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

Abstract

This thesis presents a first search for the fully leptonic decay $B^+ \to \mu^+\mu^-\mu^+\nu$ in any experiment. This search is performed using proton-proton collision data at LHCb corresponding to an integrated luminosity of 4.7 fb^{-1}. The search is carried out in the region where the minimum of the two $\mu^+\mu^-$ mass combinations is below $980\,\mathrm{MeV}/c^2$. The measurement of the branching fraction of this decay is even more interesting given that the recent theoretical prediction [1] of the branching fraction for $B^+ \to \mu^+\mu^-\mu^+\nu$ of 1.3×10^{-7} is high. Moreover, this decay is sensitive to the magnitude of the coupling strength between b and u quarks, which is of great interest given that there are some tensions in measurements of this magnitude.

The data are consistent with the background only hypothesis and a limit of 1.4×10^{-8} at 95% confidence level is set on the branching fraction in the stated kinematic region. This is therefore not consistent with the theoretical prediction made in Ref. [1].

This thesis also presents a study of the response of the detector if three muons pass through it. This study shows that correlations induced by a trimuon system in the detector are substantial and they need to be addressed properly.

Contents

Declaration of originality

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

This thesis has not been submitted for any other qualification.

Slavomira Stefkova, August 17, 2018

Copyright Declaration

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

Acknowledgements

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

To you Patrick Owen, I owe a big thank you for showing me all the support throughout my entire analysis. Without your great insight and motivation that you have kept even when you changed institute, it would have been completely different journey. Here, I would also like to thank Semileptonic Working Group team for being stimulating environment where the problems get solved.

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

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

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

To Thibaud, for everything. You know that for me you are the most special and unique person.

Lastly, I would like to thank my family. I have never heard from you that something cannot be done. You have taught me to be curious, to be independent, to be strong. We have sacrificed the most precious commodity, time spent together, in order for me to follow my interests. And below is just one example that it was well worth it.

List of Figures

12

15

List of Tables

List of abbreviations and definitions

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

IPχ^2 The IPχ^2 is the difference in the χ^2 of the fit to the primary vertex, when the track whose IPχ^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.

CKM Cabibbo-Kobayashi-Maskawa mixing matrix.

CMS Compact Muon Solenoid.

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 a 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 Čerenkov 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.

M1-M5 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 the calorimeters.

PID Particle IDentification.

PRS Pre-shower.

PS Proton Synchotron.

PSB Proton Synchotron Booster.

PV Primary Vertex, the *pp* interaction vertex.

QCD Quantum Chromodynamics.

QED Quantum Electrodynamics.

QFT Quantum Field Theory.

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

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

RICH2 Ring Imaging Čerenkov detector providing high momentum PID by using Čerenkov 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 per degree of freedom is the minimal difference in fit χ^2 (quality of the fit) to the primary vertex between fit with this track added and removed.

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

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

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

Chapter 1

Introduction

The field of particle physics aims to describe the universe we see today by decomposing everything into fundamental building blocks, which then exhibit certain behaviour according to a given set of rules. So far, the best theoretical formulation that describes the universe around us in form of these building blocks, the Standard Model (SM), was conceived last century. Some achievements of the SM do really leave us breathless, with agreement between theoretical and experimental results of ten parts in a billion.

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

This thesis describes a search for a decay which can help to shine light on some of these parameters and is organised as follows. In chapter 2 the SM of particle physics is discussed together with the theoretical and experimental motivation for fully leptonic decays, especially for the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay. In chapter 3 the tool to search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decays, the LHCb detector, is detailed. Discussion about how a trimuon

signature behaves in the detector is covered in chapter 4. The analysis of $B^+ \to \mu^+ \mu^- \mu^+ \nu$, the central theme for the thesis, is then described in three chapters: chapter 5, where the selection for signal and normalisation are given; chapter 6, where backgrounds to $B^+ \to \mu^+ \mu^- \mu^+ \nu$ are considered; and finally chapter 7, where the efficiencies and mass fits are discussed. The result, along with its implications, will then close this thesis in chapter 8.

Chapter 2

Theory

The Standard Model is without question the most powerful and tested theory of particle physics. It describes and predicts many phenomena very well even though as discussed in the previous chapter it fails to address certain known issues. In this chapter, the theoretical basis of the Standard Model is first laid out and then followed by the experimental and theoretical considerations of fully leptonic decays. The introduction in this chapter is based on Refs. [2], [3] and [4].

2.1   Review of the Standard Model

The SM of particle physics [5–12] is currently the most accurate model describing the buildings blocks of matter, particles, and their interactions via forces. In particular the SM describes all the fundamental forces but gravity. It is a quantum field theory (QFT) whereby the dynamics of the system is captured by the most general renormalisable Lagrangian density that is invariant under gauge symmetry. QFT considers particles to be excited states of an underlying field, also known as quanta. In the SM, particles and forces are the results of interactions between scalar, vector and spinor fields. In general there are two sets of particles. The first set are force-carrying particles also known as bosons, which have integer spin and are quanta of the scalar and vector fields. More specifically, there is the Higgs boson, the only elementary scalar boson in the SM, and

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

Table 2.1: Quarks and their properties such as flavour, charge and mass. Flavour is a property which distinguishes different species of quarks. Another property is the mass of the quark. The masses are taken from [13].

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

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

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

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

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

2.2 The Principle of Standard Model Building

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

$$\mathcal{L} = i\overline{\psi}\gamma^\mu\partial_\mu\psi - m\overline{\psi}\psi, \tag{2.1}$$

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

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

for $\overline{\psi}$ in Equation 2.2 the equation

$$i\gamma^\mu\partial_\mu\psi - m\psi = 0 \tag{2.3}$$

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

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

$$\mathcal{L} \to \mathcal{L} - (\partial_\mu\phi)\overline{\psi}\gamma^\mu\psi. \tag{2.4}$$

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

$$\mathcal{L} = i\overline{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\overline{\psi}\psi - q\overline{\psi}\gamma^{\mu}\psi A_{\mu} \tag{2.5}$$

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

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

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

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

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

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

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

$$\mathcal{D}_{\mu} = \partial_{\mu} + iqA_{\mu}, \tag{2.8}$$

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

2.3   Quantum Chromodynamics

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

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

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

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

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

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

Another interesting behaviour of the strong interaction is that the quarks are not observed in isolation. This is due *confinement* which can be understood within the framework of QFT theory by observing evolution of the coupling strength g as a function of energy scale, also known as β function. The β-function for a coupling

constant g in the SM takes the following form:

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

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

The full Lagrangian density for the strong interaction is

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

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

2.4 Electroweak Unification

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

$$\psi = \psi_L + \psi_R = P_L\psi + P_R\psi, \tag{2.13}$$

where $P_L = \frac{1-\gamma^5}{2}$ and $P_R = \frac{1+\gamma^5}{2}$ are known as the projection operators. By calling these operators left-handed and right-handed, there is a misconception that ψ_L is a helicity eigenstate, but this is only true given the particle in question is massless. These spinors

are known to have chirality - known as left or right-handedness. Helicity is rather the projection of the spin on the direction of the momentum.

The spinor field decomposition has an impact on electromagnetic currents, weak currents as well as the fermion mass terms. Firstly, the fermion mass term mixes both left handed and right handed spinors as

$$m\overline{\psi}\psi = m(\overline{\psi_R}\psi_L + \overline{\psi_L}\psi_R). \tag{2.14}$$

Secondly the electromagnetic current does not mix the left and right handed components, since $\overline{\psi}\gamma^\mu\psi = \overline{\psi_R}\gamma^\mu\psi_R + \overline{\psi_L}\gamma^\mu\psi_L$. Finally the weak charged current only acts on left-handed fermions as $\frac{1}{2}\overline{\psi}\gamma^\mu(1-\gamma_5)\psi = \overline{\psi_L}\gamma^\mu\gamma_L$.

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

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

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

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

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

$$Q = I_3 + \frac{1}{2}Y, \tag{2.17}$$

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

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

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

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

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

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

$$W_\mu \equiv W_\mu^-, \tag{2.19}$$

$$W_\mu^\dagger \equiv W_\mu^+.$$

The neutral bosons are obtained using W_μ^3 and B_μ in a similar fashion as

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \tag{2.20}$$

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W, \tag{2.21}$$

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

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

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

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

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

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

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

The first part of $\mathcal{L}_{NC}$ can be recognized as the electromagnetic interaction realising that $e = g\sin\theta_W$. New couplings of the Z to fermions can be seen where $g_V^i = I_3^i - 2Q_i\sin^2\theta_W$ and $g_A^i = I_3^i$.

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

2.5   The Higgs Mechanism

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

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{2.25}$$

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

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

where D_μ is given in Equation 2.18, and V is the famous Mexican hat potential where the x-axis is ϕ_1 and the y-axis is ϕ_2. It is required that $\lambda > 0$ in order for it to be a ground state.

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

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

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

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

2.6   Fermion Mass Generation

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

$$\mathcal{L}_Y = \mathcal{L}_L + \mathcal{L}_Q. \tag{2.29}$$

The leptonic term for one family of leptons using the definitions in Equation 2.15 and Equation 2.16 is

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

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

$$\mathcal{L}_Q = y^u_{ij}\overline{Q}^i_L \phi u^j_R + y^d_{ij}\overline{Q}^i_L \phi_c d^j_R + h.c., \tag{2.31}$$

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

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

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

$$\mathcal{L}_Q = \frac{v}{\sqrt{2}}(y^d_{ij}\overline{u}^i_L u^j_R + y^d_{ij}\overline{d}^i_L d^j_R + h.c)(1 + \frac{H}{v}). \tag{2.33}$$

where the quark masses are grouped into 3×3 complex matrices of up-type quark (down-type quark) $M^u_{ij} = \frac{v}{\sqrt{2}}y^u_{ij}$ ($M^d_{ij} = \frac{v}{\sqrt{2}}y^d_{ij}$). In conclusion, before the spontaneous breakdown of the electroweak symmetry, all quarks and leptons were massless. Once the Higgs scalar field acquires a vacuum expectation value implying a broken symmetry, quarks and leptons acquire mass.

The mass matrices can be diagonalised by unitary transformations $U_{\{uL, uR\}}$ and $U_{\{dL, dR\}}$ in the following way:

$$\begin{aligned} \mathcal{M}_u &= U^\dagger_{uL}M^u U_{uR} = Diag\{m_u, m_c, m_t\}, \\ \mathcal{M}_d &= U^\dagger_{dL}M^d U_{dR} = Diag\{m_d, m_s, m_b\}. \end{aligned} \tag{2.34}$$

This way of diagonalising mass matrices is the most general case of a weak basis transformation which transforms a system to a different basis without altering the physics. Such a transformation is equivalent to changing quark fields from the basis of flavour eigenstates to that of mass eigenstates.

This change into the mass eigenstate basis does not affect most of the Lagrangian. More specifically, there will be no change to the $\mathcal{L}_{NC}$ in Equation 2.24 when expressed in mass eigenstates (hence at tree-level there are no flavour changing neutral-currents in the SM), however, the charged current $\mathcal{L}_{CC}$ in Equation 2.35 is affected. Due to the diagonalisation of the mass matrices, $\mathcal{L}_{CC}$ now includes non diagonal couplings for the current as seen in the $\mathcal{L}_{CC}$ for all three fermion generations:

$$\mathcal{L}_{CC} = -\frac{g}{2\sqrt{2}}\left[W_\mu^\dagger\left[\sum_l \overline{\nu}\gamma^\mu(1-\gamma_5)l + \sum_{ij}\overline{u_i}\gamma^\mu(1-\gamma_5)V_{ij}d_j\right] + h.c\right]. \tag{2.35}$$

In this equation there is a new term $V_{ij} = V_{CKM} = U_{uL}U_{dL}^\dagger$ which is the Cabibbo-Kobayashi-Maskawa (CKM) [14] [15] mixing matrix. From Equation 2.34 it follows that $V_{CKM}^\dagger V_{CKM} = 1$, or that CKM mixing matrix is unitary by assuming that only the charged current via W will lead to a transition from the up-quark to down-type quark sector. Therefore the CKM matrix elements provide the probabilities of how a $W^\pm$ bosons decay.

2.7 The Quark Mixing Matrix

As mentioned above, from the transformation of the mass matrix using two unitary matrices one obtains the CKM matrix which exhibits a strong hierarchy in the size of the matrix elements. From the previous discussion the quark mixing matrix is a 3×3 complex unitary matrix yielding 18 parameters to start with. Unitarity of the CKM matrix implies that matrix elements are orthonormal, reducing the count of free parameters to 9. Further, 5 out of 6 quark phases can be absorbed into the redefinition of the quark field, cutting the number of parameters down to 4 parameters, three quark mixing angles and one CP (charge-parity) violating phase. There are many different parametrisations of the CKM matrix, but the standard parametrisation of the CKM matrix [16] for flavour mixing is

$$
V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.36}
$$

$$
= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\exp(-i\delta) \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\exp(i\delta) & c_{12}c_{23} - s_{12}s_{23}s_{13}\exp(i\delta) & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\exp(i\delta) & -c_{12}s_{23} - s_{12}c_{23}s_{13}\exp(i\delta) & c_{23}c_{13} \end{pmatrix}, \tag{2.37}
$$

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

A parametrisation reflecting the hierarchical nature in flavour mixing, which is an expansion in terms of the small parameter λ, was introduced by Wolfenstein [17]. The four Wolfenstein parameters are related to the standard parametrization via the following expressions:

$$
\lambda = s_{12},
$$
$$
A\lambda^2 = s_{23}, \tag{2.38}
$$
$$
A\lambda^3(\rho - i\eta) = s_{13}exp(-i\delta),
$$

$$
V_{\text{CKM}_{\text{Wolfenstein}}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\left(\lambda^4\right). \tag{2.39}
$$

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

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

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0. \tag{2.40}$$

Dividing this constraint by $V_{cd} V_{cb}^*$

$$\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*} + 1 + \frac{V_{td} V_{tb}^*}{V_{cd} V_{cb}^*} = 0. \tag{2.41}$$

Using the following relation

$$V_{ub}^* = A\lambda^3 \rho + i\eta = \frac{A\lambda^3(\bar{\rho} + i\bar{\eta})\sqrt{1 - A^2\lambda^4}}{\sqrt{1 - \lambda^2}[1 - A^2\lambda^4(\bar{\rho} + i\bar{\eta})]} \tag{2.42}$$

ensures that

$$\bar{\rho} + i\bar{\eta} = -\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*}. \tag{2.43}$$

The constraint can be pictorially represented in the $\bar{\rho}$ and $\bar{\eta}$ plane as the triangle shown in Figure 2.1. The area of the triangle is half of the Jarlskog invariant J, a quantifier
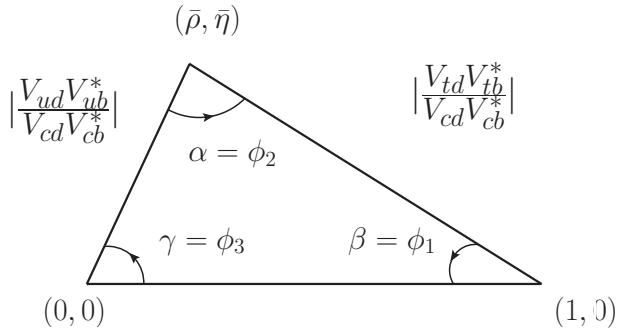
Figure 2.1: Unitarity triangle in a complex plane.

of CP violation, which is defined as $\text{Im}[V_{ij} V_{kl} V_{il}^* V_{kj}^*]$ [18]. It is interesting to notice that the SM with its parameters may or may not violate CP. Only after measuring J it is possible to determine the CP non-conservation. J vanishes only if the mixing angle $\theta_{ij} = \{0, \pi/2\}$; $\delta = \{0, \pi\}$. So measurements of J allows to verify that the CKM matrix is complex and hence different mixing for quarks and anti-quarks is obtained.

The CKM matrix elements which comprise of magnitudes and phases can be determined in different ways but the most precise option employs a global fit to all available measurements as shown in Figure 2.2. Hence, the most precise measurement of the CKM matrix magnitudes to-date [13] is

$$|V_{\text{CKM}}| = \begin{pmatrix} 0.97434^{+0.00011}_{-0.00012} & 0.22506 \pm 0.00050 & \boxed{0.00357 \pm 0.00015} \\ 0.22492 \pm 0.00050 & 0.97351 \pm 0.00013 & 0.0411 \pm 0.0013 \\ 0.00875^{0.00032}_{-0.00033} & 0.0403 \pm 0.0013 & 0.99915 \pm 0.00005 \end{pmatrix}, \qquad (2.44)$$

with non-zero Jarlskog invariant $J = (3.18 \pm 0.15) \times 10^{-5}$. Highlighted is the result for the magnitude of the V_{ub} matrix element, $|V_{ub}|$, which is the element with the highest fractional uncertainty on its value. Therefore precise measurement of this element is very important and was the original motivation for the analysis of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$. Moreover, as displayed in Figure 2.2(a)(b), the measurement of $|V_{ub}|$ (orange circle)(green circle) together with $\sin(2\beta)$ measurement (green band)(blue band) constrain the apex of the triangle. This means that these two measurements together with other measurements test the unitarity of the CKM matrix, one of the fundamental assumptions of the SM.
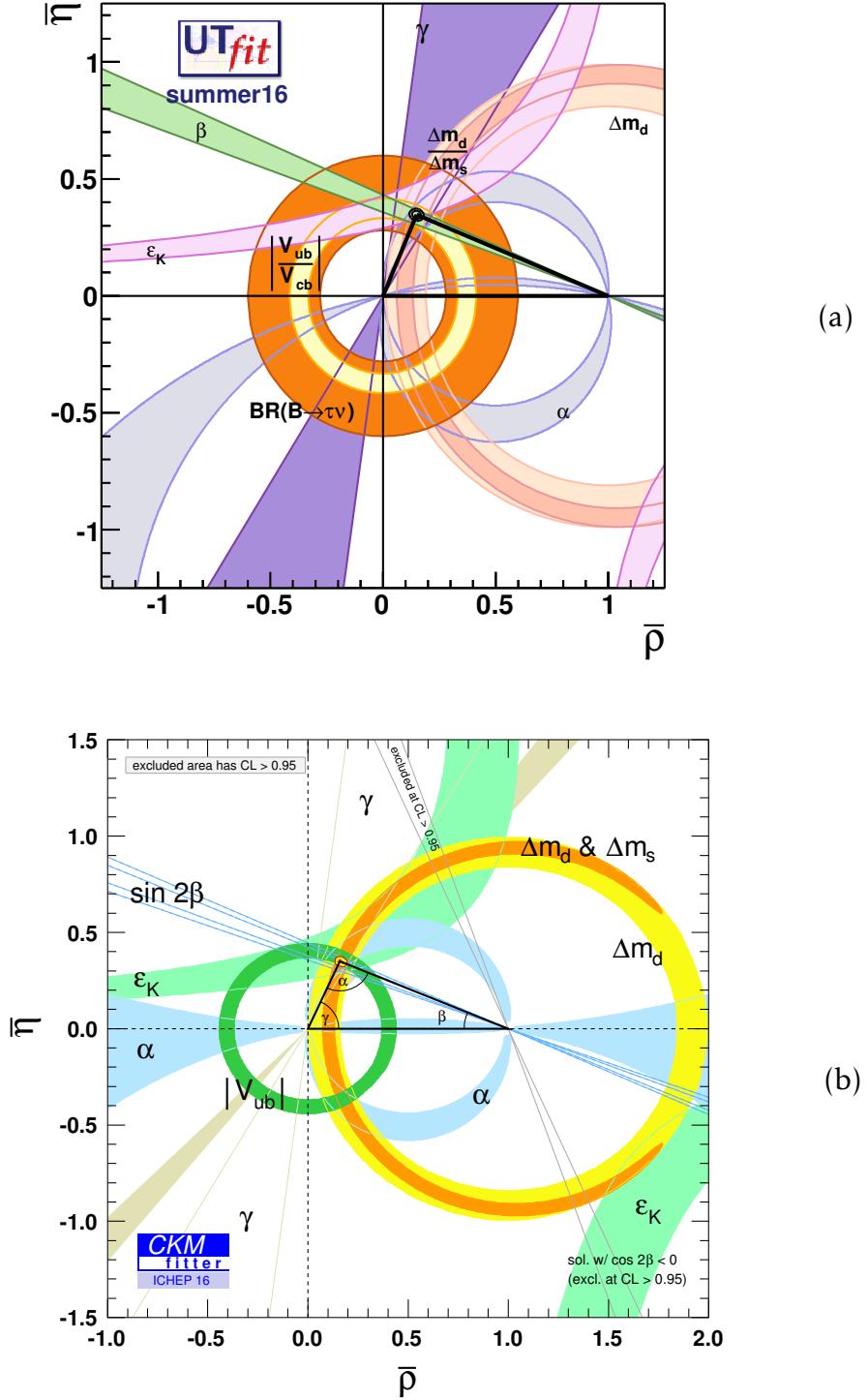
Figure 2.2: Different experimental measurements that constrain the CKM matrix elements together with the global fit results from two collaborations (a) UTFit and (b) CKMFitter as of summer 2016. These figures are taken from Refs. [19] and [20]. There is a good agreement for the results between the two different collaborations.

2.8 Fully Leptonic $P^+ \to l^+ \nu$ Decays

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

- either measurements of the CKM matrix elements,

- or measurements of leptonic decay constants,

- or measurements of new physics effects.

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

$$\Gamma(P^+ \to l^+ \nu) = \frac{G_F^2 m_{P^+} m_{l^+}^2}{8\pi} \left[1 - \frac{m_{l^+}^2}{m_{P^+}^2} \right]^2 f_P^2 |V_{q_1 q_2}|^2, \qquad (2.45)$$

where G_F is the Fermi constant, m_{P^+} and m_{l^+} are the pseudoscalar meson and lepton masses, respectively. This decay width can be compared to that of $\tau \to l \nu \bar{\nu}$ [21]

$$\Gamma(\tau \to l \nu \bar{\nu}) = \frac{G_F^2 m_\tau^5}{192\pi^3} \left[1 - f\left(\frac{m_l^2}{m_\tau^2} \right) \right], \qquad (2.46)$$

In this case $f(x) = 18x - 8x^3 + x^4 + 12x^2 \log(x)$ represents a correction due to the mass of the lepton in the final state. Corrections arising from the W propagator effects are negligible for this decay and are not considered here and nor are radiative corrections so that only the lowest order contributions are considered. As compared to Equation 2.45 the decay width is significantly higher.

So in order to measure the CKM matrix amplitude, knowledge of f_P must be inferred. f_P can be calculated using lattice QCD techniques and together with experimental determination of the decay rates provide a way to determine the amplitude squared of

the relevant CKM matrix element assuming there is no contribution from new physics. More conventionally, CKM magnitudes are determined from semileptonic decays, which are experimentally more accessible but entail larger theoretical uncertainty.

