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

Contents

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 3–5) 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, February 26, 2018

Copyright Declaration

List of Figures

List of Tables

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 discrimating variables into a single discrimating variable.

CB Crystal Ball function.

CMS Compact Muon Solenoid.

Cos(θ_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 χ^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 ÄŇerenkov 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 IPχ^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.

QED Quantum Electrodynamics.

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

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

RICH2 Ring Imaging Cherenkov providing high momentum PID by using ÄŇerenkov 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 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 accelerator 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. Its complex set of particle accelerators and detectors is shown in Figure 2.1. The process of accelerating protons starts with the source of protons. Protons are obtained from hydrogen gas bottle by applying and an electric field separating 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. Linac 2 is a tank composed of several chambers where the resonant cavity is tuned to a specific frequency creating potential differences in the cavities, which then make the protons accelerate. These 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) Super Proton Synchrotron (SPS) is the last stop. Here proton acceleration to 450 GeV is achieved.

LHC is a complex machine which accelerates beams of protons in opposite directions in $\sim$ 27km 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 physics

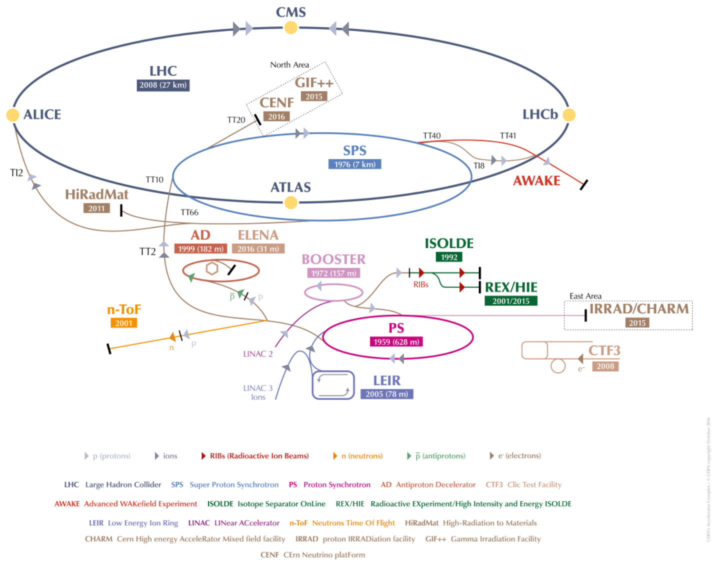


Figure 2.1: Accelerator complex at CERN. The image is taken from [1].

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 $\boxed{\text{ULRIK}}$

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 excellent particle identification (PID).

Contrary to the two general purpose detectors 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 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

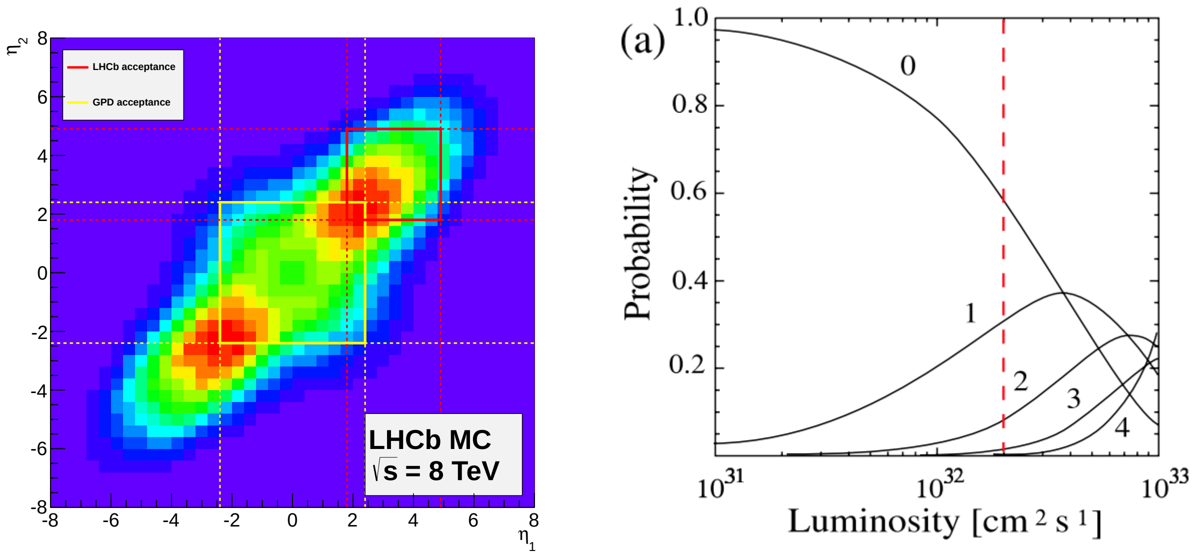


Figure 2.2: Angular production and acceptance of LHCb's $b\bar{b}$ pair (in red) as well as General Purpose Detector (in yellow). LHCb covers region with highest production cross-section at 8 TeV. These plots were produced using PYTHIA8 [2] simulation. This plot was taken from [3] (left). Probability of interaction per bunch crossing as a function of instantaneous luminosity. This figure was obtained from [4] (right).

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 2.2.

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

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

where θ is defined in Figure 2.3. LHCb detector was built to cover 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 [5] and 144±1±21 μb at 13 TeV [6], which shows that the production cross-sections scales roughly linearly with the centre-of-mass energy. Assuming design conditions of LHCb listed in Table 2.1,

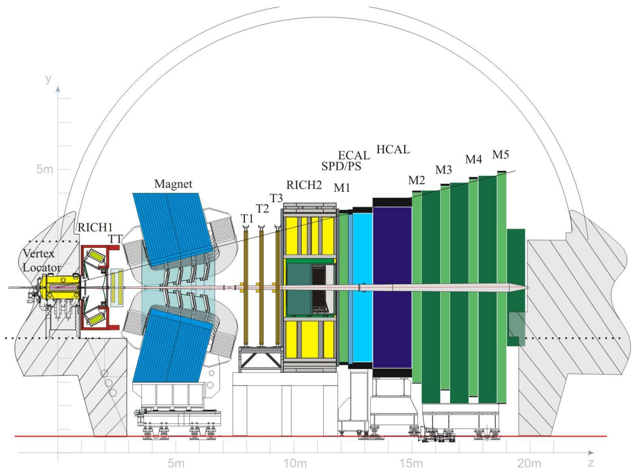


Figure 2.3: The schematic slice of LHCb detector in y,z plane where z is defined to be the direction parallel to beamline, and x,y define the plane perpendicular to the beamline. θ, the opening angle in y-z plane with $\theta = 0$ along $z-axis$. The figure was taken from [7].

$2\,\mathrm{fb}^{-1}$ of data (eqvivalent to 2012 dataset) would correspond to 10^{12} of $b\bar{b}$ pairs being produced.

Despite such impressive statistics of $b\bar{b}$ pairs available to LHCb, the bottleneck arises in 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 number of pp interaction per bunch crossing as a function of luminosity, shown in Figure 2.2, 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 is the reason for the LHCb's design luminosity. In addition, to keep the occupancy of the detector reasonable for physics analyses, global event cuts, *GEC*s, 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 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 shown in Figure 2.4. As compared to ATLAS and CMS the integrated luminosity is much lower, due to allowed luminosity 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 period (2015-2018). The summary of LHCb running conditions is provided in Table 2.1, showing the evolution of the instantaneous luminosity as well as the frequency of collisions compared to the design proposal.

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

Table 2.1: Running conditions of LHC and LHCb in different years of data-taking. The statistics of LHCb's instantaneous luminosity is extracted using run database information.

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 detectors and simulation of LHCb.

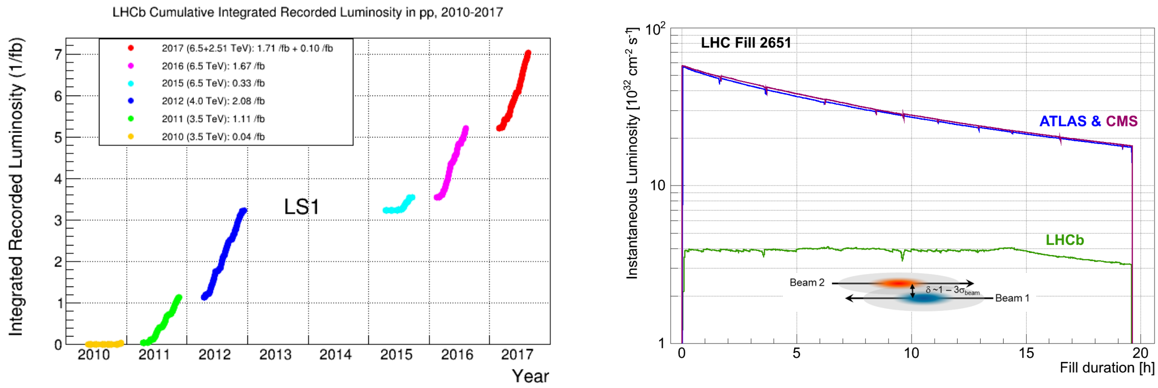


Figure 2.4: Integrated luminosity collected in different years of data-taking. This plot is taken from [8] (left). Development of the instantaneous luminosity for ATLAS, CMS and LHCb during LHC fill 2651. After ramping to the desired value of $4 \times 10^{32} cm^{-2} s^{-1}$ for LHCb, the luminosity 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 [9] (right).

2.2 VErtex LOcator ULRIK

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

.

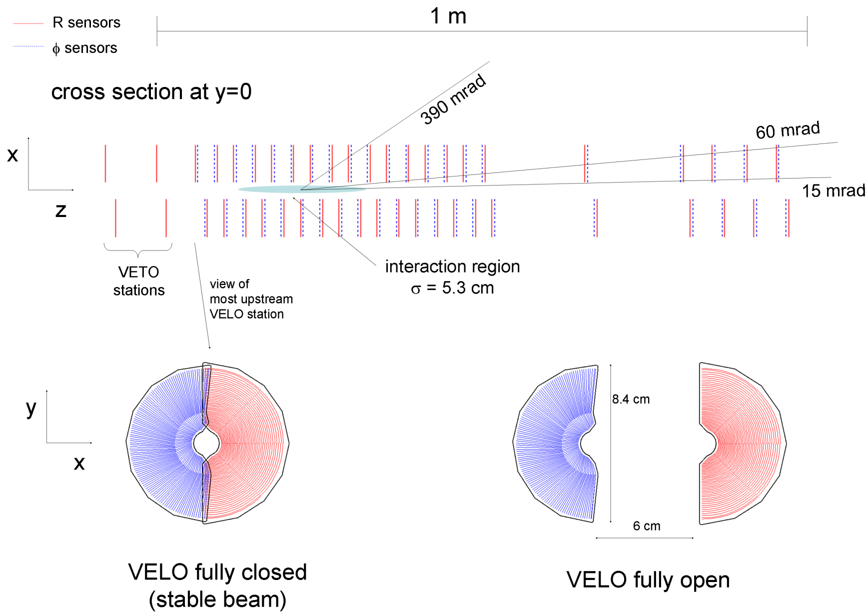


Figure 2.5: Schematic plot of VELO detector configuration along the beam pipe showing the layout as well as positions while in stable beams (discs have slight overlap) and injection. Figure taken from [10].

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 2.5. 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 (4-40 μm), which in turn allows for very high IP and very good primary vertex (PV) resolution, as seen in Figure 2.6. 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.

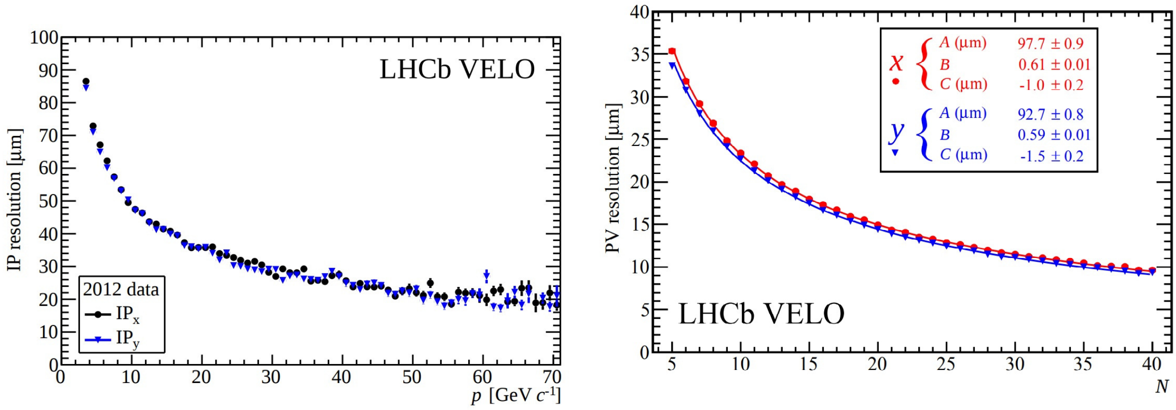


Figure 2.6: Two key variables which quantify performance of the VELO detector. IP resolution which is worse for low momentum tracks (left) and PV resolution dependent on the number of tracks forming the primary vertex N (right). Figures taken from [11].

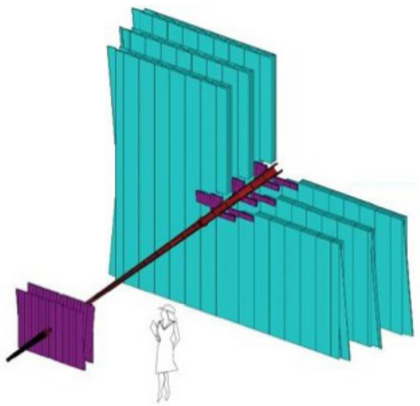
2.3 Tracking System ULRIK

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 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 2.7.

2.3.1 Tracking Algorithms $\boxed{\text{ULRIK}}$

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 2.7.



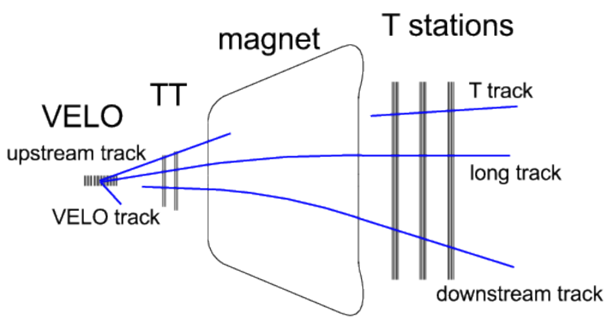


Figure 2.7: Visualisation of use of different technology with silicon technology in violet and straw-tube technology in cyan. The Figure was obtained in [12](left). Track types visualisation depending on which track stations provided hits. For the study of $B^+ \to \mu^+\mu^-\mu^+\nu$ decays only long tracks are considered as muons will travel to the end of the detector leaving the hits all along. Figure is taken from [13] (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 [14], 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 2.8. It varies logarithmically with increasing momentum.

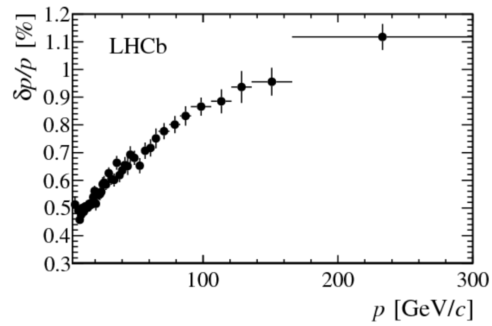


Figure 2.8: Momentum resolution of long tracks measured using *tag and probe* method at LHCb. The decay channel $J/\psi^+ \to \mu^+\mu^-$ is analysed.

2.4 Ring Imaging Detectors $\boxed{\text{ULRIK}}$

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:

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

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 $\cos\theta_c \to 1$, the particle is not producing any Cerenkov light cone. At the very high momentum, as $\cos\theta_c \to 1/n$, 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 ± 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 2.9.

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 2.9.

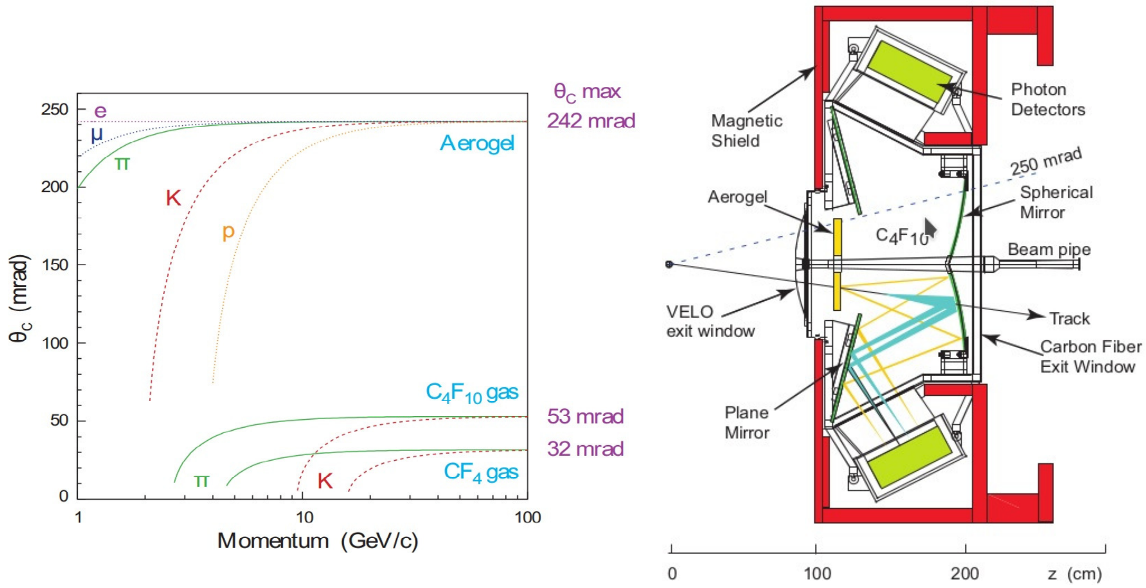


Figure 2.9: Separation power for different species of particles in momentum-Cerenkov angle plane (left). Schematic diagram of RICH1 layout (right). Both figures are taken from [10].

