Search for the rare fully leptonic decay $B^+ \to \mu^+\mu^-\mu^+\nu$ 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

Abstract

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

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

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

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

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

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

long-lived particles.

Contents

Declaration of originality **5**

Copyright declaration **6**

Acknowledgements **7**

List of Figures **9**

List of Tables **11**

List of abbreviations and definitions **12**

1 Introduction **14**

2 The LHCb detector **15**

2.1 LHCb Layout 16

2.2 VErtex LOcator 19

2.3 Tracking System 21

2.3.1 Tracking Algorithms 21

2.4 Ring Imaging Detectors 23

2.5 RICH Reconstruction and Performance 25

2.5.1 RICH performance 25

Bibliography **27**

Appendices 30

A Boosted Decision Trees 30

B The *sPlot* technique 33

Declaration of originality

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

This thesis has not been submitted for any other qualification.

Eluned Smith
September 2016

Copyright Declaration

Acknowledgements

Firstly, I would like to thank my supervisor Ulrik Egede. I would like to thank you for all the support, time and article corrections that you have invested in me. You have given me the opportunity to grow as physicist as well as you taught me to appreciate solving problems differently. You have given me the freedom to follow many different activities such as conferences, hardware, as well as detector maintance all leading to meeting many great people that I am happy to call my friends.

To you Patrick Owen, I owe a big thank you for showing me all the support throughout my entire analysis. Without your great insight and motivation that you have kept even when you changed institute, it would have been completely different journey.

I would also like to express special thanks to Mike McCainn, whose patience knows no limits, but apparently LHC's is 15 minutes. I learnt how handle many power-cycling buttons as well as croquet. Here I would also like to extend my gratitute to many of the RICH collaborators such as Antonis, Silvia, Roberta whom I have met while working with the upgrade.

And ofcourse all the Imperial crew: Dave, Andrei, Mitesh, Paula, Eluned, Will, Fede, Sophie, Matt, Felix and Malte, who all have their part in this work.

None of this would be possible without the STFC, the finacial body on behalf of British government, who financially supported my PhD and the LHCb collaboration as a whole.

Lastly but most importantly, I would like to thank my family. I have never heard from you that something cannot be done. You have taught me to be curious, to be independent, to be strong. We have sacrified the most precious commodity, time spent

together, in order for me to follow my interests. And below is just one example it was all worth it.

List of Figures

2.1 Accelerator complex at CERN. The image is taken from [1]. 16

2.2 Angular production and acceptance of LHCb's $b\bar{b}$ pair (in red) as well as General Purpose Detector (in yellow). LHCb covers region with highest production cross-section at 8 TeV. These plots were produced using PYTHIA8 [2] simulation. This plot was taken from [3]. 17

2.3 The 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 $z-axis$. The figure was taken from [6]. 18

2.4 Integrated luminosity collected in different years of data-taking. This plot is taken from [7] (left). Development of the instantaneous luminosity for ATLAS, CMS and LHCb during LHC fill 2651. After ramping to the desired value of $4 \times 10^{32} cm^{-2} s^{-1}$ 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 [8] (right). 19

2.5 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 [9]. 20

2.6 Visualisation of use of different technology with silicon technology in violet and straw-tube technology in cyan. The Figure was obtained in [10](left). Track types visualisation depending on which track stations provided hits. For the study of $B^+ \to \mu^+\mu^-\mu^+\nu$ decays only, Long track, are considered as muons will travel to the end of the detector leaving the hits all along. Figure is taken from [11] (right). 22
2.7 Momentum resolution of long tracks measured using "tag and probe" method at LHCb. The decay channel $J/\psi^+ \to \mu^+\mu^-$ is analyzed. 23
2.8 Separation power for different species of particles in momentum-Cerenkov angle plane (left). Schematic diagram of RICH1 layout (right). Boths figures are taken from [9]. 24
2.9 Invariant mass distribution of D^0 data sample (in black) overlayed with fit to both background and signal (in blue) (left). An example of kaon ID (red) and MisID (black) efficiency as a funtion of momentum under two PID hypotheses, $DLL_K > 0$ (empty) and $DLL_K > 5$ (filled) (right). Boths figures are taken from [13]. 26
2.10 $\pi^+\pi^-$ invariant mass distributions obtained using kinematic constraints only (left) and also using PID constraints (right) in order to isolate $B^0 \to \pi^+\pi^-$ peak. This figure is taken from [14]. 26

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. 31

List of Tables

List of abbreviations and definitions

ALICE A Large Ion Collider Experiment.