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

Last, but not least, is of course the measurement of presence of new physics in these decays. Especially appealing is the presence of new particles which would manifest themselves in the decay rates of heavier pseudoscalars ($D_{(s)}$ or B). Examples of such new particles are charged Higgs bosons, $H^\pm$, coming from so-called Type II Higgs-doublet models [22] [23] [24] or leptoquarks [24]. In this case, considering $B^+ \to l^+ \nu$ decay, the four-fermion interaction between the $W^\pm$ and $H^\pm$ bosons would modify the SM decay width Equation 2.45 to

$$\Gamma(B^+ \to l^+ \nu) = \frac{G_F^2 m_{B^+} m_{l^+}^2}{8\pi} \left[1 - \frac{m_{l^+}^2}{m_{B^+}^2}\right]^2 f_P^2 |V_{ub}|^2 \times r_H, \qquad (2.47)$$

where

$$r_H = [1 - \tan^2 \beta (m_{B^+}^2 / m_{H^+}^2)]^2. \qquad (2.48)$$

Here $\tan \beta = \frac{v_2}{v_1}$, where v_i are the vacuum expectation values for the Higgs doublets. In order to have an enhancing effect for the rate of the $B^+ \to l^+ \nu$ decay (to have $r_H > 1$), $\tan \beta / m_{H^\pm} > 0.27 \, \text{GeV}^{-1}$. The experimental limit presents already a strong lower bound on the charged Higgs mass $m_{H^\pm} > 600 \, \text{GeV}$ [25]. This makes most of the parameter space in $\tan \beta$ and $m_{H^\pm}$ satisfy the condition of $\tan \beta / m_{H^\pm} > 0.27 \, \text{GeV}^{-1}$.

The ratio of rates between $P \to \tau \nu$, $P \to \mu \nu$ and $P \to e \nu$ decays could also be of an interest. In the ratios the decay constant f_P cancels out making such measurements a good tool for lepton universality tests.

As seen in Equation 2.45, a purely leptonic final state going through $P \to W^* \to l \nu$ is suppressed by $\frac{m_l^2}{m_P^2}$, also known as helicity suppression. This suppression occurs as a result of angular momentum conservation. In case of $B^+ \to l^+ \nu$, the B^+ is a spin-0

particle and hence its decay products should have spin 0 combined, or in other words, be anti-aligned. Neutrinos in the SM are always produced left-handed. As the spin of the antilepton and the neutrino should be anti-aligned, the antilepton also needs to be left-handed (to have negative helicity). However, the weak current only couples to right-handed antiparticles. Therefore, the antilepton has to be boosted in order to have different helicity. For massless particles such a helicity flip is not possible making this decay impossible. The lighter the lepton, the larger the velocity and hence higher boost is necessary, making decays to lighter leptons rarer even though they have bigger kinematic phase space available.

Concentrating on the decays of $B^{\pm}$ mesons, the latest experimental measurements for rates of $B^+ \rightarrow l^+ \nu$ decays have been performed by B factories, finding evidence for $B^+ \rightarrow \tau^+ \nu$ and a first sign of $B^+ \rightarrow \mu^+ \nu$ as seen in Table 2.2. These results are to be compared with the SM predictions $\mathcal{B}(B^+ \rightarrow \tau^+ \nu) = (0.82 + 0.03 - 0.02) \times 10^{-4}$ [20] and $\mathcal{B}(B^+ \rightarrow \mu^+ \nu) = (3.80 \pm 0.31) \times 10^{-7}$ [26], which are obtained by using a $|V_{ub}|$ value resulting from other measurements and lattice calculations of f_B.

With helicity suppressed rates and very limited signatures in the detector (one charged track for muons and electrons, more charged tracks for taus, but also more missing energy depending on the reconstruction channel) searching for such decays is very challenging. In order to make measurements of the same kind (CKM precision measurements, decay constants measurements, new physics searches), fully leptonic decays with photons can be considered. This is because the rates for these processes are higher due to no helicity suppression.

Process	Experiment	Tag	$\mathcal{B}$	Significance
				$[\sigma]$
$B^+ \to \tau^+ \nu$	Belle [27]	Hadronic	$(0.72^{+0.27}_{-0.25} \pm 0.11) \times 10^{-4}$	3.0
$B^+ \to \tau^+ \nu$	Belle [28]	Semileptonic	$(1.25 \pm 0.28 \pm 0.27) \times 10^{-4}$	3.8
$B^+ \to \tau^+ \nu$	Belle [28]	Average	$(0.91 \pm 0.22) \times 10^{-4}$	4.6
$B^+ \to \tau^+ \nu$	BaBar [29]	Hadronic	$(1.83^{+0.53}_{-0.49} \pm 0.24) \times 10^{-4}$	3.8
$B^+ \to \tau^+ \nu$	BaBar [30]	Semileptonic	$(1.7 \pm 0.8 \pm 0.2) \times 10^{-4}$	2.3
$B^+ \to \tau^+ \nu$	BaBar [29]	Average	$(1.79 \pm 0.48) \times 10^{-4}$	-
$B^+ \to \mu^+ \nu$	Belle [26]	Untagged	$(6.46 \pm 2.22 \pm 1.60) \times 10^{-7}$	2.4

Table 2.2: Experimental summary of searches for $B^+ \to l^+ \nu$ that is inspired from [13]. Tag Hadronic/Semileptonic/Untagged refers to different way data is selected in Belle and BaBar factories.

2.9 Fully Leptonic $B^+ \to l^+ \nu \gamma$ Decays

The helicity suppression of $B^+ \to l^+ \nu$ decays can be lifted by considering the decay with an additional photon radiated from the B^+ meson, at the cost of the electromagnetic suppression with coupling constant α_{em}. Consequently, the branching fraction for radiative decays can be comparable or even larger than the corresponding fraction for purely leptonic decays. It has been shown that $R^\mu_B = \frac{\Gamma(B \to \mu\nu\gamma)}{\Gamma(B \to \mu\nu)} \approx (1 - 20)$ making $\mathcal{B}(B \to \mu\nu\gamma) \approx (10^{-7} - 10^{-6})$ [31].

The differential decay width with $\frac{1}{m_b}$ and radiative corrections at next-to-leading logarithmic order calculated in [32] is given by

$$\frac{d\Gamma}{dE_\gamma} = \frac{\alpha_{em} G_F^2 |V_{ub}|^2}{48\pi^2} m_B^4 (1 - x_\gamma) x_\gamma^3 [F_A^2 + F_V^2], \tag{2.49}$$

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

as

$$F_V(E_\gamma) = \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + [\xi(E_\gamma) + \frac{Q_u m_B f_B}{(2E_\gamma)^2} + \frac{Q_b m_B f_B}{2E_\gamma m_b}], \tag{2.50}$$

$$F_A(E_\gamma) = \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + [\xi(E_\gamma) - \frac{Q_u m_B f_B}{(2E_\gamma)^2} - \frac{Q_b m_B f_B}{2E_\gamma m_b} + \frac{Q_l f_B}{E_\gamma}]. \tag{2.51}$$

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

The first term in Equation 2.50 and Equation 2.51 represents the leading-power contribution in the heavy-quark expansion. Note that this term is the same for the vector and axial form factor. The second terms are $\frac{1}{m_b}$ power corrections relative to the leading term. Further corrections have been discussed in [33].

A recent measurement of the radiative $B^+ \to l^+ \nu \gamma$ decay, where l^+ is either e^+ or μ^+ was performed by Belle using hadronic tagging on their full data sample [34]. The search yielded $\mathcal{B}(B^+ \to \mu^+ \nu \gamma) < 3.4 \times 10^{-6}$ and $\mathcal{B}(B^+ \to e^+ \nu \gamma) < 6.1 \times 10^{-6}$.

2.10 Fully Leptonic $B^+ \to l^+ l^- l^+ \nu$ Decays

In LHCb, the most optimal approach due to the detector capabilities is to measure this kind of decay by converting the photon into a pair of muons, see Figure 2.3(a). If the naive expectation of only taking into account photon conversion into two muons is adopted, then the expected branching fraction for this analysis is $\mathcal{B}(B^+ \to \mu^+ \mu^- \mu^+ \nu) \approx 1.0 \times 10^{-8}$. However, such an estimate is not correct because there are other contributions to the total decay rate as shown in the first theoretical prediction for $\mathcal{B}(B^+ \to \mu^+ \mu^- \mu^+ \nu)$ in [1] based on the Vector Meson Dominance (VMD) model. This theoretical prediction yields $\mathcal{B}(B^+ \to \mu^+ \mu^- \mu^+ \nu) \approx 1.3 \times 10^{-7}$.

The VMD model was formulated to describe the interaction between photons and hadrons before QCD was formulated. It is an approximative model where the photon is treated being made of a purely electromagnetic component and a vector meson

component. This idea originates in the fact that both photon and vector mesons have the same quantum numbers $J^{PC} = 1^{--}$ and if two particles have the same quantum numbers then they mix.

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

In this publication the amplitude of $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu)$ is estimated by calculating the $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-e^+\nu)$ amplitude first and then adding a negative interference term that arises due to the identical fermions in the final state doubling the number of possible diagrams. The numerical calculation yields $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-e^+\nu) \approx 1.3 \times 10^{-7}$ and $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) \approx 1.3 \times 10^{-7}$.
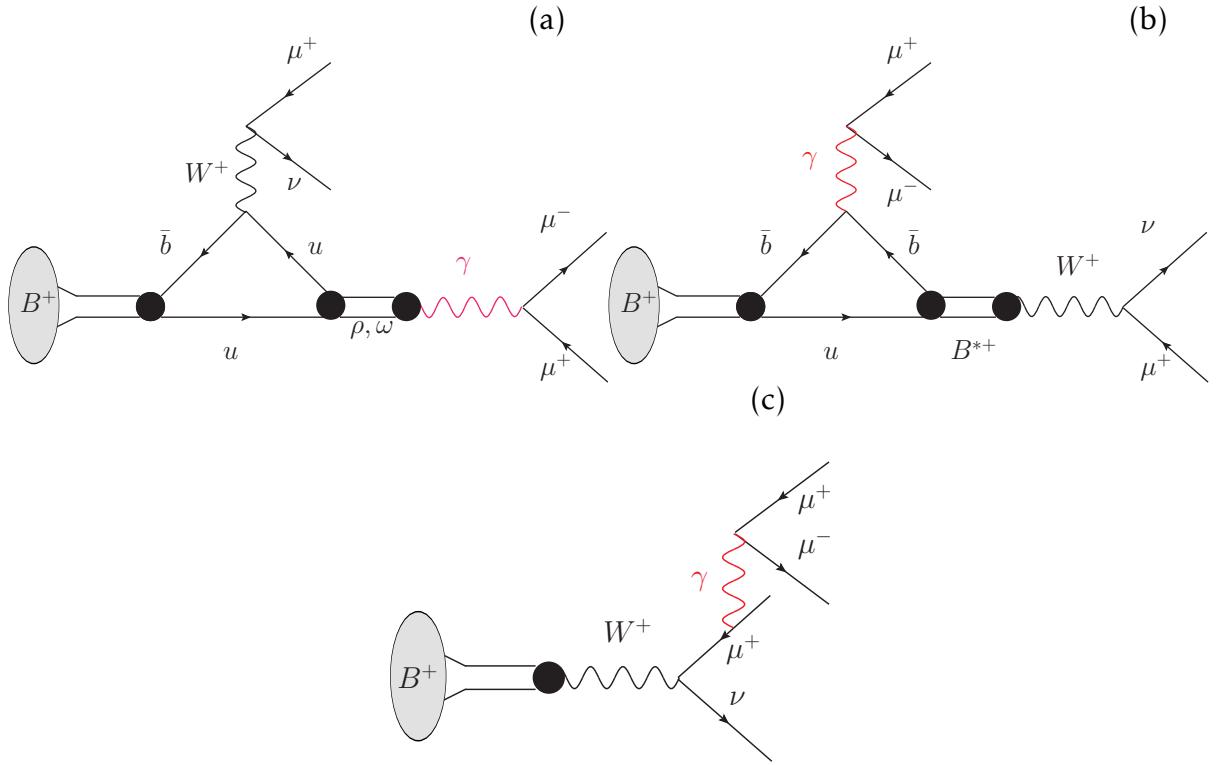
Figure 2.3: Different contributions to the $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ decay. (a) Initial u-quark state radiates off a virtual photon which decays into a pair of muons and the W^+ decays into a muon and muon neutrino. Most of the contribution to the rate comes from hadronic contribution to the photon. (b) Photon emission from b-quark and (c) finally emission from the final state muon.

2.11   The $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ Decay Model

As the search for the $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ decay is the first of its kind, a simulation that describes this type of decay was not available. There are, however, three types of decay models for $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ which were adopted and used for different purpose. More detail about their use is covered in section 5.4.

For any decay, it is possible to use a phase space model, *PHSP*, which only takes into account the kinematic constraints of the decay without taking into account any input from theoretical considerations as the matrix element is constant. This is not

satisfactory for decays where there are intermediate virtual photons or vector meson resonances.

The following decay model is developed to reflect the expected behaviour of decays shown in Figure 2.3. The decay proceeds through a virtual W decaying to $\mu^+ \nu$ and a virtual photon decaying to a muon pair. This has similar structure to the $B^+ \to (K^{*+})\mu^+\mu^-$ decay, where the K^{*+} can take the role of the virtual W decay. By using the *BTOSLLBALL* model [35], traditionally used for $B^+ \to (K^{*+})l^+l^-$ decays, but modifying the properties of the K^{*+} to those of a virtual W (having mass of 0.1 GeV/c^2 and width 50 GeV), it is possible to obtain a good approximation to the correct features of the decay. This is visible in Figure 2.4, where there is a characteristic photon pole for low $q(\mu^+, \mu^-)$, the invariant mass of the opposite muon pair, and flat distribution for $K^*(\mu^+, \nu)$, the invariant mass of the muon and neutrino pair. This decay model will be further referred to as the *INSP* model.
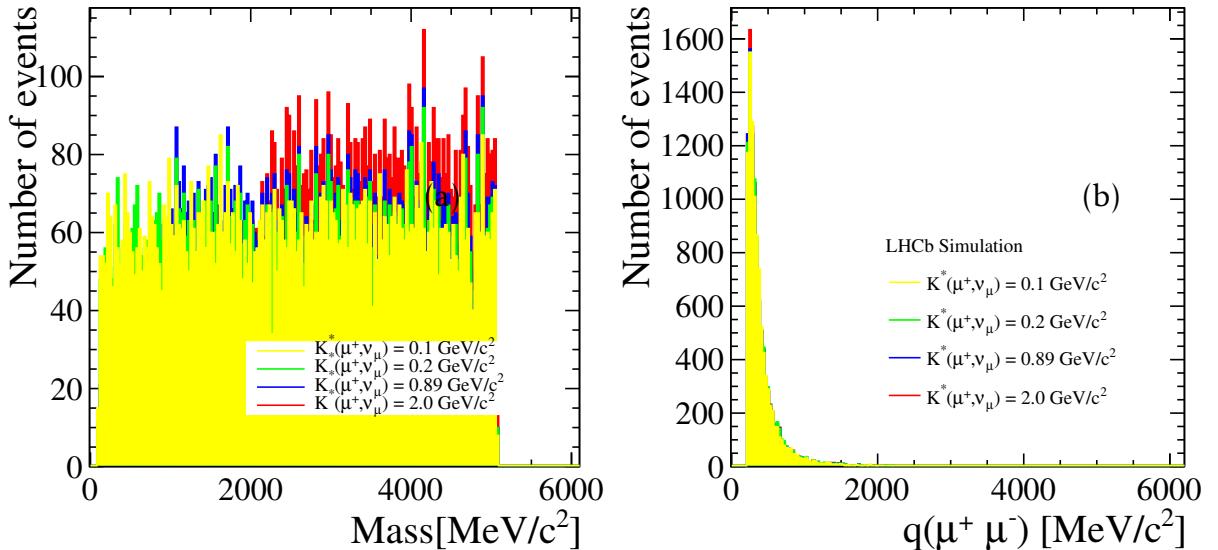
Figure 2.4: Distributions for signal simulation. (a) $K^*(\mu^+, \nu)$ (b) $q(\mu^+, \mu^-)$ distributions under different K^* mass hypotheses. The most flat distribution in $K^*(\mu^+, \nu)$ is plotted in yellow.

Finally, there is a decay model based on calculations from the VMD model, which

was written by authors of [1]. This model is denoted as *NIKI*. The distribution for $q^2(\mu^+, \mu^-)$ can be seen in Figure 2.5.
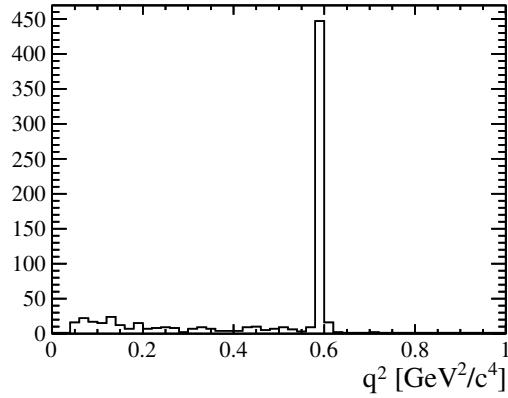
Figure 2.5: $q^2(\mu^+, \mu^-)$ distribution using the VDM model. The contribution from photon pole, ρ and ω can be seen. This Figure was produced by a collaborator.

Chapter 3

The LHCb Detector

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

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

The LHC is a complex machine which accelerates beams of protons in opposite directions in a $\sim$ 27km long circular tunnel. It is located 50-157 m below ground crossing the border between Switzerland and France. Once the desired energy is achieved proton-proton (pp) or ion collisions happen at four distinct points, where
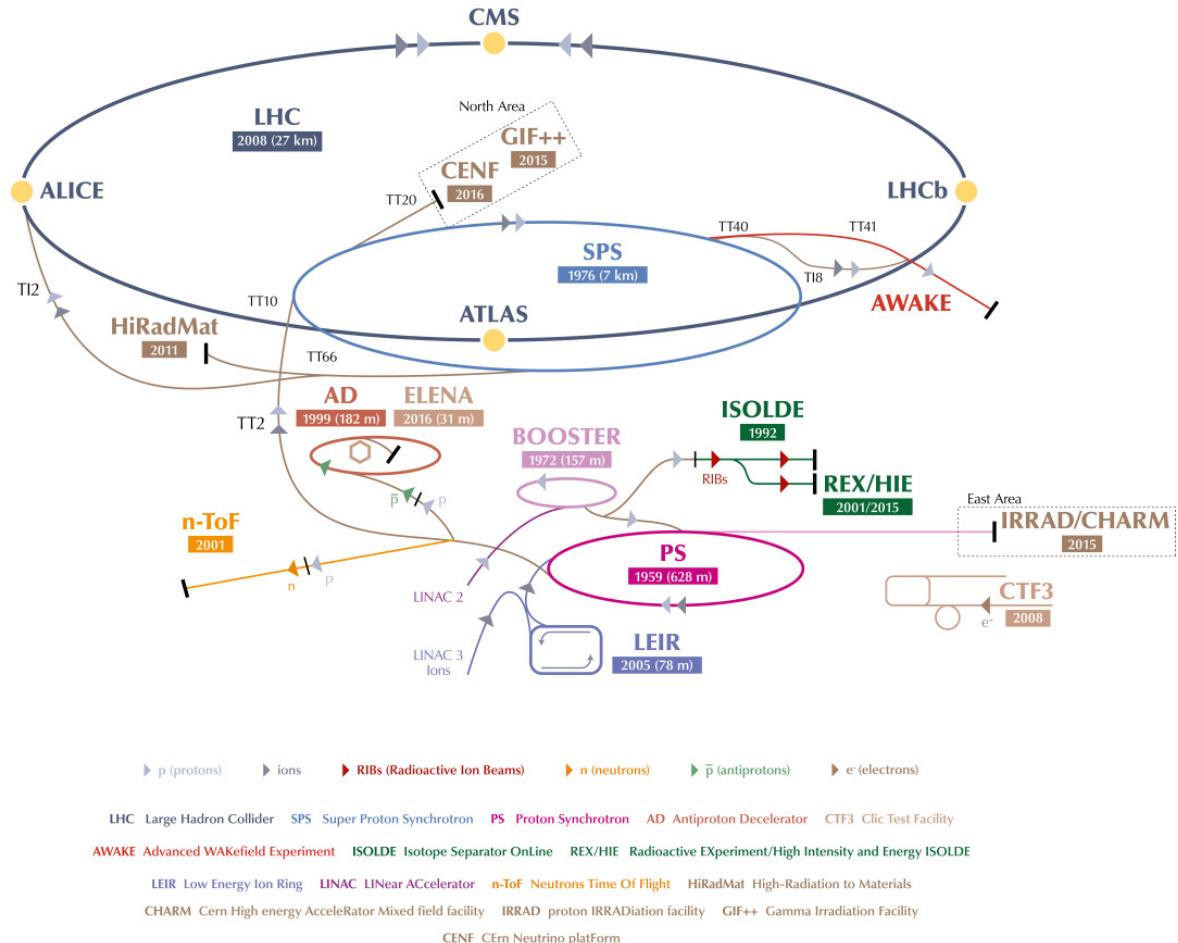
Figure 3.1: Accelerator complex at CERN. The image is taken from [36].

different detectors with different physics focus are located. These are ATLAS, CMS, ALICE and LHCb. The search for the decay $B^+ \to \mu^+ \mu^- \mu^+ \nu$ was performed using data obtained at LHCb [37].

3.1 LHCb Layout

LHCb, seen in Figure 3.2, differs from the other general purpose detectors on the LHC ring as its main aim is to study properties of heavy particles containing b or c quarks. This is possible as this experiment was designed to have a geometrical acceptance and
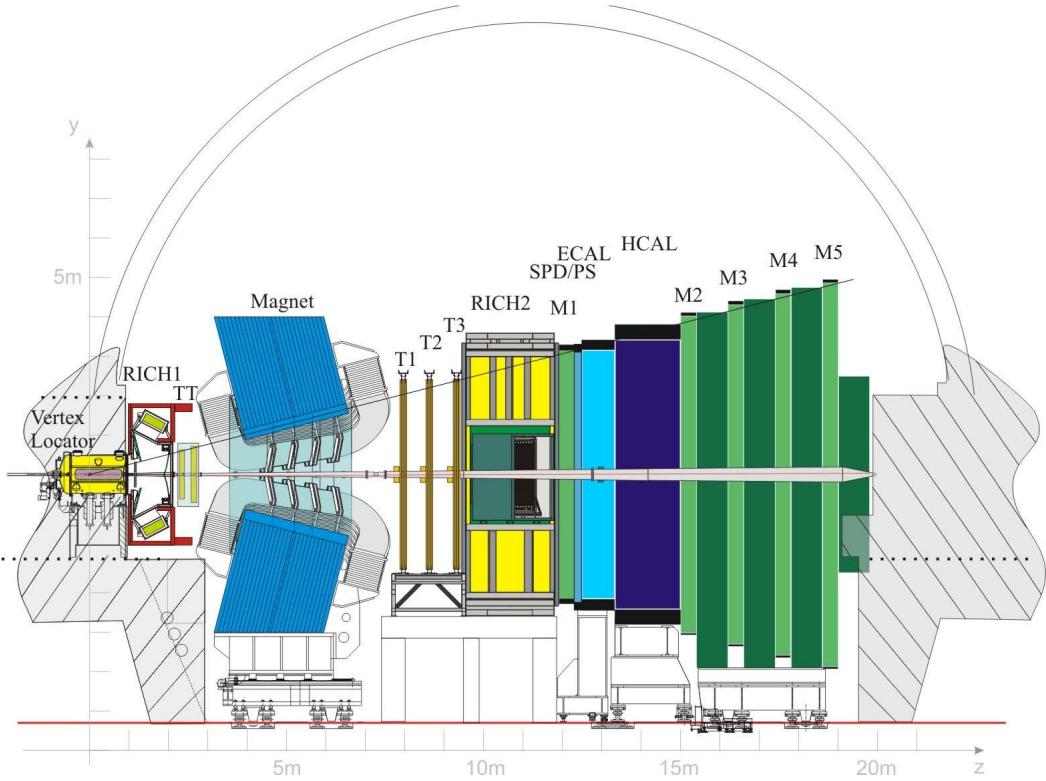
Figure 3.2: Schematic slice of LHCb detector in the y, z plane where z is defined to be the direction parallel to beamline, and x, y define the plane perpendicular to the beamline. θ, the opening polar in the y-z plane with $\theta = 0$ along the $z - axis$. Figure from [38].

unique vertex resolution, as well as excellent particle identification (PID), suitable for beautiful and charming physics.