2.5 RICH Reconstruction and Performance $\boxed{\text{ULRIK}}$

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 [15].

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,

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

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

2.5.1 RICH performance $\boxed{\text{ULRIK}}$

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 $DLLK$, identification efficiency (ID), and probability of mistakenly swapping pion identification, MisID efficiency, are summarized in

Figure 2.10. Identification probabilities of $\approx 85\%$ with misIDentification rate of $\approx 3\%$ provide invaluable discriminating separation between kaon and pion.

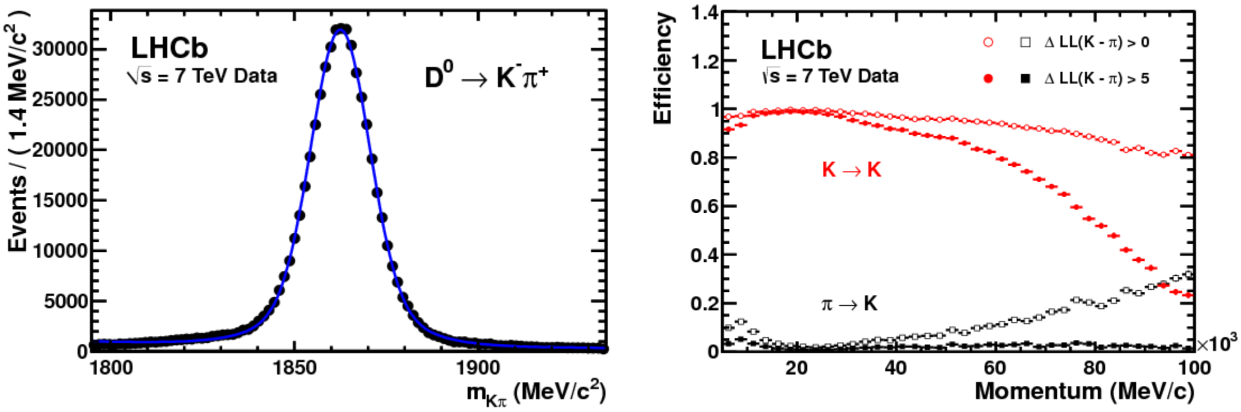


Figure 2.10: Invariant mass distribution of D^0 data sample (in black) overlaid with fit to both background and signal (in blue) (left). An example of kaon ID (red) and MisID (black) efficiency as a function of momentum under two PID hypotheses, $DLLK > 0$ (empty) and $DLLK > 5$ (filled) (right). Both figures are taken from [16].

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 2.11. These plots clearly demonstrate increase in sensitivity searching for $B^0 \rightarrow \pi^+\pi^-$ signal amongst other components.

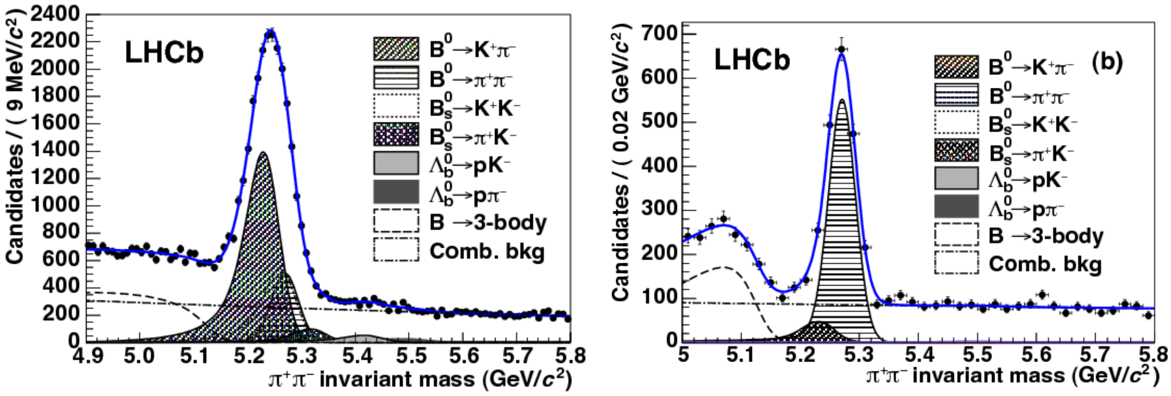


Figure 2.11: $\pi^+\pi^-$ invariant mass distributions obtained using kinematic constraints only (left) and also using PID constraints (right) in order to isolate $B^0 \to \pi^+\pi^-$ peak. This figure is taken from [17].

2.6 Calorimetry ULRIK

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 shashliktype 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 2.12.

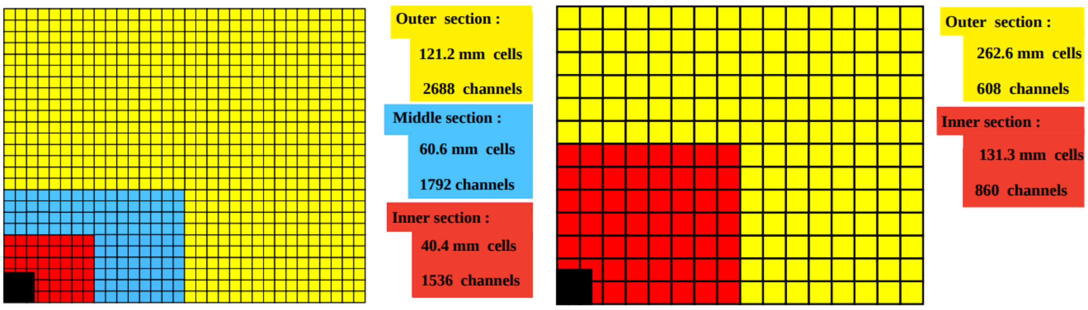


Figure 2.12: Granularity of ECAL (left) and HCAL (right) detectors. The figure was taken from [10].

2.7 Muon Stations $\boxed{\text{ULRIK}}$

Muons are considered to be of fundamental importance to many flagship analyses by LHCb, such as the search for the rare $B_s^0 \to \mu^+\mu^-$ decay . Analysis of $B^+ \to \mu^+\mu^-\mu^+\nu$ 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 calorimetry system, with first station M1 upstream of the SPD, and four stations (M2-M5) downstream of HCAL as shown in Figure 2.13. 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 $\mathrm{Ar-CO2-CF4}$ gas mixture. They are interlayed 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 ≥99% to give overall 95% trigger efficiency. Geometrical layout covers $\approx 20\%$ muons originating in semileptonic b decays.

2.7.1 Muon Identification $\boxed{\text{ULRIK}}$

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

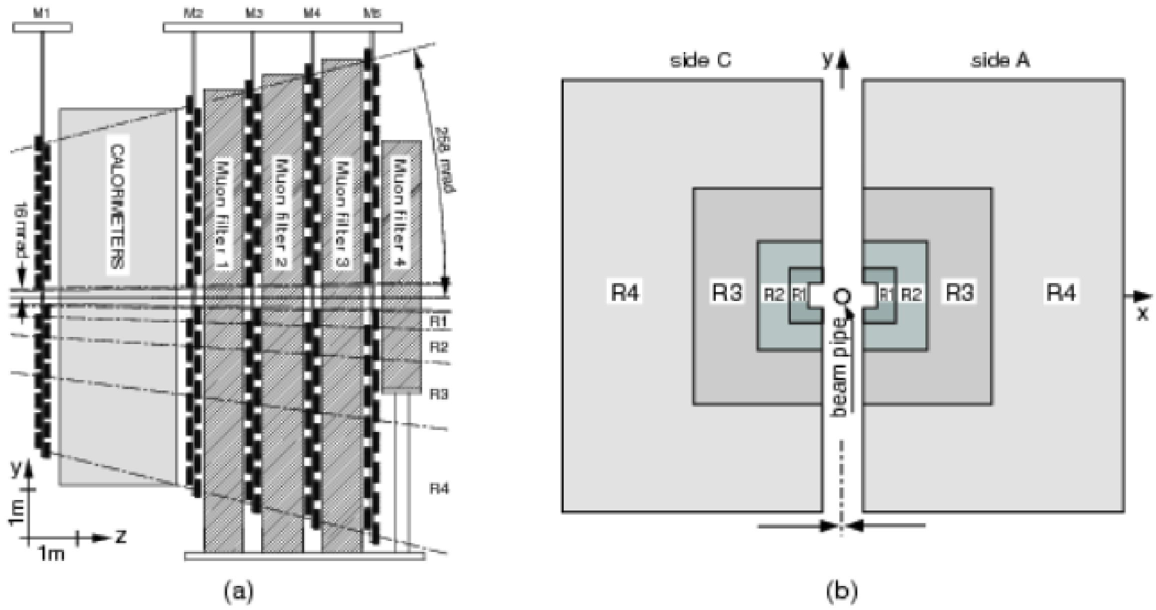


Figure 2.13: (a) Layout of the muon detector x-z plane and (b) x-y plane. This figure is taken from [18].

- **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 2.2.

- **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 GeV/c $<p<$6 GeV/c	M1 & M2
6 GeV/c $<p<$10 GeV/c	M1 & M2 & (M3 || M4)
10 GeV/c $<p$	M1, M2, M3 and M4

Table 2.2: Momentum-dependent definition `IsMuon` variable.

2.7.2 Muon Performance $\boxed{\text{ULRIK}}$

As in hadron performance measurements, muon ID is determined using high statistics decay channel $J/\psi \to \mu^+\mu^-$ 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, $D^{*+} \to D^0(K^-\pi^+)\pi^+$. The summary of `IsMuon` ID and misID rates are presented in Figure 2.14. 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.

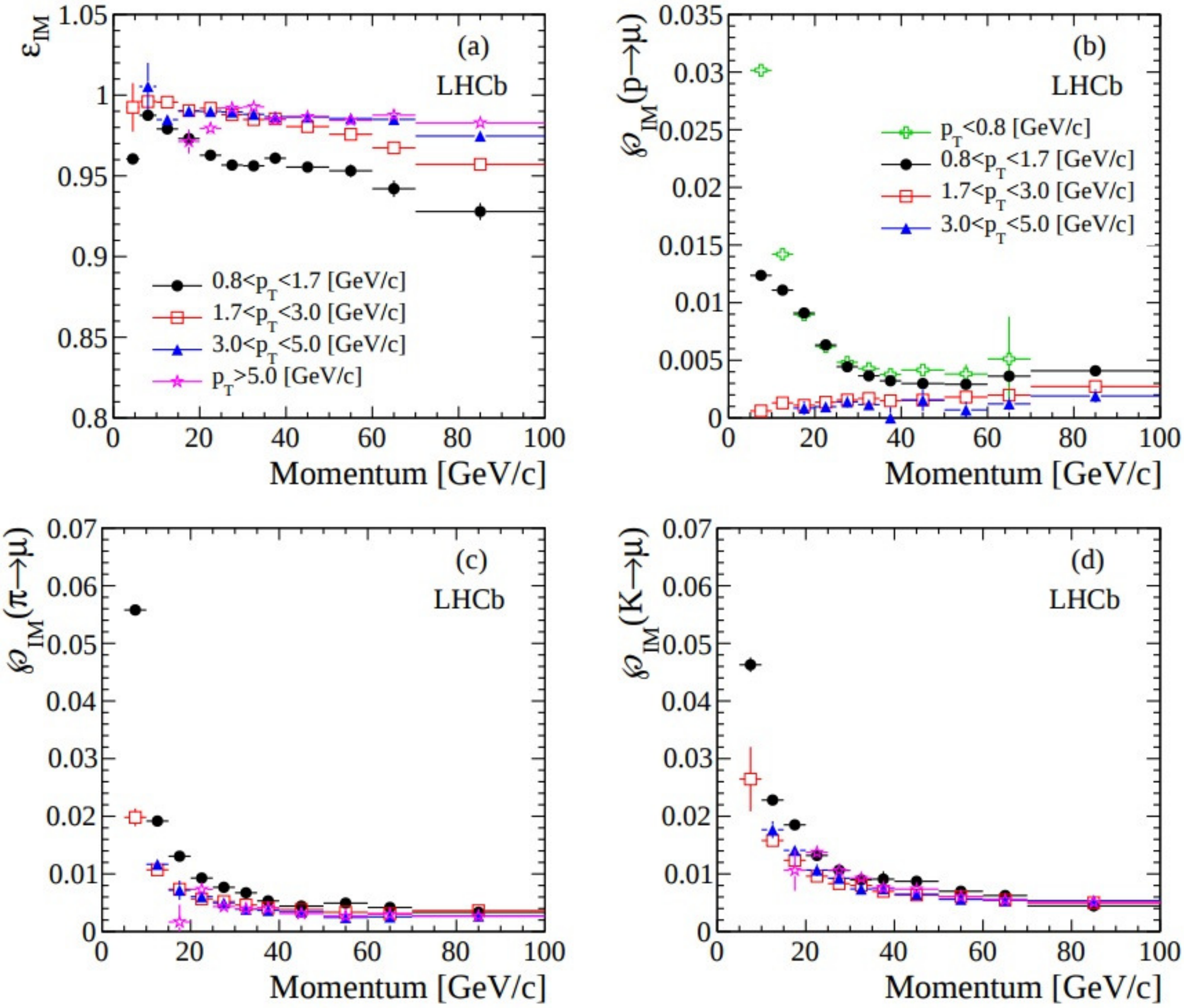


Figure 2.14: (a) Probability of correctly identifying muons as a function of momentum p in the bins of p_T for $J/\psi \to \mu^+\mu^-$ with IsMuon constraint. (c) Probability of incorrectly identifying pion (b) proton and (d) kaon as muon with IsMuon. This figure is taken from [19].

2.8 Trigger ULRIK

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 $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ decays. Hadron trigger efficiency in different decay channels can be seen in Figure 2.15.

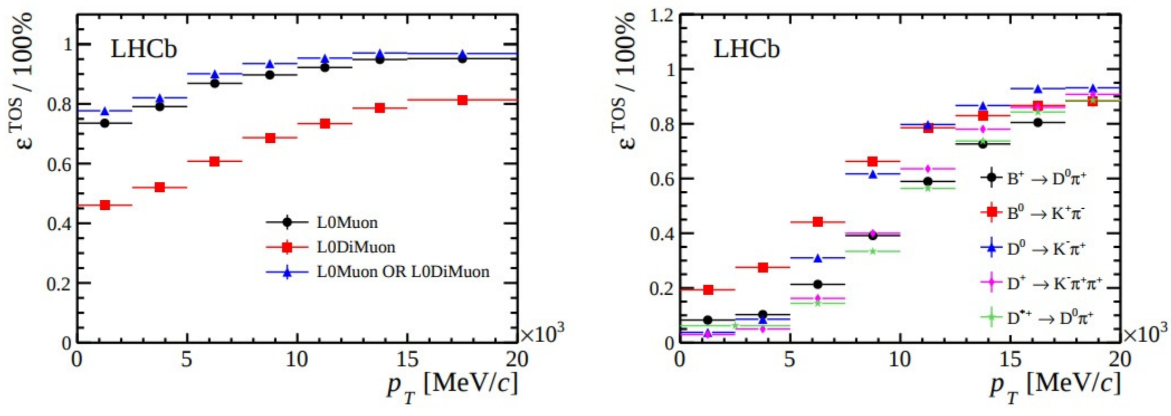


Figure 2.15: TOS efficiency as a function of p_T for muon-based decisions (left). TOS efficiency for different decays using L0 hadron trigger lines. This figure is taken from [18].

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 trakcs 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 2.16.

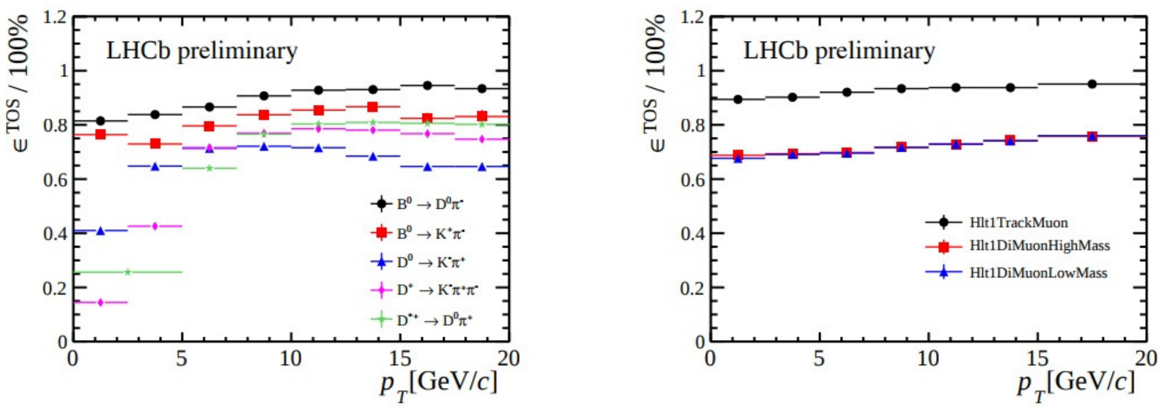


Figure 2.16: HLT1 efficiencies of the corresponding triggers using the same proxy as in Figure 2.15. This figure is taken from [20].

