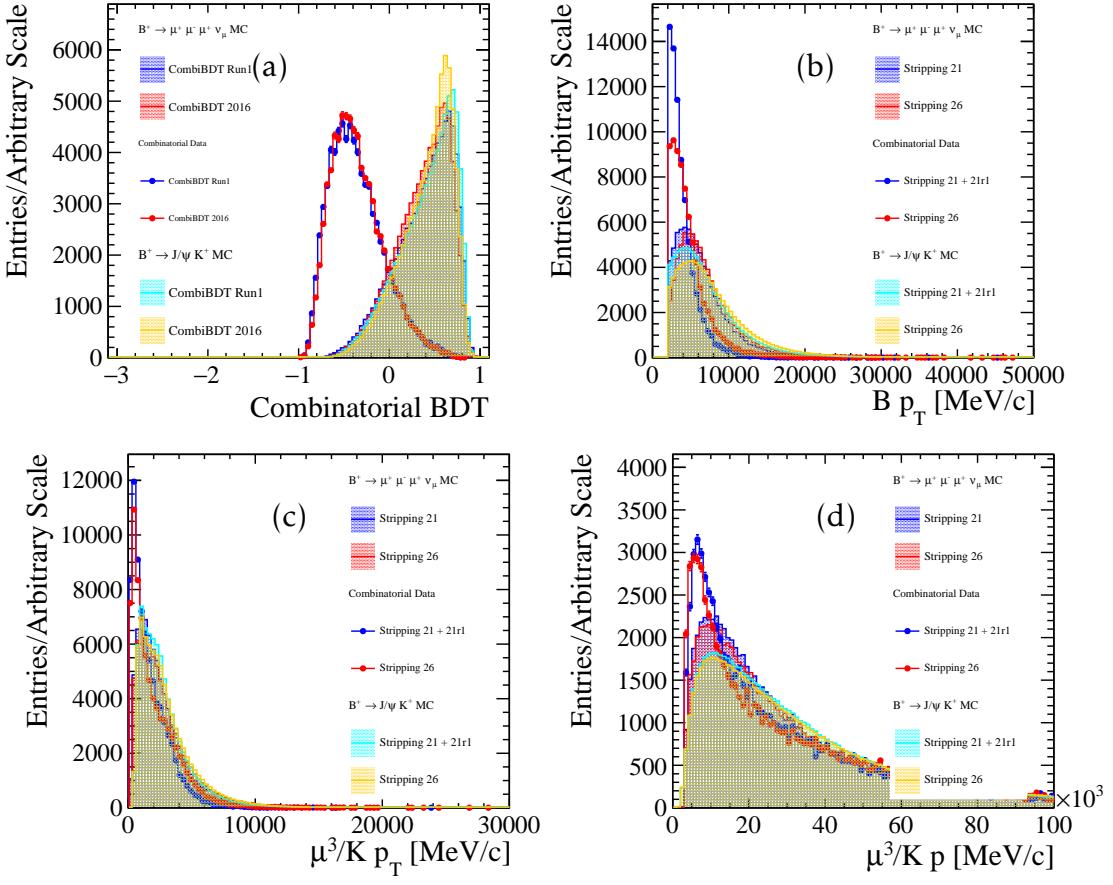


Figure 7.2: (a) Combinatorial BDT response for signal MC and upper mass sideband as well as for normalisation channel MC for Stripping 21 and Stripping 26. The most discriminative variables are (b) p_T of B , (c) muon/kaon p_T and (d) muon/kaon p .

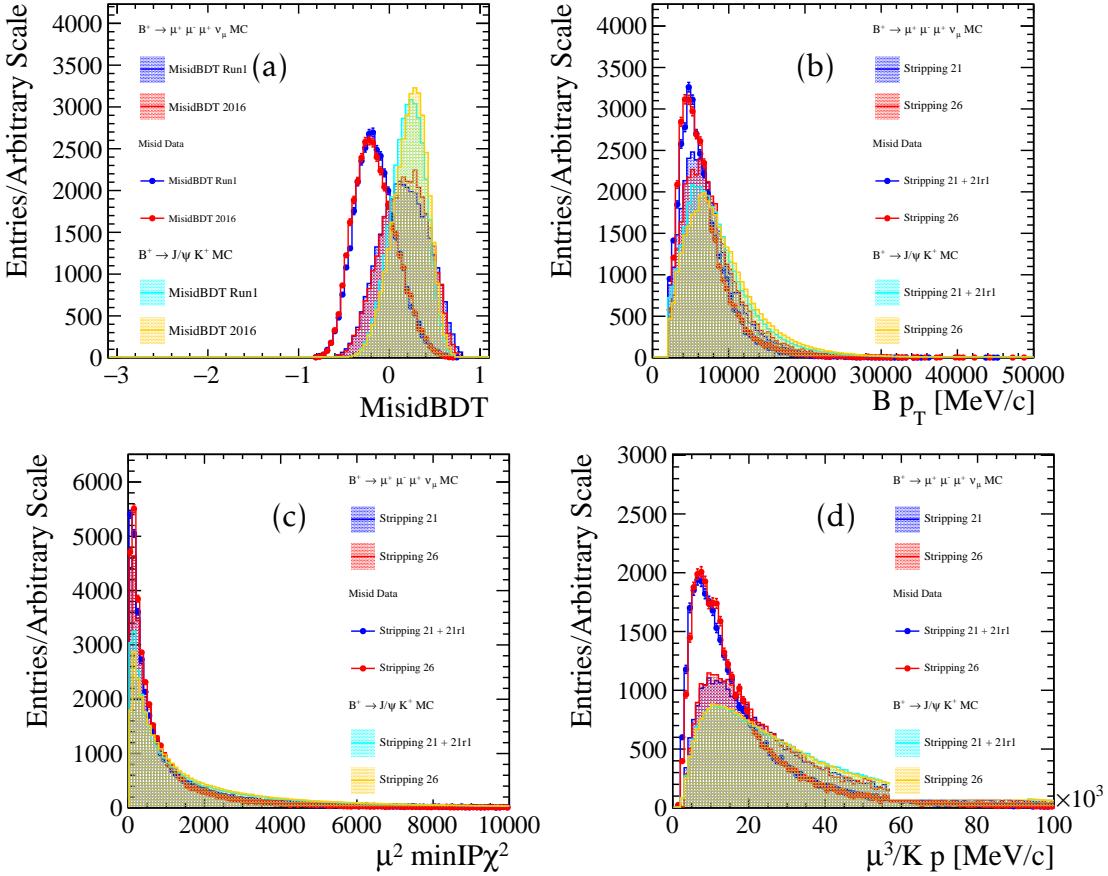


Figure 7.3: (a) Misid BDT response for signal MC and upper mass sideband as well as for normalisation channel MC for Stripping 21 and Stripping 26. The most discriminative variables are (b) p_T of B , (c) muon IP χ^2 and (d) muon/kaon p . Misid BDT responses are plotted with combinatorial BDT already applied.

7.2.6 Fitting Range Efficiency (FR)

As discussed in subsection 5.8.5, fitting region was chosen firstly in order to avoid modelling combinatorial background drop in low corrected mass region (exclusion below 4000 MeV/ c^2) and secondly in order to not include region where there are very few/no events (exclusion above 7000 MeV/ c^2) in **corrected mass**. As seen in Table 7.2 normalisation channel does not loose many candidates compared to signal channel. This is expected as the **visible mass** is more

constrained in normalisation channel from preselection stage (see [Table 5.7](#), where $5150 \text{ MeV}/c^2 < B \text{ Mass} < 5450 \text{ MeV}/c^2$) than in signal channel as seen in [Figure 7.4](#)). Hence, restricting region in **corrected mass** is does not affect normalisation channel much.

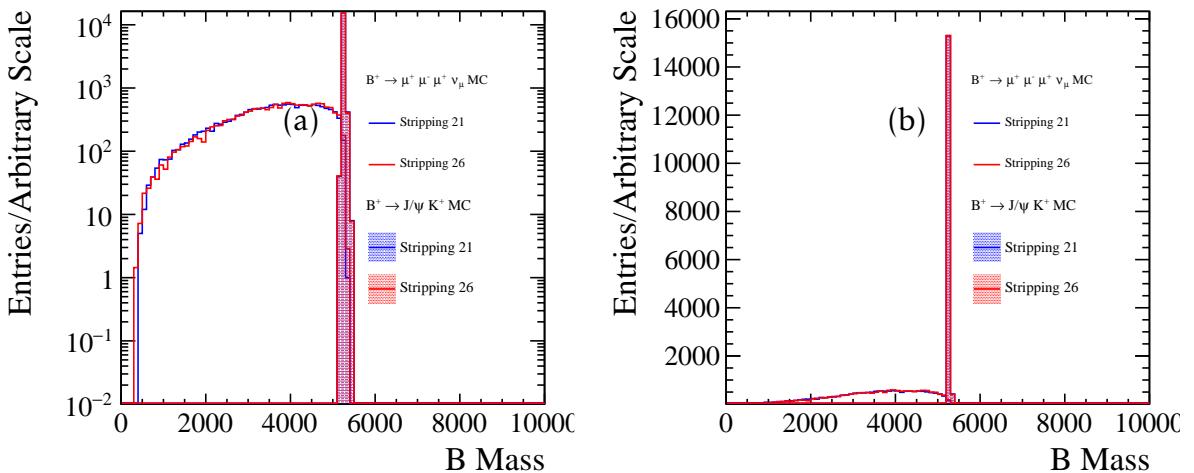


Figure 7.4: (a) Visible mass of normalisation and signal simulation. It can be seen that normalisation previous preselection has a sharp cut around B visible mass leading to much higher fitting region efficiency. (b) The corresponding logarithmic version of plot (a).

7.2.7 PID Efficiencies (PID)

As **PID** variables are not correctly modelled in simulation, mentioned in [section 3.9](#), data-driven approach of extracting PID efficiency is taken. To not introduce any biases in previous steps, especially in multivariate selection, PID efficiencies are evaluated at the end of the selection chain with **PIDCalib** package data samples.

The PID efficiency is higher for normalisation channel with all PID requirements given in [Table 7.6](#) compared to signal provided in [Table 7.5](#) due to weaker PID requirement on kaon as compared to muons.

species	2012 PID Simulation	2016 PID Simulation
muon	$\Delta LL(\mu - \pi) > 0$	$\Delta LL(\mu - \pi) > 0$
muon	$\Delta LL(\mu - K) > 0$	$\Delta LL(\mu - K) > 0$
muon	-	<code>IsMuonTight==1.0</code>
muon	<code>Nshared==0</code>	<code>Nshared<2</code>
muon	<code>Probnnmu> 0.35</code>	<code>Probnnmu> 0.35</code>
ε_{PID}	0.631 ± 0.005	0.623 ± 0.006

 Table 7.5: Signal simulation efficiency using `PIDCalib` efficiencies.

species	2012 PID Simulation	2016 PID Simulation
muon	$\Delta LL(\mu - \pi) > 0$	$\Delta LL(\mu - \pi) > 0$
muon	$\Delta LL(\mu - K) > 0$	$\Delta LL(\mu - K) > 0$
muon	-	<code>IsMuonTight==1.0</code>
muon	<code>Nshared==0</code>	<code>Nshared<2</code>
muon	<code>Probnnmu> 0.35</code>	<code>Probnnmu> 0.35</code>
kaon	$\Delta LL(K - \pi) > 0$	$\Delta LL(K - \pi) > 0$
kaon	$\Delta LL(p - K) < 5$	$\Delta LL(p - K) < 5$
ε_{PID}	0.685 ± 0.001	0.656 ± 0.001

 Table 7.6: Normalisation MC efficiency using `PIDCalib` efficiencies.

7.3 Mass Fits

In this section, firstly parametrisation of the normalisation channel is shown in [subsection 7.3.1](#). The fit to normalisation channel using the full fit model is described in [subsection 7.3.2](#). This is followed by the signal fit parametrisation outlined in ?? resulting in blinded and non-blinded data fits described in [subsection 7.3.4](#).