Studies of B mesons can happen either at positron-electron colliders or at hadron colliders. The advantage of positron-electron colliders is that the information about all the event is known, as just two B mesons and nothing else is produced in the collisions. This gives an overall constraint on collision information, unlike in the hadron collider B factory, LHCb. Contrary to the two general purpose detectors at LHC, where the collisions occur 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.

The disadvantage of not having an overall constraint on collision information is,
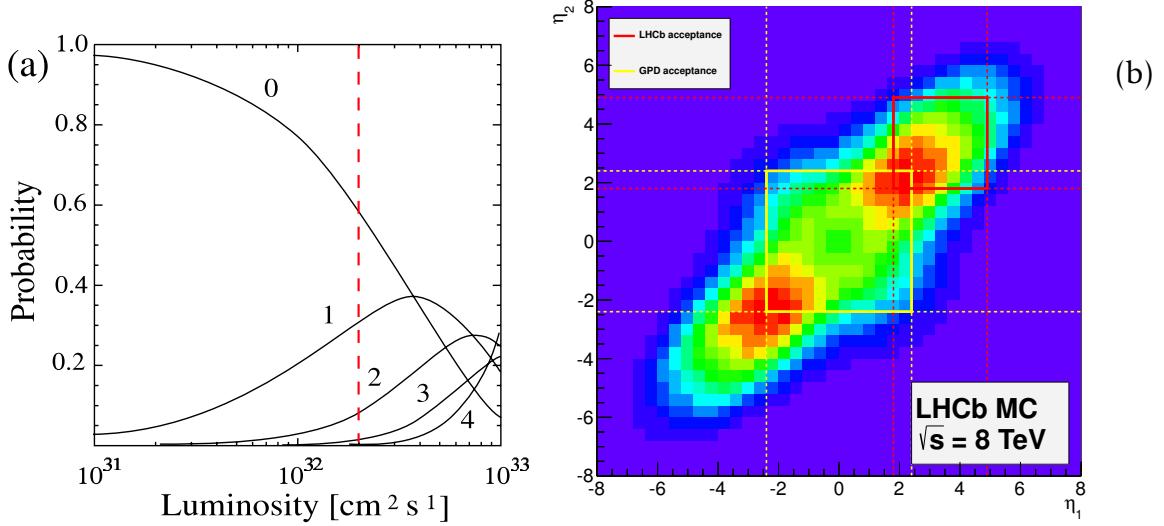
Figure 3.3: (a) Probability of interaction per bunch crossing as a function of instantaneous luminosity. Figure from [39]. (b) Angular production and acceptance of the b (x-axis) $\bar{b}$ (y-axis) pair produced from a pp collision at the LHC. The acceptance of the LHCb detector is the red box and the acceptance of the General Purpose Detector is shown in the yellow box. LHCb covers the region with highest production cross-section at 8 TeV. These plots were produced using a Pythia 8.1 [40] simulation. Figure from [41].

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

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

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

where θ is the polar angle measured from the beam axis. The LHCb detector was built to cover the region $2 < \eta < 5$. The production cross-section of the fundamental process

of $pp \rightarrow b\bar{b}X$ was measured in this region yielding, $\sigma(pp \rightarrow b\bar{b}X)=$ 75.3±5.4±13.0 μb
at 7 TeV [42] and 144±1±21 μb at 13 TeV [43], which shows that the production cross-
sections scales roughly linearly with the centre-of-mass energy. Assuming the design
conditions of LHCb, listed in Table 3.1, 2 fb^{-1} of data (eqvivalent to the 2012 dataset)
would correspond to 10^{12} $b\bar{b}$ pairs being produced in a full 4π region with 27% of these
$b\bar{b}$ pairs produced in the LHCb acceptance. The summary of LHCb running conditions
is also provided in Table 3.1. The analysis of $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ is done with the Run I and
2016 dataset.

Despite the impressive statistics of $b\bar{b}$ pairs available to LHCb, the bottleneck in
terms of data collection arises from the much more copious inelastic background. That
mostly originates from soft QCD processes which are related to the amount of pile-up,
the visible number of pp interactions in the visible events. By looking at the probability
of the number of pp interaction per bunch crossing as a function of luminosity, shown
in Figure 3.3(a), it can be noted that the maximum probability for only one pp interaction
(and hence minimizing the background) is found to be at $\sim 2 \times 10^{32}$cm^{-2}s^{-1}. This was
the reason behind the LHCb design luminosity. Subsequently it has been found that it
is more optimal to run at a higher luminosity of $\sim 4 \times 10^{32}$cm^{-2}s^{-1} but then implement
a set of global event cuts (GEC). Only events with 600 (in 7,8 TeV) and 450 (in 13 TeV)
hits and less, corresponding to the track density in the particular part of the detector,
are allowed to be processed. As the majority of the branching fractions at LHCb are
measured with respect to other branching fractions, there is no bias introduced by the
GECs.

As LHCb requires much lower luminosity compared to other LHC detectors, there
is an LHCb-specific control of luminosity known as *luminosity levelling*, shown in Fig-
ure 3.4. 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.

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

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

2011, 2012 — Run I; 2015, 2016, 2017 — Run II

Table 3.1: Running conditions of LHC and LHCb in different years of data-taking. The statistics of LHCb's instantaneous luminosity, $\mathcal{L}$ is extracted using run database information. Run II data-taking finishes in 2018.

in Figure 3.2, is presented. The vertexing at LHCb is performed with the vertex locator system, also known as the VELO, and is described in section 3.2. The tracking system at LHCb consisting of trackers before the magnet (TT), and three tracking stations behind the magnet (T1, T2, T3) is highlighted in section 3.3. The particle identification is provided by two Ring Imaging Čerenkov counters (RICH1 and RICH2), which are detailed in section 3.4. No particle physics experiment is complete without a calorimeter system, discussed in section 3.6, which consists of a Scintillator Pad Detector (SPD), Preshower (PS), an electromagnetic calorimeter (ECAL) and finally a hadronic calorimeter (HCAL). The muon system positioned at the end of the detector, consisting of five muon chambers is described in section 3.7. The trigger chain as well as the simulation chain are discussed in section 3.8 and section 3.9. Particular emphasis is given to the muon detectors and the simulation of LHCb.
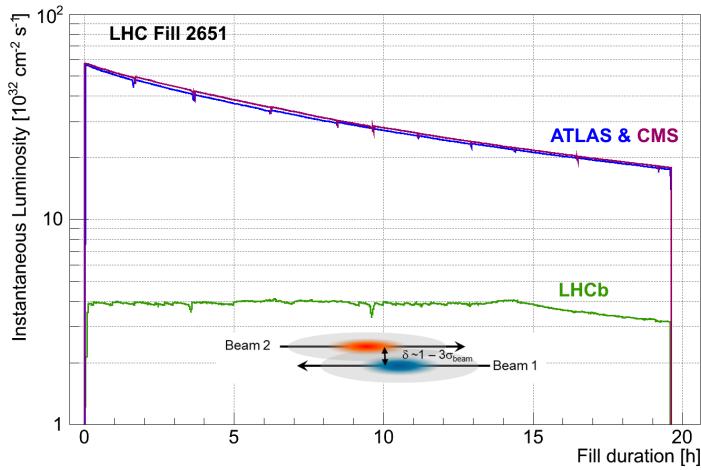
Figure 3.4: Development of the instantaneous luminosity for ATLAS, CMS and LHCb during a random representative LHC fill. After ramping to the desired value of $4 \times 10^{32} \mathrm{cm}^{-2}\mathrm{s}^{-1}$ for LHCb, the luminosity is kept stable in a range of 5% for about 15 hours by adjusting the 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 [44].

3.2 VErtex LOcator

The subdetector closest to the collision point is the 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 property of a b-hadron decay include large impact parameter (IP), the minimal distance between the track and a primary vertex, in addition to significantly higher transverse momentum, p_T. Therefore, the main tasks of this subdetector are to find:

- primary vertices

- secondary vertices of short-lived particles (heavy quark hadrons)

- tracks that did NOT originate from the primary vertex

.

R sensors
φ sensors

cross section at y=0

1 m

390 mrad

60 mrad

15 mrad

x

z

pileup
VETO
stations

view of
most upstream
VELO station

interaction region
σ = 5.3 cm

y

x

8.4 cm

6 cm

VELO fully closed
(stable beam)

VELO fully open
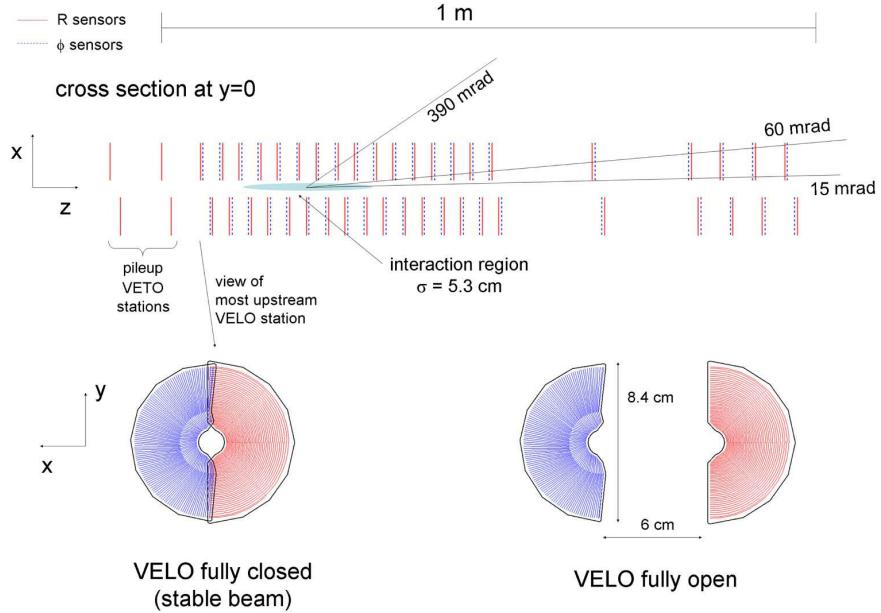
Figure 3.5: Schematic plot of the VELO detector configuration along the beam pipe showing the layout as well as positions while in stable beams (discs have slight overlap) and injection. Figure from [37].

The detector consists of two sets of 21 silicon modules positioned around the beam pipe, where each module has 2 types of half-moon-shaped discs as seen in Figure 3.5. In the first type, the strips are arranged to provide radial information (R), whereas the second type provides azimuthal (ϕ) information. As pp collisions bring a high dose of radiation to this detector, the first sensitive strip starts 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 3 cm away, perpendicular to the beam axis. For the R sensor, the individual module's strip pitch, the 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 form their own vacuum which is separated from 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 3.6(a)(b). 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 a 3 trillion Hz rate. As will be seen later, this excellent resolution is also very important for the detection of decays with neutrinos in the final state.
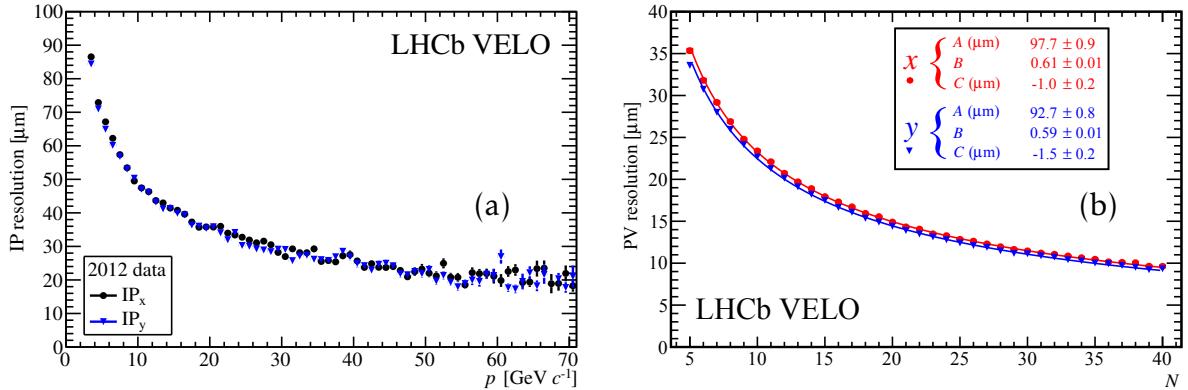
Figure 3.6: Two key variables which quantify performance of the VELO detector. (a) IP resolution which is worse for low momentum tracks and (b) PV resolution dependent on the number of tracks forming the primary vertex N. Figures from [45].

3.3 Tracking System

In addition to tracking information provided by the VELO, the trajectories of charged particles are measured by a series of tracking subdetectors. The main task of these tracking subdetectors is to provide efficient reconstruction and precise measurement of a particle's momentum. There are four tracking stations apart from VELO: Tracker Turicensis (TT), positioned upstream from the magnet, and the T1, T2 and T3 tracking stations on the other side of the magnet. The dipole magnet with ≈ 4 Tm integrated field provides strength to bend charged particles in the horizontal plane.

Two different detection technologies are used in these trackers reflecting the nature of track occupancy as a function of polar angle. The parts at small polar angles, TT station together with central region of T1, T2 and T3, also known as the Inner Tracker (IT), expect higher occupancy and make use of the silicon microstrip detection mechanism. The outer part of T1, T2 and T3 stations, also known as the Outer Tracker (OT), is made of straw-tube detectors. Straw tubes measure the trajectory of the track by measuring the drift-time of ionized electrons. Use of the two technologies is illustrated in Figure 3.7(a).

3.3.1 Tracking Algorithms

Different types of particles will leave different footprints in the detector. Charged particles will form tracks. Depending on the presence of hits in individual subdetectors, they are grouped into several categories, visualized in Figure 3.7(b).
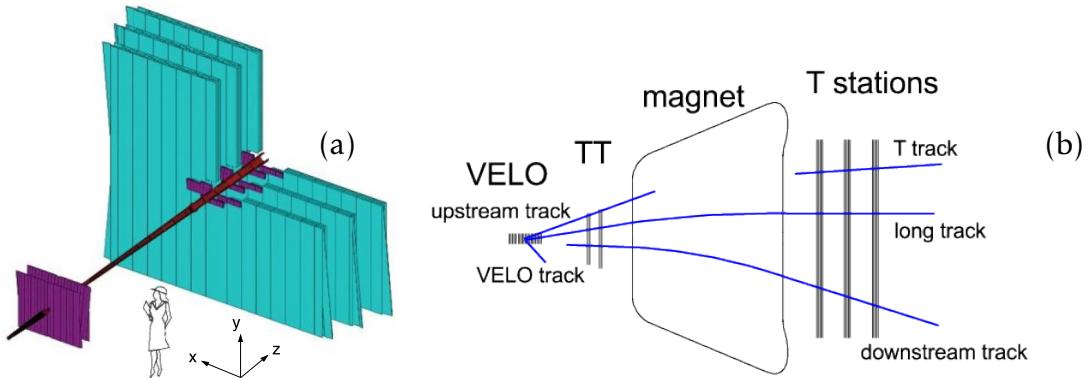
Figure 3.7: (a) Visualisation of use of different technology with silicon technology in violet and straw-tube technology in cyan. Figure from [37]. (b) Track types categorisation depending on which track stations provided hits. For the study of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decays, only long tracks are considered as muons will travel to the end of the detector leaving hits all along. Figure from [46].

Most of the physics analyses at LHCb, as it is the case for the search of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$, use only long tracks, tracks leaving hits in the VELO and T1, T2 and T3, as they give

most precise momenta measurements. There are also other types of tracks as indicated in Figure 3.7 but they are rarely used.

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 subdetectors. The long track candidates are formed and fitted with a Kalman filter [47], where, because of the material present in the detector, corrections for energy losses as well as multiple scattering are incorporated.

In LHCb there are types of tracks which are not really the trajectories of charged particles. Sometimes the *pattern recognition* may combine random hits into a track, which is then known as a *ghost track*. On the other hand, it could also happen that several tracks are sharing the same hits, known as *clone tracks*. The presence of these types of tracks are suppressed through the use of a neural network based variable (P_{ghost}), which relies on the χ^2 of the track fit, and information about missing hits along the trajectory to calculate its value.

When searching for a *b*-hadron decay, the mass of a candidate can be calculated from the 4-momenta of the decay products. Uncertainty on this mass is one of the crucial parameters to minimize as it enables a better separation between the identified signal and background. It strongly correlates with the momentum resolution that is obtained using the tracking system. The resulting relative momentum uncertainty (0.5-1.1%) on long tracks using $J/\psi \rightarrow \mu^+\mu^-$ data can be seen in Figure 3.8. It varies logarithmically with increasing momentum.
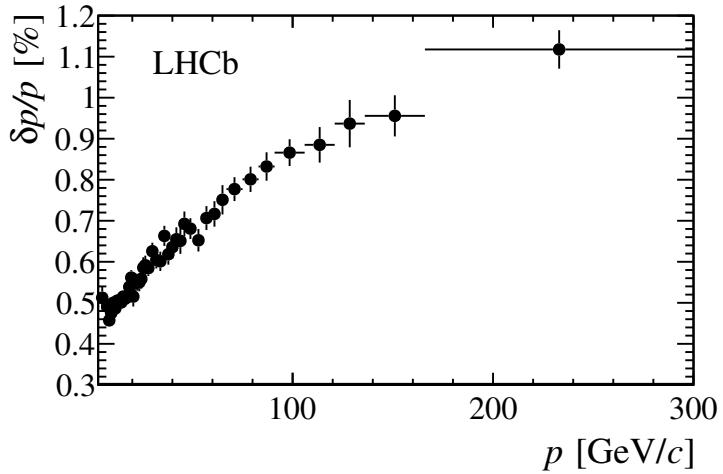
Figure 3.8: Momentum resolution of long tracks measured at LHCb. The decay channel $J/\psi \to \mu^+\mu^-$ is analysed for this purpose. Figure from [44].

3.4   Ring Imaging Čerenkov Detectors

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

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

As the momentum is not an intrinsic property of a passing particle, the momentum identification range is limited by the choice of medium, also known as radiator. For very low-momentum particle, as $\cos\theta_c \to 1$ ($p = \sqrt{\frac{m^2}{n^2-1}}$), the particle is not producing any Čerenkov light cone. At very high momentum, as $\cos\theta_c \to 1/n$, there is a saturation point as all species of particle will emit light at the same Čerenkov 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 an angular acceptance of 25-300 mrad using C_4F_{10} ($n = 1.0014$) as the radiator. RICH2 has a more limited acceptance of 15-120 mrad and uses CF_4 as the radiator, with lower $n = 1.0005$. The discrimination power between different particles can be seen in Figure 3.9(a).

Both RICH1 and RICH2 use a set of spherical primary mirrors to guide the photons onto the flat secondary mirrors which are then further focused into Čerenkov rings on the surface of a plane of Hybrid Photon Multipliers, (HPD). The schematic view of a particle passing through RICH1 can be seen in Figure 3.9(b).
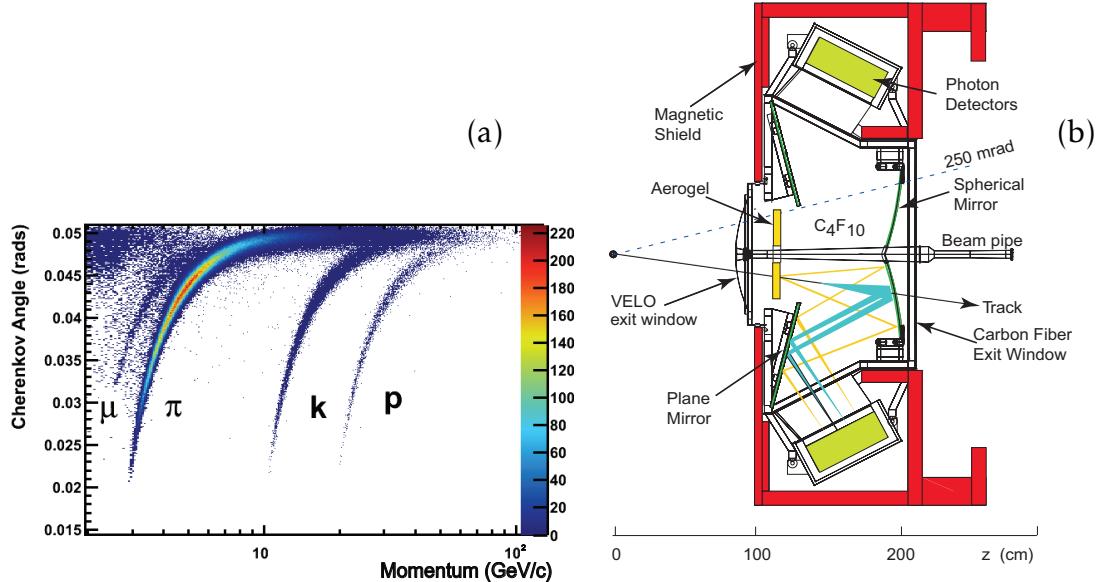
Figure 3.9: (a) Separation power for different species of particles in the momentum-Čerenkov angle plane for the C_4F_{10} radiator. Figure from [48]. (b) Schematic diagram of RICH1 layout. Figure from [37].

70

3.5   RICH Reconstruction and Performance

In order to correctly associate species of particles to a given track, the Čerenkov angle is combined with the track momentum measured by tracking. In practice, however, as RICH detectors operate in high track density environment, many Čerenkov rings will be overlapping and hence a complex pattern recognition algorithm is deployed [49].

For each event, the RICH computes a full event likelihood that is consistent with assigning a pion mass hypothesis to all tracks given the observed hit distribution read out by the HPDs. The algorithm then iterates through all other possible particle species, $(e, \mu, \pi, K, \text{proton}, \text{deuteron})$, assigning a new full event likelihood for a given track, with 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 the pion hypothesis,

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

By calculating $\text{DLLx}_1 - \text{DLLx}_2$, one can obtain discriminative strength between any two species.