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 $B \to J/\psi X$ 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 2.17. 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?).

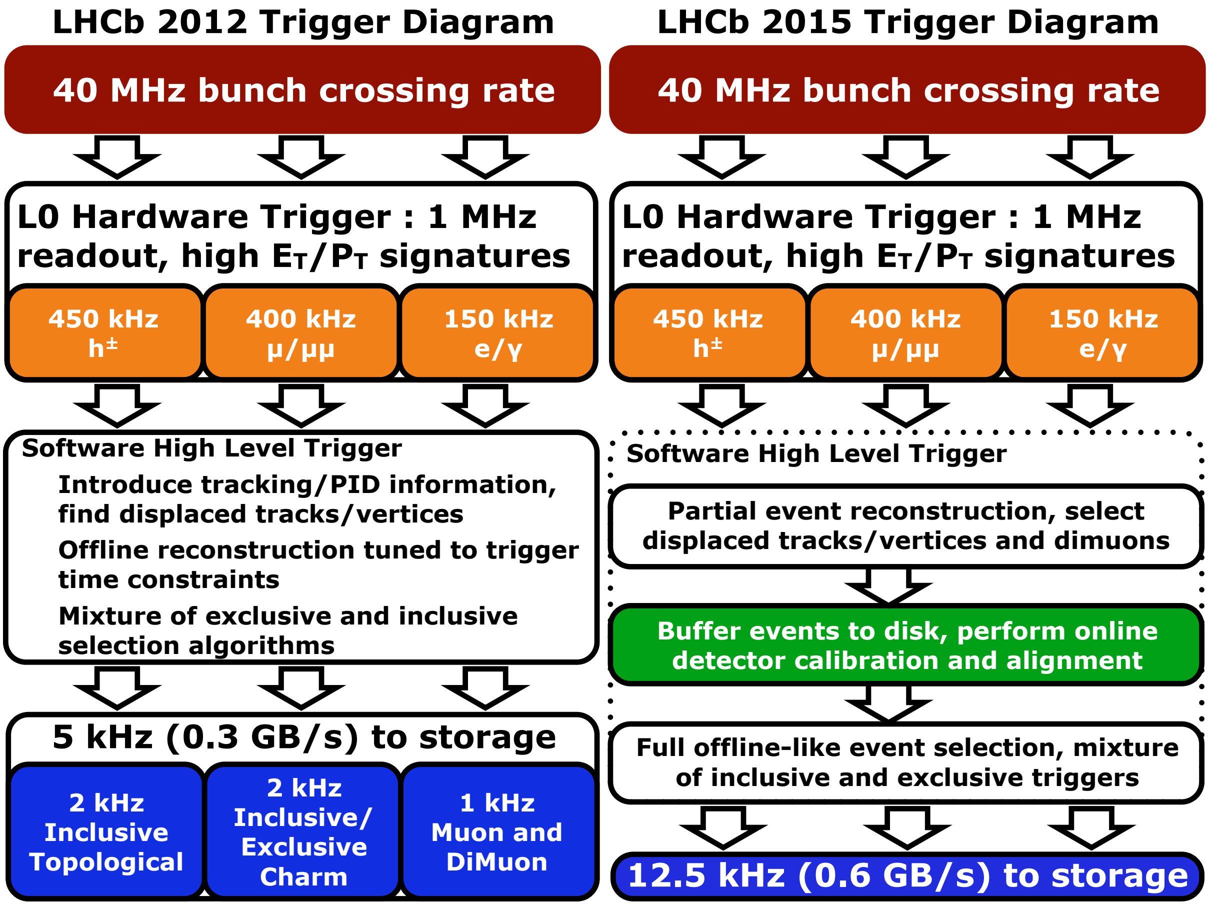


Figure 2.17: Trigger scheme differences between Run I and Run II. Figures obtained from [21]

2.9 Simulation ULRIK

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 [22] are simulated with Pythia 6.4 [23] and Pythia 8.1 [2]. 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 [24] 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 [25] algorithm can be used. All of this combined establishes *generator-level simulation* of LHCb.

In the next phase, *detector simulation*, the interactions of the all the particles with the detector, transport, as well as detector's response are simulated using the C++ GEANT4 toolkit [26], [27]. LHCb's interface to GEANT4 is detailed in [28].

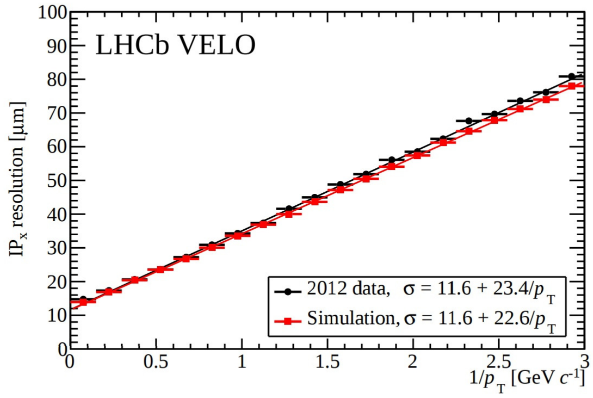
2.9.1 Differences in Simulation And Data ULRIK - not fully finished

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 then in the data due to the mismodelling of material description of VELO simulation. As shown in Figure 2.18 IP resolution does greatly differ depending 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



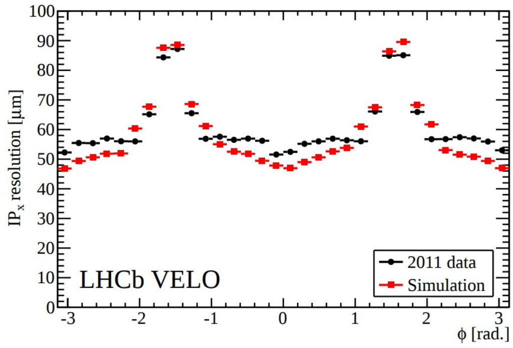


Figure 2.18: IP resolution in x-direction comparing the data and simulation output for 2012 data-taking period (left). IP resolution in x-direction comparing the data and simulation output for 2011 data-taking period as a function of angle, ϕ (right). This figures are taken from [11].

of this problem arises as a consequence of much lower estimate of low momentum 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.

Chapter 3

Handling of trimuon correlations at LHCb

Having three muons passing through the detector may result in few issues, which need to be addressed. Two collimated muons may cross through the same parts of the detector if they bear the same charge, which causes problems in resolving their individual tracks. Hence, from the tracking point of view, ghosts and clones are much more likely to occur. In LHCb, plethora of muon PID variables can be used to suppress these types of spurious tracks, however, one must be careful once estimating the relevant PID efficiencies as most efficiencies will be dependant on number of muons in the detector.

3.1 Muon PID variables ULRIK

There is a further set of muon variables apart from the ones mentioned in subsection 2.7.1 that are available for PID of muon. In this section summary of the variables used in analysis of $B^+ \to \mu^+\mu^-\mu^+\nu$ is discussed.

3.1.1 Binary Muon PID variables $\boxed{\text{ULRIK}}$

Similar to `isMuon` shown in Table 2.2, there are two more binary variables, `isMuonTight` and `isMuonLoose`, that can help with classification of muons. As their name suggests, `isMuonLoose` has weaker and `isMuonTight` has stronger conditions to satisfy as compared to `isMuon`.

In each muon station (M1-5) and in each dimension in perpendicular plane ($d = \{x, y\}$), FOI is defined as a function of momentum p in a following way:

$$FOI_d = \rho_{d,1} + \rho_{d,2} \cdot \exp\left(\frac{\rho_{d,3} \cdot p}{\mathrm{GeV}/c}\right). \tag{3.1}$$

Muon passing through the detector will leave hit in pad_d of each muon station and therefore have h_d coordinate. Extrapolation coordinates from tracking system made into the stations are denoted E_d. The hits are considered to be within the FOI for given d if they satisfy the condition that $||h_d - E_d|| < FOI_d \cdot pad_d$.

The detector information is read out in 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 3.1, the overlapping x and y *physical* pads of can be grouped into *logical* pads, which give information about x and y simultaneously. These 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.

3.1.2 Muon PID variables based on sharing hits $\boxed{\text{ULRIK}}$

Another way of identifying muon tracks is based on variable, `nShared`, which gives number of tracks with shared hits in the muon stations. For each hit within FOI of the extrapolated track, the `nShared` algorithm will check whether any other track was built using the given hit. In this case, the muon track which has bigger distance between the extrapolation coordinates and the hit coordinates is increased by 1 and the other track becomes the hit owner track. Hence this integer PID variable helps supressing

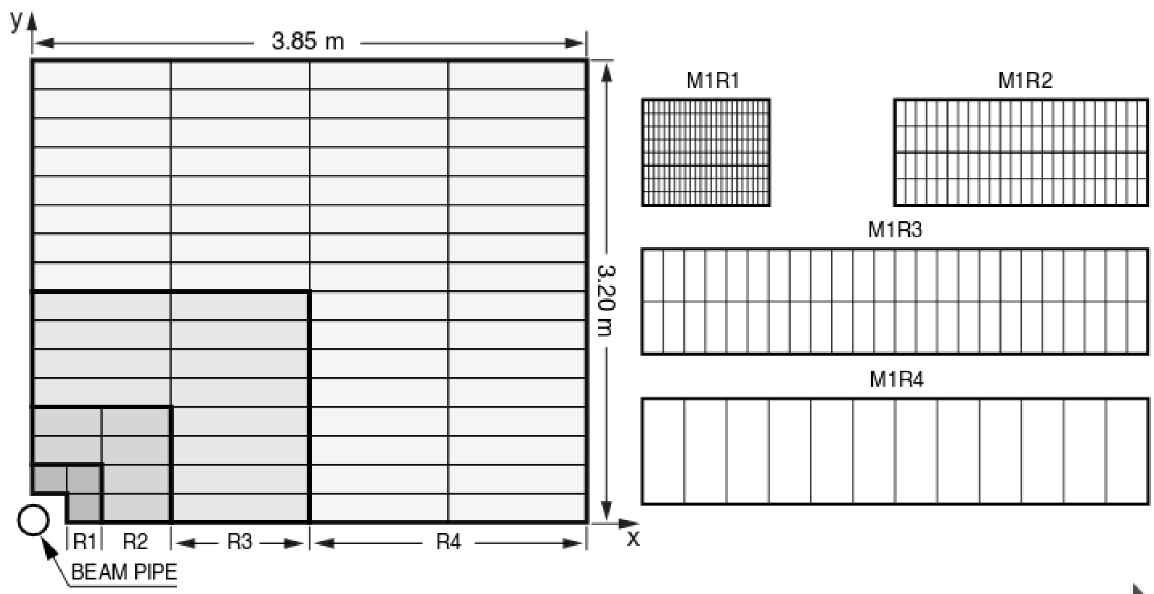


Figure 3.1: Schematic view of the muon station slicing into x-y pads. This is left quadrant of M1 station, showing decreasing granularity of the muon stations away from the beam. This figure has been taken from [18]. M1R1 is the innermost region and M1R4 is the outermost region of the M1 station.

ghost tracks and *clones* if no tracks have hits in common with the owner of the track (nShared==0).

To make sure that muons for $B^+ \to \mu^+\mu^-\mu^+\nu$ are not coming from these spurious tracks, nShared==0 for all three muons is required. These three muon tracks will not to share hits in muon stations with any other downstream or long tracks. In analysing data, there were features in the Muon ID alghorithm, software that calculates most of muon PID variables, which changed between Run I and Run II and will be discussed in greater detail, as it impacts the selection performed for $B^+ \to \mu^+\mu^-\mu^+\nu$ search.

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 extrapolated track and 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 algorithm. For analysts, this feature was present across all reconstruction software and hence it is consistent within stripping version, 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 criteria for 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 $[1,4]$ array (M2-M5 station) resulting in 5-element array where 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 M5 muon station. This obviously has effect, but again it is consistent across the reconstruction version.

The interplay between all discussed features for $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ can be seen in Figure 3.2, which see shift in distribution of `nShared` for 2016 data taking, making the muons less isolated.

This will have inevitably and impact on the PID probabilities. Using the same calibration channels as in subsection 2.7.2, misID and ID rates can be seen in Figure 3.3. 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).

3.1.3 Muon PID variables based on regression techniques $\boxed{\text{ULRIK}}$

Similar to the `DLLmu` variable in subsection 2.7.1, which combines all the information from the detector into a global likelihood, it is possible to feed all the different variables

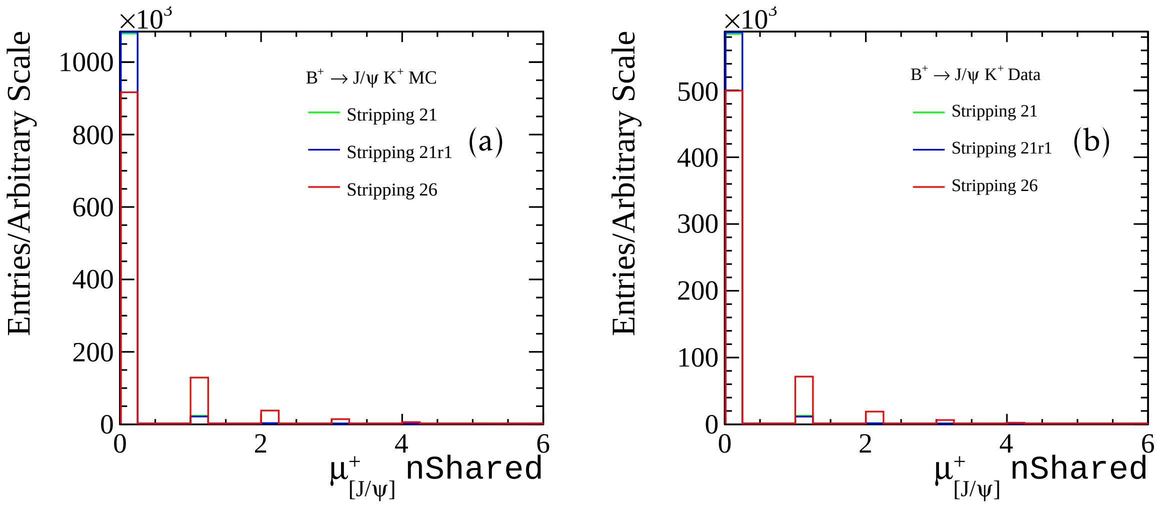


Figure 3.2: (a) The nShared variable in simulation, (b) data for $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ in different stripping versions corresponding to 2012 (*Stripping 21*), 2011 (*Stripping 21r1*), 2016 (*Stripping 26*) data-taking. There is shift of distribution in *Stripping 26* towards less isolated tracks.

to a neural network, which can then produce an output corresponding to the probability of a particle to be of a certain species. Probnn$x$, where x is the species of interest, is calculated and can be used also for muon identification. Compared to *DLLx* variables, Probnn$x$ 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 $B^+ \to \mu^+\mu^-\mu^+\nu$. As with any machine learning algorithms, the selection of the training sample is important as well as input variables. In Run I there were two tunings (trainings) introduced V2 and V3, with more input variables in V2. Depending on the species of particles, V2 or V3 performed better. In the analysis of $B^+ \to \mu^+\mu^-\mu^+\nu$ Probnn$x$_V2 is used.

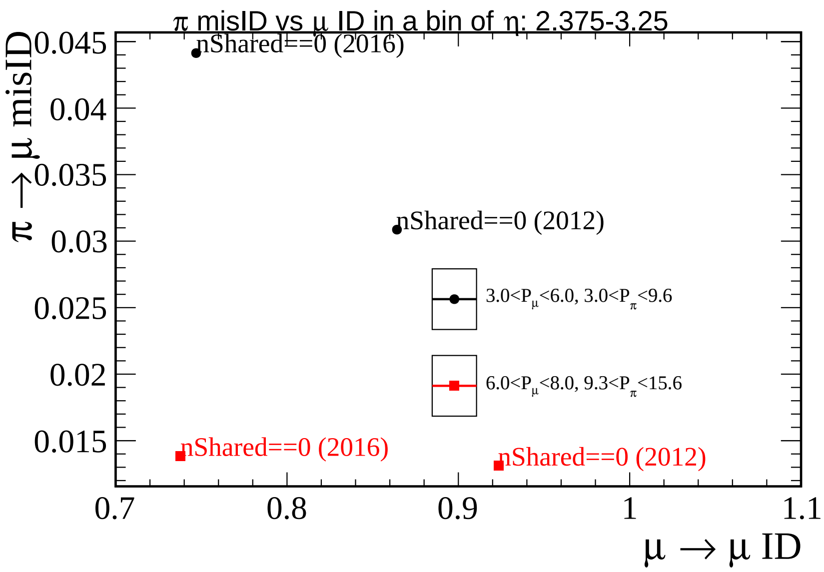


Figure 3.3: ID and misID probabilities from standard calibration datasets from 2012 (*Stripping 21*) 2016 (*Stripping 26*), binned using default 2-dimensional binning scheme in momentum p and pseudorapidity η. In this plot, ID and misID rate in central bin of η and first and second bin in p are compared. This demonstrates that for same ID efficiency, the misID rate is significantly higher in 2016.