7.3.1 Normalisation Channel Parametrisation

The obtain the $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ yield of Run I and Run II, an unbinned extended maximum likelihood fit to the invariant $\mu^+ \mu^- K^+$ data distribution in each respective Run is performed. In this section contributions to the normalisation fit model are considered.

Signal

The first component is the signal itself, which is modelled with PID-weighted simulation and can be best described by double-sided Ipatia function, detailed in [section C.2](#), where all the parameters apart from the mean μ^{IP} and width σ^{IP} are fixed from the signal simulations. These simulations passed through the same selection process as the corresponding $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ data, described in [section 5.9](#), with one exception. Instead of directly cutting on [PID](#) variables, the simulations were reweighted with the relevant PID weights, because of known mismatch between simulation and data. More on this will be covered in [subsection 7.2.7](#).

$B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)\pi^+$ Background

Secondly, since the PID requirements on kaon are very loose, there will background contribution from $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)\pi^+$. This contribution is modelled by double-sided Crystal Ball function to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)\pi^+$ simulation, where the pion track is given kaon mass hypothesis. Again, all the parameters apart from mean μ^{CB} and width σ^{CB} are fixed from the fit of this simulation. In [Figure 7.5](#), fits to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$

simulation and $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) \pi^+$ simulation from Stripping 21 are showed using different scales. For signal, Run I Stripping 21 simulation is used and for Run II (Stripping 26 - TCK 288888335) simulation sample is used, for $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) \pi^+$ Stripping 21 is used for all samples.

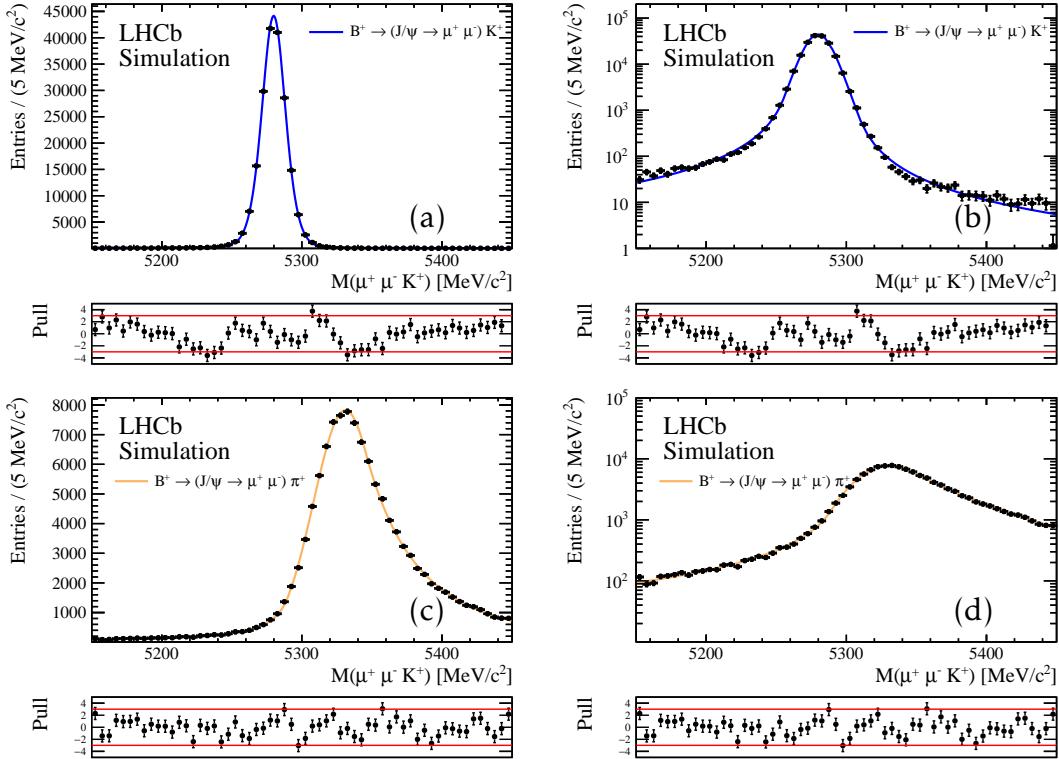


Figure 7.5: Fit to 2012 (a) $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ simulation and (c) $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) \pi^+$ simulation under kaon mass hypothesis. On right, the same plots but with logarithmic scale instead.

Combinatorial Background

Lastly, combinatorial background is modelled by exponential function, where the exponential constant is let free.

7.3.2 Normalisation Fit

The full fit model for the normalisation data fit containing description of the individual components as well as their constraints is given in [Table 7.7](#).

Fit Parameter	Status
Yields	
$N_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+}$ (Signal)	free
$N_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)\pi^+}$	free
$N_{Combinatorial}$	free
Signal Shape Parameters (double-sided Ipatia)	
$\mu_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+}^{IP}$	constrained from signal MC
$\sigma_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+}^{IP}$	constrained from signal MC
Others	Fixed from MC
$B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)\pi^+$ Shape Parameters (double-sided Crystal Ball)	
$\mu_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)\pi^+}^{CB}$	constrained from signal MC
$\sigma_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)\pi^+}^{CB}$	constrained from signal MC
Others	Fixed from MC
Combinatorial Shape Parameters	
exponential par.	Free

Table 7.7: Summary of the fit parameters and individual component constraints for the fit to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ decays.

The $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ signal yield is extracted by performing an unbinned extended maximum likelihood fit with the full fit model to the invariant $\mu^+ \mu^- K^+$ distribution in $5150 < M_{B^+} < 5450$. Fits to the $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ for Run I and II are shown in [Figure 7.6](#). Yields from fit to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ are obtained and

summarized in [Table 7.8](#).

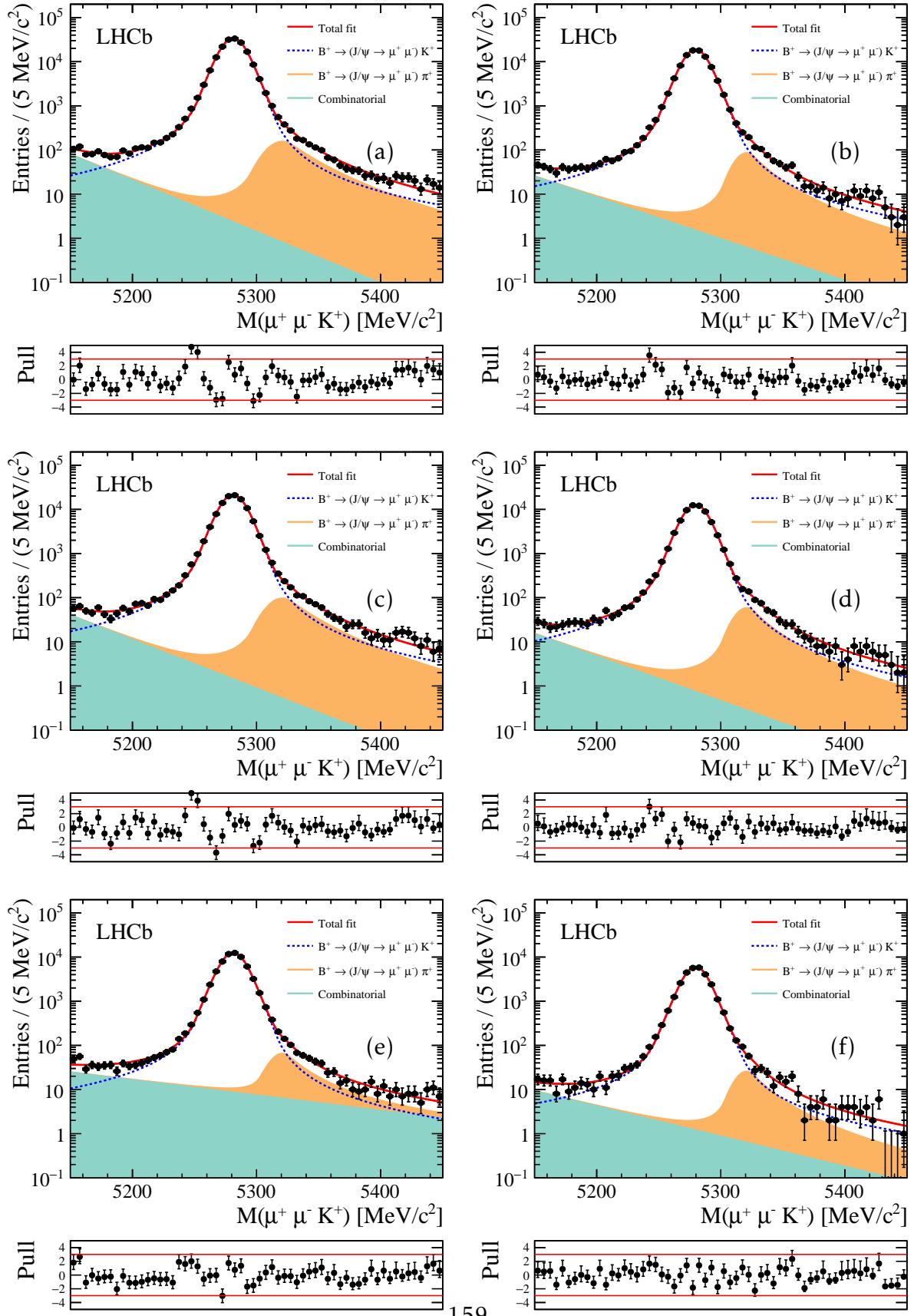


Figure 7.6: Fit results in logarithmic scale to (a) Run I (b) Run II $\mu^+ \mu^- K^+$ mass spectrum with no fractional corrected mass split, (c)(d) low FCME bin, (e)(f) high FCME bin.

Sample	Stripping	Split	Yields
$N_{B^+ \rightarrow J/\psi K^+}$	Run I	NOFCME	173422 ± 446
$N_{B^+ \rightarrow J/\psi K^+}$	Run II	NOFCME	94491 ± 313
$N_{B^+ \rightarrow J/\psi K^+}$	Run I	lowFCME	109224 ± 337
$N_{B^+ \rightarrow J/\psi K^+}$	Run II	lowFCME	64723 ± 259
$N_{B^+ \rightarrow J/\psi K^+}$	Run I	highFCME	64078 ± 257
$N_{B^+ \rightarrow J/\psi K^+}$	Run II	highFCME	29760 ± 176

Table 7.8: $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ signal yield obtained from fits to $\mu^+ \mu^- K^+$ mass spectrum shown in Figure 7.6.

7.3.3 Signal Channel Parametrisation

To fit the signal data the default fitting strategy is to use a simultaneous unbinned maximum likelihood fit to the $\mu^+ \mu^- \mu^+$ corrected mass spectrum to combined Run I and II dataset after the full selection in two bins of fractional corrected mass error, which is denoted as simultaneous fit. As a cross-check non-simultaneous fit, where no splitting in bins of corrected mass is done, is denoted as non-simultaneous fit. The parametrisation of all the components for the full signal model fit is described below.

Signal

The full fit to Run I and Run II data requires the signal shape knowledge for combined dataset. This is obtained from Run I and Run II signal simulation *cocktail*. The cocktail is created after full selection by assigning event-by-event weights, w^i , which capture the differences between Run I and Run II simulation, that were not considered by full selection.