ATLAS A Toroidal LHC ApparatuS.

CMS Compact Muon Solenoid.

HPD Photomultiplier tubes that collect Cerenkov 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.

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.

misID Probability of incorrectly identifying particle given PID requirement.

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

PID Particle IDentification.

PV Primary Vertex, the pp interaction vertex.

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

SM Standard Model.

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..

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.

Chapter 1

Introduction

The Standard Model (SM) is an effective theory which describes fundamental particles and their interactions to an impressive precision.

bla

Chapter 2

The LHCb detector

In this section, overview of accelator complex at CERN as well as physics motivation behind LHCb detector and its details will be described.

CERN built one of the most exciting laboratories to study elementary particle interactions. The complex set of particle accelerators and detectors is shown in Figure 2.1. The proccess of accelerating protons starts with the source of protons. Protons are obtained from hydrogen gas bottle by applying and an electric field seperates hydrogen into positively and negatively charged constituents. The first proton accelerator in the chain, Linac 2, accelerates the protons to the energy of 50 MeV. Is is a tank composed of several chambers where the resonant cavity is tuned to a specific frequency which creates potential differences in the cavities making accelerate the protons. These are then injected into the Proton Synchrotron Booster (PSB). Here the protons are accelerated to 1.4 GeV. The next line is the Proton Synchrotron (PS) reaching energy of 25 GeV. Before either entering the LHC or North Area (mainly used as testing facility for experiment upgrades) Super Proton Synchrotron (SPS) is the last stop. Here proton acceleration to 450 GeV is achieved.

The Large Hadron Collider (LHC) is a complex machine which accelerates beams of protons in opposite directions in $\sim$ 27km 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

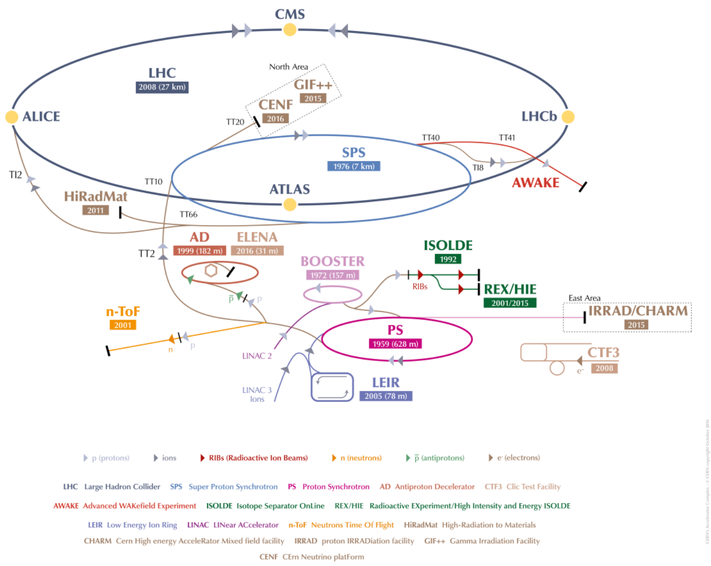

Figure 2.1: Accelerator complex at CERN. The image is taken from [1].

detectors with different physics focus are located. These are ATLAS, CMS, ALICE and LHCb. Study of $B^+ \to \mu^+\mu^-\mu^+\nu$ was performed using data obtained at LHCb.

2.1 LHCb Layout

LHCb differs from the other general purpose detectors on the LHC ring as its studies properties of heavy particles containing b or c quarks. This can be attributed to the geometrical acceptance and unique vertex resolution as well as excelent PID.