3.2 Clones ULRIK

When analysing decays with two muons of the same charge, LHCb magnet bends these two muons in two separate planes. With two muons of the same sign, it is more likely that their two tracks will be collimated. This poses more difficult task for tracking 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.

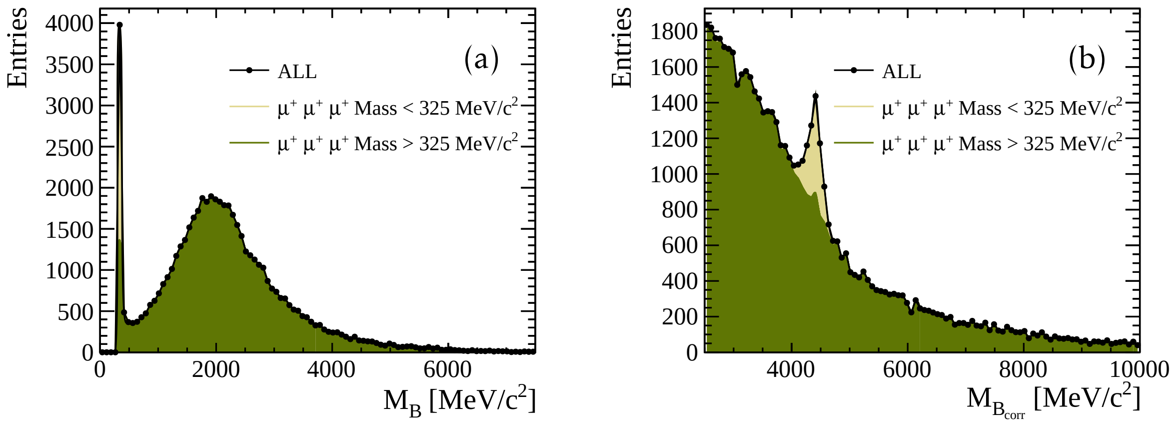


Figure 3.4: (a) Visible and (b) corrected mass of B, shows a clear peak coming from clones in 2012 same sign data sample.

In search for $B^+ \to \mu^+\mu^-\mu^+\nu$ two muons have the same charge, and hence are affected by the *clones*, which needs to be understood. In this case, sample with three muons of same sign was obtained, where the effect is even more prominent and can create potentially *fake peaks* in visible mass spectrum. *Clones* peak at well defined visible mass

$$M_B = \sqrt{(3 \times M_\mu)^2} \approx 318\,\mathrm{MeV}/c^2 \tag{3.2}$$

Once translated into corrected mass, these *fake peaks* are smeared and look like genuine resonances with resolution as seen in Figure 3.4.

The shape of genuine resonance comes as a result of LHCb many factors: 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. *Clones* can be, however, differentiated by the position of the decay vertex of B, Figure 3.5 as well as the occupancy in the tracking, OT as seen in Figure 3.6.

With this typical path for the clones there is a fixed angle of the clones trough the detector (the angle between muon momentum the z-axis), which is calculated using information from OT in a following way:

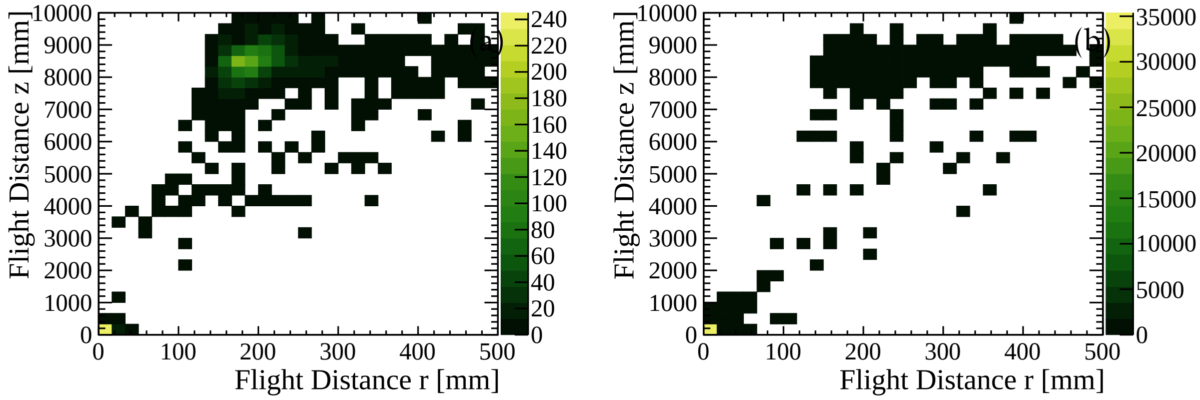


Figure 3.5: (a) Clone and (b) no clones flight distance properties. It can be seen that *clone* tracks have their decay vertex placed at the end of the detector, whereas regular good tracks will decay within VELO.

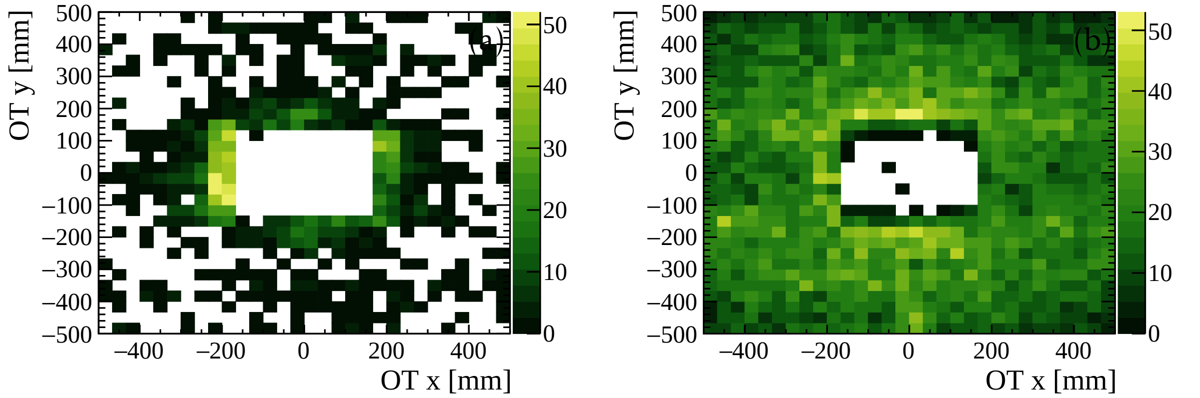


Figure 3.6: The occupancy difference in the OT detector between (a) clones and (b) real tracks in the OT at the distance 9450 mm along the LHCb. *Clones* are concentrated along the inner edge of the OT. Good muon tracks will cover most of OT evenly.

$$\arctan(\theta) = \arctan\left(\frac{FD\ radius}{FD\ distance\ along\ z}\right) = \arctan\left(\frac{200\ \text{mm (Figure 3.6)}}{8500\ \text{mm}}\right) = 0.023rad. \tag{3.3}$$

With the `L0Muon` p_T threshold of 1.76 GeV/c for 2012 [20], the typical momentum from about 75 to 120 GeV/c is yielded because

$$p = 1.76\,\text{GeV}/c/\sin\left(\arctan(\theta) = \arctan\left(\frac{200\text{ mm}}{8500\text{ mm}}\right)\right). \quad (3.4)$$

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 3.7.

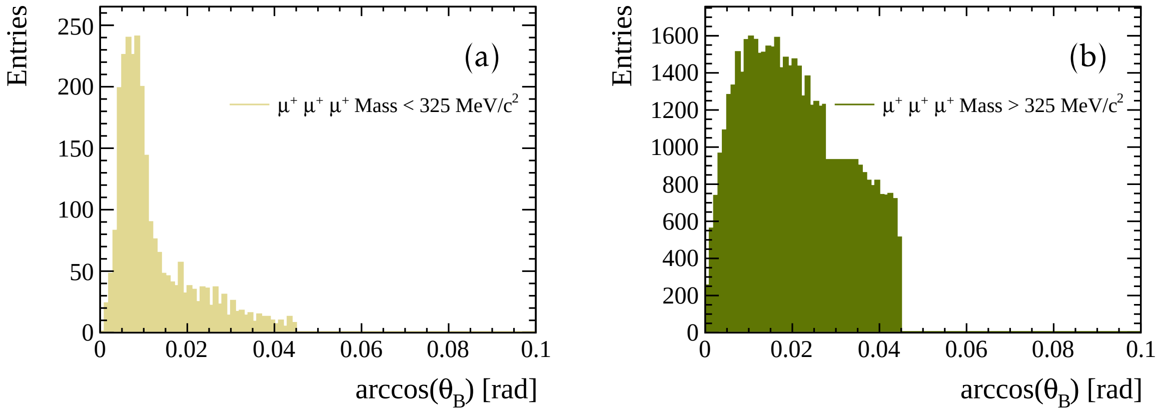


Figure 3.7: (a) Peaking clone distribution is visible as all of *clone* tracks are collinear compared to (b) smooth no clone distribution for $\cos(\theta_B)$.

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

$$p_T = 100\,\text{GeV}/c \times \sin(0.0007) = 0.7\,\text{GeV}/c. \quad (3.5)$$

Hence corrected mass $M_{corr} = \sqrt{M^2 + |p_T^2|} + |p_T| = 4.2\,\text{GeV}/c^2$, using missing p_T from Equation 3.5 and visible mass of *clones* from Equation 3.2.

In order to suppress these tracks in analysing $B^+ \to \mu^+\mu^-\mu^+\nu$, where two muons have the same sign, any distinguishing features mentioned could be used. But the most powerful PID-wise is requiring `nShared==0`, as this requirement removes all of the clones, as seen in Figure 3.8.

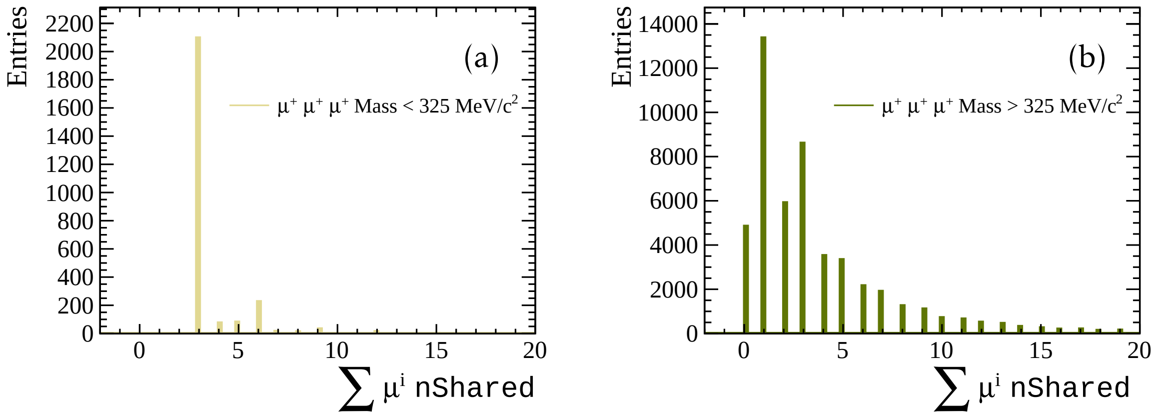


Figure 3.8: (a) Clone and (b) no clone distribution for sum of all muon `nShared`. Since in this case the clones are of each other, for clone there is clear peak at three.

3.3 Probability of $K/\pi \to \mu$ misidentification at LHCb $\boxed{\text{ULRIK}}$

Usually, in order to estimate background coming from misidentification of species into muons in the detector, data samples with *non-muon* (identified not to be muons) tracks are obtained. They are assigned parametrically probabilities of misid rates from some specific control samples, which were discussed in subsection 2.5.1. However, three muon signature will induce problems for PID variables that are correlated with the number of muons in the detector and other control samples need to be used.

3.3.1 Specific control sample for $K/\pi \to \mu$ misid rates $\boxed{\text{ULRIK}}$

At LHCb, `PIDCalib` package can usually provide misID and ID rates obtained from standard K/π control samples $D^{*+}(\to D^0(\to \underline{K^+\pi^-})\pi^+)$. 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 $D^{*+}(\rightarrow D^0(\rightarrow \underline{K^+\pi^-})\pi^+)$, where there are no muons in the final state. This influences rest of misID rates as some of the PID variables are strongly correlated with number of muons in the decay, due to the fact that the mis-identified particle can share hits with other muons in the rest of the decay. This should be reflected mostly in high momenta region, where the three particles tend to be collimated and share hits most often.

For this reason, an alternative sample that is also be statistically populated background-free, $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*(\rightarrow \underline{K^+\pi^-})$. It mimics the two real muon plus fake muon correctly and will be used to obtain pion and kaon misID probabilities.

3.3.2 Selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ $\boxed{\text{ULRIK}}$

Data samples for each year of data taking were obtained from *stripping* dedicated to look for this type of decay, but with both kaon and pion having no particle identification applied, which ensures that PID performance of different variables can be studied. Some initial selection was applied together with more stringent $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ selection including trigger (but on J/ψ candidate rather than B). Finally to remove most of the backgrounds that still pollute the signal samples following selection summarized in Table 3.1 is used.

3.3.3 Fitting Strategy for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ decay $\boxed{\text{ULRIK}}$

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 mean μ and width σ. It is fitted with double-sided Hypatia function [29] (more in section C.2).

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

Idea	Cut
ID K^*	$\vert$ m(Kπ) - m$_{PDG}(K_0)$ $\vert$ < 100 MeV/c^2
Compatible with `PIDCalib`	for K,π , p$_T$ > 250 MeV/c
Compatible with `PIDCalib`	for μ , p$_T$ > 800 MeV/c
Muon swap veto	$\vert$ m((h $\rightarrow \mu$)μ) - m$_{PDG}$(J/ψ)$\vert$ > 60 MeV/c^2
Veto $B^+ \rightarrow K^+\mu^+\mu^-$	max(m($K^+\mu^+\mu^-$)), m(($\pi^+ \rightarrow K^+$)$\mu^+\mu^-$)) < 5100 MeV/c^2
Veto $B_s^0 \rightarrow \phi\mu^+\mu^-$	m(K($\pi \rightarrow$ K)) > 1040 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 `IsMuon==0`

Table 3.1: Offline selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ decay.

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 double sided Crystal Ball function [30] (more in section C.1).

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

And finally combinatorial component is modelled by the exponential function.

In order to obtain K/π misid probabilities fit to $\mu^+\mu^-\pi^+K^-$ mass between 5150 - 5450 MeV/c^2 was performed, where all of the yields for different components are left floating. Moreover, mass of J/ψ was *constrained* to its nominal mass, known as *mass constraint*, which will yield new estimates for track parameters of the final state particles, from which a new kinematic refit is done.

Actual extraction of the misid rate was obtained using statistical subtraction of background, also known as *sPlot* technique. The same is used in `PIDCalib` package. It was cross-checked with *fit twice method*, because *sPlot* technique relies on the fact that there is no correlation between the 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 $B^0 \to J/\psi(\to \mu^+\mu^-)K^*$ before and after PID requirement in a given kinematic (p, η) bin separately. Misid probabilities are then then obtained as a ratio of signal yields arising from these two fits.

In the *sPlot* method, the invariant mass distribution is fitted once and each event is assigned *sWeights*, probabilities that a given event is a signal-like or a background-like. Then, through *sPlot* technique, background is subtracted. The misid probabilities are then obtained by looking at sum of all *sWeights* in each bin.

It was shown that these two methods yields very similar results, hence, for purposes of $B^+ \to \mu^+\mu^-\mu^+\nu$ analysis *sWeight* values will be used. Fits to Run I and Run II data can be seen in Figure 3.9.