Firstly, weights that reflect the expected difference due to increased luminosity are computed. To obtain these signal weights, following conditions must be satisfied

$$n^{2012} = \mathcal{L}^{2012} \times \sigma_{pp \rightarrow b\bar{b}}^{2012}, \quad (7.6)$$

$$n^{2016} = \mathcal{L}^{2016} \times \sigma_{pp \rightarrow b\bar{b}}^{2016}, \quad (7.7)$$

$$w^{2012} \times N^{2012} + w^{2016} \times N^{2016} = N^{2012} + N^{2016}, \quad (7.8)$$

$$\frac{w^{2012} \times N^{2012}}{w^{2016} \times N^{2016}} = \frac{n^{2012}}{n^{2016}}. \quad (7.9)$$

These constraints yield following value for event-by-event (or rather yearly) weights

$$w^{2012} = \frac{N^{2012} + N^{2016}}{N^{2012} \times (1.0 + \frac{n^{2016}}{n^{2012}})} = 0.931 \quad (7.10)$$

$$w^{2016} = \frac{N^{2016} + N^{2012}}{N^{2016} \times (1.0 + \frac{n^{2012}}{n^{2016}})} = 1.073, \quad (7.11)$$

where N^{2012}, N^{2016} number of events at *generator level*, $\mathcal{L}^{2012}, \mathcal{L}^{2016}$ are integrated luminosities, and $\sigma_{pp \rightarrow b\bar{b}}^{2012}, \sigma_{pp \rightarrow b\bar{b}}^{2016}$ are cross-sections in a given year. N^{2012}, N^{2016} number of events at *generator level* is obtained by dividing number of reconstructed events N_{REC} by the reconstruction efficiency ε_{REC} (see [Table 7.2](#)). Values for these variables are summarized in [Table 7.9](#).

Secondly, event-by-event weight which differs between Run I and II that need to be accounted is PID efficiency ε_{PID} (see [Table 7.2](#)), which depends on the kinematics of the final state particles. It is denoted as $w_{\varepsilon_{PID}^i}^{i\varepsilon\{2012,2016\}}$.

The final weight of an event depending on Run I and Run II is calculated

$$w_{total}^{i\varepsilon\{2012,2016\}} = w^{i\varepsilon\{2012,2016\}} \times w_{\varepsilon_{PID}^i}^{i\varepsilon\{2012,2016\}}. \quad (7.12)$$

	Summary	2012 Simulation	2016 Simulation
N_{REC}		1114130	1107715
\mathcal{L}		2968 pb^{-1}	1612 pb^{-1}
$\sigma_{pp \rightarrow b\bar{b}}$		1	2

Table 7.9: Signal simulation weights used to create cocktail of mixed Run I (2012) and Run II (2016) cocktail. The cross-sections listed here are not absolute numbers, but rather relative as only its ratio matters.

After obtaining combined Run I and II signal *cocktail*, fit to this weighted simulation is done with **the shape** in the corrected mass modelled by double-sided Crystal Ball function [section C.2](#). Fit and its parameters can be seen in [Figure 7.7](#). The function describing this shape is denoted as f^{sig} and hence is a function 6 parameters for non-simultaneous fit and 12 parameters for simultaneous fit.

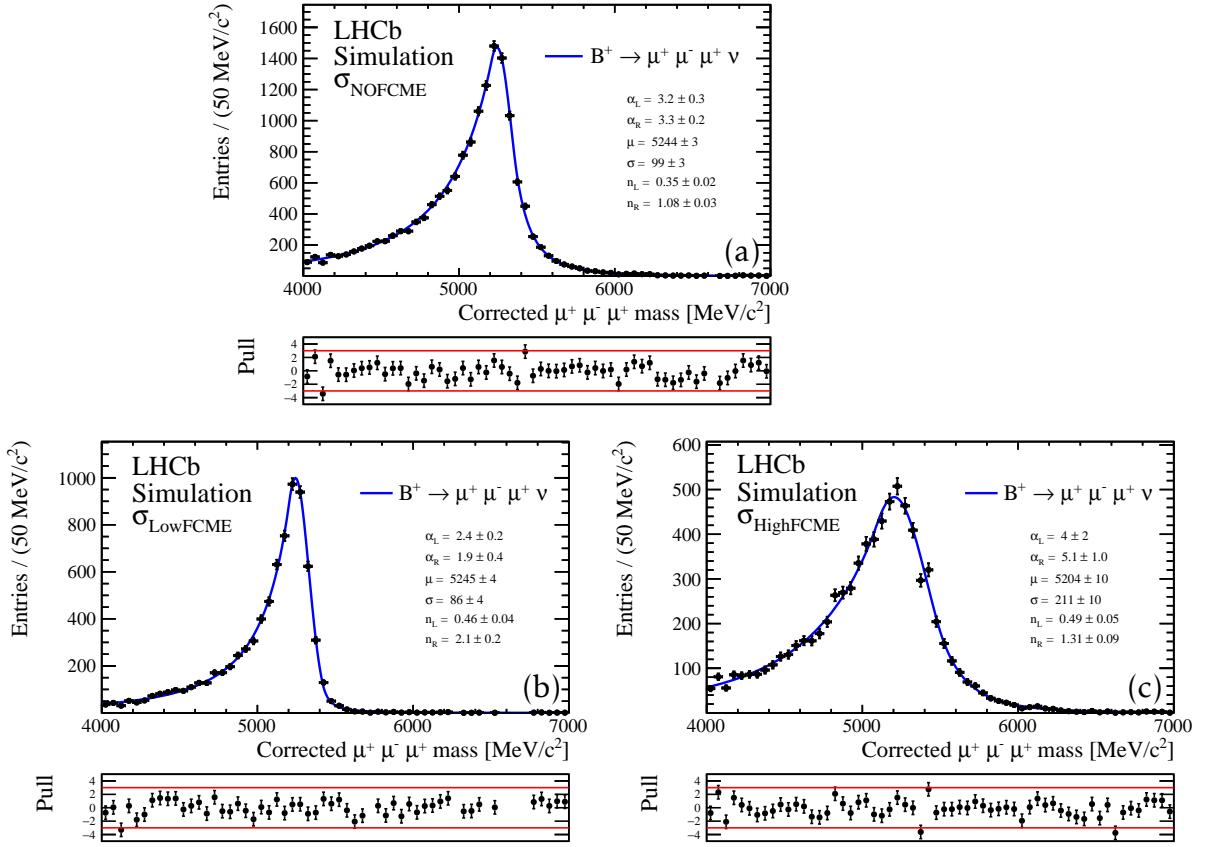


Figure 7.7: Fit to weighted combined signal *cocktail* for (a) NO FCME (b) Low FCME and (c) High FCME split.

The signal yield that is observed in the fit, $N^{sig} = N(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$, is related to the branching fraction using normalisation channel in a following way:

$$\begin{aligned}
 \mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) &= \alpha \times N(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) \\
 &= \frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+) \times \varepsilon_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+}}{N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+) \times \varepsilon_{B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu}} \times N(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu), \\
 &= \underbrace{\frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)}{N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)}}_{\alpha} \times R_{FCME} \times N(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu),
 \end{aligned} \tag{7.13}$$

where ε^x is total selection efficiency of x channel, N_x is number of x decays, $\mathcal{B}(x)$ is

the branching fraction of decay x , R_{FCME} is the relevant efficiency ratio between two decays detailed in [section 7.2](#) and finally α is the *single event sensitivity*, variable which describes the sensitivity for the search. Hence for Run I and Run II altogether for non-simultaneous fit N^{sig} is a function of six parameters $N^{sig}(R_{FCME}^{21}, R_{FCME}^{26}, N^{Run\ I}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), N^{Run\ II}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), \mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), \mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu)$. In simultaneous case $R_{FCME}^{21}, R_{FCME}^{26}, N^{Run\ I}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)$ and $N^{Run\ II}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)$ is further split into $\sigma_{lowFCME}$ and $\sigma_{highFCME}$ bins, resulting in 10 parameters for N^{sig} in the fit.

Partially Reconstructed Background

Partially reconstructed backgrounds are still non-negligible after full selection chain. In order to account for the contribution of the partially reconstructed backgrounds in the final signal fit, contaminating yield needs to be estimated and shape needs to be modelled. Simulation sample for partially reconstructed background that originates through D^0 was described in [section 6.3](#) is used for both the yield estimate and shape modelling.

In order to get estimate for **the yield** of these partially reconstructed decays, N^{PR} , $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ decays are again used as normalisation. Normalising to $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ decay channel following relationship must hold

$$\begin{aligned} \frac{N_{B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu}}{N_{B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+}} &= \frac{\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu)}{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)} \times \frac{\varepsilon^{B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu}}{\varepsilon^{B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+}} \\ &= \frac{\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu)}{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)} \times R_{FCME}, \end{aligned} \tag{7.14}$$

where ε^x is total selection efficiency of x channel and R_{FCME} is hence the efficiency ratio between the two channels, N_x is number of x decays, $\mathcal{B}(x)$ is the branching fraction of decay x . The quantity of interest, $N_{B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu}$, can be therefore calculated given the knowledge of all other terms. $N_{B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+}$ was obtained from [Table 7.8](#).

$\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu)$ is obtained by amalgamating $\mathcal{B}(D^0 \rightarrow K^+\pi^-\mu^+\mu^-) = (4.17 \pm 0.12 \pm 0.40) \times 10^{-6}$ [60] and $\mathcal{B}(B^+ \rightarrow Dl^+\nu X) = (9.8 \pm 0.7) \times 10^{-2}$ [1] yielding $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu) \approx (4.10 \pm 0.50) \times 10^{-7}$. $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ branching fraction is obtained with the same approach: multiplying $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.026 \pm 0.031) \times 10^{-3}$ [1] and $\mathcal{B}(J/\psi \rightarrow \mu^-\mu^+) = (5.961 \pm 0.0033) \times 10^{-2}$ [1] yielding

$$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+) = (6.12 \pm 0.19) \times 10^{-5}. \quad (7.15)$$

All the relevant total selection efficiencies are obtained from the full simulation sample and are shown in Table 7.10. Due to the usage of a proxy simulation for this partially reconstructed decays (using a pion rather than a muon in one case), as discussed in section 6.3, trigger efficiency ε_{TRG} cannot be obtained from simulation for partially reconstructed decays, because the HLT2 trigger (see Table 5.3) would make positive decision only either because of finding dimuon pair or two or three-body decays, hence the trigger ratio $\frac{\varepsilon_{TRG}^{B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu}}{\varepsilon_{TRG}^{B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+}}$ is assumed to be 1, which is rather a conservative estimate (overestimate) but makes sure that other partially reconstructed backgrounds are accounted for. Other efficiency that was not accounted for because of the same reason is PID efficiency, ε_{PID} . Moreover as this proxy simulation was not accessible for Run II, the same ratio of efficiencies as in Run I is used.

The summary of expected yield is summarized in Table 7.10. For the non-simultaneous final fit there are hence 5 parameters $N^{PR}(R_{FCME}^{21}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu), N^{Run\ I}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), N^{Run\ II}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), \mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), \mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu))$ and for simultaneous data fit with FCME splitting 8 parameters. The total yield expected for this type of background is very low compared to other expected backgrounds.

The shape for partially reconstructed backgrounds, f^{PR} , is also obtained from the simulation proxy after all the selection. The shape is best described with sum of two Crystal Ball functions, more in section C.1, with free means μ^1, μ^2 and widths σ^1, σ^2 as seen in Figure 7.8. Because the shape of this background suggests that the majority

Properties	Run I	Run II
$\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	$(4.10 \pm 0.50) \times 10^{-7}$	$(4.10 \pm 0.50) \times 10^{-7}$
$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	$(6.12 \pm 0.19) \times 10^{-5}$	$(6.12 \pm 0.19) \times 10^{-5}$
$\varepsilon_{total}^{B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu}$	$(1.87 \pm 0.04) \times 10^{-4}$	Using 2012
$\varepsilon_{total}^{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+}$	$(5.80 \pm 0.01) \times 10^{-3}$	Using 2012
$R_{FCME}^{21}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	$(3.22 \pm 0.07) \times 10^{-2}$	Using 2012
$\sigma_{lowFCME}$		
$N_{B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu}$	19.8 ± 2.6	11.7 ± 1.5
$\sigma_{highFCME}$		
$N_{B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu}$	17.0 ± 2.2	7.9 ± 1.0
σ_{NOFCME}		
$N_{B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu}$	37.3 ± 4.8	20.3 ± 2.6

Table 7.10: Summary of number of events that comes from partially reconstructed backgrounds in different bins of FCME, assuming 2012 efficiencies but extrapolating to all samples.

contamination is below $5000 \text{ MeV}/c^2$, it is one of the least dangerous backgrounds. In the non-simultaneous fit f^{PR} is a function of 9 parameters and for simultaneous fit 18 parameters coming from sum of two Crystal Ball functions.