Contrary to the two general purpose detectors where the collisions are occuring 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 means that information about products outside of its scope are not known, meaning that there is no overall constraint on collision information, unlike in other flavour experiments. This is compensated by production mechanism of $b\bar{b}$ and $c\bar{c}$ in pp interactions, which occurs via gluon-gluon fusion. In this process, each gluon will carry part of proton's momentum.

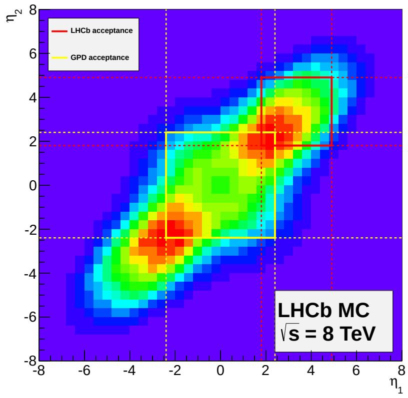

Figure 2.2: Angular production and acceptance of LHCb's $b\bar{b}$ pair (in red) as well as General Purpose Detector (in yellow). LHCb covers region with highest production cross-section at 8 TeV. These plots were produced using PYTHIA8 [2] simulation. This plot was taken from [3].

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 2.2.

The angular coverage of LHCb is formally defined using pseudorapidity η,

$$\eta = -\ln(\tan\frac{\theta}{2}) \tag{2.1}$$

where θ is defined in Figure 2.3. LHCb detector, hence, covers the region $2 < \eta < 5$. The production cross-section of the fundamental process of $pp \to b\bar{b}X$ was measured in this region yielding, $\sigma(pp \to b\bar{b}X)$= 75.3±5.4±13.0 μb at 7 TeV [4] and 144±1±21 μb at 13 TeV [5], which shows that the production cross-sections scales roughly linearly with the centre-of-mass energy.

To limit the background coming from especially soft QCD processes (due to hadronic collision enviroment) global event cuts, *GEC*s, are put in the place. To limit the occupancy of the detector only events with 600 (in 7,8 TeV) and 450 (in 13 TeV) tracks and less are allowed to be processed. In order to achieve these occupancies, μ_{vis}, the

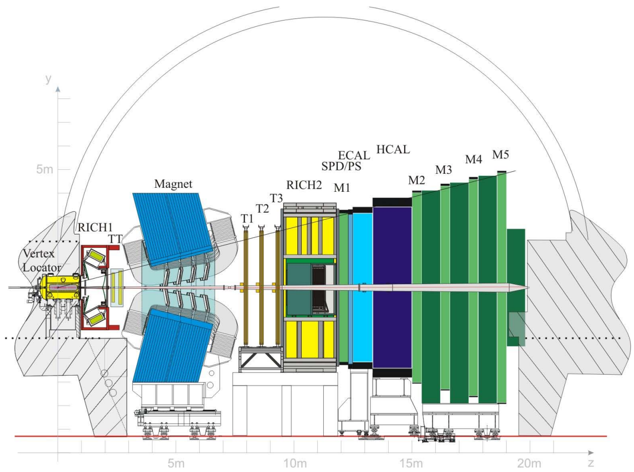

Figure 2.3: The schematic slice of LHCb detector in y,z plane where z is defined to be the direction parallel to beamline, and x,y define the plane perpendicular to the beamline. θ, the opening angle in y-z plane with $\theta = 0$ along $z-axis$. The figure was taken from [6].

average number of visible pp interactions per bunch-crossing is kept below 1, so that the pile-up, the visible number of pp interaction in the visible events, is limited. This LHCb-specific control of luminosity is achieved by *luminosity levelling*. This procedure achieves stable instantenous luminosity by controlling the transverse overlap of the beams at collision point. 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, with collected integrated luminosity shown in Figure 2.4. As compared to ATLAS and CMS the integrated luminosity is much lower, due to allowed pile-up conditions. In 2017, there were two pp collision energies at which the data was taken: at $\sqrt{(s)} = 13$ and 5 TeV. Run I data-taking (2010-2012) was paused by Long Shutdown 1 (LS1) and followed with Run II data-taking (2015-2018).