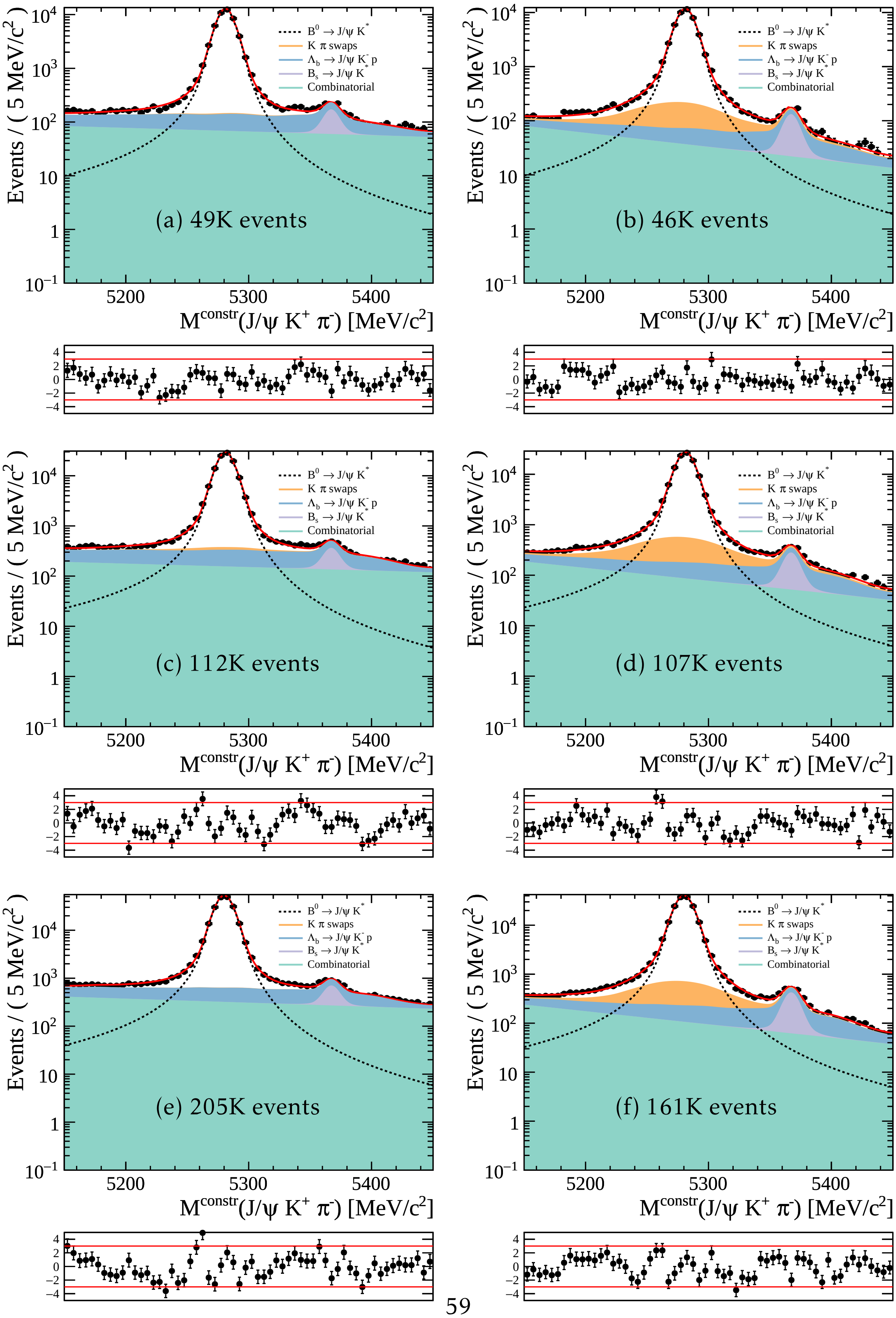


Figure 3.9: Fit to constrained $J/\psi(\to \mu^+\mu^-)K^*(\to \pi^+K^-)$ mass with all the components for (a)(b) 2011, (c)(d) 2012, (e)(f) 2016. On the left, fit to data with pion ID (giving kaon misid probabilities), on right data with kaon ID (pion misid rates).

3.3.4 Results of $B^0 \to J/\psi(\to \mu^+\mu^-)K^*$ control sample for $K/\pi \to \mu$ misid rates $\boxed{\text{ULRIK}}$

Using the *sWeight* method, misID rates for K/π can be obtained. In order to rule out that any disagreement of PID performance is caused by the rest of the event (such as the different nature of the tracks) following check is performed. For both samples 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 3.2 that the ID performance is the same across nearly all of the momentum range for kaons, showing that these kaons are good proxy for misID probability studies. The same check was done 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 3.10, $\pi \to \mu$ 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. On the right, the ratio of the misID probabilities between two samples show particular trend.

In general the agreement is good in the low momenta regions between the two samples. These pions are softer and hence they will not be collimated, causing less interference with other two real muons in decay. However, in high momenta region, particles will tend to be more collimated. 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 with 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 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.

p range [MeV/c]	$D^{*+}(\to D^0(\to \underline{K^+\pi^-})\pi^+)$	$B^0 \to J/\psi(\to \mu^+\mu^-)K^*(\to \underline{K^+\pi^-})$	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 3.2: `K_InMuonAcc==1.0` shows the interpolation of K tracks into muon chambers. It can be seen that both samples agree with each other, meaning that the rest of the event information is the same for both samples. This measurement is in a bin $1.5 < \eta < 5.0$.

Even though this disagreement is decreased, it can be still noted that for the high momenta region $\pi \to \mu$ Figure 3.10 and $K \to \mu$ Figure 3.11 rate is 2 to 3 times higher with additional two real muon tracks, which is significant.

Due to the different PID definitions of `nShared` between Run I and II, different PID requirement are tested. Results for $\pi \to \mu$ and $K \to \mu$ are summarized in Figure 3.12 and Figure 3.13. The misID probabilities in 2016 for also show the same momentum dependent trend as in 2012.

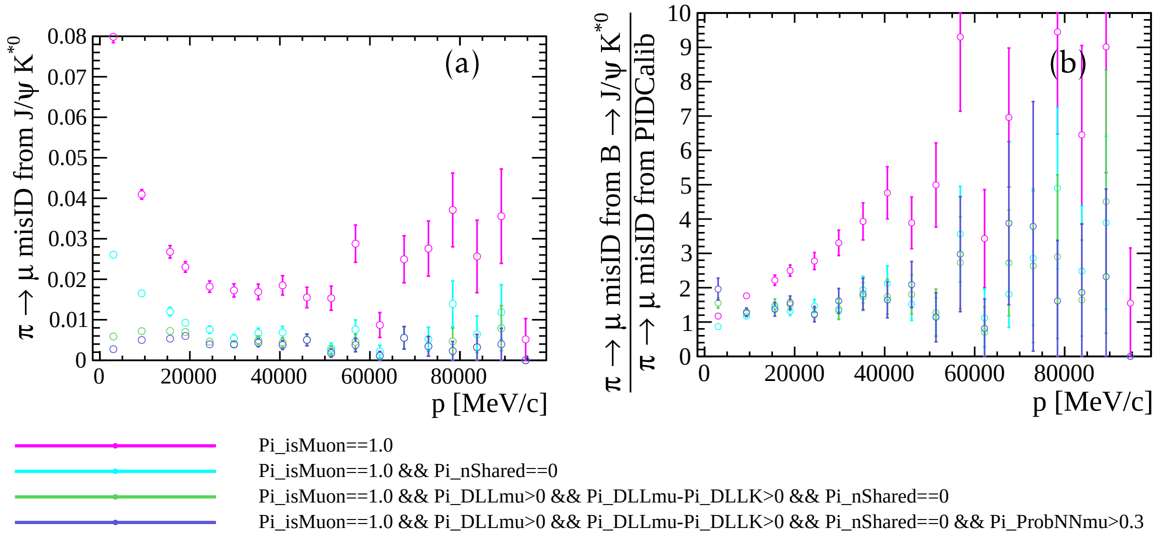


Figure 3.10: (a) $\pi \to \mu$ misID probability for different PID requirements obtained using $B^0 \to J/\psi(\to \mu^+\mu^-)K^*(\to K^+\pi^-)$ for 2012 data. (b) This is compared to the standard `PIDCalib` $D^{*+}(\to D^0(\to K^+\pi^-)\pi^+)$ sample.

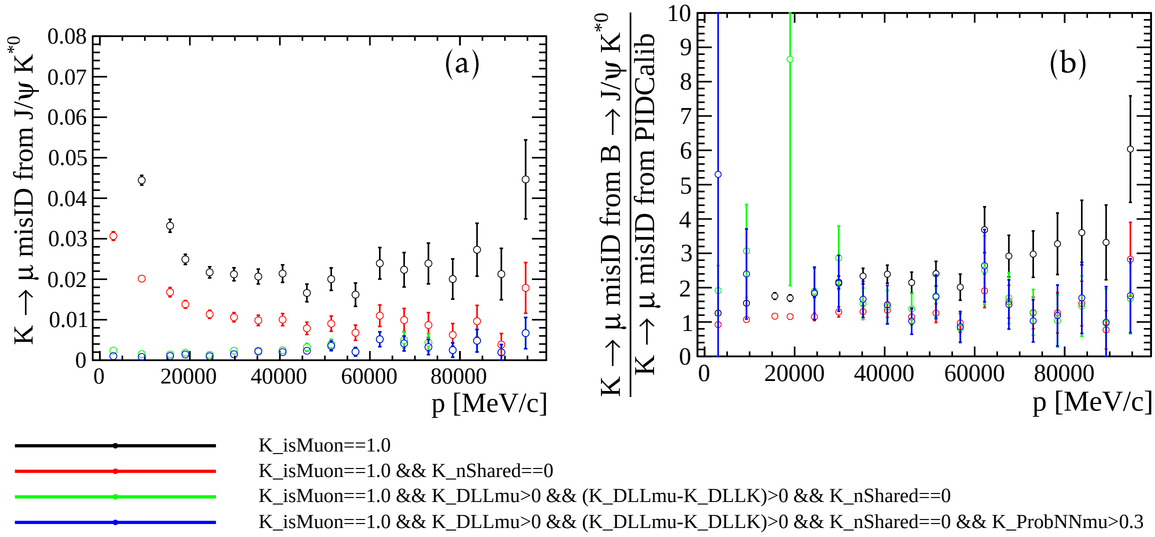


Figure 3.11: (a) $K \to \mu$ misID probability for different PID requirements obtained using $B^0 \to J/\psi(\to \mu^+\mu^-)K^*(\to K^+\pi^-)$ for 2012 data. (b) This is compared to the standard `PIDCalib` $D^{*+}(\to D^0(\to K^+\pi^-)\pi^+)$ sample.

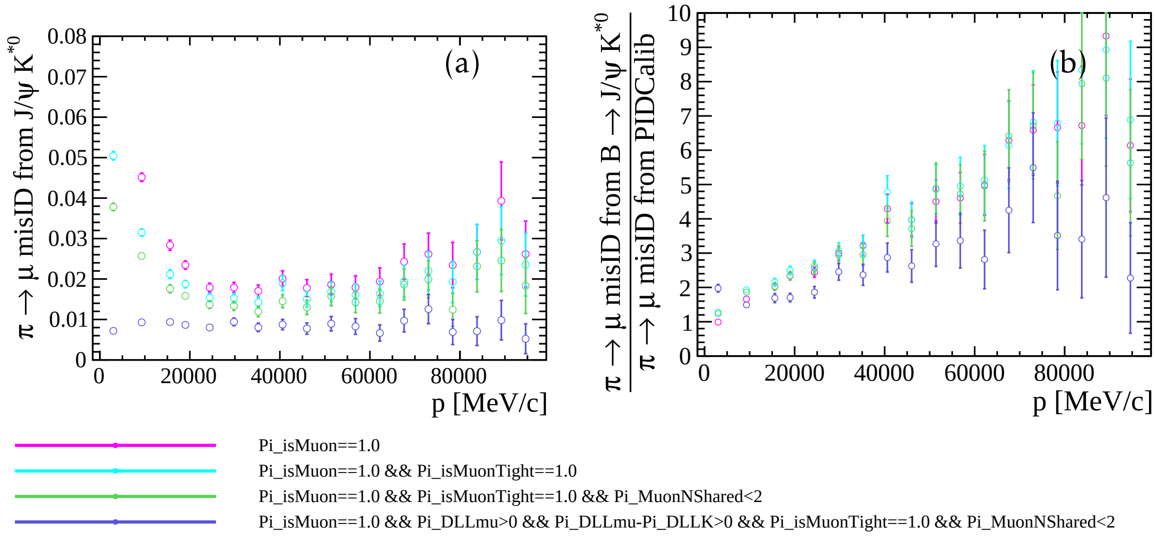


Figure 3.12: (a) $\pi \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*(\rightarrow K^+\pi^-)$ for 2016 data. (b) This is compared to the standard `PIDCalib` $D^{*+}(\rightarrow D^0(\rightarrow K^+\pi^-)\pi^+)$ sample.

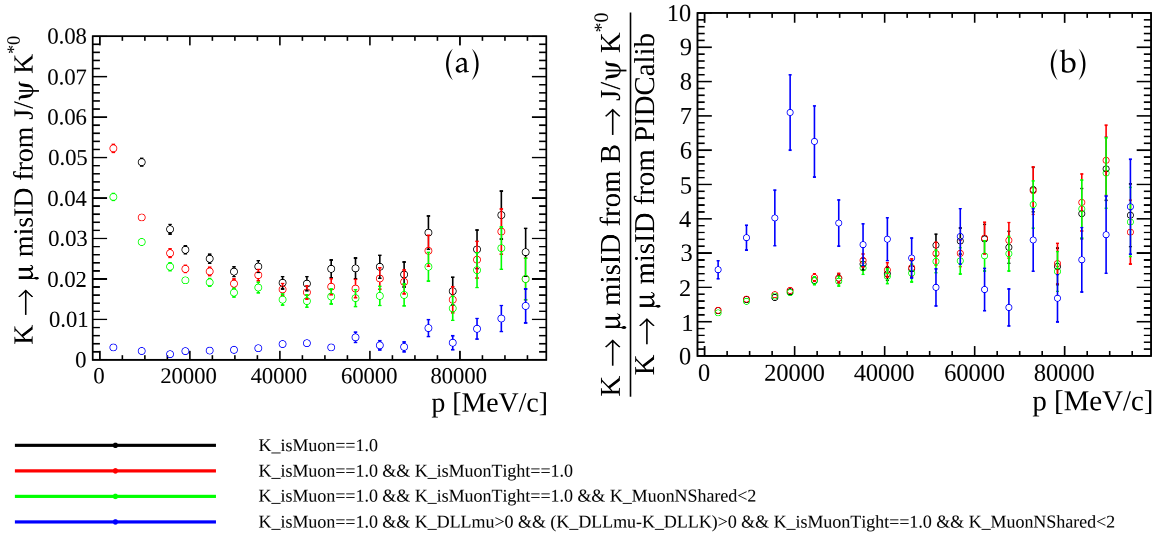


Figure 3.13: (a) $K \rightarrow \mu$ misID probability for different PID requirements obtained using $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*(\rightarrow K^+\pi^-)$ for 2016 data. (b) This is compared to the standard `PIDCalib` $D^{*+}(\rightarrow D^0(\rightarrow K^+\pi^-)\pi^+)$ sample.

Chapter 4

Discovering (Setting Limit for) $B^+ \to \mu^+\mu^-\mu^+\nu$ 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 $B^+ \to \mu^+\mu^-\mu^+\nu$ 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 4 will concentrate on characterisation of backgrounds as well as selection that is performed in order to reduce these backgrounds.

4.1 Topology of at LHCb $B^+ \to \mu^+\mu^-\mu^+\nu$ at LHCb $\boxed{\text{ULRIK}}$

Upon hadronisation of $b\bar{b}$ pair $B^\pm$ particle will travel less than a millimetre in the laboratory frame of reference before it decays into its decay products. This allows reconstruction of a primary vertex PV and its decay vertex, *secondary vertex* SV. By joining these vertices, direction as well as length of the $B^\pm$ existence, also known as flight distance (FD), can be established. In order to infer information about 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 $B^\pm$, neutrino is assigned all missing momentum transverse to the direction of the flight of $B^\pm$. The schematic diagram can be seen in Figure 4.1.

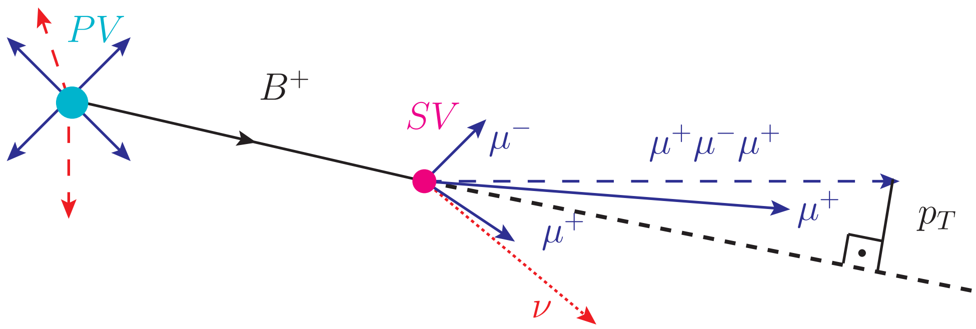


Figure 4.1: Schematic view of $B^+ \to \mu^+\mu^-\mu^+\nu$ decay. At pp interaction point, or PV, $b\bar{b}$ pair hadronizes into $B^\pm$. $B^\pm$ flies some distance before decaying into three muons and neutrino. All charged tracks (in filled-blue) seen can be combined into four-vector representing 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 $B^\pm$. Neglecting momentum component parallel to the direction of flight for neutrino, transverse component of momentum is given.

Altogether available information allows for reconstruction of a quantity, *corrected mass*, that plays similar role to invariant mass in fully reconstructed decays. Invariant mass is usually used in LHCb as fitting distribution from which physics results are extracted. This particular quantity is used as it distinguishes well signal and background shapes with minimal modelling assumptions.