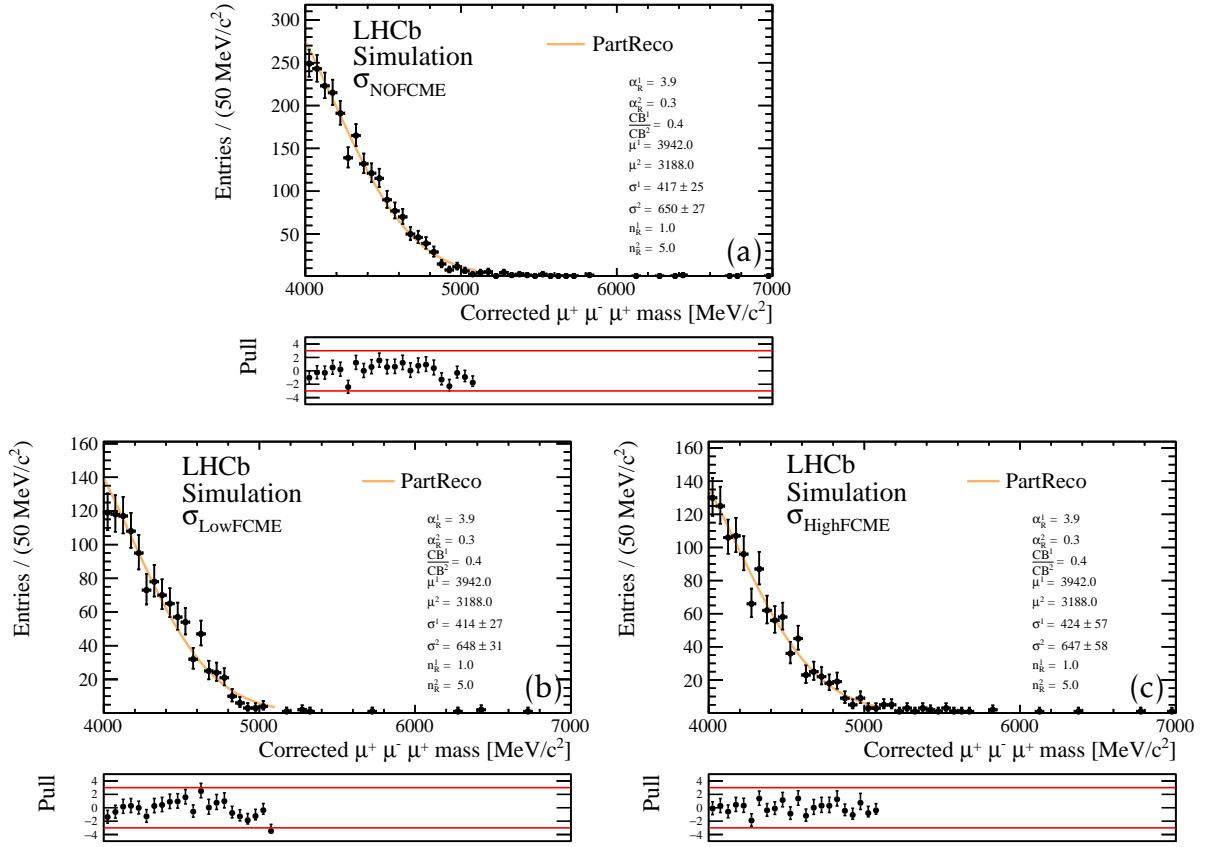


Figure 7.8: Fit to weighted combined partially reconstructed background simulation proxy for (a) NO FCME (b) Low FCME and (c) High FCME split.

MisID background

The level and the shape of misID background is determined by fitting the misID data samples obtained using the method described in section 6.2. A binned χ^2 fit is used to extract the shape and yields parameters. The reason for usage of the binned χ^2 fit is that the misID samples are low-statistics weighted samples and the shape and yield needs be propagated to the final data fit while preserving the fit parameter correlations. Since there is prescale factor of 1% at stripping level, to obtain the correct yield, the final number needs to be multiplied by to counteract the prescale.

The misID weights obtained from kinematically binned $B^0 \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)(K^* \rightarrow$

$\pi^+K^-)$ samples (see [subsection 4.3.1](#)) have uncertainties associated with them as shown in [Figure 4.10](#), [Figure 4.11](#), [Figure 4.12](#), [Figure 4.13](#). These uncertainties are accounted for in the fit by Gaussian variation of the weights within the uncertainty in a given kinematic bin of p, η for each particle species and then folded in to the misID calculation. In this case 100 variations were used. Each variation results in different template for misID shape. This misID template is then subsequently binned in 15 bins of corrected mass. From each corrected mass bin, mean μ_{var} and error σ_{var} from gaussianly distributed number of misID events is obtained.

The total uncertainty due to the weight, σ_{tot} , for given bin of corrected mass is calculated using $\sqrt{\sigma_{var}^2 + \sqrt{\sum w_i^2}}$, where $\sigma_{par} = \sqrt{\sum w_i^2}$ is the associated error per bin and σ_{var} is the standard deviation obtained from variation of misID weights. Finally, The binned χ^2 fit is made to the misID samples with the total uncertainty. **The number of misID events**, N^{MisID} , for different species-regions after all selections are seen in [Table 7.11](#). Hence for non-simultaneous fit this just add 1 parameter and for simultaneous fit there will be two parameters describing the total yield of misID.

Also it can be seen in [Table 7.11](#), crossfeedweight is only considered for kaon-like and pion-like *SS misID* samples. This arises as a consequence of two characteristics of the misID crossfeedweight procedure. First the convergence criteria makes unbalanced samples (one sample very high in misID events and other sample very small in number of misID events) hard to satisfy (case for most of *OS misID* samples). Secondly, proton-like region samples are very sparse and hence it is not necessary to account for crossfeed.

The binned χ^2 fits using Crystal Ball function to different bins of FCME is performed as seen in [Figure 7.9](#). This means that **the shape** of this background, f^{MisID} , is a function of 4 parameters in non-simultaneous case and a function of 8 parameters in simultaneous case. Both full weight error σ_{tot} and partial weight error σ_{par} are plotted. The difference between the two is the error due to uncertainty on the weight σ_{var} . Results of the fits are propagated into the signal data fits preserving correlations between parameters, which are hence set as multidimensional gaussian constraints in the signal data fits. This means that all uncertainties due to misID will be directly

accounted for in the signal data fits.

Sample	Region	PID	weight	Cummulative misID count
Run I <i>SS misID</i>	Kaon-region	Run I PID	crossfeedweight	198
Run I <i>SS misID</i>	Pion-region		crossfeedweight	301
Run I <i>SS misID</i>	Proton-region		no-crossfeedweight	307
Run I <i>OS misID</i>	Kaon-region		no-crossfeedweight	310
Run I <i>OS misID</i>	Pion-region		no-crossfeedweight	352
Run I <i>OS misID</i>	Proton-region		no-crossfeedweight	353
Run II <i>SS misID</i>	Kaon-region	Run II PID	crossfeedweight	489
Run II <i>SS misID</i>	Pion-region		crossfeedweight	565
Run II <i>SS misID</i>	Proton-region		no-crossfeedweight	565
Run II <i>OS misID</i>	Kaon-region		no-crossfeedweight	573
Run II <i>OS misID</i>	Pion-region		no-crossfeedweight	618
Run II <i>OS misID</i>	Proton-region		no-crossfeedweight	619

Table 7.11: The final misid template is constructed by summing the contribution from Run I and II kaon, pion and proton-like regions for both *SS* and *OS misID* contributions. The last column adds cummatively the contributions with respect to the previous row.

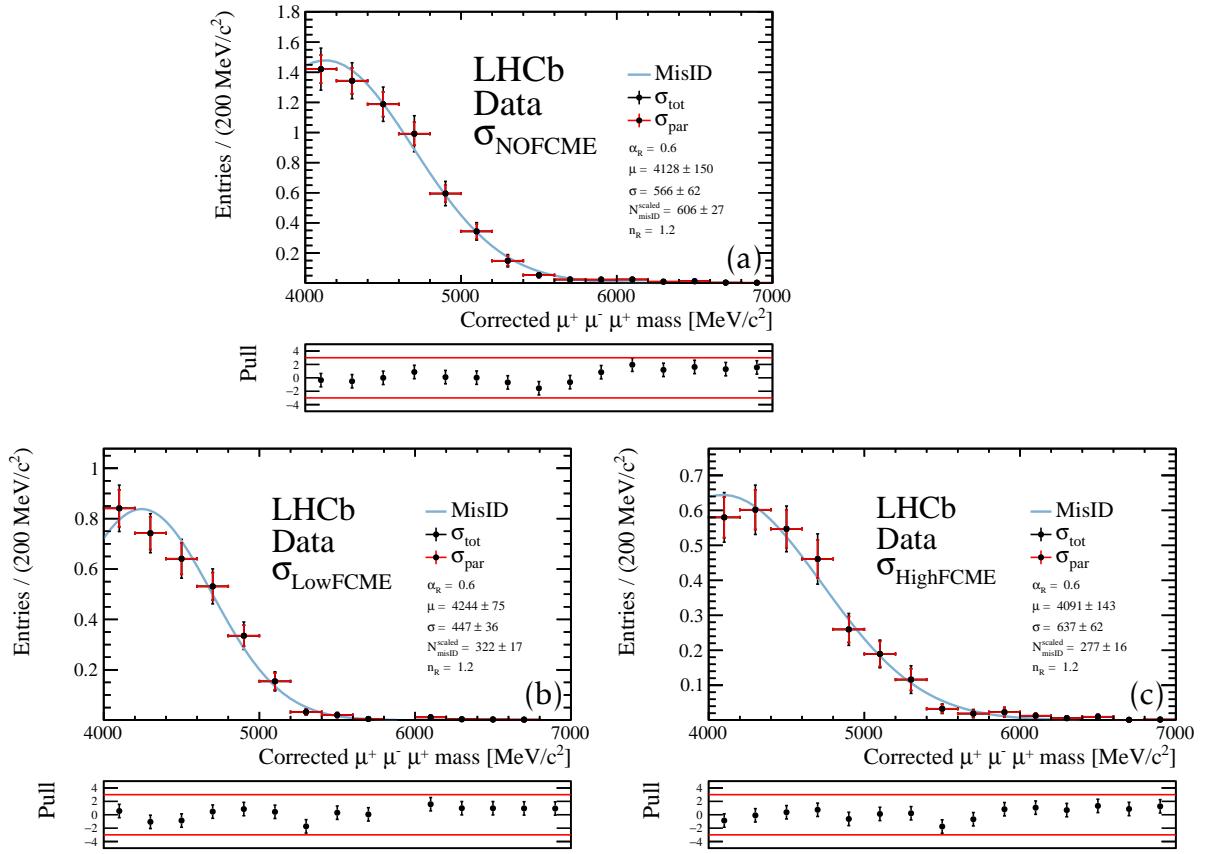


Figure 7.9: Binned χ^2 to misid template with no FCME split (b) Low FCME (c) High FCME. In high FCME bin, the distribution of misid is pollutes signal window more than in the low FCME bin. Both full weight error σ_{tot} and partial weight error σ_{par} can be seen.

Combinatorial Background

Signal data fit so far includes components for signal component, partially reconstructed background component, and misid background component. The only component left to estimate is the contamination of combinatorial background. To model the combinatorial component exponential function is left floating and fit to data, which was motivated in section 6.1. Hence **the shape**, f^{Combi} , has one(two) parameter(s) for non-simultaneous(simultaneous) fit, which is the exponential constant. And **the yield**,

N^{Combi} , is also parametrised just with one (two) parameter(s) for non-simultaneous (simultaneous) fit.