In the following sections, brief discussion of different subdetectors is presented. Both hardware and software overview will be presented with particular emphasis given to Muon Station and Simulation of LHCb.

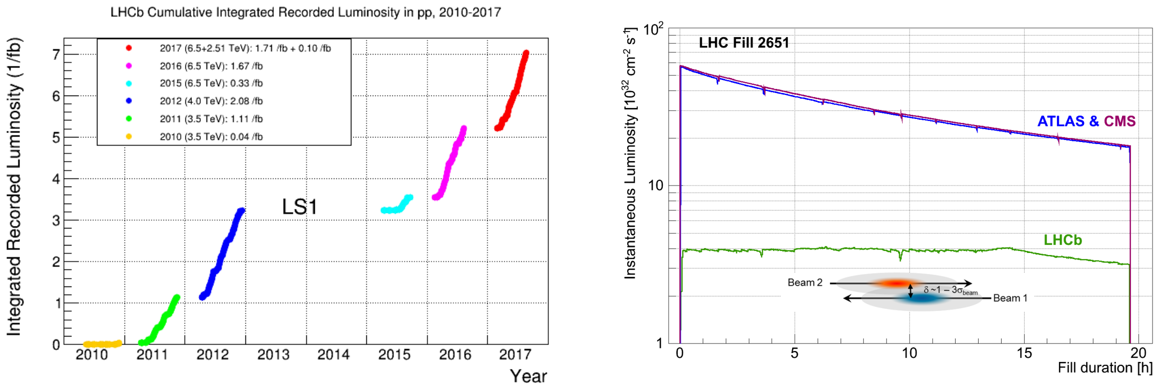

Figure 2.4: Integrated luminosity collected in different years of data-taking. This plot is taken from [7] (left). Development of the instantaneous luminosity for ATLAS, CMS and LHCb during LHC fill 2651. After ramping to the desired value of $4 \times 10^{32} cm^{-2} s^{-1}$ for LHCb, the luminosity 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 [8] (right).

2.2 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 for to distinguish events from promt 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.

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

.

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. In the first disc 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. This setup, which can be seen in Figure 2.5, brings outstanding hit resolution (4-40μm).

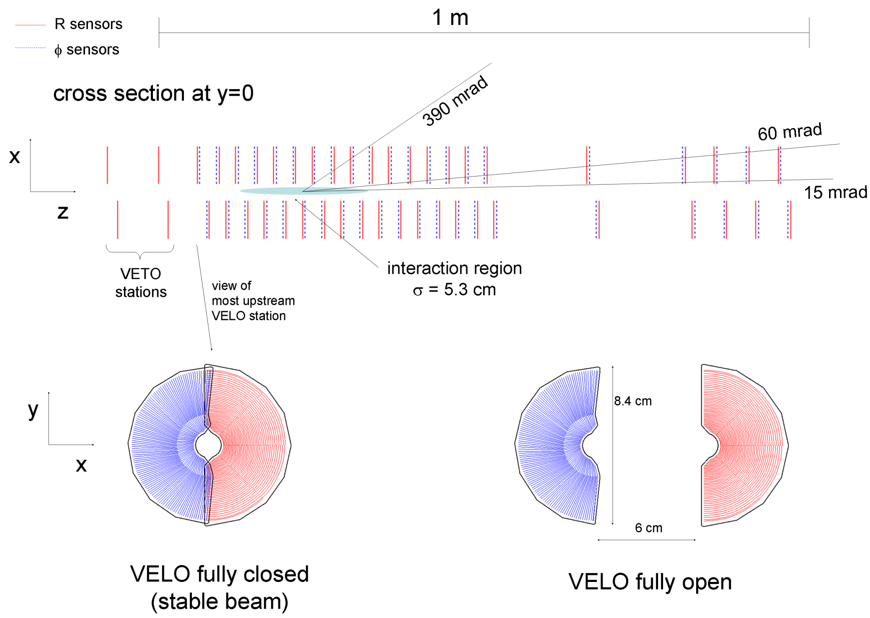

Figure 2.5: Schematic plot of VELO detector configuration along the beam pipe showing the layout as well as positions while in stable beams (discs have slight overlap) and injection. Figure taken from [9].