3.5.1   RICH Performance

In order to measure the performance of the PID computed by a 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 the D^0 become proxies for the evaluation. The invariant mass for the D^0 candidates can be seen in Figure 3.10(a). The probability of correctly identifying a kaon given a certain constraint on DLLK, the identification efficiency (ID), and the

probability of mistakenly swapping pion identification, the misidentification efficiency (misID), are summarized in Figure 3.10(b). Identification probabilities of $\approx 85\%$ with a misID rate of $\approx 3\%$ provide invaluable discriminating separation between kaons and pions.
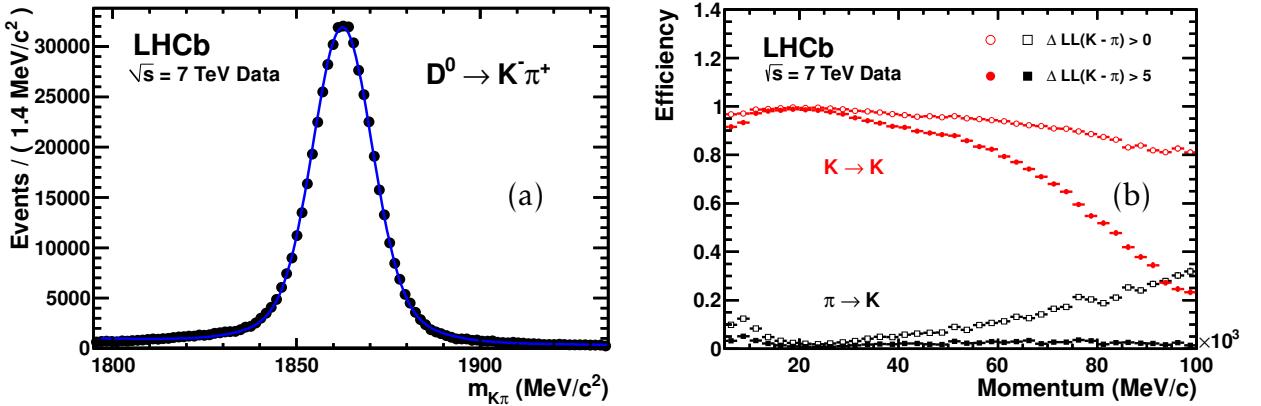
Figure 3.10: (a) Invariant mass distribution of D^0 data sample (in black) overlaid with fit to both background and signal (in blue). (b) An example of kaon ID (red) and misID (black) efficiency as a function of momentum under two PID hypotheses, DLLK > 0 (empty) and DLLK > 5 (filled). Both Figures from [48].

3.6 Calorimetry

As many other particle physics detectors, LHCb is equipped with series of subdetectors providing separation between electrons, pions and photons. This separation is achieved because different particles interact differently with the material, producing differently shaped showers. This part of the detector is not only integral to the way the LHCb trigger system works but it also provides a measurement of the energies of these objects. All the subcomponents discussed here operate on the same principle. Particles passing through the material emit light. The light from the scintillating material, which is created by absorbing the energy of the particle and re-emitted it in the form of light, is guided to photomultiplier tubes by wavelength shifting fibres.

Electrons, pions and photons firstly encounter two planes of scintillating tiles: the Scintillating Pad Detector (SPD), and the Preshower Detector (PRS) intersected by a wall of lead. The SPD senses the passage of charged particles as they emit light whereas neutral particles do not, making this subdetector able to distinguish between electrons and 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 Electromagnetic Calorimeter (ECAL) in LHCb is based on a sampling shashlik-type technology, where scintillating tiles are alternated with 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 the ECAL is $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%$, where E is in GeV.

On the other hand, the Hadronic Calorimeter HCAL sandwiches iron instead of lead as the absorber with a thickness of 5.6 interaction length only, achieving a 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, which is the hadron trigger. Away from the beampipe the granularity of cells is coarser to mirror the track occupancy as seen in Figure 3.11(a)(b).
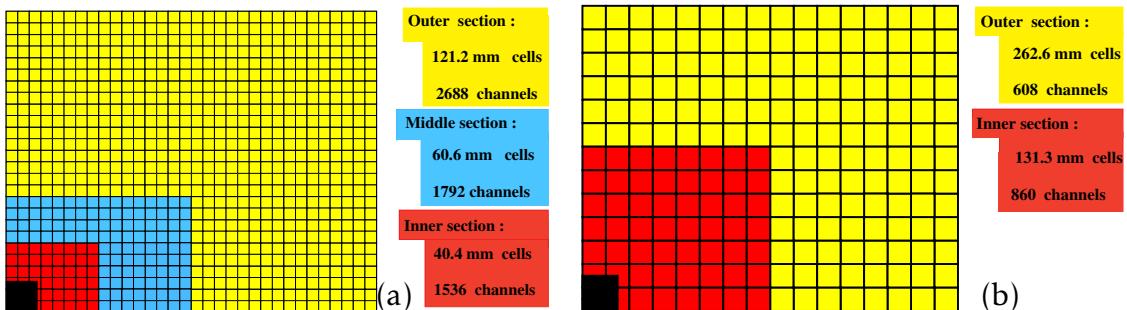
Figure 3.11: Granularity of (a) ECAL and (b) HCAL detectors. This is just a quarter view and that the black region is where the beam pipe is located. Figure from [37].

3.7 Muon Stations

Muons are considered to be of fundamental importance to many flagship analyses by LHCb, such as the search for the rare $B_s^0 \to \mu^+\mu^-$ decay [50]. Analysis of $B^+ \to \mu^+\mu^-\mu^+\nu$ of course relies heavily on a good performance of this part of the detector. Muon stations are positioned at the end of the detector, taking advantage of the fact that muons penetrate material better than any other particle type.

LHCb's five rectangular muon stations M1-M5 are positioned before and after the calorimetry system, with the first station M1 upstream of the SPD, and four stations (M2-M5) downstream of HCAL as shown in Figure 3.12. 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 measurement of p_T in the hardware trigger. The M2-M5 stations each consist of 276 multi-wire proportional chambers (MWPCs) filled with an Ar, CO_2, CF_4 gas mixture. They are interlayered with $0.8\,$m iron walls, to provide a stopping target for all particles, other than muons with momentum higher than 6 GeV/c.

Each half of a muon station is 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 the magnet's horizontal bending plane.

Both GEM and MWPCs operate on the same principle. In each station, the 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 a positive binary decision. This means the efficiency of each station must be $\geq$99% to give an overall 95% trigger efficiency.
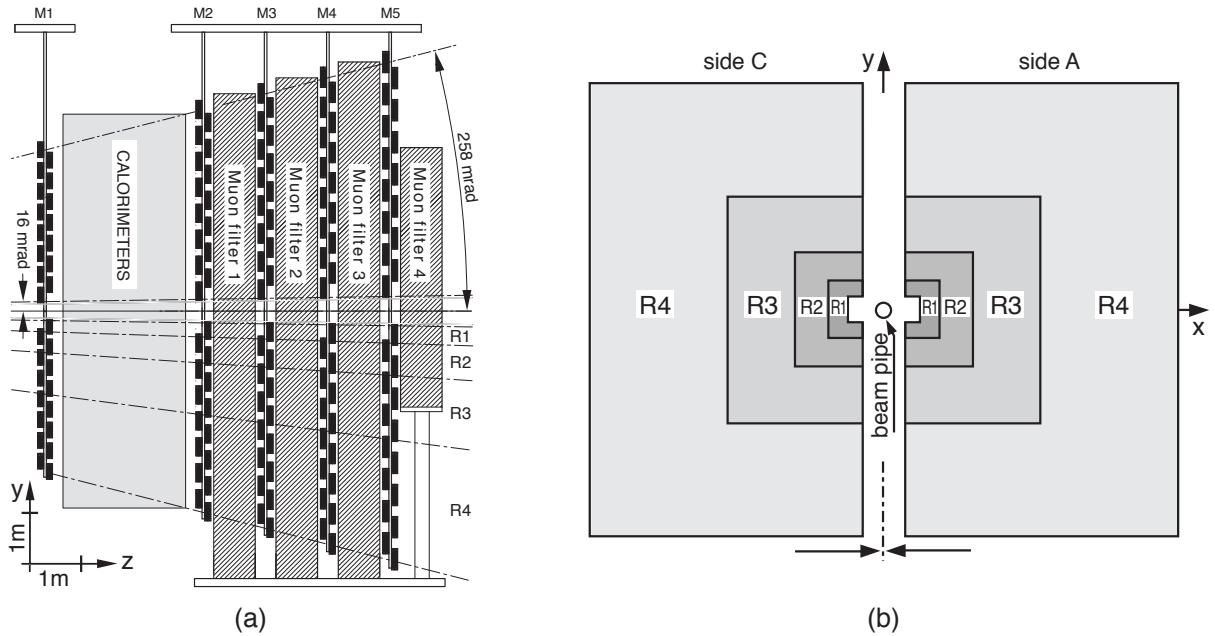
Figure 3.12: (a) Layout of the muon detector x-z plane and (b) x-y plane. Figure from [51].

3.7.1 Muon Identification

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

- **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 the acceptance. For each station, a search for hit information within an elliptical area defined by momentum, a field of interest (FOI), is performed. The hit requirements are summarized in Table 3.2.

- **muDLL**: Difference in log likelihoods computed using a muon and non-muon hypothesis. These hypotheses are based on the prox imity/distance, D^2, of the track extrapolation into the muon stations and corresponding closest sensed hits

75

in those stations. Muon-like particles will tend to have a sharper distribution in D^2 as compared to other species. Protons were chosen to be the other species for the calibration purposes. They give a broader distribution as they originate either as punch-through protons (protons coming from showers not fully contained in the HCAL), protons having the same hit position as true muon, or random hits.

- **DLLmu**: For each track the same global likelihood is produced, by combining the muon and non-muon likelihoods from **muDLL**, with the RICH different mass hypothesis likelihoods, and the calorimetry likelihood exploiting information about the energy deposits. Like in the RICH likelihoods, the default hypothesis corresponds to separation between the muon and pion hypotheses.

In the $B^+ \to \mu^+\mu^-\mu^+\nu$ analysis the variables IsMuon and DLLmu are used to identify muons. In addition, other variables that are used for muon identification in the search for $B^+ \to \mu^+\mu^-\mu^+\nu$, are described in section 4.1. The use of several variables for muon identification is done as they are mostly complimentary, exploiting different information from different parts of the detector.

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 3.2: Momentum-dependent definition IsMuon variable.

3.7.2 Muon Identification Performance

As for hadron performance measurements, the muon ID performance is determined using the high statistics decay channel $J/\psi \to \mu^+\mu^-$ with a *tag and probe* method. MisID rates for kaons and pions are computed using the same decay channels, which were

used for the identification of hadrons, $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$. The summary of `IsMuon` ID and misID rates are presented in Figure 3.13. A very high ID rate (above 90%) for relatively low misID probability (below 10%) is key to analyses with muons in the final state. The identification rate for the low p_T muons suffers because these muons can end up outside of the LHCb acceptance. MisID rates for kaon and pions are significantly higher in the low momenta region as the dominant process for this occurence is muons from decay-in-flight.
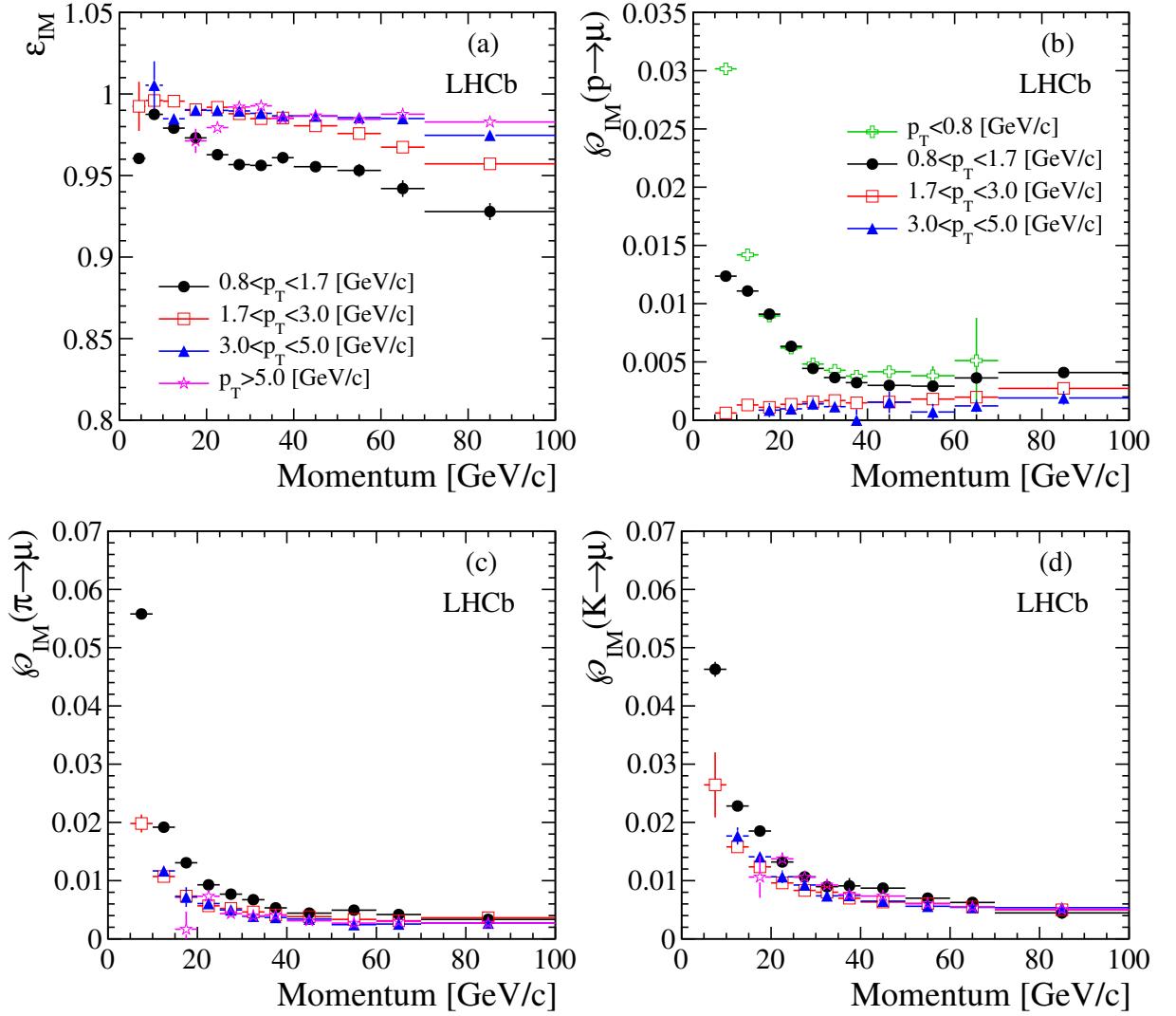
Figure 3.13: (a) Probability of correctly identifying muons as a function of momentum in bins of p_T for $J/\psi \to \mu^+\mu^-$ with an IsMuon constraint. (c) Probability of incorrectly identifying a pion (b) proton and (d) kaon as a muon with IsMuon. This figure is taken from [52].

3.8 Trigger

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 a series of decisions, which is known as the trigger. The 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 a hardware-based level 0 trigger (L0) and a software-based high level trigger (HLT).

In the end, selected events have their trigger decisions categorized. An event where the signal candidate caused the trigger to fire is known as Trigger on Signal (TOS). An event where it is a non-signal like particle causing the trigger decision to occur is labelled as Trigger Independent of Signal (TIS). Finally, if only a combination of signal particle(s) together with other particles in the event produces an 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. The first three lines make a decision using calorimeter information about the transverse energy, E_T, and whether it is a photon, electron or hadron causing the shower energy deposit. Two other lines read out information from the muon system by looking for p_T, of muon and dimuon (two muon tracks) objects. The efficiencies of the L0 muon triggers are evaluated using $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ decays and can be seen in Figure 3.14(a). The hadron trigger efficiency in different decay channels can be seen in Figure 3.14(b).
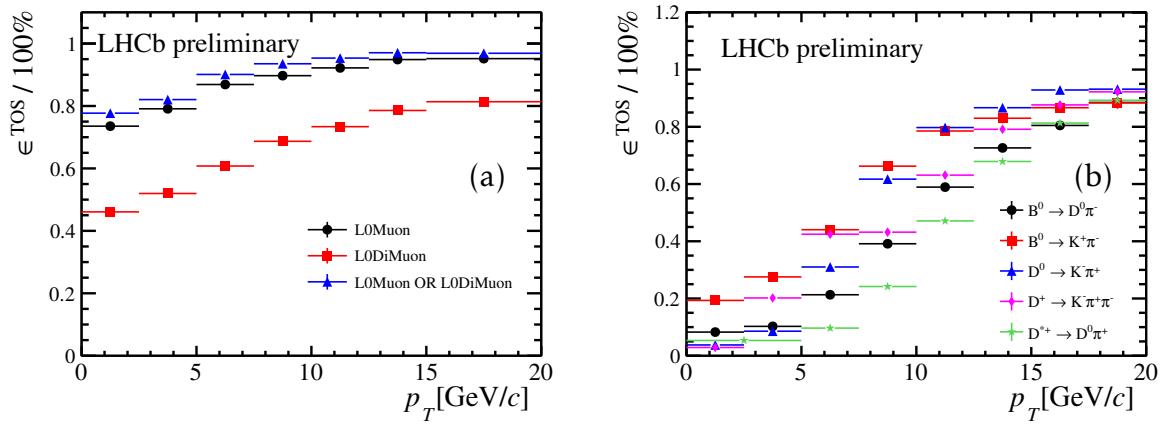
Figure 3.14: (a) TOS efficiency as a function of p_T for muon-based decisions. (b) TOS efficiency for different decays using L0 hadron trigger lines. Figures from [53].

The software-based HLT then further reduces the rate from 1 MHz down to 5

kHz which can be recorded to long-term storage. The first stage of the HLT, (HLT1), performs limited track reconstruction and hence makes a decision based on the presence of charged particles in the event. HLT1 uses VELO hits to reconstruct PVs and VELO tracks by using 3D pattern recognition. As LHCb's primary mission is to study decays of hadrons containing b and c quark, HLT1 will make a decision based on the track being displaced (having a 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, where the magnetic field of the dipole will not make them bend. 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 a fast Kalman filter providing the χ^2 of the fit. The corresponding performance of the HLT1 trigger lines are shown in Figure 3.15(a)(b).
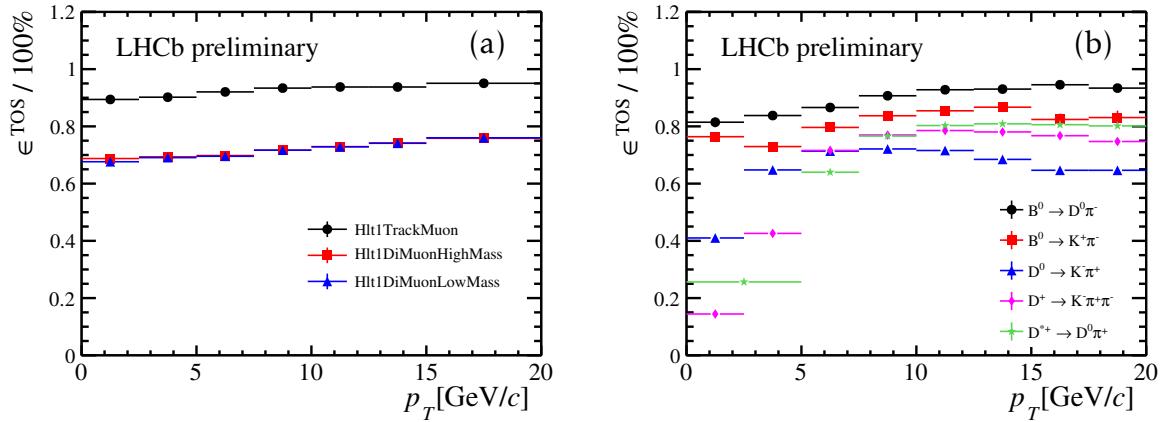
Figure 3.15: HLT1 efficiencies of the corresponding triggers using the same proxy as in Figure 3.14. Figures from [53].

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 that are well separated from the PV are reconstructed. To account for decays where a final state particle is not fully reconstructed, the corrected mass (will be defined in Equation 5.1) 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 there are *Exclusive triggers* concentrating on selecting events with D mesons. They perform a selection which is very similar to the offline selection but without PID cuts.

Between the Run I and Run II period there has been a change in how the software trigger operates, which can be seen in Figure 3.16. As more computing resources were introduced for both HLT1 and HLT2, LHCb took advantage in upgrading the trigger system by introducing an update of the calibration and alignment constants of the relevant subdetectors before the data is sent to permanent disk. *Online reconstruction*, defined as being produced at the trigger farm, became the same as the *offline reconstruction*, defined as reconstruction made when data reached the permanent disk. Hence, there is an enhancement of the available information, such as the PID in the HLT, which can then be used at the trigger level.

LHCb 2012 Trigger Diagram

40 MHz bunch crossing rate

L0 Hardware Trigger : 1 MHz
readout, high E_T/P_T signatures

| 450 kHz h± | 400 kHz μ/μμ | 150 kHz e/γ |

Software High Level Trigger
- Introduce tracking/PID information, find displaced tracks/vertices
- Offline reconstruction tuned to trigger time constraints
- Mixture of exclusive and inclusive selection algorithms

5 kHz (0.3 GB/s) to storage

| 2 kHz Inclusive Topological | 2 kHz Inclusive/Exclusive Charm | 1 kHz Muon and DiMuon |

LHCb 2015 Trigger Diagram

40 MHz bunch crossing rate

L0 Hardware Trigger : 1 MHz
readout, high E_T/P_T signatures

| 450 kHz h± | 400 kHz μ/μμ | 150 kHz e/γ |

Software High Level Trigger

Partial event reconstruction, select displaced tracks/vertices and dimuons

Buffer events to disk, perform online detector calibration and alignment

Full offline-like event selection, mixture of inclusive and exclusive triggers

12.5 kHz (0.6 GB/s) to storage

Figure 3.16: Trigger scheme differences between Run I and Run II. Figures from [54].

3.9 Simulation

In order to optimise the event selections, determine 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 the LHCb configuration [55] are simulated with Pythia 6.4 [56] and Pythia 8.1 [40]. 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, the B^+ meson is the hadron of interest. Hadrons are then further decayed using EVTGEN [57] into the chosen decay products. At this stage, different physics models or inputs from theory can be configured. In order to account for the effects of QED radiative corrections, the PHOTOS [58] algorithm can be used. All of this combined establishes *the generator-level simulation* of LHCb.

In the next phase, *detector simulation*, the interactions of all the particles with the detector, transport, as well as detector's response are simulated using the C++ GEANT4 toolkit [59], [60]. LHCb's interface to GEANT4 is detailed in Ref [61].

3.9.1   Differences in Simulation and Data

Despite the complexity and best intention of the LHCb simulation, there are several shortcomings that require corrections. 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 a better trend in the simulation then in the data due to the mismodelling of the material description in the VELO simulation. As shown in Figure 3.17(a)(b) the IP resolution does greatly differ depending on the variation of material density of VELO. Around $\phi = \pm\pi/2$, where the two VELO parts overlap, the material difference causes the discrepancy. It can be corrected either by reweighting to data or by smearing the resolution with a Gaussian distribution.