Corrected mass is defined as

$$M_{corr} = \sqrt{M^2 + |p_T^2|} + |p_T|, \tag{4.1}$$

where the M^2 is the invariant visible mass squared and p_T^2 is the missing momentum squared transverse to the direction of B^+ flight. Usually the corrected mass of $B^\pm$ 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 $B^{\pm}$ 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) is

$$p_T^2 = |\vec{p} - (\vec{x}_{SV} - \vec{x}_{PV})\frac{\vec{p}\cdot(\vec{x}_{SV} - \vec{x}_{PV})}{|(\vec{x}_{SV} - \vec{x}_{PV})|^2}|^2 \tag{4.2}$$

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

$$\langle f^2 - \langle f \rangle^2 \rangle = \langle f(x+\delta x, y+\delta y, z+\delta z)^2 - f(\langle x \rangle, \langle y \rangle, \langle z \rangle)^2 \rangle \tag{4.3}$$

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

$$\begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix} \begin{bmatrix} \delta x^2 & \delta x \delta y & \delta x \delta z \\ \delta y \delta x & \delta y^2 & \delta y \delta z \\ \delta z \delta x & \delta z \delta y & \delta z^2 \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix} \tag{4.4}$$

So now assuming that $x = \vec{x}_{PV}$, $y = \vec{x}_{SV}$ and $z = \vec{\boldsymbol{p}} = \{E, p_x, p_y, p_z\}$,

$$\nabla_{x_{PV}}^T \mathrm{COV}_{\mathrm{x_{PV}}} \nabla_{\mathrm{x_{PV}}} + \nabla_{\mathrm{x_{SV}}}^{\mathrm{T}} \mathrm{COV}_{\mathrm{x_{SV}}} \nabla_{\mathrm{x_{SV}}} + \nabla_{\mathrm{p}}^{\mathrm{T}} \mathrm{COV}_{\mathrm{p}} \nabla_{\mathrm{p}} \tag{4.5}$$

where COV is the covariance matrix.

In conclusion in order to calculate error on *corrected mass*, δ_{corrm}

$$\delta_{corrm} = \sqrt{\langle f^2 - \langle f \rangle^2 \rangle} = \sqrt{\nabla_{x_{PV}}^T \mathrm{COV}_{\mathrm{x_{PV}}} \nabla_{\mathrm{x_{PV}}} + \nabla_{\mathrm{x_{SV}}}^{\mathrm{T}} \mathrm{COV}_{\mathrm{x_{SV}}} \nabla_{\mathrm{x_{SV}}} + \nabla_{\mathrm{p}}^{\mathrm{T}} \mathrm{COV}_{\mathrm{p}} \nabla_{\mathrm{p}}} \tag{4.6}$$

which can be calculated analytically (method used for all the plots) or using numerical approximation of first derivative of *finite differences*.

4.2 Sources of Backgrounds $\boxed{\text{ULRIK}}$

The largest background that can be will be looking similar to signal comes from *cascade decays*, where semileptonic $b \to c \to s$ or ($\bar{b} \to \bar{c} \to \bar{s}$) transition occurs. A typical example of this background in hadronic terms is $B^+ \to (\bar{D}^0 \to (K^+ \to \mu^-\nu)\mu^+\nu)$, where K^+ is misidentified as muon. Because K^+ is misidentified as muon, this type of background is denoted as misID background.

In fact, any other particle species that is misidentified, belongs to the misID background category. 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. However, *OS misID* background is expected to have smaller rate as the misidentified particle would have to proceed via decays with additional particles or if coming as product from other b hadronization.

The presence of other B-hadron from $b\bar{b}$ pair does create its own decay chain and hence it is possible to combine one of its muon tracks with two muons from the "signal" B. This is denoted as combinatorial background.

Then presence of 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 either they are charged but they are soft, 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 $B^+ \to D\mu^+\nu$ type partially reconstructed backgrounds where $B^+ \to (D^0 \to K^-\pi^+\mu^+\mu^-)\mu\nu$, where $\mathcal{B}(D^0 \to K\pi^+\mu^+\mu^-) \approx 4.17 \times 10^{-6}$ and $B^+ \to D^0\mu\nu \approx 10\%$. This predicts $\mathcal{B}(B^+ \to K^+\pi^-\mu^+\mu^-) = 1 \times 10^{-7}$.

4.3 Analysis strategy

The analysis of the $B^+ \to \mu^+\mu^-\mu^+\nu_\mu$ 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 1 dataset as well $\approx$ 1.7 fb^{-1} Run 2 data (not using 2015 dataset due to very low sensitivity (high muon trigger thresholds)). Additionally the search will be conducted in a particular minq^2 = min$(q^2(\mu_1^+,\mu^-), q^2(\mu^-,\mu_2^+))$ region.
To perform the search for $B^+ \to \mu^+\mu^-\mu^+\nu_\mu$, a specific preselection was applied to form potential signal candidates. To reconstruct the mass of the B^+ with missing information about the neutrino, a corrected mass variable $M_{B_{corr}} = \sqrt{M_{3\mu}^2 + |p_\perp^2|} + |p_\perp|$, where $M_{3\mu}^2$ is the invariant visible mass squared and $p_\perp^2$ is the missing momentum squared transverse to the direction of flight of B^+, is introduced. A simulation sample that mimics the decay of the $B^+ \to \mu^+\mu^-\mu^+\nu_\mu$ passing through preseletion 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 **??**.
Combinatorial background, which arises as random combinations of tracks passing the preselection, is taken from the upper corrected $\mu^+\mu^-\mu^+$ mass side band, $M_{Bcorr} > 5.5$ GeV, where very few signal candidates are expected.

4.4 Samples $\boxed{\text{ULRIK}}$

4.4.1 Data Samples $\boxed{\text{ULRIK}}$

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

4.4.2 Simulation Samples $\boxed{\text{ULRIK}}$

For signal simulation, three different decay models were exploited and are summarized in Table 4.1.

Channel	Year	Pythia	EVTGEN	Size	Stage
		Simulation used for fitting mass shapes			
$B^+ \to \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [23]	PHSP	0.5M	*generator-level+detector*
$B^+ \to \mu^+\mu^-\mu^+\nu$	2012	Pythia 8.1 [2]	PHSP	0.5M	*generator-level+detector*
$B^+ \to \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [23]	MINE	0.5M	*generator-level+detector*
$B^+ \to \mu^+\mu^-\mu^+\nu$	2012	Pythia 8.1 [2]	MINE	0.5M	*generator-level+detector*
$B^+ \to \mu^+\mu^-\mu^+\nu$	2016	Pythia 8.1 [2]	MINE	1.0M	*generator-level+detector*
		Simulation used for evaluating *generator-level* efficiencies			
$B^+ \to \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [23]	PHSP	25000	*generator-level*
$B^+ \to \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [23]	MINE	25000	*generator-level*
$B^+ \to \mu^+\mu^-\mu^+\nu$	2012	Pythia 8.1 [2]	MINE	25000	*generator-level*
		Simulation used for ratification of minq^2 selection			
$B^+ \to \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [23]	NIKI	25000	*generator-level*

Table 4.1: 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.

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 and hence angular momentum is disregarded.

In order to produce simulation with decay model which is more representative of the spin structure involved, 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^+ is decayed to $\mu^+\nu$. *BTOSLLBALL* model [31], traditionally used for $B \to (K, K^*)l^+l^-$ decay, with the form factor calculations can be used to simulate $B^\pm \to W^\pm l^+l^-$ decay. After that, W^+ is decayed to $\mu^+\nu$ using *PHSP*. For semileptonic $b \to sl^+l^-$ 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 0.1GeV/c^2, and width, corresponding to $\tau = 1.3 \times 10^{-17}$ nanoseconds as can be seen in Figure 4.2. This procedure was also applied for the charge conjugate case. This model is denoted as *MINE* and is used as default in mass fits and efficiency calculations.

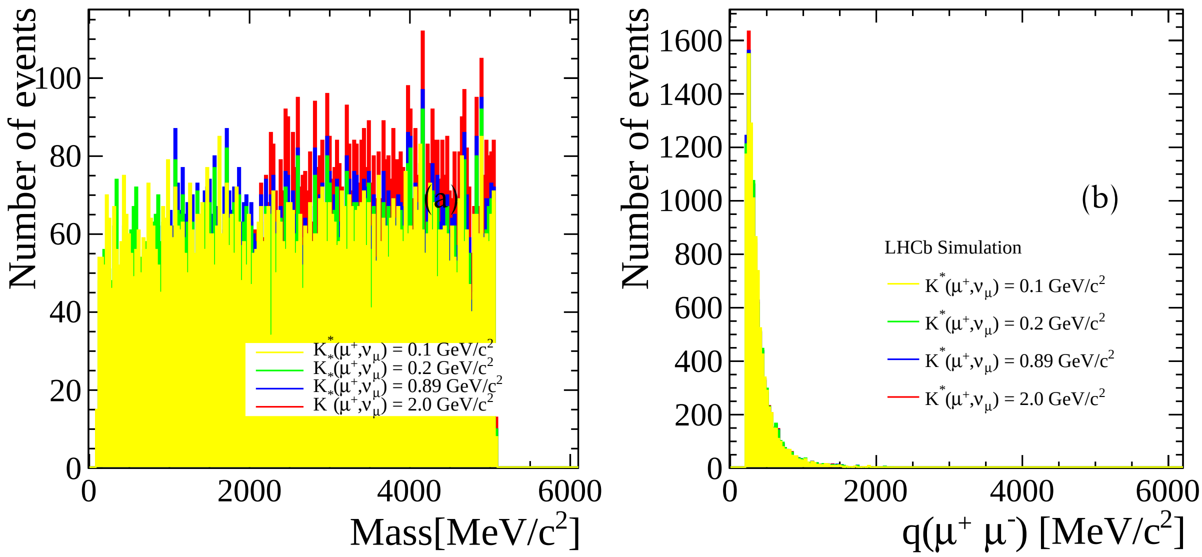


Figure 4.2: Distributions for signal MC in using Pythia 6.4 [23] conditions. (a) $K^*(\mu^+, \nu_\mu)$ (b) $q(\mu^+, \mu^-)$ distributions under different K^* mass hypotheses. The most flat distribution in $K^*(\mu^+, \nu_\mu)$ is plotted in yellow.

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.

4.5 Preselection for $B^+ \to \mu^+\mu^-\mu^+\nu$ $\boxed{\text{ULRIK}}$

Set of initial identification for signal $B^+ \to \mu^+\mu^-\mu^+\nu$ summarized in Table 4.2, also known as *stripping* selection was develloped in order to improve signal to background ratio.

Firstly, all three muon tracks are required to have a significant IP with respect to the primary vertex. Minimum Impact Parameter χ^2, (min IPχ^2), gives the minimum significance of a particles's trajectory to the primary vertex. Hence by requiring min IPχ^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), (track χ^2/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 muon than pion or kaon using global DLLmu variables defined in subsection 2.7.1. 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 χ^2/ndof), the χ^2 of the trimuon vertex per degree of freedom fit required to be small. This decreases the contamination from *cascade decays* where the particle with the c quark content from $b \to \underline{c} \to s$, such as D, would have non-negligible lifetime leading to higher vertex χ^2/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

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

Table 4.2: Selection of events based on muon and the B^+ candidate requirements. *Stripping selection* for the signal decay $B^+ \to \mu^+\mu^-\mu^+\nu_\mu$ is the same for both Run1 and 2016 data.

translates into a well reconstructed event, which minimizes combinatorial background, where random track makes this pointing worse. Putting bounds on mass window, whether it is *visible* or *corrected* mass, also supresses combinatorial events. This is because of on average higher momentum combinatorial muon.

4.6 Trigger Selection $\boxed{\text{ULRIK}}$

In order to obtain triggered data, $B^+ \to \mu^+\mu^-\mu^+\nu_\mu$ candidates are required to pass certain set of trigger decisions at L0, HLT1 and HLT2 levels summarized in Table 4.3. It can be noted that the decision is applied at the mother $B^\pm$ level. In particular, `Bplus_L0MuonDecision_TOS decision`, means that one of the muons from $B^\pm$ in an event has triggered and made positive decision.

Trigger Selection	
`Bplus_L0MuonDecision_TOS`	
`Bplus_Hlt1TrackMuonDecision_TOS`	
`Bplus_Hlt2TopoMu2BodyBBDTDecision_TOS`	OR
`Bplus_Hlt2TopoMu3BodyBBDTDecision_TOS`	
`Bplus_Hlt2DiMuonDetachedDecision_TOS`	
`Bplus_Hlt2DiMuonDetachedHeavyDecision_TOS`	

Table 4.3: Trigger selection applied on both signal and normalisation samples

As discussed in section 2.8 `L0MuonDecision` decides on whether an event is accepted depending on the p_T of 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 GeV/c [32] and 1.76 GeV/c [20]. Run I SPD rate only accepts events 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.

`Hlt1TrackMuonDecision` accepts events where mother particle has certain lifetime (which is case for B, D mesons), by requiring certain $IP\chi^2$ of the track with respect to all of its PVs. 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 [32].

At HLT2 level, the candidates are reqiured to pass through at least one of the four decisions. `Hlt2TopoMu[2,3]BodyBBDTDecision` belong to the *topological triggers* category with extra requirement of muon being idendified by `isMuon` decision. `Hlt2DiMuonDetachedDecision` and `Hlt2DiMuonDetachedHeavyDecision` reconstruct decays with two muons in a final state, dimuon. The two lines differ in selection that is optimized either for heavy or light dimuon pair. For example, `Hlt2DiMuonDetachedDecision` accepts events with dimuon p_T above 1.5 GeV/c and with mass above 1 GeV/c^2, whereas `Hlt2DiMuonDetachedHeavyDecision` 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 $IP\chi^2$.

4.7 q^2 selection $\boxed{\text{ULRIK}}$

In $B^+ \to \mu^+\mu^-\mu^+\nu$, 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, $minq^2 = min[q^2(\mu_1,\mu_2), q^2(\mu_2,\mu_3)]$ and $maxq^2 = max[q^2(\mu_1,\mu_2), q^2(\mu_2,\mu_3)]$. This measurement is made in region where $\sqrt{minq^2} = minq < 980$ MeV/c^2 because of two main reasons: most of the contributions to the amplitude of the decay is below this value and combinatorial background is greatly reduced if $minq^2 < 1$ (GeV/c^2)^2, see Figure 4.3.

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 4.4 for more details.

Veto	$minq$ [MeV/c^2]	$maxq$ [MeV/c^2]
J/ψ	!(2946.0 < $minq$ < 3176.0)	!(2946.0 < $maxq$ < 3176.0)
Ψ(2S)	!(3586.0 < $minq$ < 3766.0)	!(3586.0 < $maxq$ < 3766.0)

Table 4.4: Vetoes for J/ψ and $\Psi(2S)$ contributions.

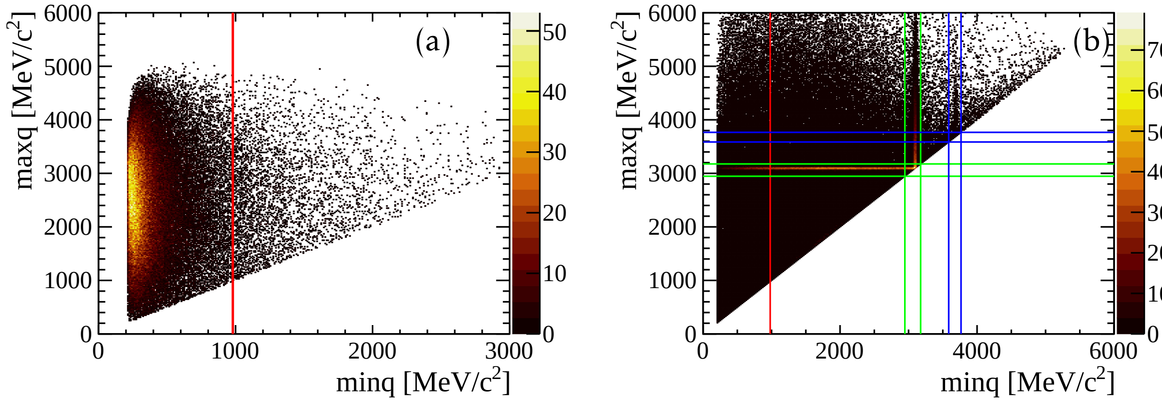


Figure 4.3: (a) Signal simulation sample distribution in *minq* and *maxq* variables. Values below 980 MeV/c^2 (red line) are accepted. (b) Combinatorial data sample after *stripping* selection with no other cuts shows clearly the J/ψ (green) and $\Psi(2S)$ (blue) resonances which are vetoed and the measurement region (red).

4.8 Further Selection

Further selection was performed as seen in Table 4.5. This selection further suppresses backgrounds but requires different treatement in Run I and Run II due to the different definitions of variables as seen in subsection 3.1.2.

Idea	Object	Run I Selection	
Clean	Muon	-	I
Clone and ghost	Muon	`Nshared==0`	
Fit Region	B	4000 MeV/c^2 $< M_{B_{corr}} <$ 7000 MeV/c^2	4000 Me
Bkg Removal	event	Combinatorial BDT selection	Comb
Bkg Removal	event	Misid BDT selection	M
Optimize FOM $\boxed{\text{INTRODUCE FOM}}$	Muon	`PID`μ`>0.35`	

Table 4.5: Offline selection performed after stripping. Differences can be seen between Run I and Run II datasets

The sections below will outline multivariate techniques used in this analysis, the isolation BDT subsection 4.8.1, combinatorial BDT subsection 4.8.2 and misidBDT section 4.9.

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 classfier 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 is 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 misidBDT is known as *k-folding* technique [33].