2.3 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 strenght to bend charged particles with p of 200 GeV/c^2.

Two different dectection 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-T3 (Inner Tracker (IT) expects higher occupancy, making use of the 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. It measures the hit position by measuring the drift-time of ionized electrons. This two technologies are seen in Figure 2.6.

2.3.1 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 2.6.

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, meaning that they 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 ouside of the VELO acceptance and hence will produce no hits until

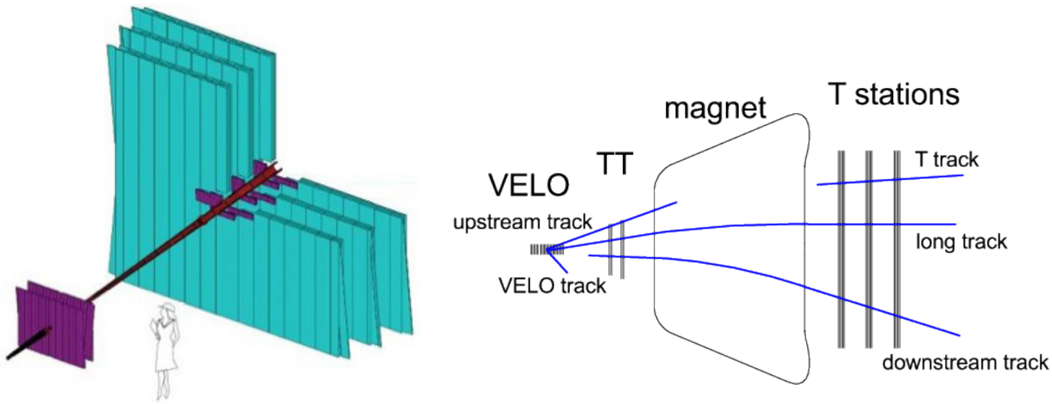

Figure 2.6: Visualisation of use of different technology with silicon technology in violet and straw-tube technology in cyan. The Figure was obtained in [10](left). Track types visualisation depending on which track stations provided hits. For the study of $B^+ \to \mu^+\mu^-\mu^+\nu$ decays only, Long track, are considered as muons will travel to the end of the detector leaving the hits all along. Figure is taken from [11] (right).

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. Mass uncertainty 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 $J/\psi \to \mu^+\mu^-$ using *tag and probe* can be seen in Figure 2.7. It varies logarithmically with increasing momentum.

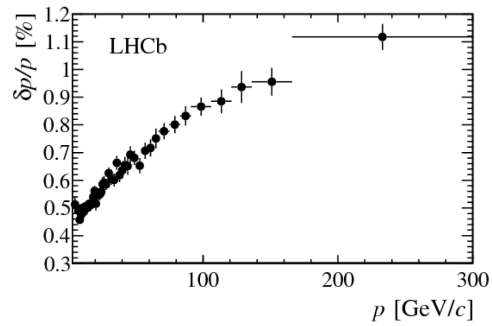

Figure 2.7: Momentum resolution of long tracks measured using "tag and probe" method at LHCb. The decay channel $J/\psi^{+} \rightarrow \mu^{+}\mu^{-}$ is analyzed.

2.4 Ring Imaging Detectors

Particle identification, PID, at LHCb relies heavily on two dedicated Ring Imaging subdetectors. 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 knowlegde of refractive index of the medium, n, and momentum p that is measured using tracking, mass m of the particle can be obtained through:

$$\cos\theta_c = \frac{\sqrt{m^2 + p^2}}{pn}. \tag{2.2}$$

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 $\cos\theta_c \rightarrow 0$, 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 ± 25-300 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 ±15-120 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 2.8.

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 schmematic view of a particle passing through RICH1 can be also be seen in Figure 2.8.