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

The most critical problem that needs to be addressed in the presented analysis is the inaccuracies of the PID variables, which are mismodelled in the simulation. This problem arises as a consequence of the much lower estimate of low momentum tracks in the detector, making the photoelectron background underestimated. This results in better separation in simulation and is corrected using a data calibration.

Therefore the PID efficiency is usually obtained from the data. More specifically, this
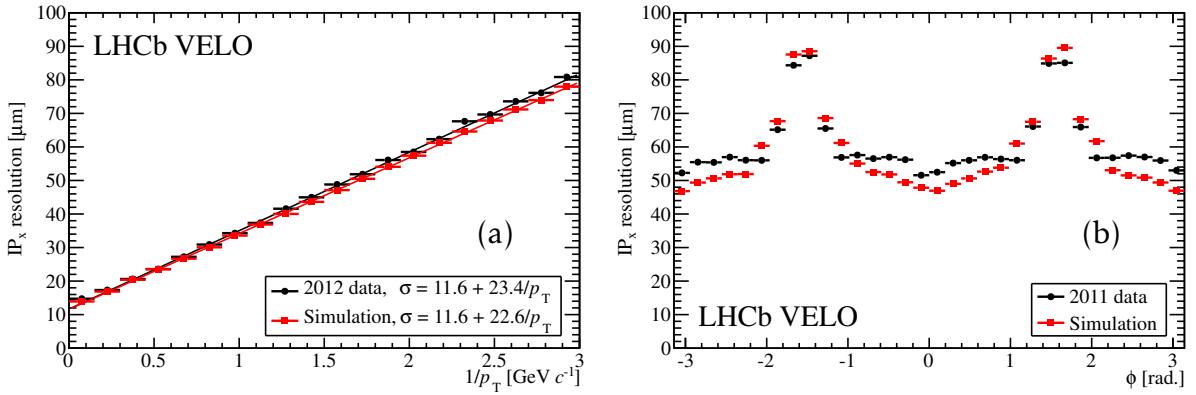
Figure 3.17: (a) IP resolution in the x-direction comparing the data and simulation for the 2012 data-taking period. (b) IP resolution in the x-direction comparing the data and simulation for the 2011 data-taking period as a function of angle, ϕ. Figures from [45].

is done by using high-yield and relatively background-free calibration channels, where the species of the particle can be deduced from kinematics of the decay. A standard set of these channels are "housed" in a PIDCalib package [62]. With this package, the PID efficiency can be computed in a given kinematic region of interest.

Chapter 4

Handling of Trimuon Correlations at LHCb

This chapter discuss issues associated with three muons passing through the detector. Two collimated muons may traverse through the same parts of the detector if they have the same charge, causing problems in resolving their individual tracks. Therefore, ghosts and clones are much more likely to occur. In LHCb, a plethora of muon PID variables are used to suppress these types of spurious tracks. However, the usage of PID variables in an analysis in LHCb brings its own challenges. As the simulation is not able to estimate PID efficiencies correctly, most of the PID efficiencies are taken from control samples. New control samples for $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ are considered as the PID efficiencies depend strongly on the number of muons in the detector and in the standard misID control samples there is just a single muon in each event.

4.1   Muon PID Variables

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

4.1.1 Binary Muon PID Variables

Similar to `isMuon` shown in Table 3.2, there are more binary variables, such as `isMuon-Tight`, that can help with the classification of muons. As its name suggests, `isMuon-Tight` has stronger conditions to satisfy as compared to `isMuon`.

In each muon station (M1-M5) a field of interest, FOI is defined as

$$FOI_{x,y} = \rho_{x,y}^0 + \rho_{x,y}^1 \cdot \exp\left(\frac{\rho_{x,y}^2 \cdot p}{\text{GeV}/c}\right),\qquad(4.1)$$

where x,y are the dimensions perpendicular to the direction of the beam, p is the momentum of the muon, $\rho_{x,y}^i$ are three dimensional parameters tuned to give the best performance, by maximizing efficiency versus misID rate. This FOI can be thought of as a cone whose radius depends on the p.

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

The detector information is read out in the x and y direction separately. The pad slicing according to this read-out scheme is known as *physical* slicing of pads. However, as seen in Figure 4.2, the overlapping x and y *physical* pads can be grouped into *logical* pads, which give information about x and y simultaneously. This leads to two groups of hits according to pad type: uncrossed hits - registered within *physical* pads only, and crossed hits - given by *logical* pads. Whereas `isMuon` only requires a positive decision from uncrossed hits, `isMuonTight` requires a positive decision based on crossed hits.

4.1.2 Muon PID Variables Based on Sharing Hits

Another way of identifying muon tracks is based on the variable, `nShared`, which identifies the number of tracks with shared hits in the muon stations. For each hit within the FOI of an extrapolated track, the `nShared` algorithm will check whether
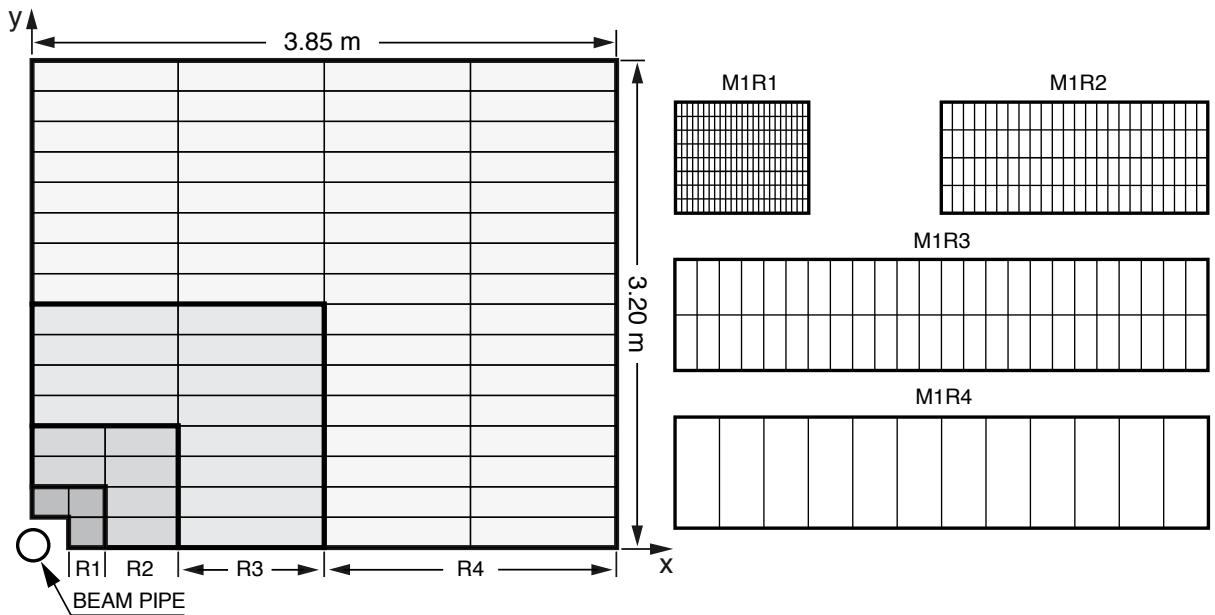
Figure 4.1: Schematic view of the muon station slicing into x-y pads. This is the left quadrant of the M1 station, showing decreasing granularity of the muon stations away from the beam pipe. This figure has been taken from [51]. M1R1 is the innermost region and M1R4 is the outermost region of the M1 station.
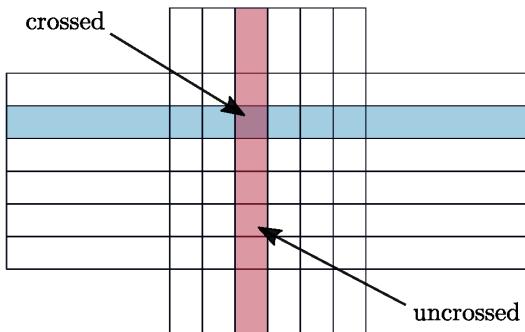
Figure 4.2: Difference between crossed and uncrossed hits. A hit in a muon station is considered a crossed hit if it is registered both by a horizontal and a corresponding vertical strip. If a hit is only seen by either, it is considered uncrossed. This figure has been taken from [63].

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

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

The first feature that is different between Run I and Run II arises from the calculation of the distance between the extrapolation and the hit in the `nShared` algorithm. In *Stripping 21* (where *stripping* is a preselection) used for 2012 and *21r1* used for 2011 data, it was discovered that the distance between an extrapolated track and a hit was wrongly calculated. This mistake was corrected before *Stripping 23*, used for analysing 2015 data.

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

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

Further in the process, however, this information is used to calculate the sum and average of distances per station between the hits and extrapolations. This algorithm

again iterates over $[0, 3]$ arrays, meaning that no information is used from the M5 muon station. This obviously has an effect, but again it is consistent across the versions of the reconstruction software used for the processing of Run II data.

The interplay between all these features for $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ decays can be seen in Figure 4.3, which sees a shift in distribution of nShared for 2016 data taking, making the muons less isolated.
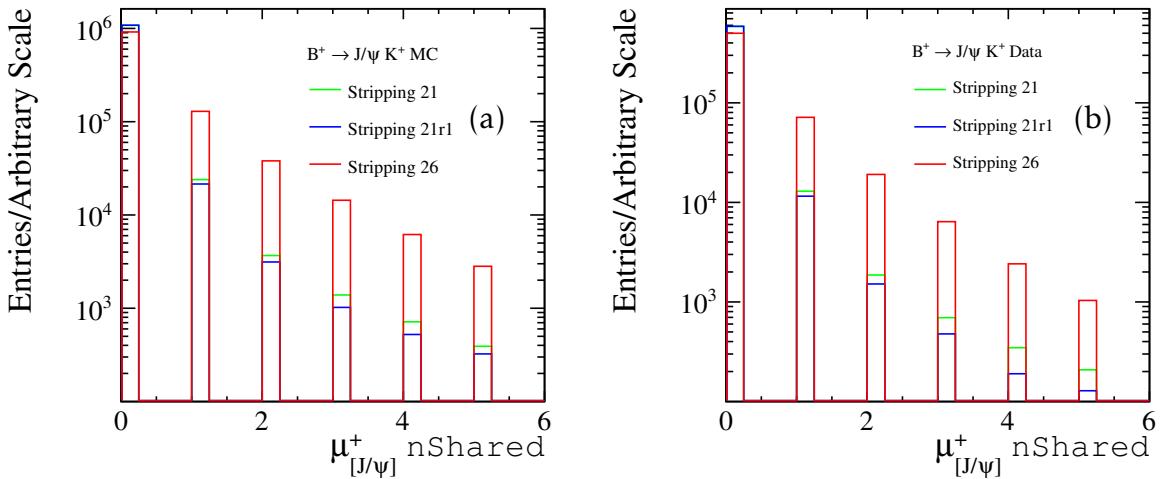
Figure 4.3: (a) nShared variable distribution for the positive muon in $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ decays in (a) simulation and (b) data. Different stripping versions corresponding to 2012 (*Stripping 21*), 2011 (*Stripping 21r1*), 2016 (*Stripping 26*) data-taking are shown. The distributions are normalised to have the same area. There is a shift of distribution in *Stripping 26* towards less isolated tracks. The proportion of muon tracks that share no other hits with other tracks is smaller, whereas the proportion of the tracks sharing hits with other muon track is increased.

Using the same calibration channels as in subsection 3.7.2, misID and ID rates can be seen in Figure 4.4. As the tracks tend to be less isolated in *Stripping 26* used for 2016 data, typical of non-signal like events, the misID rate is expected to be higher for the same working point (ID efficiency). While the issues highlighted here can be fixed with a reprocessing of the data, this is not expected to happen before 2019 or 2020.
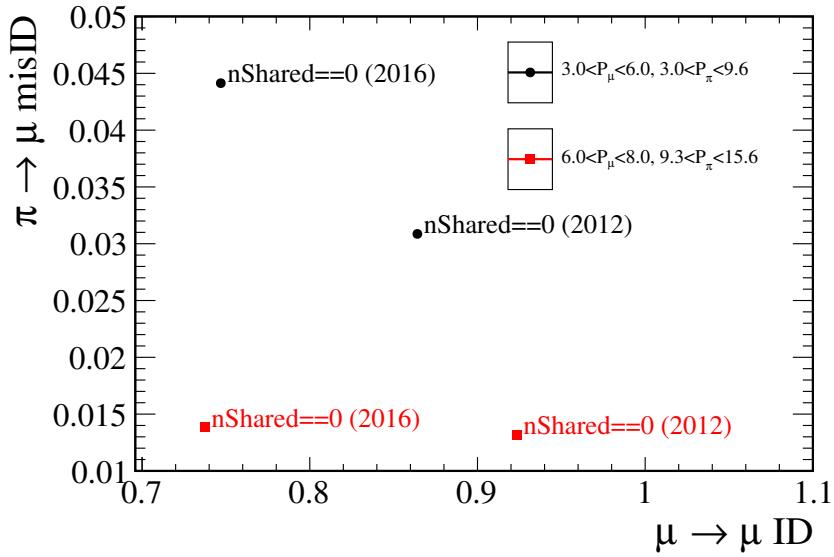
Figure 4.4: ID and misID probabilities from standard calibration datasets from 2012 (*Stripping 21*) and 2016 (*Stripping 26*), binned using the default 2-dimensional binning scheme in momentum p and pseudorapidity η. In this plot, ID and misID rates in the central bin of η, $2.375 < \eta < 3.25$, and the first and second bin in p are compared. This demonstrates that for the same pion ID efficiency, the misID rate is significantly higher in 2016 data.

4.1.3   Muon PID Variables Based on Regression Techniques

Similar to the DLLmu variable in subsection 3.7.1, which combines all the information from the detector into a global likelihood, it is possible to feed all the different variables to a neural network, which can then produce an output corresponding to the probability of a particle to be of a certain species. 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 $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$. As with any machine learning algorithm, the selection of both the training sample and the input variables are important. In Run I, there were two tunings (trainings) introduced V2 and V3, with more input variables in V2. Depending on the species of

particle, V2 or V3 performed better. In the analysis of $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$, Probnn$x$_V2 is used.

4.2 Clones

When analysing decays with two muons of opposite charge, one can rely on the fact that LHCb magnet bends these two muons in two opposite directions. With two muons of the same sign, the muons will instead bend in the same direction and can stay close together in both the tracking system and the muon detectors. This is a problem for the tracking algorithm as it distinguishes these two tracks less well. It is even possible that these two same sign muon tracks are not genuine tracks, but rather subtracks or a copy of another track, *clone tracks*. Two tracks are clones if they share at least 70% of the hits in the VELO and at least 70% of the hits in the other T-stations. Of course, once it is established that two tracks share this percentage of hits, it has to be established which track is the clone track. This decision is based on the total number of hits and the track χ^2 per number of degrees of freedom of the fit (ndof) (track χ^2/ndof) comparison of the two tracks.
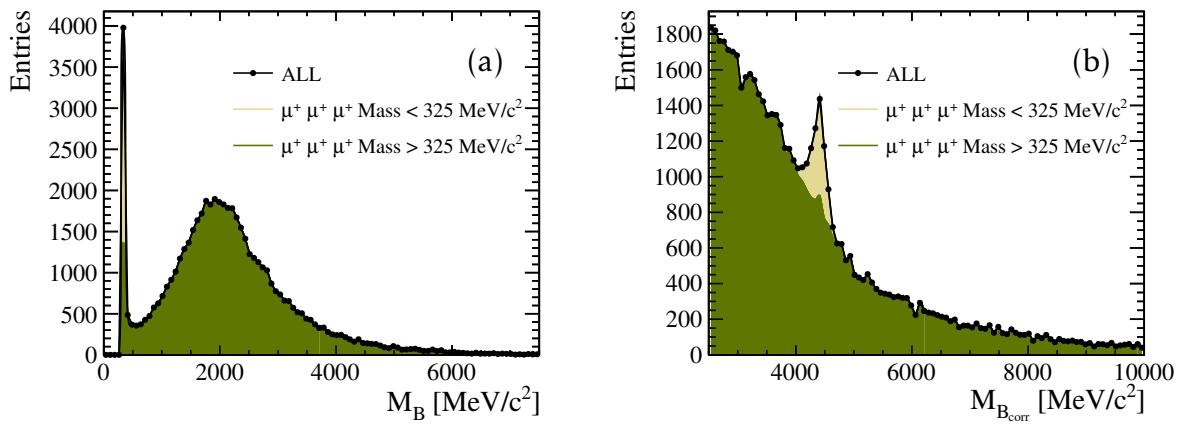
Figure 4.5: (a) Visible and (b) corrected mass of $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ candidates in 2012 data where all the muons have the same charge. Clear fake peaks, arising from the correlation of several effects in the detector can be seen.

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

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

Once translated into corrected mass (will be defined in Equation 5.1), these *fake peaks* are smeared and look like genuine resonances with a resolution as seen in Figure 4.5. The procedure which results in creating *fake peaks* from clones is described below illustrated with Figures Figure 4.5– Figure 4.9, where 2012 data control samples are plotted.

The shape emulating a genuine resonance arises as a collective effect from vertexing, tracking and trigger selection. As there are three parallel tracks, the vertex of the system is not well defined. However, the vertex fitting of the PV and secondary (decay) vertex (SV) is functional and vertex χ^2/ndof (the χ^2 of the vertex per degree of freedom in a vertex fit) is good as these tracks are subtracks of each other. The distance between the PV and the SV is defined as the flight distance (FD). However, *clones* can be differentiated by the position of the decay vertex of the B, Figure 4.6 as well as by the transverse position of the track in the tracking, OT as seen in Figure 4.7.

With this typical path for the clones there is a fixed angle of the clones through the detector (the angle between the muon momentum and the z-axis), which is calculated using information from OT as

$$\arctan(\theta) = \arctan\left(\frac{FD \ radius}{FD \ distance \ along \ z}\right) = \arctan\left(\frac{200 \text{ mm (Figure 4.7)}}{8500 \text{ mm}}\right) = 0.023 \text{rad.} \tag{4.3}$$

With the LOMuon p_T threshold of 1.76 GeV/c for 2012 [53], a typical momentum from about 75 to 120 GeV/c results because
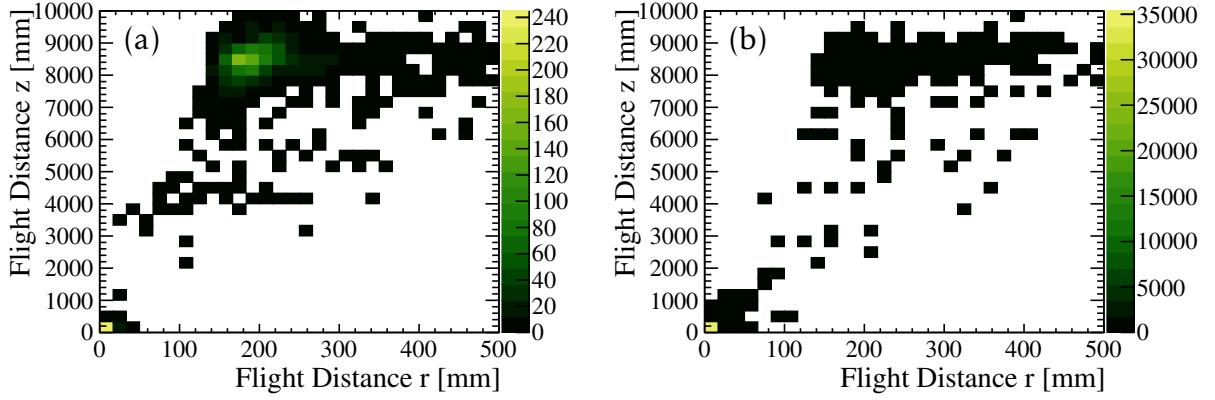
Figure 4.6: (a) Clone and (b) no clones *B* candidate flight distance properties. It can be seen that *clone* tracks have their decay vertex placed at the end of the detector, whereas regular good tracks will decay within the VELO.
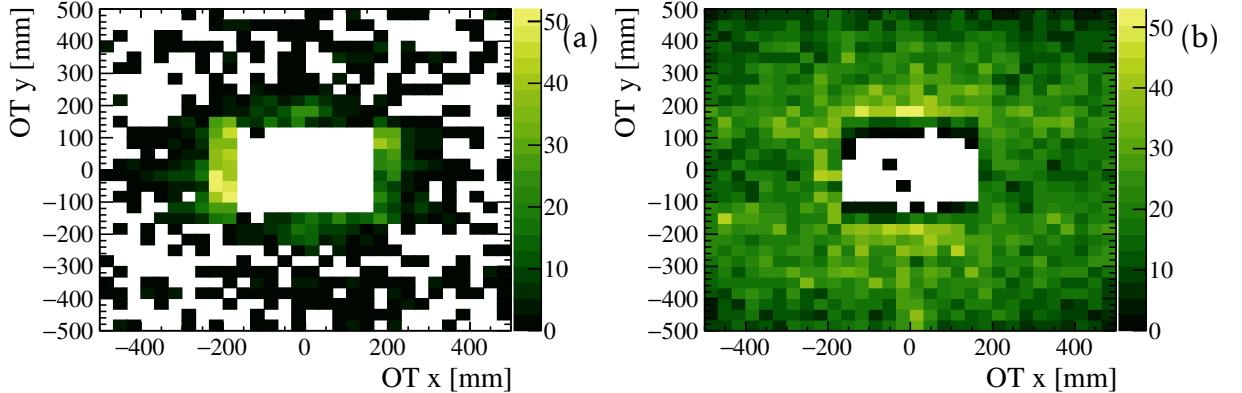
Figure 4.7: Transverse position in the OT detector for (a) clones and (b) real tracks at the distance 9450 mm along LHCb. *Clones* are concentrated along the inner edge of the OT. Good muon tracks will cover most of the OT evenly.

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

The angle between the *B* flight direction and trimuon momentum vector, $\cos(\theta_B)$, will also be fixed and has a typical value of 0.7 mrad as seen in Figure 4.8.