7.3.4 Signal Data Fits

Signal fit models

The full signal fit model described by probability density function (PDF) consists of all the mentioned components. As mentioned in [subsection 7.3.3](#) there are two fit models that are used to fit data: non-simultaneous, $f^{NS}(x, y^i)$, and simultaneous fit model, $f^S(x, z^k)$, where the latter is the one which is used for limit setting. In this case x is the corrected mass, and $y^i(z^k)$ are all the parameters for non-simultaneous (simultaneous) fit. The total fit models hence can be written

$$f^{NS}(x, y^i) = N^{sig} \times f^{sig} + N^{MisID} f^{MisID} + N^{PR} f^{PR} + N^{Combi} \times f^{Combi}, \quad (7.16)$$

$$f^S(x, y^k) = \sum_{j \in \sigma} (N_j^{sig} \times f_j^{sig} + N_j^{MisID} f_j^{MisID} + N_j^{PartReco} f_j^{PartReco} + N_j^{Combi} \times f_j^{Combi}), \quad (7.17)$$

where σ is the bin of fractional corrected mass. The shared parameters in the two bins of fractional corrected mass error for the simultaneous fit are the branching fractions $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu), \mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu), \mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$.

Signal fit model hypotheses

As the strength of the signal is not a priori known there are two types of hypotheses that are established: one where there is a presence of signal, also known as signal+background hypothesis model which takes exactly form of [Equation 7.16](#) and then background only hypothesis, where $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = 0 \rightarrow N^{sig} = 0$.

Signal data types

There are two types of signal datasets as mentioned in [section 5.3](#). Firstly, there is **blinded signal data** to which the simultaneous and non-simultaneous fits are per-

formed in order to evaluate the expected sensitivity. Since these datasets are blinded only mass regions $4000\text{ MeV}/c^2 < M_{B_{\text{corr}}} < 4500\text{ MeV}/c^2$ and $5500\text{ MeV}/c^2 < M_{B_{\text{corr}}} < 7000\text{ MeV}/c^2$ are visible in the fits. These fits are shown in [subsection 7.3.6](#). Secondly fits to **full signal data** are performed and these are used in order to set the limit on the $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$. In these fits full mass spectrum is visible.

7.3.5 Signal Fit Systematics

Systematics studies are performed in order to account for possible shortcomings of the methods used in order to measure $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$. The summary of systematic uncertainties is provided in [Table 7.12](#). Most of this systematic uncertainties directly affect efficiency ratio between signal and normalisation channel. In that case, fits account for this uncertainty directly by incorporating the uncertainty into efficiency ratios by adding them in quadrature to the statistical error. For the simultaneous fit these uncertainties are assumed to be 100% correlated between the bins of FCME.

Systematic source	Run I/%	Run II/%	Error Overall
$\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$	0.6	0.6	0.6
$\mathcal{B}(B^+ \rightarrow J/\psi K^+)$	3.0	3.0	3.0
Signal Decay Model	4.6	4.6	4.6
Trigger Data/Sim	5.1	1.3	3.2
Trigger HLT2	-	1.5	1.5
Kaon interaction probability	2.0	2.0	2.0
Kinematic Reweighting	1.0	2.0	1.5
Simulation Statistics	1.3	0.7	0.8
Fit bias	1.0	1.0	1.0
Total	8.0	6.6	7.1
Statistical source	Run 1/%	2016/%	Error Overall
Error on $B^+ \rightarrow J/\psi K^+$ yield	0.3	0.3	0.2

Table 7.12: Summary of systematic uncertainties.

Signal Decay Model

Then the largest systematic uncertainty arises due to the choice of decay model for the signal channel. The nominal signal model creates a photon pole, increasing the branching fraction in the low invariant dimuon mass region. The associated systematic uncertainty is estimated by replacing this decay model with one determined purely by phase-space and by comparing the selection efficiencies, resulting in 4.6% systematic uncertainty on the efficiency ratio.

Trigger

The second most important systematic effect affecting the efficiency ratio is due to incorrect modelling of the trigger in simulation. This can be split into two effects: the efficiency difference between used simulation and data and incorrect emulation of HLT2 on the simulation in 2016 as only one TCK was considered for this trigger selection. This

uncertainty is assigned by comparing the difference between the trigger efficiency of $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ decays in simulation and data using TISTOS method.

TISTOS method allows measuring the trigger efficiency, whereby the events that are **TIS** and **TOS** are assumed to be independent. Hence, the efficiency for selection of **TOS** candidates can be calculated by

$$\epsilon_{TOS} = \frac{N_{TIS\&TOS}}{N_{TIS}}. \quad (7.18)$$

In [Equation 7.18](#) $N_{TIS\&TOS}$ is the number of **TIS** and **TOS** events passing the trigger requirement and N_{TIS} is the number of **TIS** events which passed the trigger requirement.

The efficiency difference of the full trigger chain between $B^+ \rightarrow J/\psi K^+$ decays in used simulation and *Sweighted* data is 3.2%. Incorrect emulation of HLT2, which is calculted by comparing HLT2 trigger efficiency of the chosen TCK simulation and all the *Sweighted* data, yields 1.5%. Altogether the incorrect modelling of the trigger in simulation results in 3.5% systematic uncertainty.

Kaon Interaction Probability

One difference between the signal and the normalisation channels that can have effect on the efficiencies is that the kaon in the decay $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ can interact with the detector at a probability proportional to the amount of material traversed. The uncertainty on this amount of material leads to a 2% systematic uncertainty, which follows the procedure outlined in Ref. [34].

Kinematic Reweighting

Another source of systematic uncertainty is caused by differences in the B^+ production kinematics in the simulation. Corrected variables are momentum, p , transverse momentum p_T and vertex χ^2 of B . The kinematic weights in a given kinematic bin are calculated using normalised histograms of $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ data, and normalized

$B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ simulation in a following way

$$w_{(bin)} = \frac{w_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+ data, bin}}{w_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+ simulation, bin}}. \quad (7.19)$$

Obtained correction weights are then applied to correct both $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$, $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ simulation. The difference between uncorrected and kinematically corrected efficiency ratios is assigned as a systematic uncertainty yielding 1.5% systematic uncertainty on the efficiency ratios.

Signal fit bias and coverage

In this section systematic uncertainty due to signal bias is evaluated. The pull is calculated using pseudo-experiments where the data is generated for signal branching fraction $\mathcal{B} = 1.0 \times 10^{-8}$ corresponding to ≈ 17 signal events for Run I and II data. These pseudo-datasets are then re-fitted with floating \mathcal{B} and corresponding number of fitted signal events is obtained. The pull is defined as the difference between number of fitted signal events and the number events that pseudo-experiment was created with, divided by the error on the number of signal events given by the fit:

$$\frac{N_{sig}^{orig} - N_{sig}^{fit}}{\sigma^{fit}}. \quad (7.20)$$

The pull distributions then can be fitted with gaussian function and the fit bias is calculated as a shift of the mean from the 0. The overcoverage/undecoverage of the fit is established by shift of the standard deviation of the gaussian function from 1.

For this study 10000 pseudo-experiments were created testing the bias and coverage of both the extended non-simultaneous fit and extended simultaneous fit and the pull distributions can be seen in [Figure 7.10](#).

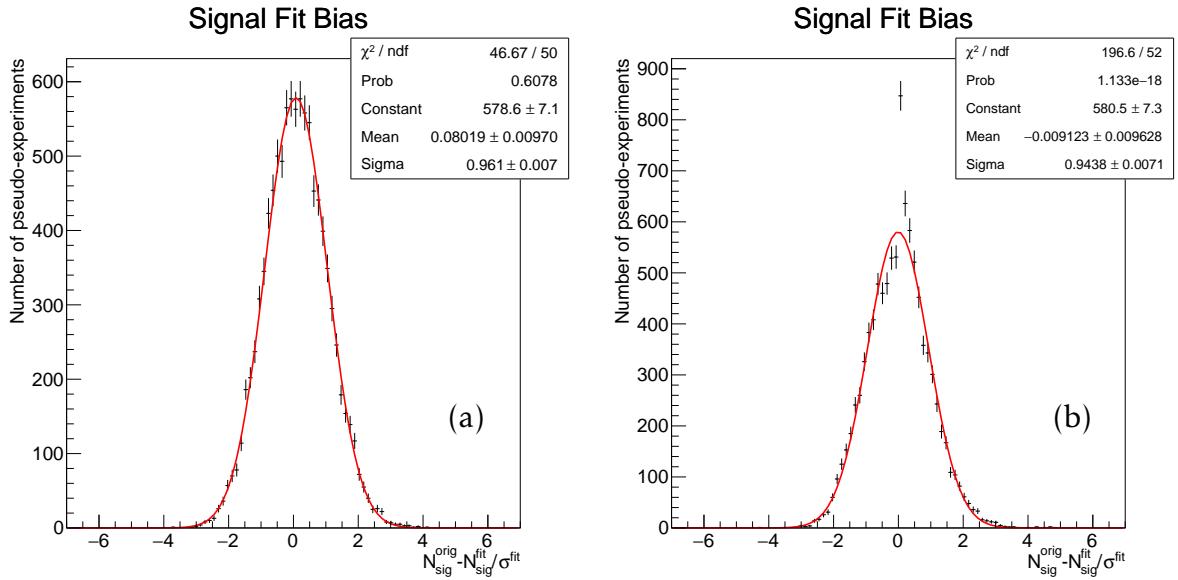


Figure 7.10: (a) Non-simultaneous fit, (b) simultaneous fit pulls.

As it can be seen, for non-simultaneous fit there is overcoverage of 4 % but since the statistical error is of the order of 100% this is insignificant. The bias shows 8% preference for lower signal yields, and hence this will be added as a systematic uncertainty.

Fit	\mathcal{B}	Statistical Error	Overcoverage	Bias
Not Simultaneous	1.0×10^{-8}	$\approx 100\%$	4%	8%
Simultaneous	1.0×10^{-8}	$\approx 100\%$	6%	1%

Table 7.13: Signal bias estimate from 10 000 pseudo-experiments for both simultaneous and not simultaneous fit.

Others

Other smaller systematic uncertainties are assigned due to the finite size of the simulation and the branching fraction of the decay $J/\psi \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$.

7.3.6 Blinded Data Fits

In order to be able to get the expected sensitivity for this search a simultaneous unbinned maximum likelihood extended fit to the blinded data of corrected mass after the full selection in two bins of FCME is performed. As a crosscheck, also non-simultaneous fit with no FCME split is done. The summary of all the components adding to the total PDF for the blinded signal data fit, their modelling and constraints are shown in [Table 7.14](#).

Most of the parameters in these fits are fixed and if their are not fixed they are constrained with their range allowed to be within $\pm 5\sigma$ of the constraint. Error propagation from the parametrisations of different components is dealt with by using two types of constraints: gaussian constraints and multivariate gaussian constraints. The gaussian constraint, *gaussian*, when imposed has central value of the fitted parameter and as width the error of the fitted parameter. Multivariate gaussian constraint, *mvg_gaussian*, is generalisation of gaussian constraint to higher dimensions and is used for misID parametrisation as the correlations between parameters' errors need to be propagated to the signal fit.

The maximum unbinned likelihood fit to blinded signal data after all selection is shown in [Figure 7.11](#) for all categories of FCME. As the signal region is blinded in this case the fit model uses background-only hypothesis. The total number of expected background events, N_b , can be then obtained by integrating the total PDF in the signal region.