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 two consecutive BDTs, number of folds used is k=10.

Both of the BDT classifiers use AdaBoost algorithm and same set of variables listed in Table 4.6.

After classifiers to reduce combinatorial and misid backgrounds are trained and applied, further PID selection is performed. This can be done as the preselection has relatively loose DLL selection and hence it is possible to improve the performance by using cuts on additional PID variables. In the optimisation, different hypotheses were tested, such as cuts on Probnnmu, Probnnpi, and ProbnnK variables and their

$B^+ p$	min IPχ^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 IPχ^2 of all three muons	Isolation variable	subsection 4.8.1
B^+ lifetime			

Table 4.6: BDT variables used in both combinatorial and misID Run I and Run II BDTs

combinations. Optimisation was performed in a such a way as to optimize Punzi figure of merit (`FOM`) [34], defined in Equation 4.7, in a blinded signal region, by performing full blinded data fit make reference to this once it is written, but with Run I and Run II fits separately.

Throughout the analysis the Punzi `FOM` is used to find optimal working point. It is defined,

$$\texttt{FOM} = \frac{\varepsilon_S}{\sqrt{B} + \sigma/2}, \tag{4.7}$$

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. The advantage of this `FOM` to more conventional ones, is is insensitive to the knowledge of exact number of signal events, which is the case as the branching fraction is unknown.

In 2016, misID templates above 5800 MeV/c^2 are not populated. Hence final 6 bins of corrected mass (5800 MeV/c^2-7000 MeV/c^2) are merged into 1. For each PID hypothesis, FOM was calculated. In both cases, in I and II `Probnnmu` > 0.35 yielded highest `FOM`.

4.8.1 The Isolation Boosted Decision Tree

Vast majority of the backgrounds that share the possibility of contaminating $B^+ \to \mu^+\mu^-\mu^+\nu$ 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

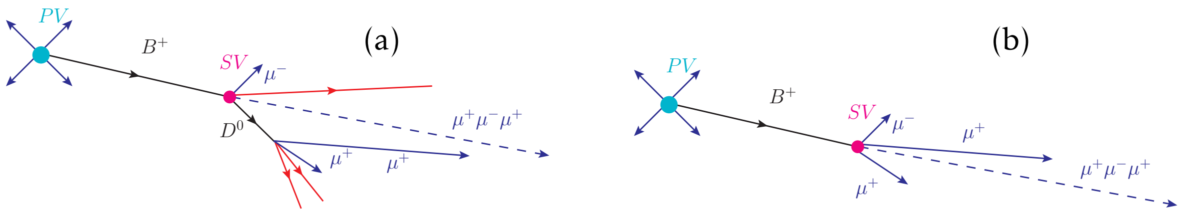


Figure 4.4: (a) An example of decay topology for background and (b) signal.

how *isolated* the signal trimuon vertex is as compared to background trimuon vertex as seen in Figure 4.5

The isolation quality of the vertex is determined using TMVA's [35] implementation of Boosted Decision Tree (BDT) based on the AdaBoost algorithm. 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 $\Lambda_b^0 \to p\mu^-\bar{\nu}$ simulation sample, where all tracks apart from the $p\mu^-$ signal tracks are taken into the account. Background sample was formed with $\Lambda_b^0 \to (\Lambda_c \to p)\mu^-\bar{\nu}$ tracks, also disregarding the $p\mu^-$ tracks. The isolation BDT is based on these samples yield weights, which are computed in [36], and are result of other's people work. These weights are however then applied parametrically on $B^+ \to \mu^+\mu^-\mu^+\nu$ signal and background proxies, as they share similar topology with respect to isolation propreties.

The *isolation variables* for include track p_T, the opening angle between track's momentum and momentum of the combined signal/background visible system, the track χ^2/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 4.5 for both types. Backrounds 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 variables for the for other BDTs.

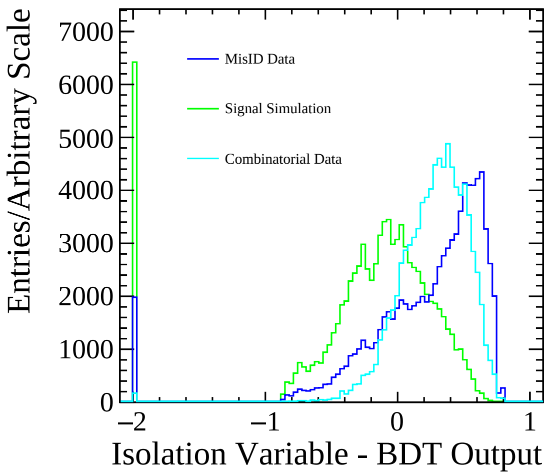


Figure 4.5: Isolation score for signal and backgrounds using 2016 samples. If isolation fails to find any other track in the event, by default it gives value -2.

4.8.2 The Combinatorial Boosted Decision Tree

One of the most prominent background is combinatorial background and to reduce its contamination while keeping signal as high as possible combinatorial BDT is trained. To obtain combinatorial BDT discriminant, a simulated sample for signal and upper mass sideband data ($M_{B_{corr}} > 5.5$ GeV/c^2) for background is used as training and testing samples. These samples have passed through all previous pre-selection, trigger and selection and are using variables mentioned in Table 4.6. As the branching fraction and hence the number of signal events is unknown, the Punzi `FOM`, defined in Equation 4.7, is used to find optimal working point. 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 4.6.

The `FOM` is computed in mass region, $4.5\,\text{GeV}/c^2 < M_{B_{corr}} < 5.5$ GeV/c^2, which also known as blinded region. To estimate the number of background candidates in blinded region, final fit strategy described in $\boxed{\text{add reference to fit}}$ 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

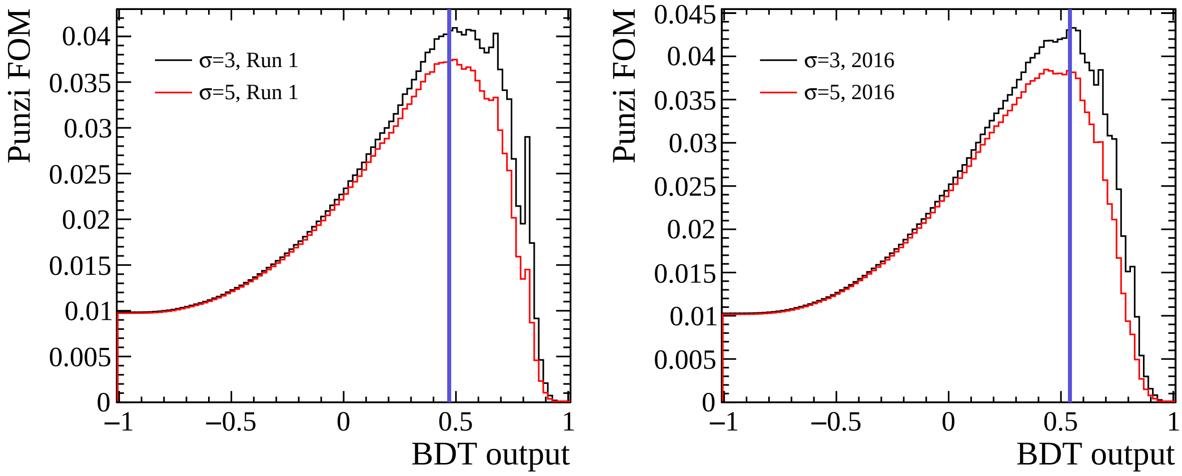


Figure 4.6: Punzi FOM have the optimum working point at 0.47 for Run I and 0.54 for Run II as seen in both figures with violet line for $\sigma = 3$ and $\sigma = 5$.

P.D.F in the blinded region.

In order to accomodate 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 trainings can be seen in Figure 4.7. Different properties (such as number of trees) and variables (2-particle vertex) have been added to see whether improvement in discrimination is possible but here are the results that prove to be most optimal.

In both Combinatorial BDTs the most discriminating variables are the isolation variable subsection 4.8.1, B^+ vertex χ^2/ndof, min IPχ^2 of the muons and p_T of B^+. Combinatorial muon comes more from somewhere else in the event and hence its min IPχ^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. It also tends to have higher momenum and hence p_T of B^+ is higher. Distributions for these different variables can be seen in Figure 4.8.

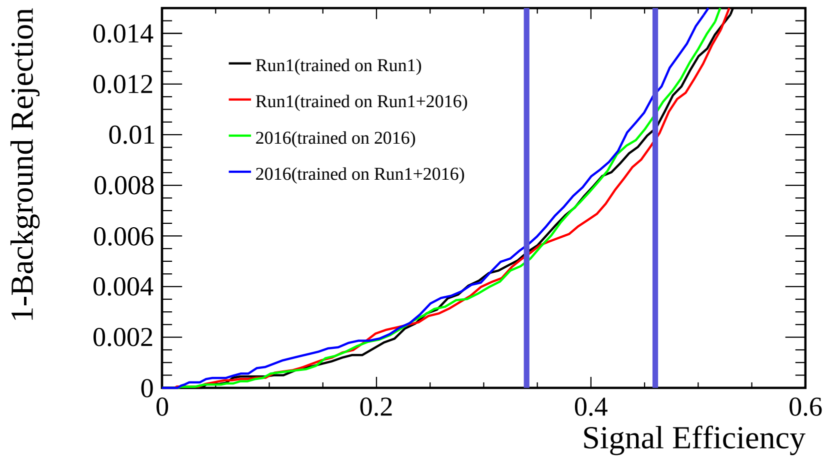


Figure 4.7: Comparison of separate and combined training samples and performance on different datasets. Two vertical violet lines represent optimal cut points at signal efficiency, for Run I (0.47) and for 2016 (0.34) where the working point of the two BDTs are chosen. Separate training provide greater rejection power in 2016. In Run I training on both datasets provide comparable performance for given optimal signal efficiency. Taking into the account the fact that offline selection slightly differs for 2016, it is advantagous to keep training separately.

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 this value stays constant then the background will not be skewed, which is the case for Run II as seen Figure 4.9. This is also the case in Run I.

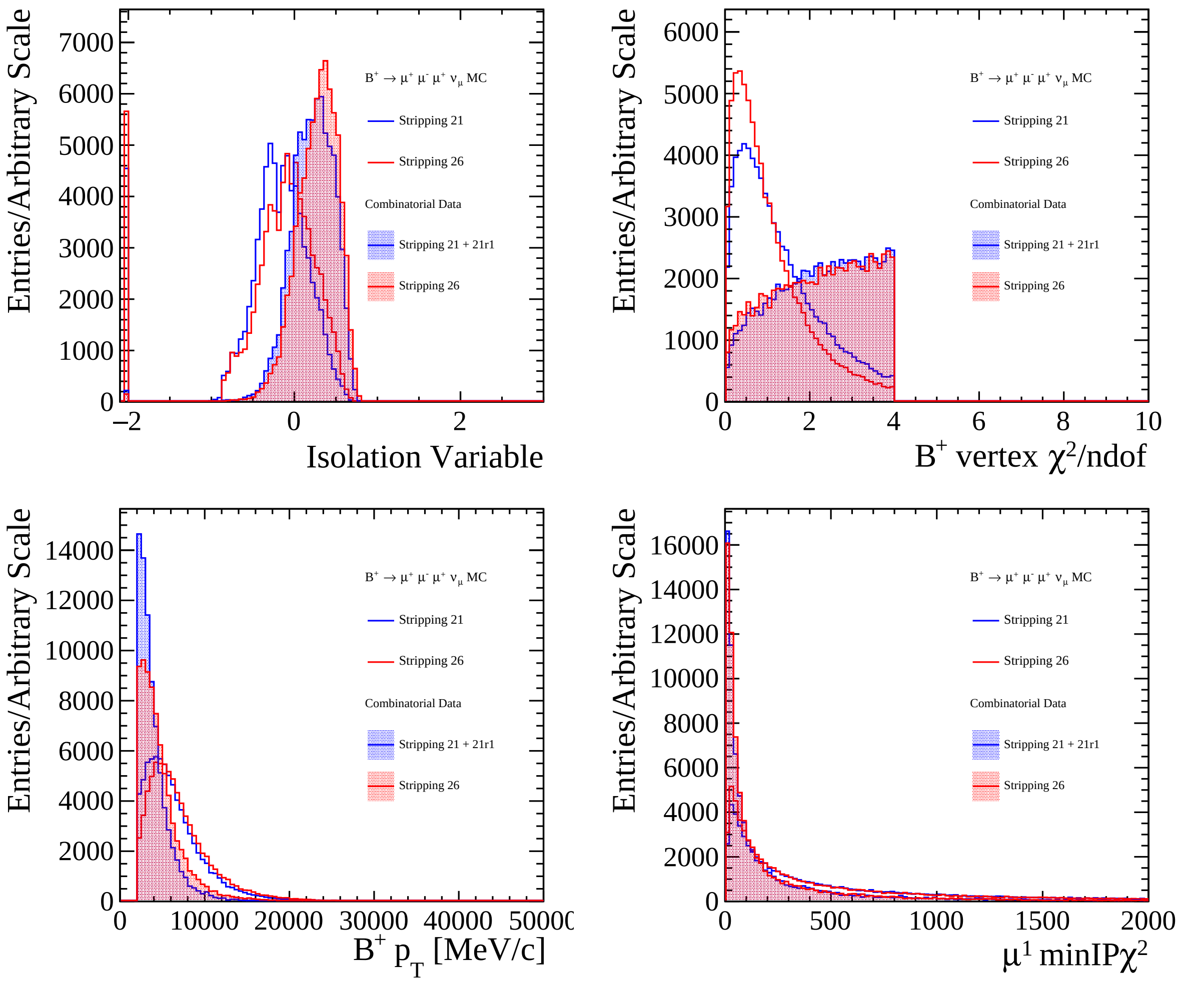


Figure 4.8: The variable with the most discriminative power for both Run I and 2016.

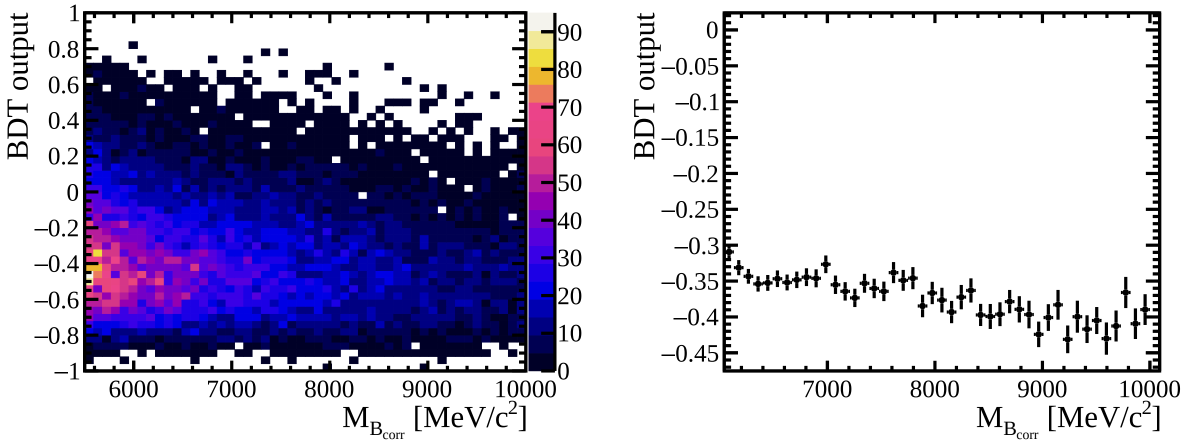


Figure 4.9: Study of linear correlation between BDT output and $M_{B_{corr}}$ and BDT value for each bin of $M_{B_{corr}}$ in 2016 shows that Combinatorial BDT is relatively flat as a function $M_{B_{corr}}$.

4.9 Misid Boosted Decision Tree

In the same way, boosted decision tree was used to yield a classifier that distinguishes well between signal and misid background. The misid sample, that is used for training and testing, was obtained the same way as signal but with one of the muons not identified as 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 Figure 4.10. Optimisation metric for this classifier was again was Punzi `FOM` in a signal region and hence number of background events need to be obtained. *k-folding* technique was applied again to prevent overtraining.

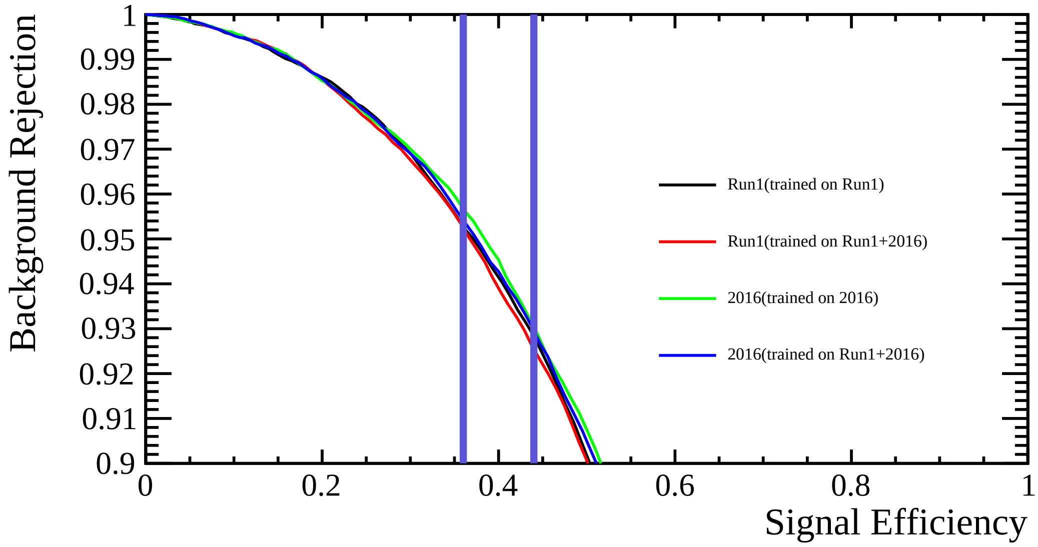


Figure 4.10: Comparison of separate and combined training samples and performance on different datasets. Optimal working point is chosen, see Figure 4.14 and its corresponding signal efficiency in Run I is 0.44 and for 2016 0.37 denoted with violet line. As the performance is better for 2016 when the training is performed separately, the training is kept separately also to be consistent with previous methodology.