Hence, the missing p_T in the direction of the flight can be calculated using $\cos(\theta_B)$
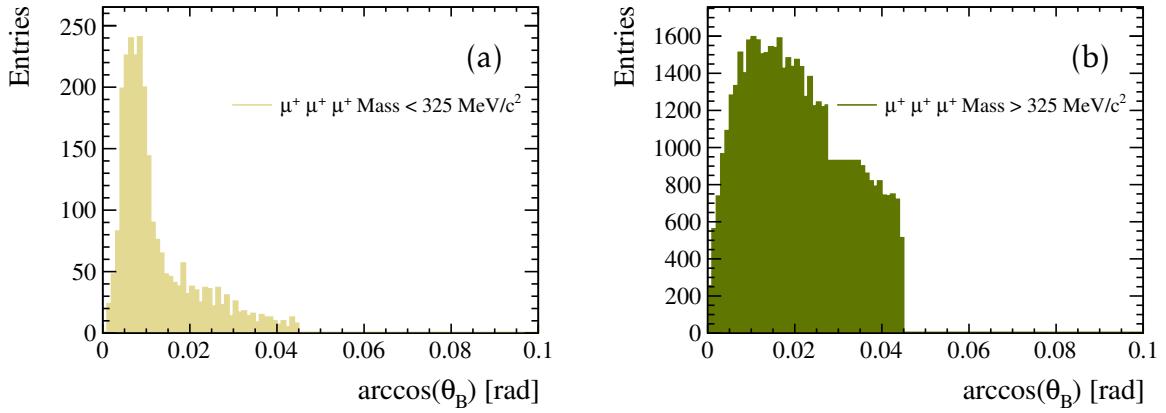
Figure 4.8: (a) Peaking clone distribution is visible as all of *clone* tracks are collinear compared to (b) smooth no clone distribution for $\cos(\theta_B)$.

and typical p,

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

Finally, corrected mass $M_{\text{corr}} = \sqrt{M^2 + |p_T^2|} + |p_T| = 4.2\,\text{GeV}/c^2$ is calculated using missing p_T from Equation 4.5 and visible mass M of *clones* from Equation 4.2 and was shown in Figure 4.5(b).

In order to suppress these tracks in analysing $B^+ \to \mu^+ \mu^- \mu^+ \nu$, where two muons have the same sign, any of the distinguishing features mentioned could be used. But the most powerful PID-wise is requiring nShared=0 in Run I, as this requirement removes all of the clones, as seen in Figure 4.9. For Run II, due to the bugs introduced, such a strong requirement would harm the signal efficiency too much so a combination of nShared<2 and isMuonTight=1 is applied instead.
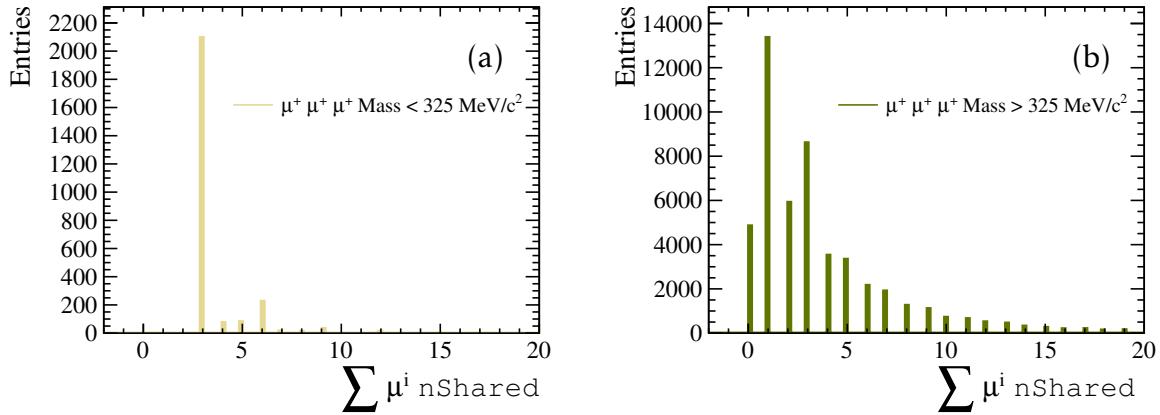
Figure 4.9: (a) Clone and (b) no clone distribution for sum of all muon `nShared`. Since in this case the clones are of each other, for the clones there is clear peak at three.

4.3 Probability of $K/\pi \to \mu$ Misidentification at LHCb

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

4.3.1 Specific Control Sample for $K/\pi \to \mu$ MisID Rates

A platform that LHCb analysts usually use to obtain the misID and ID efficiencies, as described in subsection 3.5.1, is known as the `PIDCalib` package [62]. It contains samples where the identity of the particle is known purely from kinematics. In this `PIDCalib` package, such a control same for K/π is obtained from $D^{*+}(\to D^0(\to \underline{K^+\pi^-})\pi^+)$ decays. These statistically populated background-free *sWeighted* samples [64], 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 misID background component, which is two real muon tracks with an additional *fake* muon track is very different to the `PIDCalib` sample $D^{*+}(\to D^0(\to \underline{K^+\pi^-})\pi^+)$.

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

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

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

Idea	Cut
ID K^*	$\mid M(K\pi)$ - $M_{PDG}(K_0)\mid < 100$ MeV/c^2
Muon swap veto	$\mid M((h \to \mu)\mu)$ - $M_{PDG}(J/\psi)\mid > 60$ MeV/c^2
Veto $B^+ \to K^+\mu^+\mu^-$	$\max(M(K^+\mu^+\mu^-)), M((\pi^+ \to K^+)\mu^+\mu^-)) < 5100$ MeV/c^2
Veto $B^0_s \to \phi\mu^+\mu^-$	$M(K(\pi \to K)) > 1040$ MeV/c^2
ID muons	`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 4.1: Offline selection for $B^0 \to J/\psi(\to \mu^+\mu^-)K^*$ decay.

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

In order to obtain misID and ID rates using the $B^0 \to J/\psi(\to \mu^+\mu^-)K^*$ channel, the number of signal events N needs to be obtained. The shape for the signal component, $B^0 \to J/\psi K^*$, is obtained by fixing the shape from simulation apart from the mean μ and the width σ. It is fitted with a double-sided Ipatia function [65] (more in section A.1). In addition, the residual background after selection needs to be modelled.

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

It is also possible that kaons and pions are swapped between themselves. Background coming from $K \leftrightarrow \pi$ swaps is modelled from simulation where the mass hypotheses were swapped. Its distribution is fitted with a double sided Crystal Ball function [66] (more in section A.2).

Misidentified background comes from the decay $\Lambda_b \to K^- p \mu^+ \mu^-$ where the proton is misidentified as a pion. This background is modelled from simulation and fitted with a RooKeys probability density function (PDF) (more in section A.3).

Finally a combinatorial component is modelled by an exponential function.

The mass of the J/ψ was *constrained* to its nominal mass, a procedure also known as a *mass constraint*. It yields new estimates for track parameters of the final state particles, from which a new kinematic refit is done.

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

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

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

The actual determination of the misID rate was obtained using a statistical method of background subtraction, known as the *sPlot* technique [64], as the samples are not fully background-free. The same method is also used in the PIDCalib package. In the *sPlot* method, the invariant mass distribution is fitted with no PID applied and each event is assigned *sWeights*, probabilities that a given event is signal-like or a background-like. Then, through the *sPlot* technique, background is subtracted. The signal component can then be calculated by summing all the *sWeights* for all the candidates. The misID

probabilities are finally obtained by dividing signal component sum of *sWeights* with PID applied and by the sum with no PID applied. This misID probabilities are then considered within some kinematic partitioning, bins of p, η.

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

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

It was shown that these two methods yield very similar results, hence, for purposes of the $B^+ \to \mu^+\mu^-\mu^+\nu$ analysis the *sWeight* values will be used. Fits to Run I and 2016 data for both kaon and pion misID studies can be seen in Figure 4.10.

100
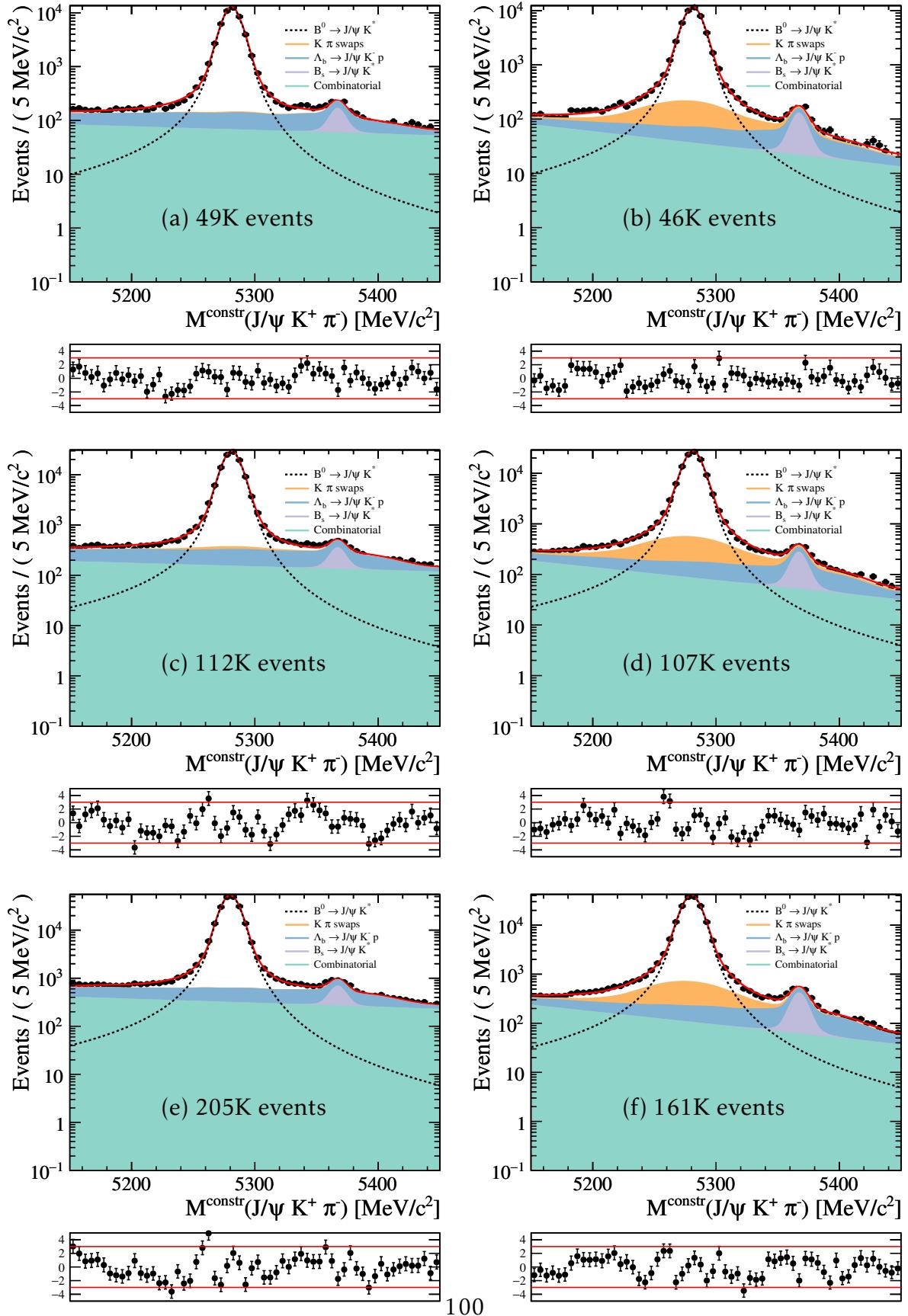
Figure 4.10: Fit to constrained $J/\psi(\to \mu^+\mu^-)K^*(\to \pi^+K^-)$ mass with all the components for (a)(b) 2011, (c)(d) 2012, (e)(f) 2016. On the left, fit to data with pion ID (giving kaon misID probabilities), on right data with kaon ID (pion misID rates).

4.3.4 Results of $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ Control Sample for $K/\pi \rightarrow \mu$ MisID Rates

Using the *sWeight* method, misID rates for kaons and pions can be obtained. Two control samples that are analysed and compared are $D^{*+}(\rightarrow D^0(\rightarrow \underline{K^+\pi^-})\pi^+)$ events (standard PIDCalib sample), where there are no other muons in the decay, and $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*(\rightarrow \underline{K^+\pi^-})$ events, where there are two real muons along with kaons or pions. Selecting only tracks that are within the muon fiducial region for both pions and kaons allows to perform study of the misID probabilities within the two calibration samples. In Figure 4.11(a), the $\pi \rightarrow \mu$ misID probability for different PID hypotheses from the $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*(\rightarrow K^+\pi^-)$ sample is studied. As it can be noticed, the more stringent the muon selection on the pion track, the lower the probability of misidentification.

In general the agreement between the two samples is good in the low momentum regions as shown in Figure 4.11(b). These pions are softer and hence they will spread out more in the magnetic field, causing less interference with two other real muons in decay. However, in the high momentum region, the pion will follow a path through the muon system that is more similar to the path of the muon of the same charge in the $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ decay. The influence of the two other real muons in a high momenta region will lead to bigger disagreement as these two real muons leave hits in the muon chambers close to the collimated pion track, making the rate of IsMuon==1.0 (pink) 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.

Even though this disagreement is decreased, for the high momenta region the $\pi \rightarrow \mu$ (Figure 4.11) and $K \rightarrow \mu$ (Figure 4.12) rates are 2 to 3 times higher with an additional two real muon tracks. Such disagreement is significant and if the misID rates from the
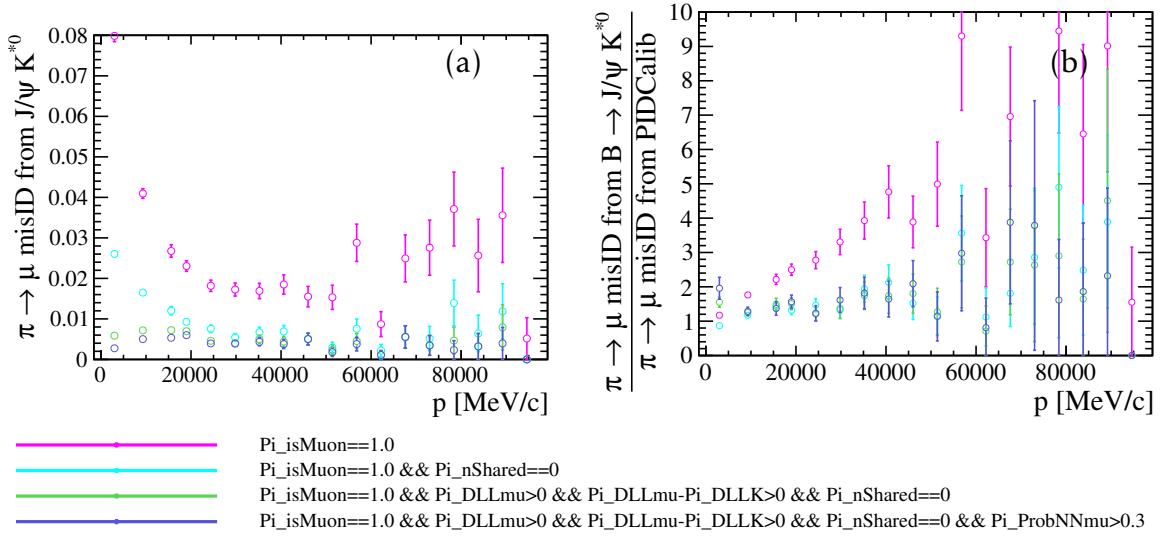
Figure 4.11: (a) $\pi \to \mu$ misID probability for different PID requirements obtained using $B^0 \to J/\psi(\to \mu^+\mu^-)K^*(\to K^+\pi^-)$ for 2012 data. (b) This is compared to the standard PIDCalib $D^{*+}(\to D^0(\to K^+\pi^-)\pi^+)$ sample. The errors shown are statistical.

standard control samples were used to estimate the misID background, there would be an underestimate the misID component by the same factor.

In conclusion, it was shown that the standard misID samples are not good proxies for estimating the misID probabilities as there is interference from the two other muons in the event. Instead, the misID probabilities that are used in calculations for the misID background for $B^+ \to \mu^+\mu^-\mu^+\nu$ are obtained from *Sweighted* $B^0 \to J/\psi(\to \mu^+\mu^-)K^*$ events. This means only Figure 4.12(a)– Figure 4.14(a) are used for calculating the misID contamination, which is discussed in detail in section 6.2. Remaining effects of taking this sample for calibration are considered as a systematic uncertainty, with more details in section 7.3.3.

Due to the different PID definitions of nShared between Run I and 2016, different PID requirements are tested. Results for $\pi \to \mu$ and $K \to \mu$ are summarized in Figure 4.13 and Figure 4.14. The misID probabilities in 2016 also show the same momentum dependent trend as in 2012.
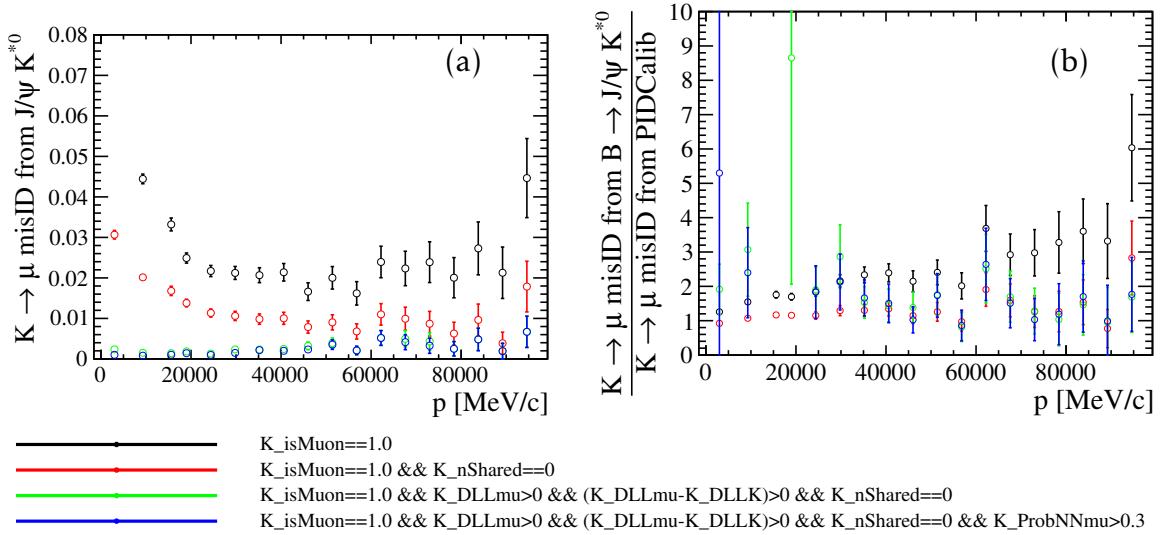
Figure 4.12: (a) $K \to \mu$ misID probability for different PID requirements obtained using $B^0 \to J/\psi(\to \mu^+\mu^-)K^*(\to K^+\pi^-)$ for 2012 data. (b) This is compared to the standard PIDCalib $D^{*+}(\to D^0(\to K^+\pi^-)\pi^+)$ sample. The errors shown are statistical. In the (b) plot, there is a big uncertainty associated with some of the bins. This is due to the lack of the statistics in PIDCalib samples. This is of no concern as the misID rates used in the analysis are coming only from (a).
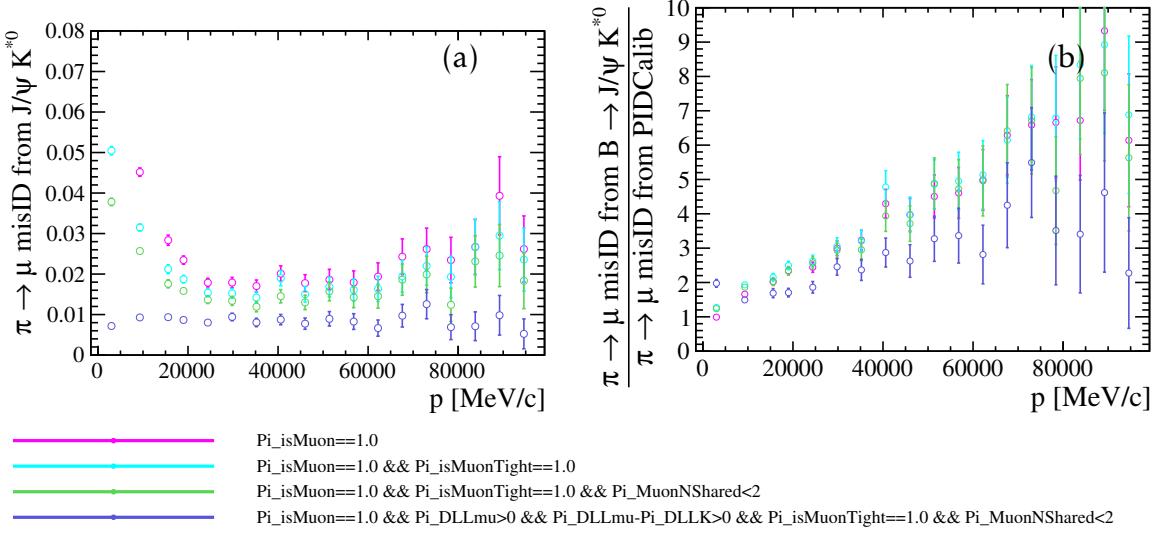
Figure 4.13: (a) $\pi \to \mu$ misID probability for different PID requirements obtained using $B^0 \to J/\psi(\to \mu^+\mu^-)K^*(\to K^+\pi^-)$ for 2016 data. (b) This is compared to the standard `PIDCalib` $D^{*+}(\to D^0(\to K^+\pi^-)\pi^+)$ sample. The errors shown are statistical.
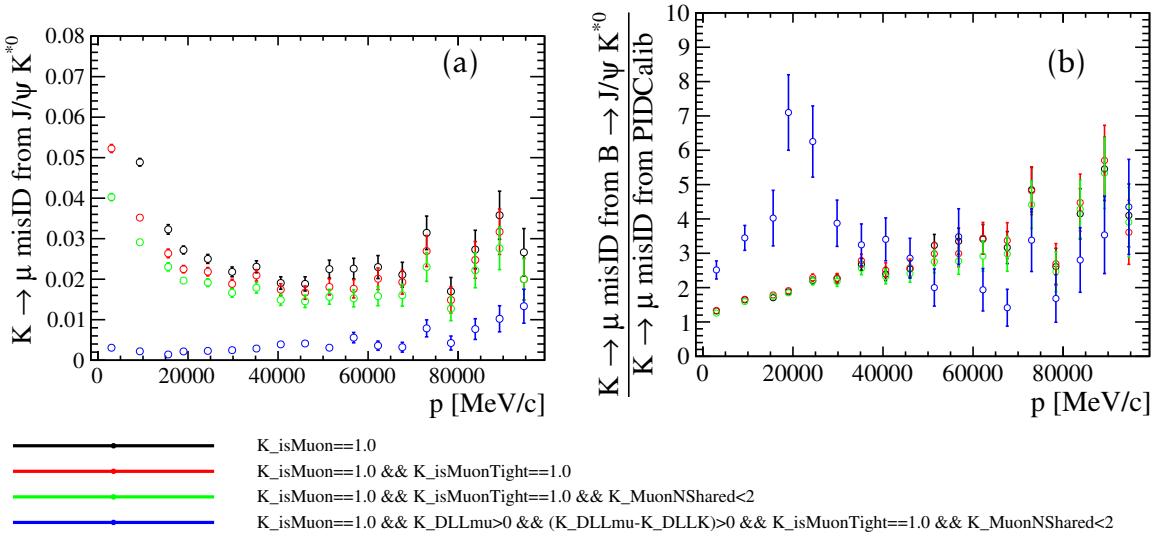
Figure 4.14: (a) $K \to \mu$ misID probability for different PID requirements obtained using $B^0 \to J/\psi(\to \mu^+\mu^-)K^*(\to K^+\pi^-)$ for 2016 data. (b) This is compared to the standard `PIDCalib` $D^{*+}(\to D^0(\to K^+\pi^-)\pi^+)$ sample. The errors shown are statistical.