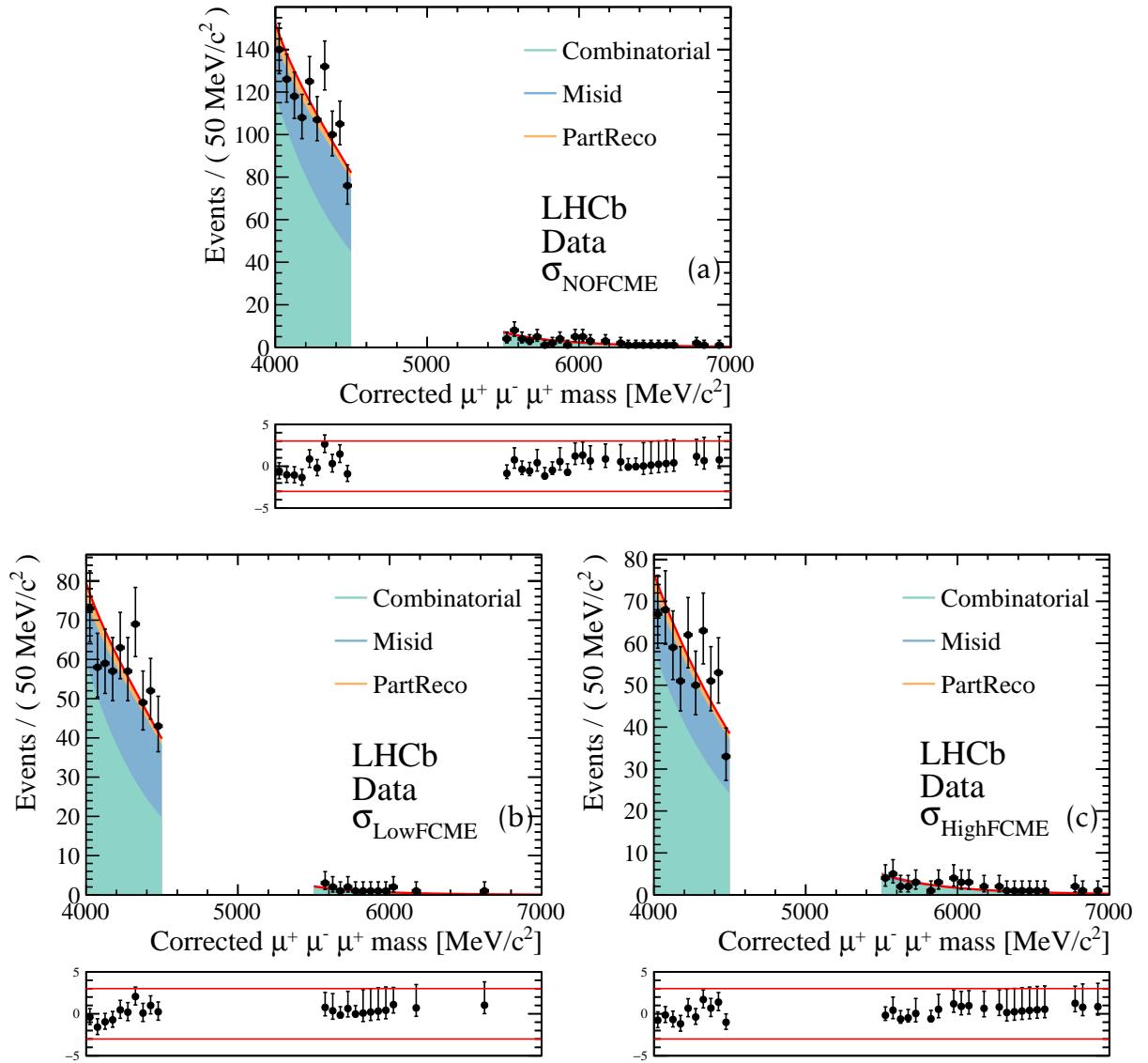


Figure 7.11: (a) Unbinned maximum likelihood fit to the blinded data in one bin of FCME. Simultaneous unbinned maximum likelihood fit to blinded data after full selection chain in two bins of FCME, with (b) fit to σ_{low} bin, (c) σ_{high} bin.

Fit Parameter	Status	Constraint	Obtained in
Yields			
N^{sig}	$\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	Free	-
	$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	Constrained	gaussian
	$R_{FCME}^{21}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	Constrained	gaussian
	$R_{FCME}^{26}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	Constrained	gaussian
	$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)^{RunI}_{FCME}$	Constrained	gaussian
	$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)^{RunII}_{FCME}$	Constrained	gaussian
N^{PR}	$\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	Constrained	gaussian
	$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	Constrained	gaussian
	$R_{FCME}^{21}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	Constrained	gaussian
	$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)^{RunI}_{FCME}$	Constrained	gaussian
	$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)^{RunII}_{FCME}$	Constrained	gaussian
N^{Misid}	Constrained	mvg_gaussian	Figure 7.9
N^{Combi}	Free	-	This fit
MisID Shape Parameters (Crystal Ball function)			
μ_{misid}	Constrained	mvg_gaussian	Figure 7.9
σ_{misid}	Constrained	mvg_gaussian	Figure 7.9
Others	Fixed	fixed to values seen in	Figure 7.9
PartReco Shape Parameters (sum of two Crystal Ball functions)			
All	Fixed	fixed to values seen in	Figure 7.8
Signal Shape Parameters (Double-sided Crystal Ball function)			
All	Fixed	fixed to values seen in	Figure 7.7
Combinatorial Shape Parameters (exponential function)			
β	Free	-	This fit

 Table 7.14: For all constrained variables the range is set to be within $\pm 5\sigma$.

7.3.7 Expected Sensitivity

The CL_s method [61] is used to produce expected upper limits from the blinded fit results, where extrapolation to the signal region is performed. This means that for a given branching fraction assumption \mathcal{B} :

- 1000 pseudo datasets with no signal injected are generated (background only hypothesis) and are denoted as data_{bkg} .
- 1000 pseudo datasets with a given signal branching fraction hypothesis injected are generated (signal+background hypothesis) and are denoted as data_{sig} .
- Likelihood ratio $-2\log \frac{\mathcal{L}(s+b)}{\mathcal{L}(b)}$ is defined, where $\mathcal{L}(s+b)$ is the likelihood function under signal and background hypothesis and $\mathcal{L}(b)$ is likelihood under background only hypothesis.
- Fits with background only hypothesis and signal+background hypothesis are performed to both data_{bkg} and data_{sig} . By extracting the likelihood ratio from these fits likelihood ratio distributions for signal, ΔLL_{sig} , and background, ΔLL_{bkg} , are obtained.
- The median of background ΔLL_{bkg} is identified.
- CL_b is calculated by integrating ΔLL_{bkg} from the tail of the distribution to this median. In this case, CL_b represents the probability to obtain a result less compatible with the signal than the observed one in the background-only hypothesis.
- CL_{s+b} is calculated by integrating ΔLL_{sig} from the tail to this median. CL_{s+b} then represents the probability to obtain a result which is less compatible with the signal than the observed result, assuming the signal+background hypothesis.
- CL_s then can be calculated as CL_{s+b}/CL_b .

The expected exclusion limits are shown in ??.

7.3.8 Unblinded Data Fits

After full strategy was reviewed the signal datasets were unblinded, observing no significant signal. The simultaneous unbinned extended maximum likelihood fit to all the data was performed. As expected with no excess signal, background-only hypothesis fit described the data very well, which can be seen in Figure 7.12.

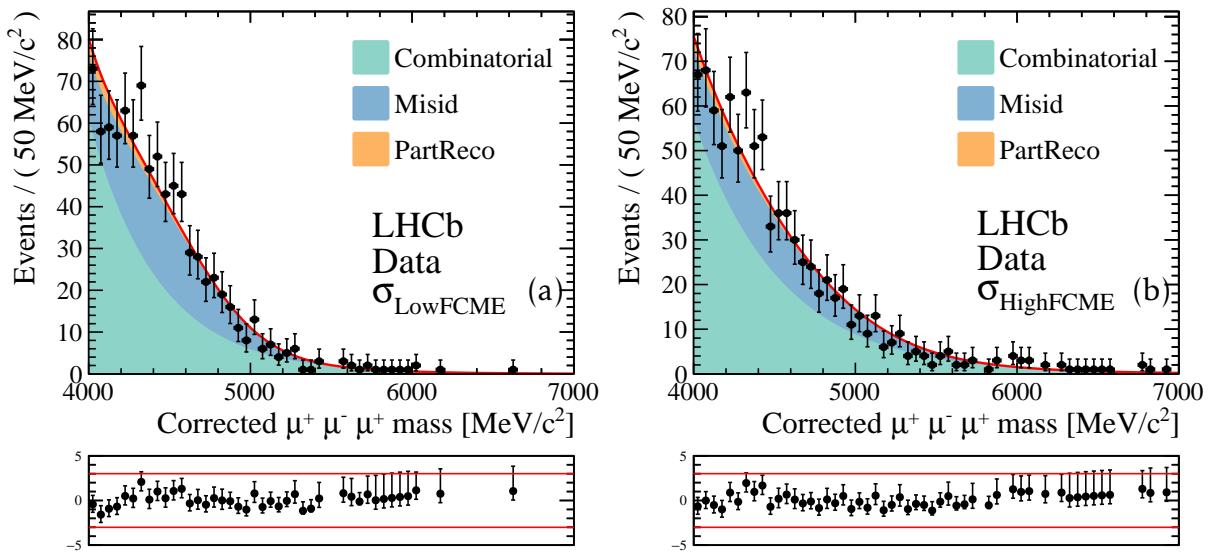


Figure 7.12: Simultaneous unbinned extended maximum likelihood fit to unblinded data after full selection chain in two bins of FCME and with $\mathcal{B} = 0$, with (b) fit to σ_{low} bin, (c) σ_{high} bin.

In order to perform fit with signal and background hypothesis good range of the \mathcal{B} needs to be established in order to aid the fit convergence. This is done by maximizing the log likelihood values for different values of \mathcal{B} , also known as profiling the \mathcal{B} . The results of the profiling can be seen in Figure 7.13. This shows that the minimum is at $\mathcal{B} \sim -2.0 \times 10^{-8}$, hence the fit prefers negative values of \mathcal{B} .

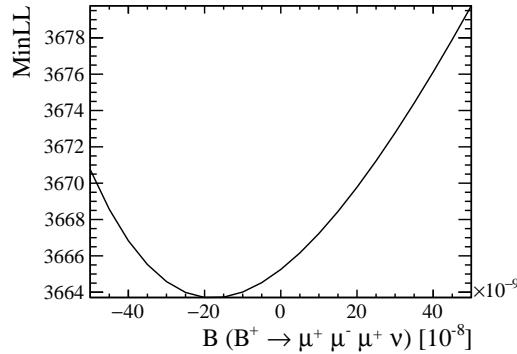


Figure 7.13: Minimized -log likelihood value at different \mathcal{B} .

The simultaneous unbinned extended maximum likelihood fit to all the data with signal+background hypothesis (with the \mathcal{B} floating) can be seen in Figure 7.14 converging at value of $\mathcal{B} = -1.78 \times 10^{-8}$. This hence represents a downward fluctuation. The full list of all free and constrained parameters in the fit with its values are shown in Table 7.15. This fit results then can be translated into yields in all the fitting range and specifically blinded region by integrating the resulting PDF in a relevant region as shown in Table 7.16.

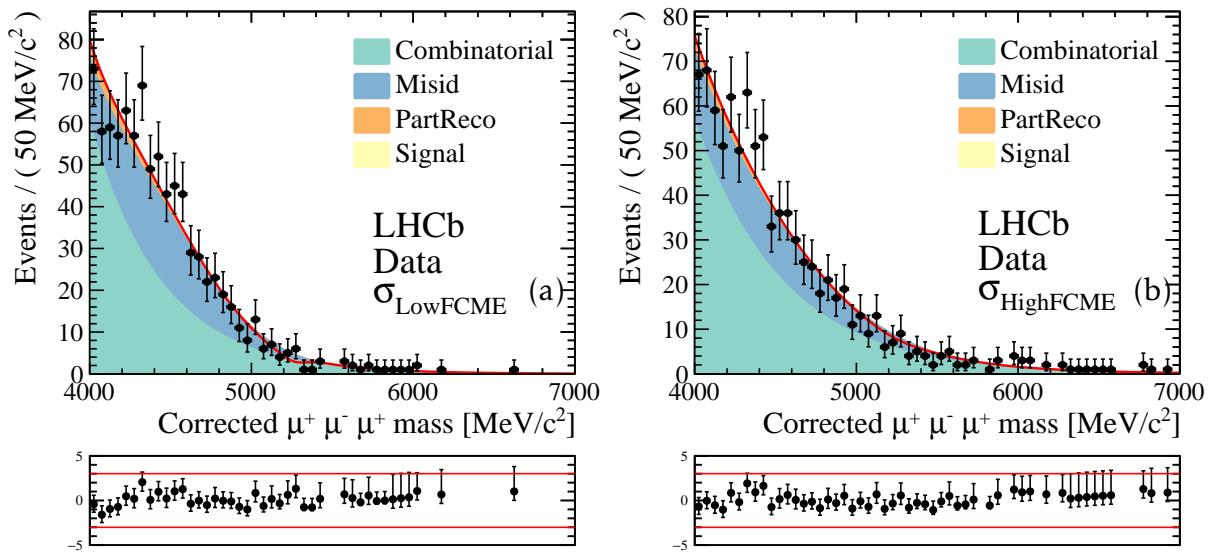


Figure 7.14: Simultaneous unbinned maximum likelihood fit to unblinded data after full selection chain in two bins of FCME, with (b) fit to σ_{low} bin, (c) σ_{high} bin.