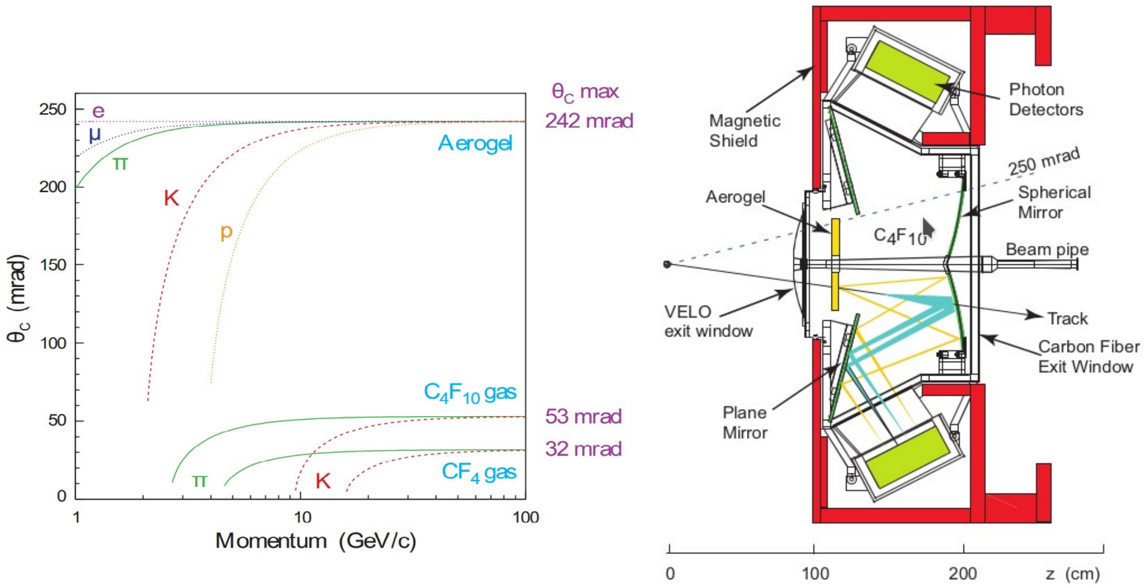

Figure 2.8: Separation power for different species of particles in momentum-Cerenkov angle plane (left). Schematic diagram of RICH1 layout (right). Boths figures are taken from [9].

2.5 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 [12].

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, DLL_x, that quantify the strenght of the chosen species hypothesis against pion hypothesis,

$$DLL_x = log(\mathcal{L})_x - log(\mathcal{L})_\pi \quad x \in e, \mu, K, proton, deuteron. \tag{2.3}$$

By calculating $DLL_{x_1} - DLL_{x_2}$, one can obtain discriminative strenght between any two species.

2.5.1 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, $D^{*+} \to D^0(K^-\pi^+)\pi^+$ backround-substracted samples are used, whereby the daughter tracks of D^0 become proxies for evaluation. The probability of correctly identifying kaon given certain constraint on DLL_K, identification efficiency (ID), and probability of mistakenly swapping pion identification, MisID efficiency, are summarized in

Figure 2.9. Identification probabilities of $\approx 85\%$ with misidentification rate of $\approx 3\%$ provide invaluable discriminating separation between kaon and pion.