Chapter 5

Looking for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ Decays at LHCb

In this chapter, the selection for the search for $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ decays at LHCb is presented. This search is challenging because of the rareness of its occurrence as well as the different backgrounds that can mimic its signature in the detector. Moreover, the presence of the invisible neutrino in the decay induces uncertainties into the reconstruction. This chapter concentrates on the data selection which reduces background contamination. The normalisation channel along with its selection is also discussed. In the end, a method to improve the sensitivity is introduced.

5.1 Analysis Strategy

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

region, described in section 5.7.

To perform the search for $B^+ \to \mu^+\mu^-\mu^+\nu$, a specific preselection was applied to form potential signal candidates as described in section 5.5. A simulation sample that mimics the decay of the $B^+ \to \mu^+\mu^-\mu^+\nu$ was created, as mentioned in section 2.11. This simulation together with the background proxies discussed in chapter 6 are used to further develop a discriminating selection that would maximise the separation between signal and background. Most of the rest of this chapter is dedicated to this discriminating selection. For more details about the signal simulation samples see section 5.4.

After the selection, the $B^+ \to \mu^+\mu^-\mu^+\nu$ decays are normalised to the $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ decays, where the selection for the normalisation channel, detailed in section 5.9, is kept as similar as possible to that of the signal channel to minimize the amount of systematics on the resulting relative efficiencies between these two channels. The relative efficiencies are computed in section 7.1. The fit to $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ data is described in subsection 7.3.2. The relative efficiencies, results of the normalisation channel fit and the branching fraction of $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ are used to parametrise the expected signal yield, as described in subsection 7.3.3.

Throughout the analysis there was a blinding procedure put in place for this search in order not to bias the result. The corrected mass of the $B^\pm$ ($M_{B_{corr}}$), detailed in Equation 5.1, is the fit variable for this search. The signal region $4500\,\text{MeV}/c^2 < M_{B_{corr}} < 5500\,\text{MeV}/c^2$ was blinded until the full strategy and sensitivity was evaluated. The signal dataset corresponding to this selection is known as **blinded signal dataset**. After unblinding, data with the full mass spectrum was obtained and is known as **full signal dataset**.

In subsection 7.3.4, two fitting strategies for the signal data are described. One is more sensitive than the other, where the fitting strategy that provides the best sensitivity makes use of simultaneous fits to two bins of resolution, increasing signal separation from background. More information about this split can be found in section 5.10. Both fitting strategies were applied on blinded signal datasets in order to compute the expected exclusion limit on the branching fraction, shown in subsection 7.3.7. In

both cases, the branching fraction of $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ decay, the quantity of interest, is directly fitted for. Systematic uncertainties that affect this measurement are presented in subsection 7.3.5.

Upon unblinding, no significant signal was observed and a stringent limit $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) < 1.4 \times 10^{-8}$ at 95% confidence level was set using the CL_s method with a simultaneous fit. All information about the result as well as the implication can be found in chapter 8.

5.2 Topology of the $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ Decay at LHCb

Upon hadronisation from a $b\bar{b}$ pair, a $B^\pm$ particle will travel around a centimetre in the laboratory frame of reference before it decays. This allows reconstruction of a primary vertex PV and its decay vertex SV. By joining these vertices, the direction as well as flight distance FD, can be established. In order to infer information about the kinematic properties of the $B^\pm$ meson, the decay products are studied. All three muons are used to reconstruct the visible four-momentum. By conservation of momentum, the neutrino is assigned all missing momentum transverse to the direction of the flight of the $B^\pm$ meson. A schematic diagram of the decay topology can be seen in Figure 5.1.

Combining all information allows for reconstruction of the corrected mass that plays a similar role to invariant mass in fully reconstructed decays. It is defined as

$$M_{\text{corr}} = \sqrt{M^2 + |p_T^2|} + |p_T|, \tag{5.1}$$

where the M^2 is the invariant visible mass squared and p_T^2 is the missing momentum squared transverse to the direction of the B^+ meson flight. It is defined in Equation 5.2. The corrected mass of the $B^\pm$ meson will be denoted as $M_{\text{B}_{\text{corr}}}$. M_{corr} is equal to the true mass if the missing part of the decay has zero mass and has no momentum along the $B^\pm$ flight direction. Otherwise, M_{corr} is below the $B^\pm$ mass.

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 (the uncertainty
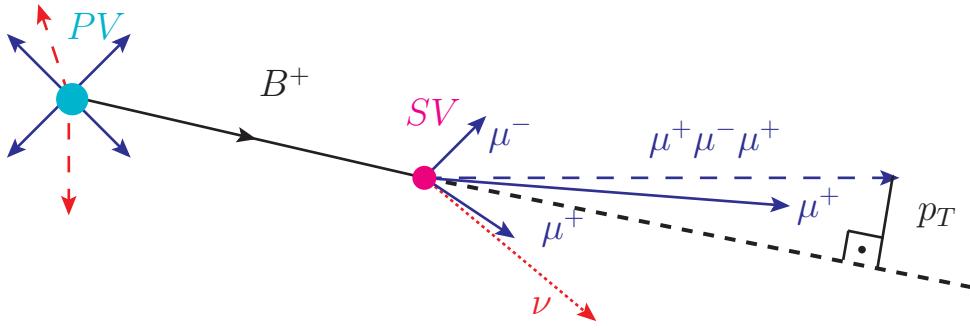
Figure 5.1: Schematic view of the $B^+ \to \mu^+\mu^-\mu^+\nu$ decay in 2D. All charged particle tracks (in solid-blue) are combined into a four-vector representing the visible part of the decay (dashed-blue). Information about the invisible neutrino (dashed-red) is deduced from the conservation of momentum with respect to the direction of the flight of the $B^\pm$ meson.

of this quantity) hence becomes a critical quantity that needs to be understood. As the method of reconstruction of corrected mass relies heavily on the knowledge of the $B^\pm$ meson flight direction, the resolution of PV position and SV vertex is crucial. Let $\vec{x}_{PV} = \{x_{PV}, y_{PV}, z_{PV}\}$, $\vec{x}_{SV} = \{x_{SV}, y_{SV}, z_{SV}\}$ be PV and SV vertex position and $\vec{p} = \{p_x, p_y, p_z\}$ be the visible trimuon momentum. Then the missing transverse momentum to the direction of the flight p_T (momentum of the neutrino) as shown in [67] is

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

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

$$\langle f^2 - \langle f \rangle^2 \rangle = \langle f(x + \delta x, y + \delta y, z + \delta z)^2 - f(\langle x \rangle, \langle y \rangle, \langle z \rangle)^2 \rangle. \tag{5.3}$$

Using a first order Taylor expansion of variance and rewriting it into matrix form gives

$$\begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix} \begin{bmatrix} \delta x^2 & \delta x \delta y & \delta x \delta z \\ \delta y \delta x & \delta y^2 & \delta y \delta z \\ \delta z \delta x & \delta z \delta y & \delta z^2 \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z}. \end{bmatrix} \tag{5.4}$$

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

With $x = \vec{x}_{PV}$, $y = \vec{x}_{SV}$, $z = p_{3\mu}$ and COV being the covariance matrix, the error (square root of variance) on the corrected mass, δ_{corr} is

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

It was shown in [67] that δ_{corr} is mostly dominated by the vertex position terms $\vec{x}_{PV}$ and $\vec{x}_{SV}$.

5.3 Sources of Backgrounds

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

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

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

The presence of a neutrino in a final state introduces uncertainty regarding the information of the fourth decay product. If some of the tracks of the decays are not reconstructed, either because they are neutral, or they are charged but with too low momentum to be found by the tracking algorithm, it means that the missing information may be attributed to the neutrino. *Missing tracks* will hence create partially reconstructed background. Some of the most dangerous are $B^+ \to D\mu^+\nu$ type partially reconstructed backgrounds, where $D^0 \to K^-\pi^+\mu^+\mu^-$.

Decays that proceed via hadronic resonances such as $B^+ \to \rho/\omega\mu^+\nu$, followed by $\rho/\omega \to \mu^+\mu^-$ are part of the signal as mentioned in section 2.10 and thus not a background.

Detailed information concerning these types of backgrounds are discussed in chapter 6.

5.4 Signal Simulation Samples

For signal simulation three different decay models are used for different purposes as summarized in Table 5.1. The physics rationale behind these models was discussed in section 2.11.

The *INSP* model is used as the default model for mass fit shapes and efficiency calculations. *PHSP* decay model is used for alternative efficiency calculations. Finally the *NIKI* model is used for validation purposes.

Channel	Year	Pythia	EVTGEN	Size	Stage
Simulation used for fitting mass shapes					
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [56]	PHSP	0.5M	*generator-level+detector*
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2012	Pythia 8.1 [40]	PHSP	0.5M	*generator-level+detector*
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [56]	INSP	0.5M	*generator-level+detector*
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2012	Pythia 8.1 [40]	INSP	0.5M	*generator-level+detector*
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2016	Pythia 8.1 [40]	INSP	1.0M	*generator-level+detector*
Simulation used for evaluating *generator-level* efficiencies					
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [56]	PHSP	25000	*generator-level*
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [56]	INSP	25000	*generator-level*
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2012	Pythia 8.1 [40]	INSP	25000	*generator-level*
Simulation used for cross-checking of *minq* selection					
$B^+ \rightarrow \mu^+\mu^-\mu^+\nu$	2012	Pythia 6.4 [56]	NIKI	25000	*generator-level*

Table 5.1: Summary of signal simulation samples used in this analysis with different decay models. In all cases, the daughters of the $B^\pm$ meson are required to be within the LHCb acceptance. All of these samples are a mixture under the two magnetic polarity conditions.

5.5 Preselection

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

Firstly, all three muon tracks are required to have a significant IP with respect to the primary vertex. Minimum Impact Parameter χ^2, min IPχ^2, gives the minimum χ^2 distance of a particle's trajectory to any primary vertex. Hence requiring min IP$\chi^2 > 9$ for muons is consistent with the hypothesis that the muon is 3σ away from any primary

vertex and hence can be well differentiated. In addition, the change in the χ^2 if the PV and SV vertices are fitted separately as opposed to a common vertex fit, FD χ^2, suppresses prompt backgrounds.

Each muon track is required to have good 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 muon stations with the binary isMuon decision. Secondly, muon candidate is chosen such that it is more likely that the candidate is a muon than a pion or kaon using global DLLmu variables defined in subsection 3.7.1. This reduces the background from misidentified muons.

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

A requirement that B^+ direction points in the same direction as the line from PV to SV is done by making sure that $\cos(\theta_B)$, angle between the two vectors, is close to unity. This translates into a well reconstructed event, which minimizes combinatorial background, where a random track makes this pointing worse. Putting bounds on the mass window, whether it is *visible* or *corrected* mass (defined in Equation 5.1), also suppresses combinatorial events.

Candidate	Stripping Selection	
muon	min IP$\chi^2 > 9$	} track
muon	track χ^2/ndof< 3	
muon	DLLmu > 0	} PID
muon	DLLmu $-$ DLLK > 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 5.2: Selection of events based on the muon and the B^+ candidate requirements. *Stripping selection* for the signal decay $B^+ \to \mu^+\mu^-\mu^+\nu_\mu$ is the same for both Run I and 2016 data.

5.6 Trigger Selection

In order to obtain triggered data, $B^+ \to \mu^+\mu^-\mu^+\nu_\mu$ candidates are required to pass a certain set of trigger decisions at L0, HLT1 and HLT2 as summarized in Table 5.3. It can be noted that the decision is applied at the mother $B^\pm$ level. In particular, a positive Bplus_L0MuonDecision_TOS decision means that one of the muons from the $B^\pm$ in an event triggered L0Muon.

As discussed in section 3.8 L0MuonDecision decides on whether an event is accepted depending on the p_T of a muon and the number of hits in the SPD. Run I can be split into 2011 and 2012 conditions where, in 2011 the most used thresholds for positive decisions are 1.48 GeV/c [68] and 1.76 GeV/c [53]. Run I SPD rate only accepts events

Trigger Selection
Bplus_L0MuonDecision_TOS
Bplus_Hlt1TrackMuonDecision_TOS
Bplus_Hlt2TopoMu2BodyBBDTDecision_TOS
Bplus_Hlt2TopoMu3BodyBBDTDecision_TOS
Bplus_Hlt2DiMuonDetachedDecision_TOS
Bplus_Hlt2DiMuonDetachedHeavyDecision_TOS

(The last four rows are grouped with a brace labelled **OR**.)

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

below 600 hits. In 2016, 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 hits.

Hlt1TrackMuonDecision accepts events where at least one identified muon has to pass thresholds on IPχ^2, p_T and p. This favours muons arising from b- and c-hadron decays. There has to be at least one muon (isMuon==true) with certain kinematic thresholds on p and p_T. For example, in 2011 the identified muons that triggered a positive decision had to have p above 8 GeV/c [68].

At HLT2 level, the candidates are required to pass through at least one of the four following decisions. The Hlt2TopoMu[2,3]BodyBBDTDecision belongs to the *topological triggers* category with an extra requirement of a particle being identified by the isMuon decision. The Hlt2DiMuonDetachedDecision and Hlt2DiMuonDetachedHeavyDecision reconstruct decays with two muons in the final state. The two lines differ in that they are optimised for heavy and light dimuon pairs respectively. For example, Hlt2DiMuonDetachedDecision accepts events with a 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 IPχ^2.

5.7 q^2 Selection

In the $B^+ \to \mu^+\mu^-\mu^+\nu$ decay, two pairs of opposite sign muons can be formed, namely $q^2(\mu_1,\mu_2)$ and $q^2(\mu_2,\mu_3)$ where $\mu_1 = \mu^+, \mu_2 = \mu^-, \mu_3 = \mu^+$. From the two invariant mass squared pairs one can define, $minq^2 = min[q^2(\mu_1,\mu_2), q^2(\mu_2,\mu_3)]$ and $maxq^2 = max[q^2(\mu_1,\mu_2), q^2(\mu_2,\mu_3)]$. This measurement is made in a region where $minq = \sqrt{minq^2} < 980$ MeV/c^2 for few reasons: most of the contributions to the amplitude of the decay are below this value, combinatorial background is greatly reduced if $minq < 1$ GeV/c^2, see Figure 5.2. The choice of 980 MeV/c^2 was made in order to avoid possible backgrounds from $\phi \to \mu^+\mu^-$.

In order to remove backgrounds that proceed via resonant J/ψ and $\Psi(2S)$ contributions, vetoes in invariant mass are placed on the corresponding regions, see Table 5.4 for more details.
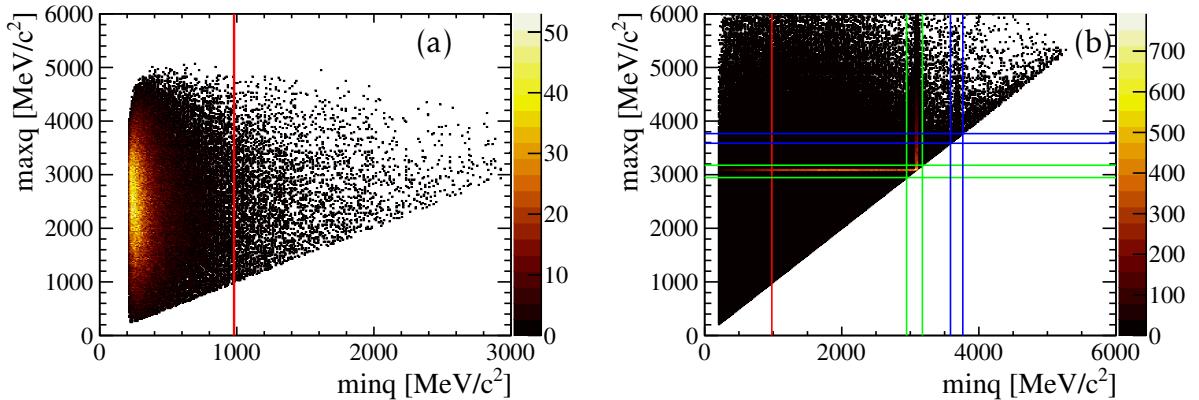
Figure 5.2: (a) Signal simulation sample distribution in *minq* and *maxq* variables. Values below 980 MeV/c^2 (red line) are accepted. (b) Combinatorial data sample after *stripping* selection with no other cuts shows clearly the J/ψ (green) and $\Psi(2S)$ (blue) resonances which are vetoed and the measurement region (red).

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

Table 5.4: Vetoes for J/ψ and $\Psi(2S)$ resonances. As $minq < 980\,\mathrm{MeV}/c^2$, these vetoes apply to the *maxq* combination only.

5.8 Further Selection

Further selection is performed with the executive summary in Table 5.5. This selection aims to further suppress backgrounds with different treatment in Run I and 2016 due to the different definitions of variables, as shown in subsection 4.1.2. The sections below comment on the more exact features of this further selection.

5.8.1 General Features of Multivariate Selections

All the multivariate classifiers in the search for the $B^+ \to \mu^+\mu^-\mu^+\nu$ decay use TMVA's [69] implementation of Boosted Decision Tree (BDT) with the AdaBoost algorithm. The multivariate selections used in the search for the $B^+ \to \mu^+\mu^-\mu^+\nu$ decay are the isolation BDT detailed in subsection 5.8.2, the combinatorial BDT detailed in subsection 5.8.3 and the misid BDT detailed in subsection 5.8.4.

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

Cross-validation is one of the useful methods used within MVAs which improves

Idea	Object	Run I Selection	2016 Selection
Clean	Muon	-	`IsMuonTight==1.0`
Clone and ghost	Muon	`Nshared==0`	`Nshared<2`
		in subsection 4.1.2	
Fit Region	B	$4000 < M_{B_{corr}} < 7000\,\mathrm{MeV}/c^2$	Same as Run I
		in subsection 5.8.5	
Bkg Removal	event	Combinatorial BDT selection	Combinatorial BDT selection
		in subsection 5.8.3	
Bkg Removal	event	Misid BDT selection	Misid BDT selection
		in subsection 5.8.4	
Optimize `FOM`	Muon	`Probnnmu>0.35`	Same as Run I
		in subsection 5.8.6	

Table 5.5: Offline selection performed after *stripping*. Differences can be seen between Run I and 2016 datasets. `FOM` is defined int Equation 5.6.

the chance of good performance of the predictive model on an independent dataset. In this way, biases due to a naive 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 the combinatorial BDT and misid BDT is known as the *k-folding* technique [70].

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 subsample. This process is repeated k-times for all possible combinations, hence the name of the cross-validation. In the last step, the results for all the samples are produced by assigning the BDT values obtained

from the tested k subsamples. In the combinatorial and misid BDT, there are 10 folds used. Both of the BDT classifiers use the same set of variables listed in Table 5.6.

$B^+ p$	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
B^+ lifetime		(subsection 5.8.2)

Table 5.6: Variables used in both the combinatorial and misID BDTs in Run I and 2016 BDTs.

5.8.2 The Isolation Boosted Decision Tree
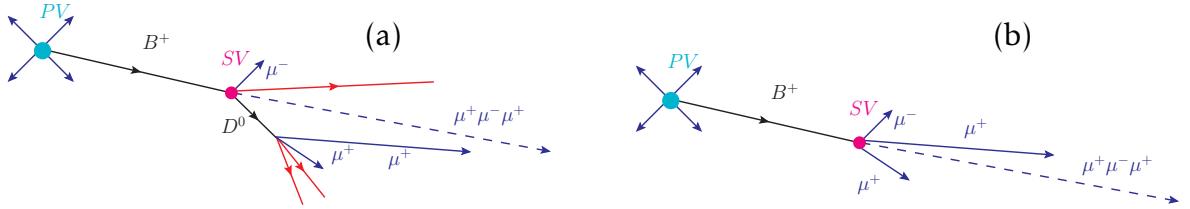
Figure 5.3: An example of decay topology for (a) background and (b) signal.

The 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 how *isolated* the signal trimuon vertex is as compared to a background trimuon vertex as seen in Figure 5.3.

The isolation quality of the vertex is determined with a BDT. This regression algorithm classifies the event to be more signal-like or background-like according to different track and vertex properties, the *isolation variables*. The *isolation variables* include track p_T, the opening angle between a track's momentum and momentum of

the combined visible system, the track $\chi^2/$ndof, the ghost probability of the track P_{ghost}, $IP\chi^2$ of the track with respect to the SV and PV.

The signal proxy for the isolation BDT was trained and tested with a $\Lambda_b^0 \to p\mu^-\bar{\nu}$ simulation sample, where all tracks apart from the $p\mu^-$ signal tracks are taken into account. The background sample consists of tracks from a Λ_c vertex from $\Lambda_b^0 \to (\Lambda_c \to pX)\mu^-\bar{\nu}$ decays, disregarding the $p\mu^-$ tracks (i.e the X). The isolation BDT is based on the weights obtained from these samples, which are computed in [71]. These weights are applied to the $B^+ \to \mu^+\mu^-\mu^+\nu$ signal and background proxies, as they share similar topology with respect to isolation properties.

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). An event is considered signal-like if all other tracks apart from the $p\mu^-$ tracks in the event are isolated from the $p\mu^-$ vertex. The output of this BDT for both types is shown in Figure 5.4. Backgrounds shown include combinatorial background and misID type background. In the analysis, there is no explicit selection on this variable, but it is used as one of the input variables for the combinatorial and misid BDTs.
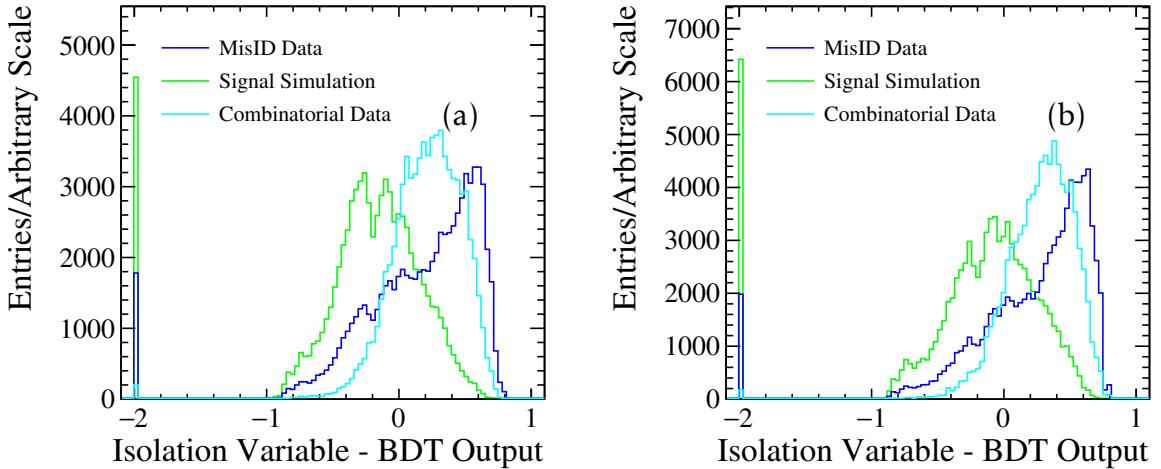
Figure 5.4: Isolation score for signal and backgrounds using (a) Run I (b) 2016 samples. If isolation fails to find any other track than $p\mu^-$ tracks in the event, by default it gives the value -2.