Fit Parameter	Value
$\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	$-(1.78 \pm 0.936) \times 10^{-8}$
$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	$(6.11 \pm 0.185) \times 10^{-5}$
$\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	$(4.12 \pm 0.5) \times 10^{-7}$
$\mu_{misID_{high}}$	4100 ± 123
$\mu_{misID_{low}}$	4330 ± 56
$\sigma_{misID_{high}}$	619 ± 56
$\sigma_{misID_{low}}$	401 ± 29.9
$R_{high}^{26}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	1.97 ± 0.0788
$R_{low}^{26}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	3.31 ± 0.1
$R_{high}^{21}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	0.0397 ± 0.00116
$R_{low}^{21}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	0.0271 ± 0.000794
$R_{high}^{21}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	1.96 ± 0.131
$R_{low}^{21}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	3.15 ± 0.233
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)_{high}^{2016}$	29800 ± 173
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)_{low}^{2016}$	64700 ± 254
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)_{high}^{Run1}$	64100 ± 253
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)_{low}^{Run1}$	109000 ± 331
$N_{misID_{high}}^{scaled}$	278 ± 15.2
$N_{misID_{low}}^{scaled}$	331 ± 16.3
β_{high}	$-0.00183 \pm 9.19 \times 10^{-5}$
β_{low}	-0.00226 ± 0.000129
$N_{combi_{high}}$	620 ± 34.3
$N_{combi_{low}}$	531 ± 35.4

Table 7.15: Fit results for all floating variables in the unblinded data fit. The $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = -1.78 \times 10^{-8}$. Variables R_K^S are the efficiency ratios obtained by normalising the decays to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ decays where S stands for stripping and K for the FCME. $N_{misID_K}^{scaled}$ is the number of misID events, N_{combi_K} is the number of combinatorial events, β_K is the exponential constant, μ_{misID_K} and σ_{misID_K} are the mean and the σ of the CB function.

Component	All ($4 - 7 \text{ GeV}/c^2$)	Signal region ($4.5 - 5.5 \text{ GeV}/c^2$)
$N_{misID_{low}}$	331	139
$N_{sig_{low}}$	-15.8	-14.4
$N_{combi_{low}}$	531	154
$N_{partreco_{low}}$	31.8	6.39
$N_{misID_{high}}$	279	122
$N_{sig_{high}}$	-14.0	-11.0
$N_{combi_{high}}$	620	209
$N_{partreco_{high}}$	25.1	5.17

Table 7.16: Resulting yields for different components from the corrected mass fit with $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = -1.78 \times 10^{-8}$.

Chapter 8

Results

Bibliography

- [1] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, **Chin. Phys. C40** (2016), no. 10 100001.
- [2] Y. H. Ahn, H.-Y. Cheng, and S. Oh, *Wolfenstein parametrization at higher order: Seeming discrepancies and their resolution*, **Physics Letters B 703** (2011), no. 5 571 .
- [3] C. Jarlskog, *Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Violation*, **Phys. Rev. Lett. 55** (1985) 1039.
- [4] UTfit, M. Bona *et al.*, *Results as of summer 2016, updated results and plots available at: <http://utfit.org/UTfit/WebHome>*, **JHEP 10** (2006) 081, [arXiv:hep-ph/0606167](https://arxiv.org/abs/hep-ph/0606167).
- [5] CKMfitter Group, J. Charles *et al.*, *Results as of up to ICHEP 2016, updated results and plots available at: <http://ckmfitter.in2p3.fr>*, **Eur. Phys. J. C41** (2005), no. 1 1, [arXiv:hep-ph/0406184](https://arxiv.org/abs/hep-ph/0406184).
- [6] W. J. Marciano and A. Sirlin, *Electroweak Radiative Corrections to tau Decay*, **Phys. Rev. Lett. 61** (1988) 1815.
- [7] W.-S. Hou, *Enhanced charged Higgs boson effects in $B^- \rightarrow \tau^- \bar{\nu}_\tau$, $\mu^- \bar{\nu}_\mu$ and $b \rightarrow \tau^- \bar{\nu}_\tau + X$* , **Phys. Rev. D48** (1993) 2342.
- [8] A. G. Akeroyd and S. Recksiegel, *The Effect of H^\pm on $B^\pm \rightarrow \tau^\pm \bar{\nu}_\tau(\tau)$ and $B^\pm \rightarrow \mu^\pm$ muon neutrino*, **J. Phys. G29** (2003) 2311, [arXiv:hep-ph/0306037](https://arxiv.org/abs/hep-ph/0306037).
- [9] B. A. Dobrescu and A. S. Kronfeld, *Accumulating evidence for nonstandard leptonic decays of D_s mesons*, **Phys. Rev. Lett. 100** (2008) 241802, [arXiv:0803.0512](https://arxiv.org/abs/0803.0512).

- [10] A. Arbey, F. Mahmoudi, O. Stal, and T. Stefaniak, *Status of the Charged Higgs Boson in Two Higgs Doublet Models*, *Eur. Phys. J.* **C78** (2018), no. 3 182, [arXiv:1706.07414](#).
- [11] Belle, I. Adachi *et al.*, *Evidence for $B^- \rightarrow \tau^- \bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle*, *Phys. Rev. Lett.* **110** (2013), no. 13 131801, [arXiv:1208.4678](#).
- [12] Belle, B. Kronenbitter *et al.*, *Measurement of the branching fraction of $B^+ \tau^+ \nu_\tau$ decays with the semileptonic tagging method*, *Phys. Rev.* **D92** (2015), no. 5 051102, [arXiv:1503.05613](#).
- [13] BaBar, J. P. Lees *et al.*, *Evidence of $B^+ \rightarrow \tau^+ \nu$ decays with hadronic B tags*, *Phys. Rev.* **D88** (2013), no. 3 031102, [arXiv:1207.0698](#).
- [14] BaBar, B. Aubert *et al.*, *A Search for $B^+ \rightarrow \ell^+ \nu_\ell$ Recoiling Against $B^- \rightarrow D^0 \ell^- \bar{\nu} X$* , *Phys. Rev.* **D81** (2010) 051101, [arXiv:0912.2453](#).
- [15] Belle, A. Sibidanov *et al.*, *Search for $B^- \rightarrow \mu^- \bar{\nu}_\mu$ Decays at the Belle Experiment*, [arXiv:1712.04123](#).
- [16] G. Burdman, J. T. Goldman, and D. Wyler, *Radiative leptonic decays of heavy mesons*, *Phys. Rev.* **D51** (1995) 111, [arXiv:hep-ph/9405425](#).
- [17] M. Beneke and J. Rohrwild, *B meson distribution amplitude from $B \rightarrow \gamma l \nu$* , *Eur. Phys. J.* **C71** (2011) 1818, [arXiv:1110.3228](#).
- [18] Y.-M. Wang, *Subleading power corrections in radiative leptonic B decay*, *Nucl. Part. Phys. Proc.* **285-286** (2017) 75, [arXiv:1609.09813](#).
- [19] Belle, A. Heller *et al.*, *Search for B^{++} decays with hadronic tagging using the full Belle data sample*, *Phys. Rev.* **D91** (2015), no. 11 112009, [arXiv:1504.05831](#).
- [20] A. Danilina and N. Nikitin, *Rare four-leptonic B -decays with light leptons in the final state in the Standard Model*, *EPJ Web Conf.* **158** (2017) 03005.

- [21] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A Comparative study of the decays $B \rightarrow (K, K^*)\ell^+\ell^-$ in standard model and supersymmetric theories*, *Phys. Rev.* **D61** (2000) 074024, [arXiv:hep-ph/9910221](https://arxiv.org/abs/hep-ph/9910221).
- [22] *Image of the cern accelerator complex from <https://cds.cern.ch/record/2225847>*, Accessed in 2017.
- [23] *Image of the lhcb detector from <http://cds.cern.ch/record/1087860>*, Accessed in 2017.
- [24] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, [arXiv:0710.3820](https://arxiv.org/abs/0710.3820).
- [25] LHCb collaboration, *Image of the acceptance of using simulation taken from https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_productionangles.html*, Accessed in 2017.
- [26] G. Raven, *LHCb: Status and prospects*, *Nucl. Phys. Proc. Suppl.* **163** (2007) 153, [,153(2007)].
- [27] LHCb collaboration, R. Aaij *et al.*, *Measurement of $\sigma(pp \rightarrow b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region*, *Phys. Lett.* **B694** (2010) 209, [arXiv:1009.2731](https://arxiv.org/abs/1009.2731).
- [28] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b -quark production cross-section in 7 and 13 TeV pp collisions*, *Phys. Rev. Lett.* **118** (2017) 052002, Erratum *ibid.* **119** (2017) 169901, [arXiv:1612.05140](https://arxiv.org/abs/1612.05140).
- [29] *Image of the luminosity overview from <https://lbgroups.cern.ch/online/operationsplots>*, Accessed in 2017.
- [30] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, *Int. J. Mod. Phys.* **A30** (2015) 1530022, [arXiv:1412.6352](https://arxiv.org/abs/1412.6352).
- [31] The LHCb collaboration, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

- [32] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, **JINST** **9** (2014) 09007, [arXiv:1405.7808](#).
- [33] *Image of the Tracker System taken from the LHCb public web page*, Accessed in 2017.
- [34] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, **JINST** **10** (2015) P02007, [arXiv:1408.1251](#).
- [35] R. Hierk, M. Merk, M. Needham, and R. Van der Eijk, *Performance of the LHCb 00 track fitting software*, Tech. Rep. LHCb-2000-086, CERN, Geneva, Aug, 2000.
- [36] LHCb, R. Forty, *RICH pattern recognition for LHCb*, **Nucl. Instrum. Meth.** **A433** (1999) 257.
- [37] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, **Eur. Phys. J.** **C73** (2013) 2431, [arXiv:1211.6759](#).
- [38] LHCb collaboration, R. Aaij *et al.*, *Measurement of b-hadron branching fractions for two-body decays into charmless charged hadrons*, **JHEP** **10** (2012) 037, [arXiv:1206.2794](#).
- [39] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, **JINST** **8** (2013) P02022, [arXiv:1211.1346](#).
- [40] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, **JINST** **8** (2013) P10020, [arXiv:1306.0249](#).
- [41] LHCb HLT project, J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, **J. Phys. Conf. Ser.** **513** (2014) 012001, [arXiv:1310.8544](#).
- [42] *Images of different trigger schemes taken from the LHCb speaker's bureau web page*, Accessed in 2018.

- [43] LHCb, I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, *J. Phys. Conf. Ser.* **331** (2011) 032047.
- [44] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [arXiv:hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175).
- [45] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152.
- [46] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, *Eur. Phys. J.* **C45** (2006) 97, [arXiv:hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026).
- [47] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, *IEEE Trans. Nucl. Sci.* **53** (2006) 270.
- [48] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: a simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250.
- [49] LHCb, M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, *J. Phys. Conf. Ser.* **331** (2011) 032023.
- [50] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.
- [51] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, *Nucl. Instrum. Meth.* **A764** (2014) 150, [arXiv:1312.5000](https://arxiv.org/abs/1312.5000).
- [52] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, PhD thesis, Cracow, INP, 1986.
- [53] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, *Nucl. Instrum. Meth.* **A555** (2005) 356, [arXiv:physics/0402083](https://arxiv.org/abs/physics/0402083).
- [54] U. Egede, P. Owen, and W. Sutcliffe, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, .