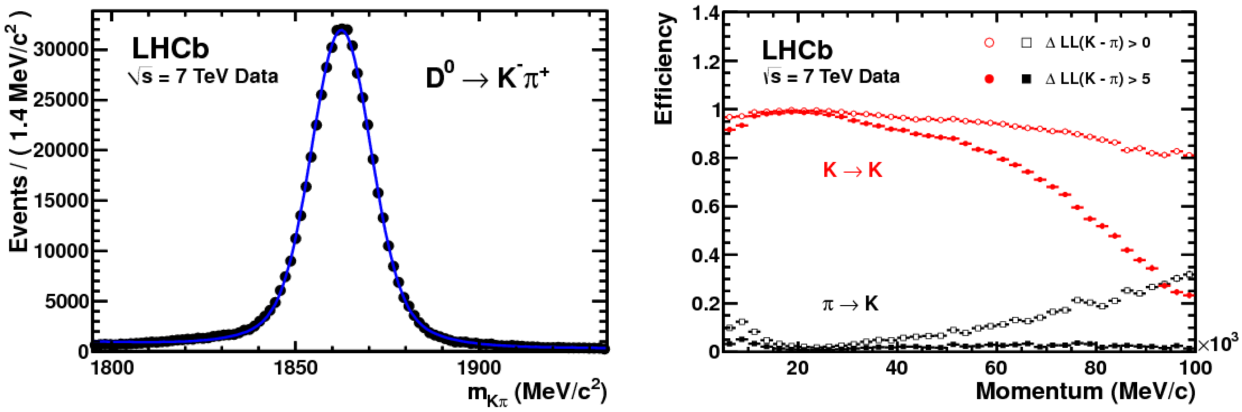

Figure 2.9: Invariant mass distribution of D^0 data sample (in black) overlayed with fit to both background and signal (in blue) (left). An example of kaon ID (red) and MisID (black) efficiency as a funtion of momentum under two PID hypotheses, $DLL_K > 0$ (empty) and $DLL_K > 5$ (filled) (right). Boths figures are taken from [13].

$\pi^+\pi^-$ invariant mass spectra in $B^0 \to h^+h^-$ with and without PID requirements, Figure 2.10, demonstrates clear increase in sensitivity searching for $B^0 \to \pi^+\pi^-$ signal.

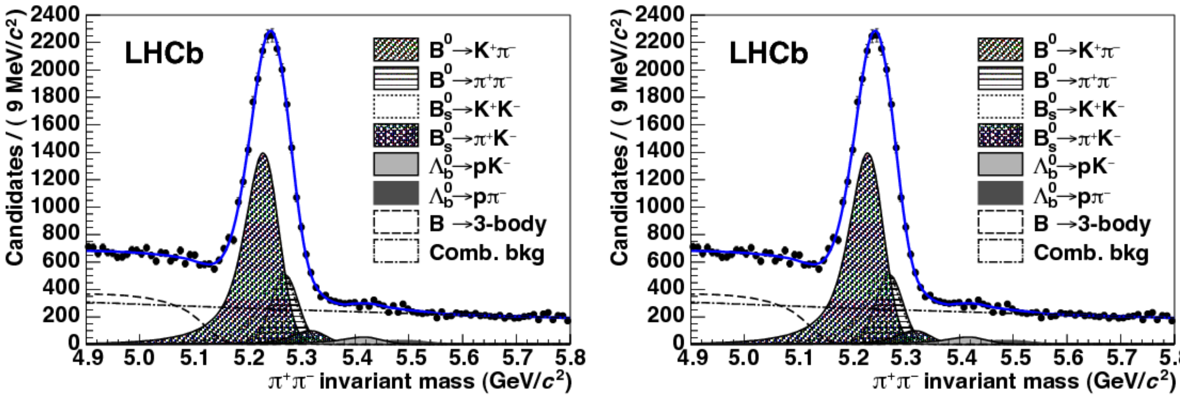

Figure 2.10: $\pi^+\pi^-$ invariant mass distributions obtained using kinematic constraints only (left) and also using PID constraints (right) in order to isolate $B^0 \to \pi^+\pi^-$ peak. This figure is taken from [14].

Bibliography

[1] *Image of the cern accelerator complex from* https://cds.cern.ch/record/2225847, Accessed in 2017.

[2] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[3] 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.

[4] LHCb collaboration, R. Aaij *et al.*, *Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region*, Phys. Lett. **B694** (2010) 209, arXiv:1009.2731.

[5] 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.

[6] *Image of the lhcb detector from* http://cds.cern.ch/record/1087860, Accessed in 2017.

[7] *Image of the luminosty overview from* https://lbgroups.cern.ch/online/operationsplots Accessed in 2017.

[8] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[9] The LHCb collaboration, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[10] *Image of the Tracker System taken from the LHCb public web page*, Accessed in 2017.