To obtain number of background events, the default fitting strategy for misID 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 is used as these samples are weighted and need to be directly used for the blinded data fit, which requires correct knowledge of the correlations of the fit results. The results of the binned χ^2 fits to the misID templates are summarized in Figure 4.11 yielding in Run I 2400 unparametrized candidates for misid events are expected to be polluting the signal window in the prescaled sample, and in Run II 2200 candidates are obtained.

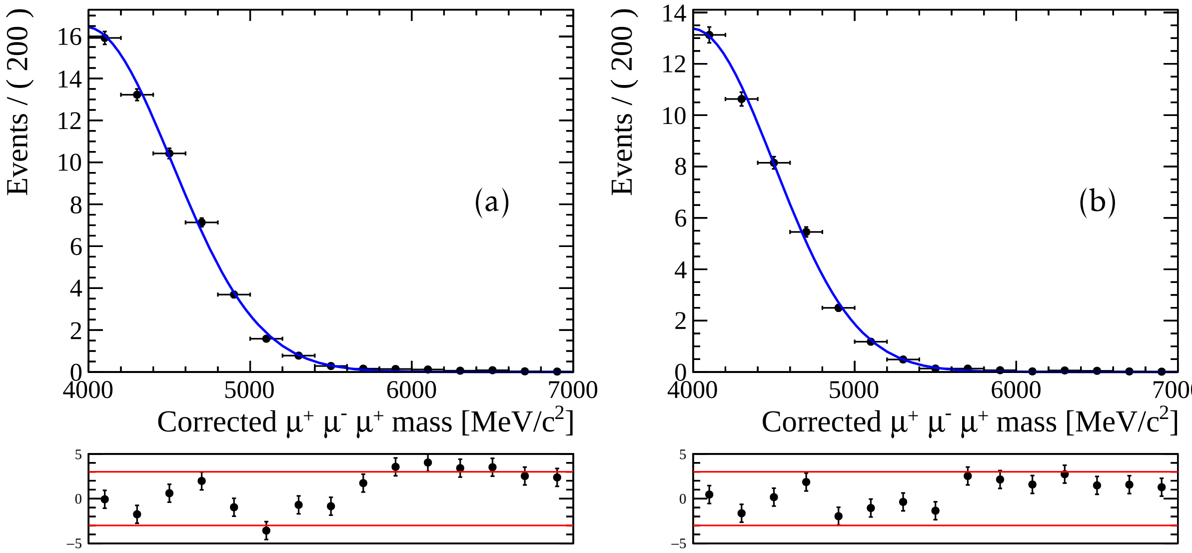


Figure 4.11: (a) Run1 (b) 2016 binned χ^2 fit to misID sample yielding estimates for the number of background events.

Misid background can proceed also through combination of random muon and hence by applying the combinatorial BDTs on the misid samples, this component is reduced and misid sample that is left should consist of true cascade decays, as seen in Figure 4.12,

The performance of the two misID BDTs classifier are shown in . The most powerful variables that distinguish the signal from misid background are the kinematic properties of the misidentified muon, namely p_T, p and min IPχ^2. Misidentified muons tend to be softer than in the signal case as they come from cascades via D^0 and its excited states.

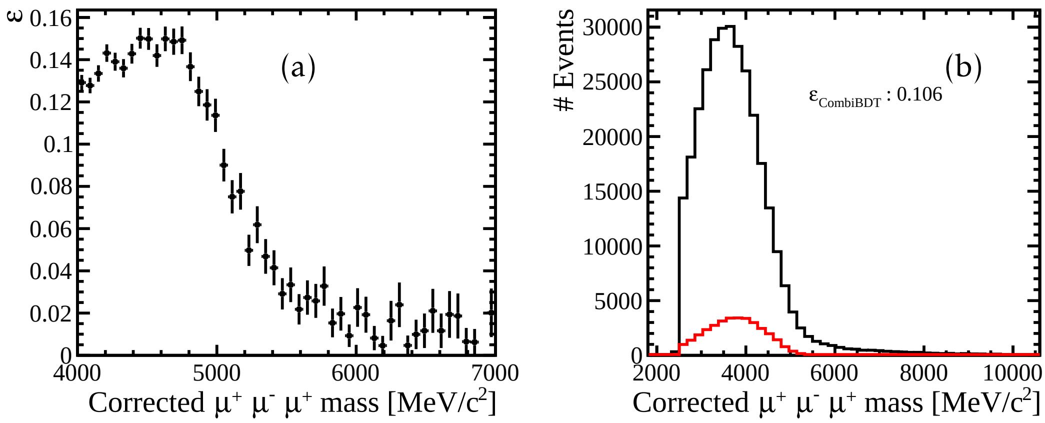


Figure 4.12: (a) The efficiency of applying 2016 combinatorial BDT at optimal working point on the 2016 SS misID sample. It can be seen that (b) combinatorial component of the misID sample has been significantly reduced, where red curve is the distribution after applying the cut.

The min IPχ^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 4.13.

The Punzi FOM for Run I and Run II as a function of BDT cut can be seen in Figure 4.14 for both significances of $\sigma = \{3, 5\}$.

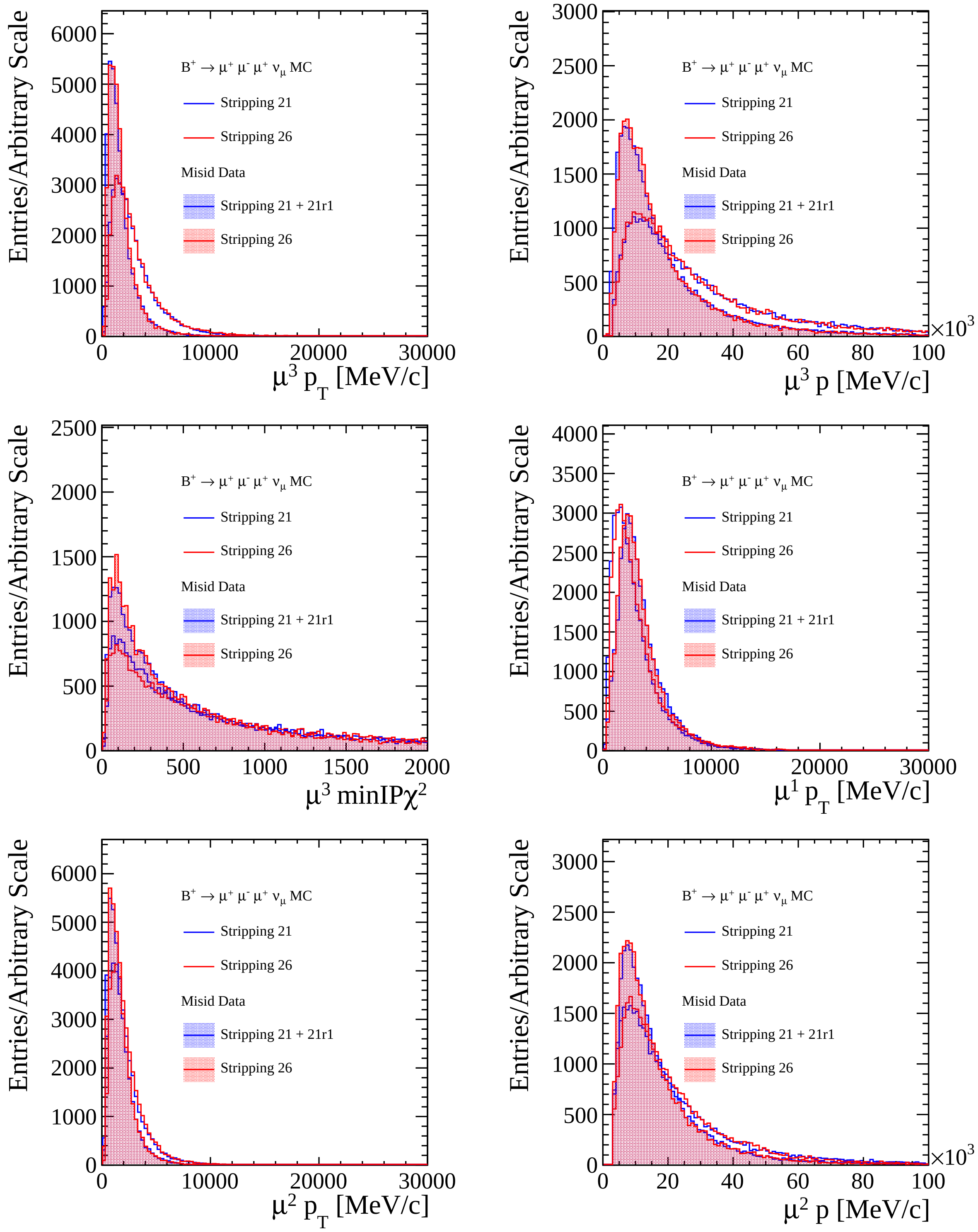


Figure 4.13: The variables with the most discriminative power for both misid Run I and II, mostly kinematic propeties of different muons.

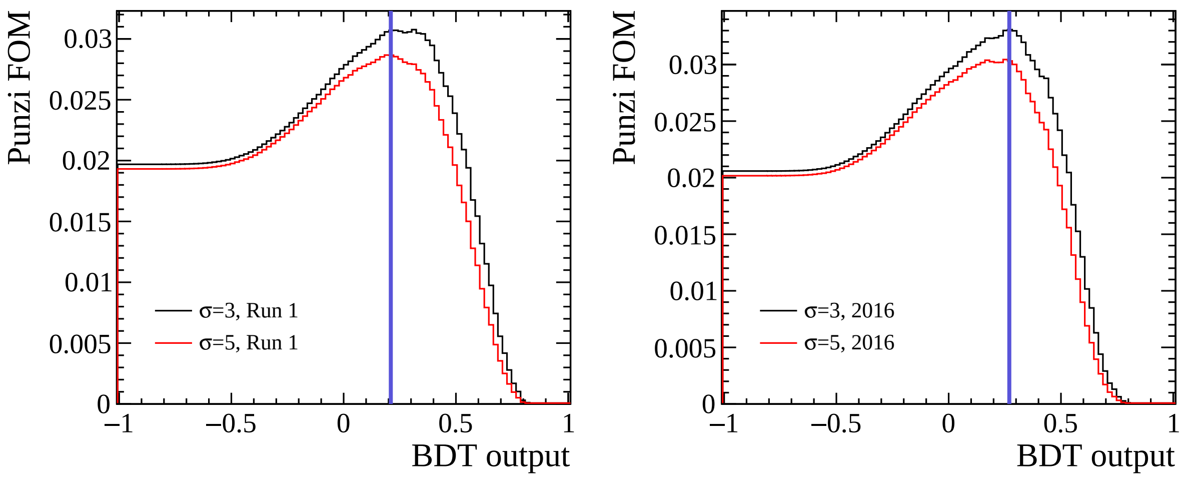


Figure 4.14: Punzi FOM have the optimum working point at 0.21 for Run I and 0.27 for Run II as seen in both figures with violet line for $\sigma = 3$ and $\sigma = 5$.

Chapter 5

Results

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] G. Raven, *LHCb: Status and prospects*, Nucl. Phys. Proc. Suppl. **163** (2007) 153, [,153(2007)].

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

[6] 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.

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

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

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

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

[11] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) 09007, arXiv:1405.7808.

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

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

[14] 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.

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

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

[17] 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.

[18] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, arXiv:1211.1346.

[19] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[20] 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.

[21] *Images of different trigger schemes taken from the LHCb speaker's bureau web page*, Accessed in 2018.

[22] 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.

[23] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[24] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[25] 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.

[26] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[27] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[28] LHCb, M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[29] D. MartÃŋnez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. **A764** (2014) 150, arXiv:1312.5000.

[30] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, PhD thesis, Cracow, INP, 1986.

[31] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A Comparative study of the decays $B \to (K, K^{*})\ell^+\ell^-$ in standard model and supersymmetric theories*, Phys. Rev. **D61** (2000) 074024, arXiv:hep-ph/9910221.

[32] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[33] 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.

[34] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, arXiv:physics/0308063.

[35] P. Speckmayer, A. Hocker, J. Stelzer, and H. Voss, *The toolkit for multivariate data analysis, TMVA 4*, J. Phys. Conf. Ser. **219** (2010) 032057.

[36] LHCb, R. Aaij *et al.*, *Determination of the quark coupling strength* $|V_{ub}|$ *using baryonic decays*, Nature Phys. **11** (2015) 743, arXiv:1504.01568.

Appendices

Appendix A

Boosted Decision Trees

Many rare decay analyses make extensive use of BDTs and they are important in the $\Lambda_b^0 \to p\pi^-\mu^+\mu^-$ 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

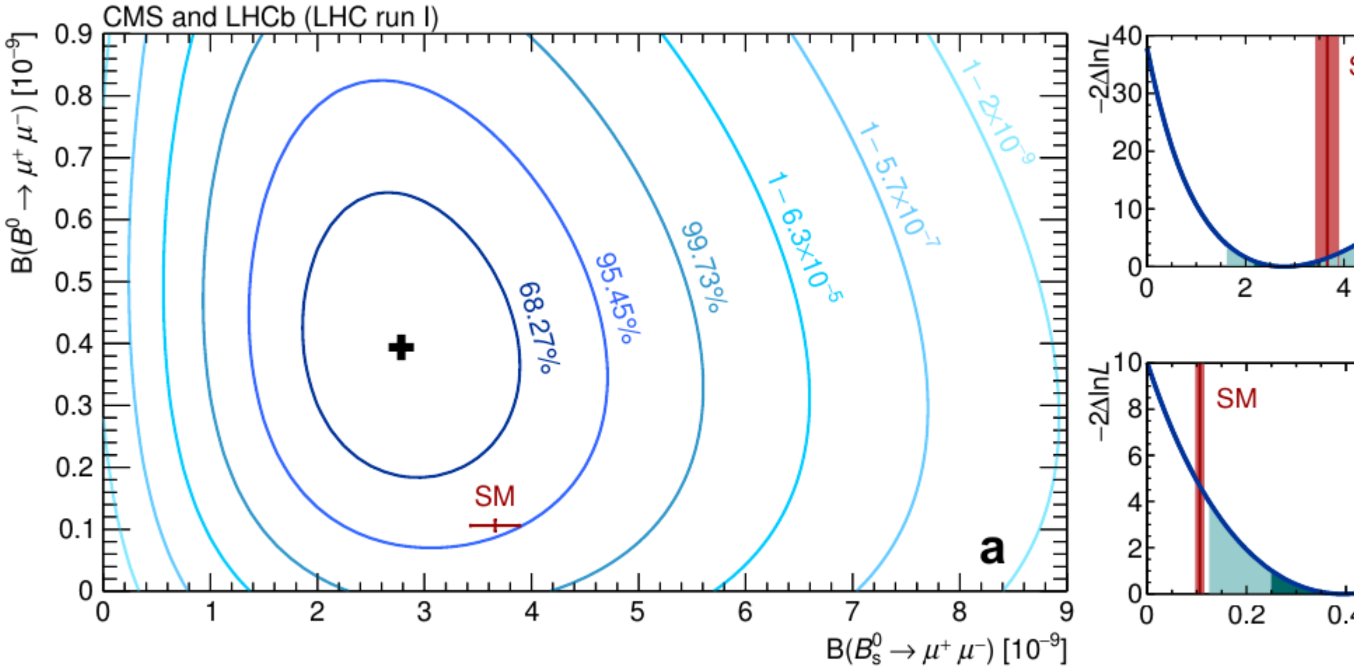


Figure A.1: An example decision tree. The S and B stand for 'Signal-like' and 'Background-like'. The β_i variables refer to the cut values chosen by the machine learning algorithm after the tree has been trained on signal and background samples. The blue ovals represent final nodes called leafs, which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

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

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.

Appendix C

Fitting functions

C.1 Crystal Ball function

Crystal Ball (CB) function [30] 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

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

where A, B and N are all constants that depend on $\alpha, n, \overline{\mu}, \sigma$ ensuring correct normalisation and continuity of the first derivative. Thus, if α is positive, the tail, $A \cdot \left(B - \frac{(x-\overline{\mu})}{2\sigma}\right)^{-n}$, 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 $f(x;\alpha,n,\overline{\mu},\sigma)$:

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

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, [29].

C.3 `Rookeys` function from `ROOFIT` 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.