- [55] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, [arXiv:1211.3055](https://arxiv.org/abs/1211.3055).
- [56] P. Speckmayer, A. Hocker, J. Stelzer, and H. Voss, *The toolkit for multivariate data analysis*, TMVA 4, J. Phys. Conf. Ser. **219** (2010) 032057.
- [57] A. Blum *et al.*, *Beating the hold-out: Bounds for k-fold and progressive cross-validation*, in *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, , COLT 1999 (New York USA) pp. 203-208.
- [58] LHCb, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, Nature Phys. **11** (2015) 743, [arXiv:1504.01568](https://arxiv.org/abs/1504.01568).
- [59] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, [arXiv:physics/0308063](https://arxiv.org/abs/physics/0308063).
- [60] LHCb, R. Aaij *et al.*, *First observation of the decay $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ in the ρ^0 - ω region of the dimuon mass spectrum*, Phys. Lett. **B757** (2016) 558, [arXiv:1510.08367](https://arxiv.org/abs/1510.08367).
- [61] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. **G28** (2002) 2693, [,11(2002)].

Appendices

Appendix A

Boosted Decision Trees

Many rare decay analyses make extensive use of BDTs and they are important in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis. Firstly, the concept of a decision tree is introduced followed by a brief explanation of boosted decision trees.

A decision tree, in the context of data mining, is a supervised machine learning method which allows for the prediction of the value of a target variable based on several input variables. In particle physics, the purpose of the decision tree is to classify an event as being either signal or background, based on the event's input variables. The input variables, $\{x_i\}$, are various physics parameters. Each cut point in the tree is referred to as a node and the final nodes are referred to as leaves. A very simple example is shown in [Figure A.1](#). The purity, P , of a leaf refers to the fraction of the weight of a leaf due to signal events, e.g. if a leaf had 20 signal events and 15 background events it would have a purity of 0.75. If a leaf has a purity larger than 0.5 it is deemed to correspond to signal and if lower, to background.

A decision tree is constructed by a process called training. For this, samples of known signal and background events are used. These samples could be either simulation or data. For each x_i the best dividing point is decided, that is, the cut that gives the best separation between signal and background. This optimum point is decided by using the Gini index defined as

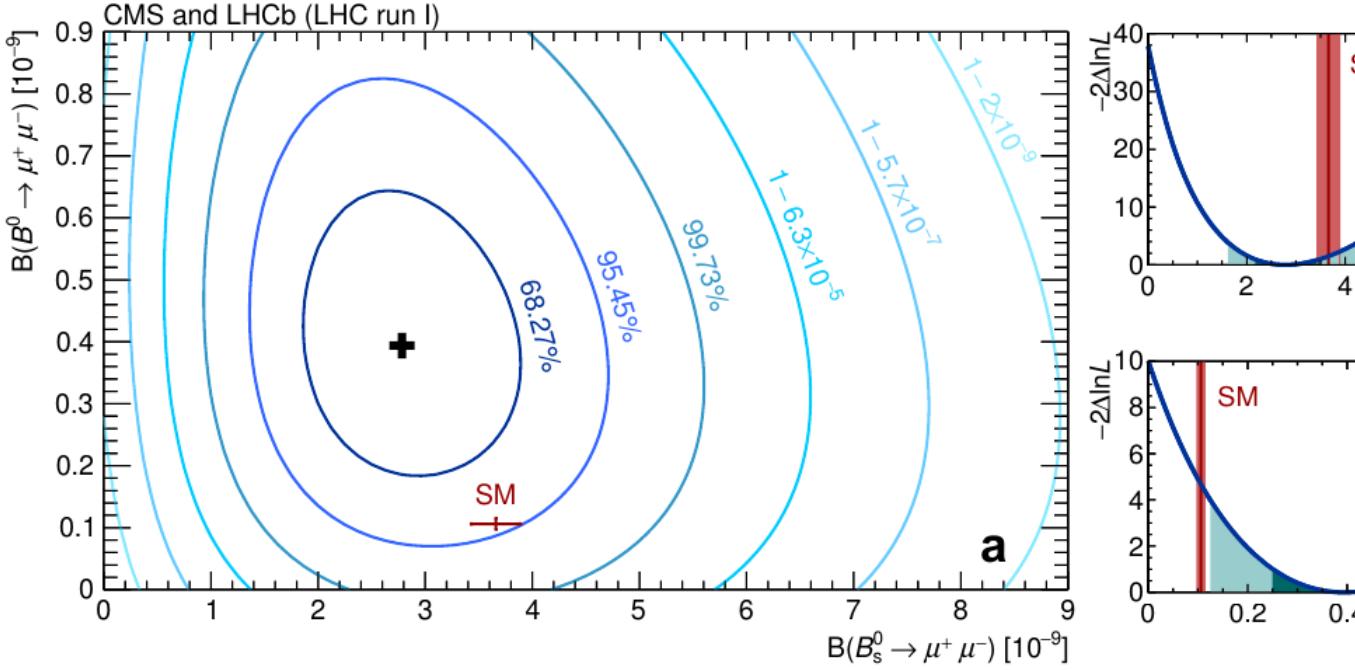


Figure A.1: An example decision tree. The S and B stand for ‘Signal-like’ and ‘Background-like’. The β_i variables refer to the cut values chosen by the machine learning algorithm after the tree has been trained on signal and background samples. The blue ovals represent final nodes called leafs, which each leaf having an associated purity, i.e. the fraction of the weight of a leaf due to signal events.

$$Gini = \sum_{i=1}^n W_i P(1 - P), \quad (\text{A.1})$$

where W_i is weight of the i^{th} event, which would generally be unity for the case of a non-boosted decision tree. The cutting point is then found by maximising the separation, Δ , between the Gini index of the parent node and the combined Gini index of the child nodes, as given in [Equation A.2](#)

$$\Delta = Gini_{parent} - Gini_{child_1} - Gini_{child_2}. \quad (\text{A.2})$$

The depth of a tree (the maximum number of cuts or nodes) is normally a number

specified before the training begins.

Boosting a decision tree involves training many trees ($O \sim 1000$) and giving misclassified events a higher weight. A misclassified event is defined as a known signal event being placed on a background leaf and vice versa. By giving the events which are difficult to classify more weight, the next tree to be trained will effectively have to work harder in order to classify events correctly.

The total score on an event is deduced by following an event through from tree to tree and, for the algorithms used in this thesis, is simply given by the weighted sum of the scores over the individual trees.

Data sets are split into two (or more) sub samples, where one half is used for training the tree and the other is used for testing the tree, and the distributions of the event scores (the BDT output) for training and testing samples are compared for signal and background. Cases where the training sample performs better than the testing sample are referred to as over-trained trees, which is often due to the BDT becoming sensitive to the statistical fluctuations of the training sample.

The distribution of events scores for a given dataset can then be cut on in order to increase the fraction of signal events.

Appendix B

The *sPlot* technique

The *sPlot* technique is used extensively throughout this thesis. It is used in cases when there is a merged dataset which consists of data from different sources of data species, namely background and signal. These datasets are assumed to have two different sets of variables associated with the events they contain. Discriminating variables are those whose distributions are known for background and signal. Control variables are those whose distributions are unknown, or are assumed to be unknown.

The *sPlot* technique allows the distribution of the control variables for each data species to be deduced by using the species discriminating variable. This method relies on the assumption that there is no correlation between the discriminating variable and the control variable. The discriminating variable used in this thesis is always the mass distribution. The full mathematical description of the *sPlot* technique can be found in Ref [24] , the key points are outlined here.

An unbinned extended maximum likelihood analysis of a data sample of several species is considered. The log-likelihood is expressed as

$$\mathcal{L} = \sum_{e=1}^N \left\{ \ln \sum_{i=1}^{N_s} N_i f_i(y_e) \right\} - \sum_{i=1}^{N_s} N_i, \quad (\text{B.1})$$

where N is the total number of events considered, N_s is the number of species of event (i.e. two - background and signal), N_i is the average number of expected events for

the i^{th} species, y represents the set of discriminating variables, $f_i(y_e)$ is the value of the Probability Density Function (PDF) of y for event e for the i^{th} species and the control variable, x , does not appear in the expression of \mathcal{L} by definition.

For the simple (and not particularly practical) case of the control variable x being a function of y , i.e. completely correlated, one could naively assume that the probability of a given event of the discriminating variable y being of the species n would be given by

$$\mathcal{P}_n(y_e) = \frac{N_n f_n(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}. \quad (\text{B.2})$$

The distribution for a control variable x for the n^{th} species, $M_n(x)$, can be deduced by histogramming in x and applying $\mathcal{P}_n(y_e)$ as a weight to event e . In this scenario the probability, $\mathcal{P}_n(y_e)$, would run from 0 to 1.

In the case considered in this thesis, where x is entirely uncorrelated with y , it can be shown that $\mathcal{P}_n(y_e)$ can be written as

$$\mathcal{P}_n(y_e) = \frac{\sum_{j=1}^{N_s} V_{nj} f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}, \quad (\text{B.3})$$

where V_{nj} is the covariance matrix between the species n and the j^{th} species. The inverse of this covariance matrix is given by the second derivative of $-\mathcal{L}$ in [Equation B.1](#).

The quantity in [Equation B.3](#) is denoted as the sWeight. In this thesis the species, n , in [Equation B.3](#) is always the signal. Because of the presence of the covariant derivative the sWeight of an event can be both positive and negative. The more negative an event is, the more likely it is to be background and vice versa for positive sWeights. The signal distribution for the control variable x , $M_s(x)$, can again be deduced by histogramming events in x , applying the sWeight to each event.

Appendix C

Fitting functions

C.1 Crystal Ball function

Crystal Ball (**CB**) function [52] is usually used for fitting of signal mass peaks in the invariant mass distributions. The **CB** function consists of Gaussian function (which usually describes mass peak) with a power-law tail below a certain threshold. Its PDF is defined as

$$f(x; \alpha, n, \bar{\mu}, \sigma) = N \cdot \begin{cases} e^{-\frac{(x-\bar{\mu})^2}{2\sigma^2}}, & \text{if } \frac{(x-\bar{\mu})}{\sigma} > \alpha \\ A \cdot \left(B - \frac{(x-\bar{\mu})}{\sigma}\right)^{-n}, & \text{otherwise} \end{cases} \quad (\text{C.1})$$

where A, B and N are all constants that depend on $\alpha, n, \bar{\mu}, \sigma$ ensuring correct normalisation and continuity of the first derivative. Thus, if α is positive, the tail, $A \cdot \left(B - \frac{(x-\bar{\mu})}{2\sigma}\right)^{-n}$, will start below the mean, usually arising from the photon-radiating decay products (left tail) and vice versa for the case where α is negative, arising from non-Gaussian resolution effects (right tail).

If one has to deal with different per-event uncertainties on the mass, one way is to model this by a sum of two Crystal Ball functions, where then each uncertainty on the event, would correspond to sum of two delta functions. Hence, double-sided Crystall Ball is defined as a linear combination of $f(x; \alpha, n, \bar{\mu}, \sigma)$:

$$g(x; \alpha, n, \bar{\mu}, \sigma, f_{cb}) = f_{cb} \cdot f(x; \alpha, n, \bar{\mu}, \sigma) + (1 - f_{cb}) \cdot f(x; \alpha, n, \bar{\mu}, \sigma). \quad (\text{C.2})$$

C.2 Double-sided Ipatia function

Generalisation of (double-sided) Crystal Ball function where per-event uncertainty is taken into account, known as (double-sided) Ipatia function, [51]. Hence it has the same number of parameters and is usually denoted as $I(m, \mu_{IP}, \sigma_{IP}, \lambda, \zeta, \beta, a_1, n_1, a_2, n_2)$.

C.3 Rookeys function from R00FIT package

A non-parametric function that is composed of superposition of Gaussians with equal surface, but with different widths σ , which are established by data at a given point.