[11] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

[12] LHCb, R. Forty, *RICH pattern recognition for LHCb*, Nucl. Instrum. Meth. **A433** (1999) 257.

[13] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[14] 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.

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 particular 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

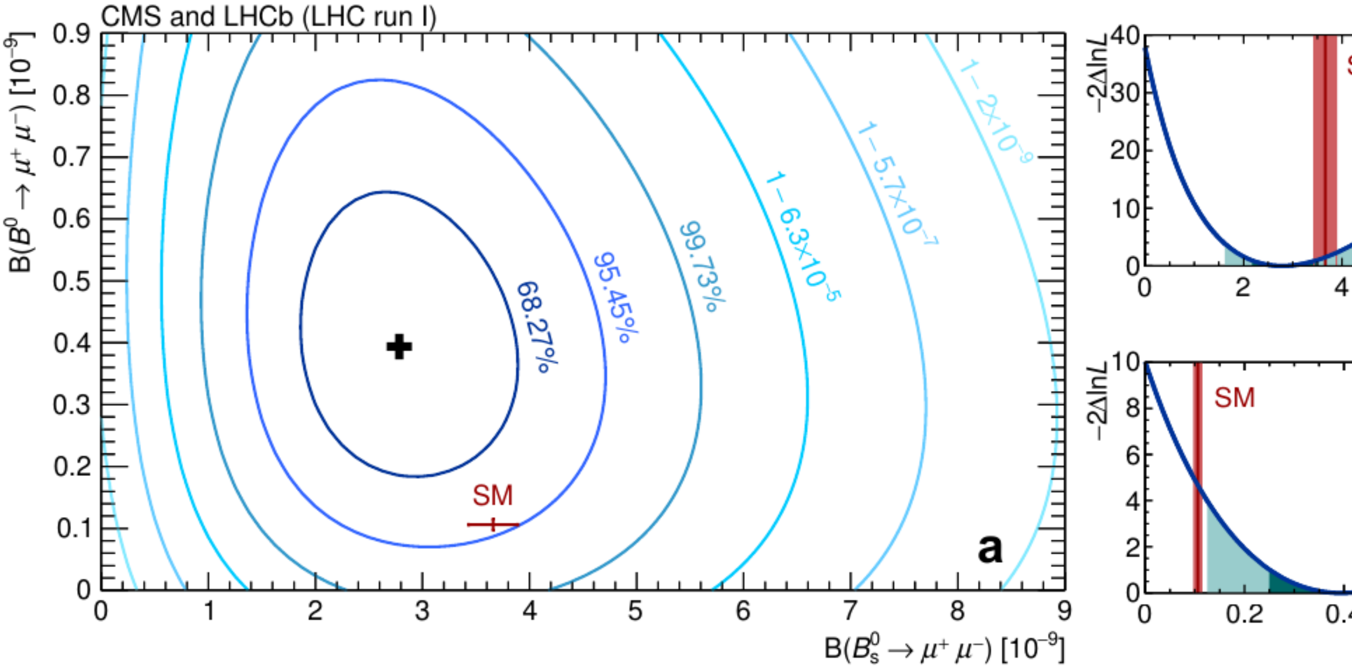

Figure A.1: An example decision tree. The S and B stand for 'Signal-like' and 'Background-like'. The β_i variables refer to the cut values chosen by the machine learning algorithm after the tree has been trained on signal and background samples. The blue ovals represent final nodes called leafs, which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

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

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

$$\Delta = Gini_{parent} - Gini_{child_1} - Gini_{child_2}. \tag{A.2}$$

The depth of a tree (the maximum number of cuts or nodes) is normally a number specified before the training begins.

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

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

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

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

Appendix B

The *sPlot* technique

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

The *sPlot* technique allows the distribution of the control variables for each data species to be deduced by using the species discriminating variable. This method relies on the assumption that there is no correlation between the discriminating variable and the control variable. The discriminating variable used in this thesis is always the mass distribution. The full mathematical description of the *sPlot* technique can be found in Ref [2] , 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

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

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

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

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

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

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 donated 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.