5.8.3   The Combinatorial Boosted Decision Tree

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

As the branching fraction and hence the number of signal events is unknown, the metric known as the Punzi figure of merit (FOM) [72], is used to find an optimal working point. It is defined as

$$\mathrm{FOM} = \frac{\varepsilon_S}{\sqrt{B} + n/2}, \tag{5.6}$$

where ε_S is the signal efficiency of the selection, B refers to the number of background candidates and n is the significance that the analysis aims to minimally achieve. In this case, the significance 3σ is used, but it was checked that there is no change to the optimal working point if it is varied to 5σ, as seen in Figure 5.5.

The FOM is computed in the blinded mass region, $4.5\,\mathrm{GeV}/c^2 < M_{\mathrm{B_{corr}}} < 5.5$ GeV/c^2 as this region is most sensitive to the signal. From Equation 5.6 it is necessary to compute the signal efficiency and number of background candidates. To estimate the number of background candidates in the blinded region, the final fit strategy described in subsection 7.3.4 is used to fit the data. This yields around 10000 and 9000 combinatorial candidates in Run I and 2016. The yields are extracted from blinded fits to data by integrating the combinatorial part of the total background PDF in the blinded region.

In order to accommodate different selections between Run I and 2016, separate BDTs are trained for these periods. Combined training of all of the datasets was also performed but it does not lead to any improvement in background rejection. Results of the comparison between separate and combined training can be seen in Figure 5.6.
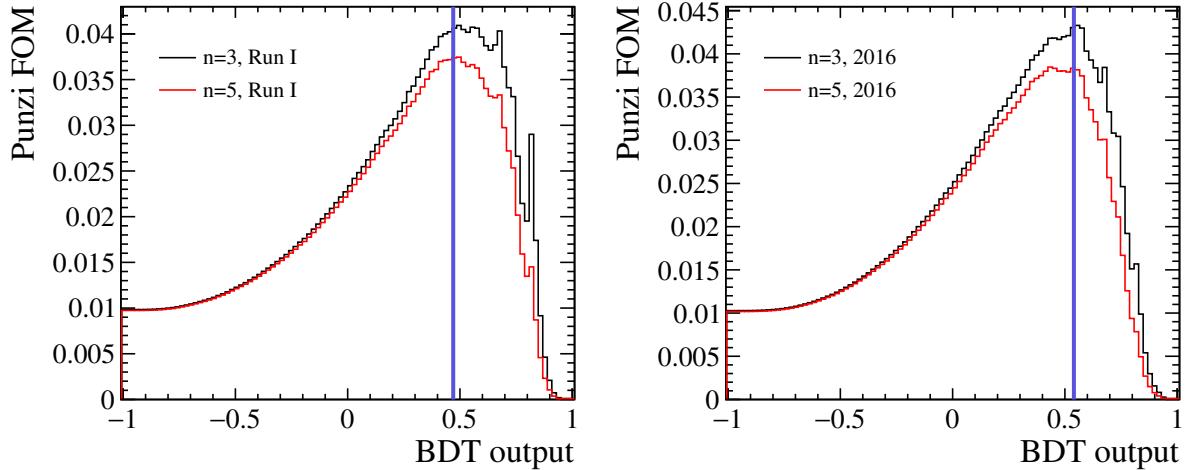
Figure 5.5: Punzi FOM shows the optimum working point at 0.47 for Run I and 0.54 for 2016 as seen in both figures with a violet line for $n = 3$ and $n = 5$. This FOM is for Combinatorial BDTs.

Different intrinsic properties (such as the number of trees used) and variables (such as two-particle vertices) have been explored but no improvement in discrimination of the BDT was achieved.

In both Combinatorial BDTs, the most discriminating variables are the isolation variable (described in subsection 5.8.2), B^+ vertex χ^2/ndof, min IPχ^2 of the muons and p_T of the B^+ meson. A Combinatorial muon comes more from somewhere else in the event and hence the min IPχ^2 is worse as compared to the signal, making the B^+ vertex χ^2/ndof worse. Moreover, as this combinatorial muon comes from somewhere else, other tracks may accompany it making the isolation variable a good discriminant. The combinatorial muon also tends to have higher momentum and hence p_T of the B^+ is higher. Distributions for these different variables can be seen in Figure 5.7.

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 the BDT output in different bins of $M_{B_{corr}}$. If the BDT value stays flat then the background will not be
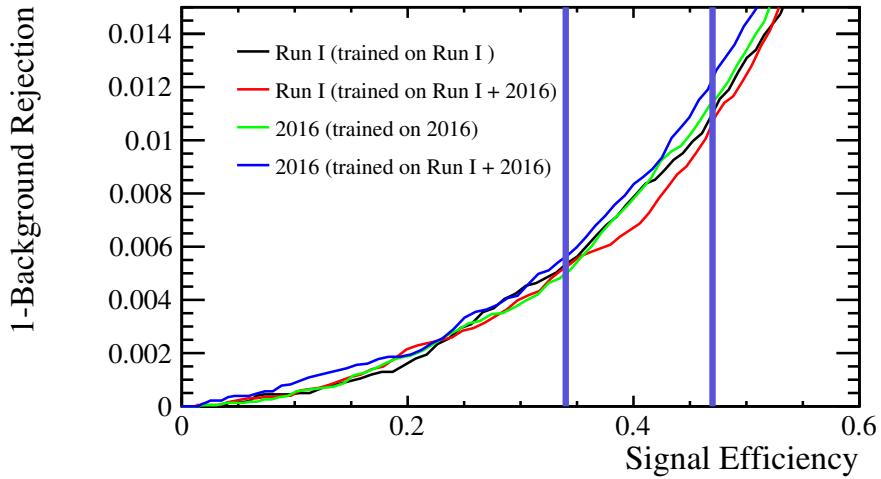
Figure 5.6: Comparison of separate and combined training samples and performance on different datasets. Two vertical violet lines represent optimal points in the signal efficiency, for Run I (0.47) and for 2016 (0.34) where the working point of the two BDTs are chosen. Separate training provides greater rejection power in 2016. In Run I training on both datasets provides comparable performance for a given optimal signal efficiency. Taking into the account the fact that selection slightly differs for 2016, it is advantageous to keep the BDTs separate.

skewed, which is the case for 2016 as seen in Figure 5.8. This is also the case for Run I.

5.8.4 The Misid Boosted Decision Tree

In the same way, the classifier that distinguishes between signal and misID background was developed. The misID sample, that is used for training and testing, was obtained the same way as the signal but with one of the muons not identified as the muon. Rather, this third particle will be identified either as a proton, pion or kaon. More about the parametrisation of this background can be found in section 6.2. These samples went through all the previous selection including the application of Combinatorial BDTs. As before, Run I and 2016 are trained and used separately on the relevant datasets, as shown in Figure 5.9.
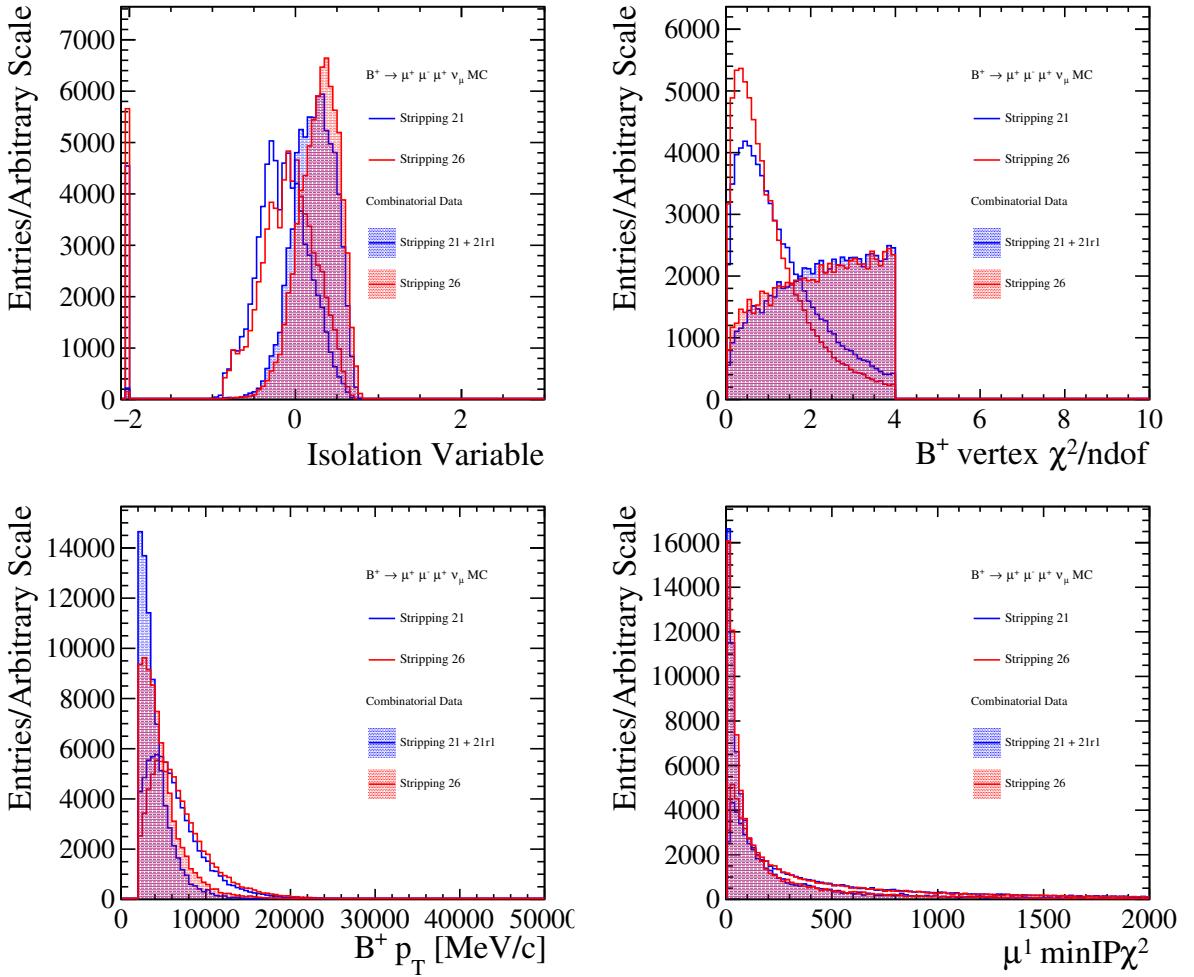
Figure 5.7: The variables with the most discriminative power for both Run I and 2016 Combinatorial BDTs. In these plots μ^1 is one of the muons with a charge that agrees with the mother B, so if the mother is B^+ then μ^1 is one of the positively charged muons (μ^+).

The optimisation metric for this classifier was again the Punzi FOM in a blinded region. The Punzi FOM for Run I and 2016 as a function of BDT cut can be seen in Figure 5.10 for both significances of $n = \{3, 5\}$.

To obtain the number of background events, the default fitting strategy for misID is used in section 7.3.3, where the total yield need to be multiplied by 100 in order to counter balance the prescale used at the pre-selection stage. To obtain the yield,
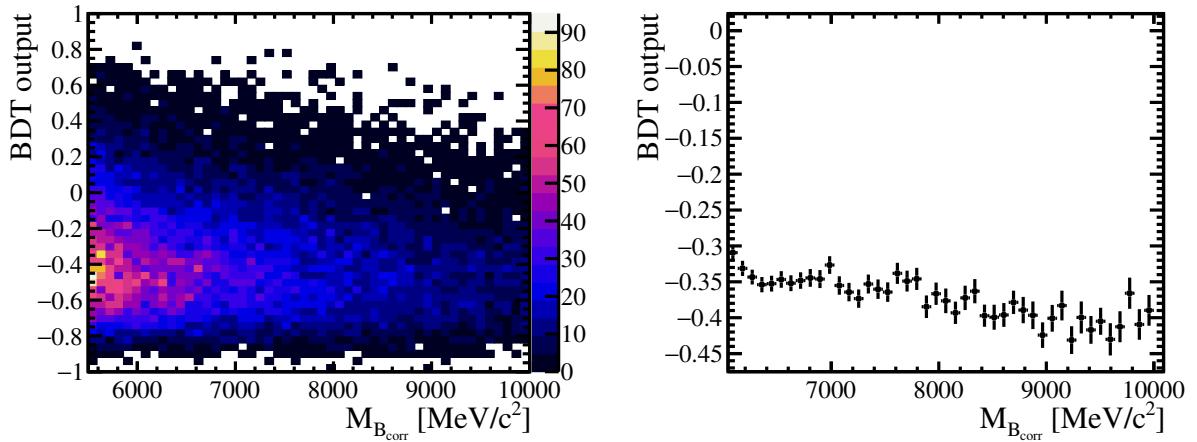
Figure 5.8: Study of linear correlation between BDT output and $M_{B_{corr}}$ and BDT value for each bin of $M_{B_{corr}}$ in 2016 shows that the Combinatorial BDT is relatively flat as a function $M_{B_{corr}}$. The right plot shows the mean and uncertainty on the mean of the 2016 Combinatorial BDT in bins of the $M_{B_{corr}}$. The full correlation matrix with all other variables is listed in Figure B.1.

a binned χ^2 fit is performed. The binned χ^2 fits to the misID templates are shown in Figure 5.11 yielding 2400 unparametrized misID candidates in Run I polluting the signal window in the prescaled sample, and 2200 in 2016.

The misID background can proceed also through combination with a random muon and hence by applying the combinatorial BDTs on the misID samples, this "combinatorial" component in the misID samples should be reduced and misID samples that are left should consist of true cascade decays. This can be seen in Figure 5.12.

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 for the signal as they come from cascades via D^0 decays and its excited states. The min IPχ^2 distribution is also different as the misidentified muon can proceed from D^0 decays, whereas the signal muon comes directly from the B meson. The kinematic distributions are also different for the two real muons between signal and misID background samples. The real muon that has
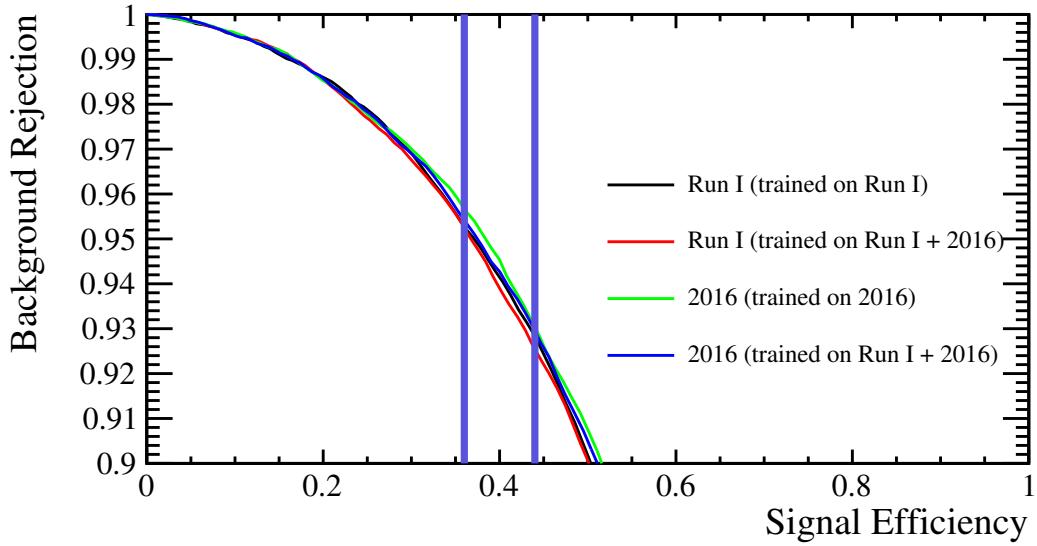
Figure 5.9: Comparison of separate and combined training samples and performance on different datasets. The optimal working point is chosen, see Figure 5.10 and its corresponding signal efficiency in Run I is 0.44 and for 2016 0.37 denoted with a violet line. As the performance is better for 2016 when the training is performed separately, the training is kept separate also to be consistent with the previous methodology.

the same charge as the B meson tends to be softer for the signal case whereas the real muon that has opposite charge proceeding via the D meson will be harder, as seen in Figure 5.13.

5.8.5 Fitting Region Selection

Because the signal mass distribution is expected to be in a more narrow window around the B^+ peak in corrected mass and the exponential description of the combinatorial background is not correct below 4000 MeV/c^2, as will be shown in section 6.1, the region in which the measurement will be made is selected as 4000 MeV/$c^2 < M_{B_{corr}} <$ 7000 MeV/c^2.
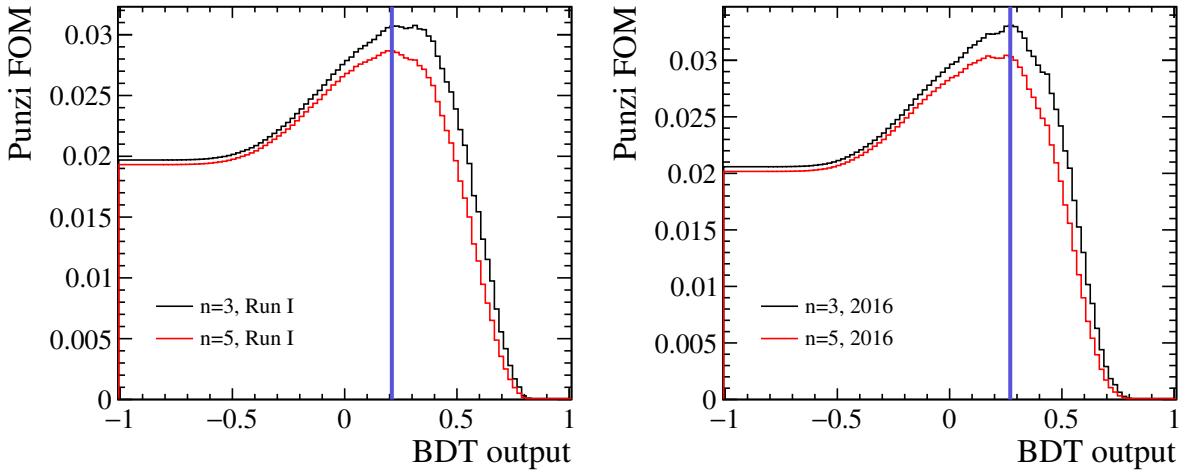
Figure 5.10: Punzi FOM have the optimum working point at 0.21 for Run I and 0.27 for 2016 as seen in both figures with a violet line for $n = 3$ and $n = 5$. These FOM values are for misid BDTs.
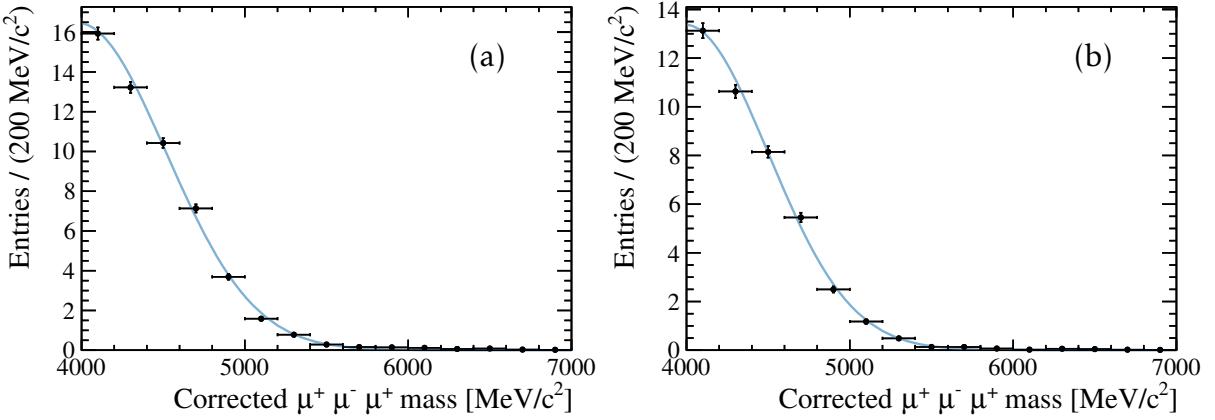
Figure 5.11: (a) Run I (b) 2016 binned χ^2 fit to misID sample yielding estimates for the number of background events.

5.8.6 Further PID Selection

After classifiers to reduce combinatorial and misID backgrounds are trained and applied and the fitting region is defined, further PID selection is performed. This can be done as the preselection had relatively loose DLL requirements and hence it is possible to
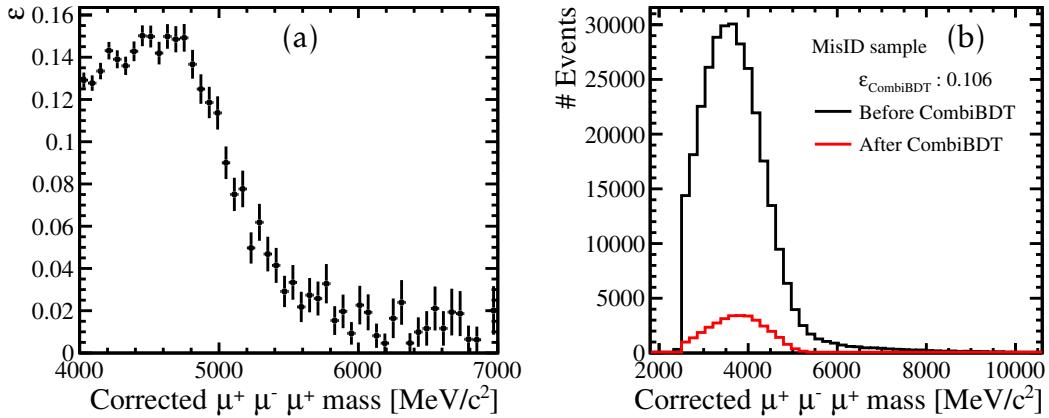
Figure 5.12: (a) The efficiency of applying 2016 combinatorial BDT at the optimal working point on the 2016 SS misID sample. It can be seen in (b) that the combinatorial component of the misID sample has been significantly reduced, where the red curve is the distribution after applying the cut.

improve the performance by using cuts on additional PID variables. In the optimisation procedure, different hypotheses were tested, such as cuts on Probnnmu, Probnnpi, and ProbnnK variables and their combinations. The optimisation was performed in such a way as to optimize the Punzi FOM with $n = 3$ in a blinded signal region, by performing the full blinded data fit (more information will be given in subsection 7.3.6), but with fits to Run I and 2016 data separately.

For each PID hypothesis, the Punzi FOM was calculated. In both cases, in Run I and 2016 Probnnmu > 0.35 yielded the highest Punzi FOM. This is the final step in selection.

The overall selection results in 1198 candidates when the signal region is blinded. There are no events with multiple candidates. The total efficiency for selecting the signal is about 0.1%. The blinded data can be seen in Figure 5.14.
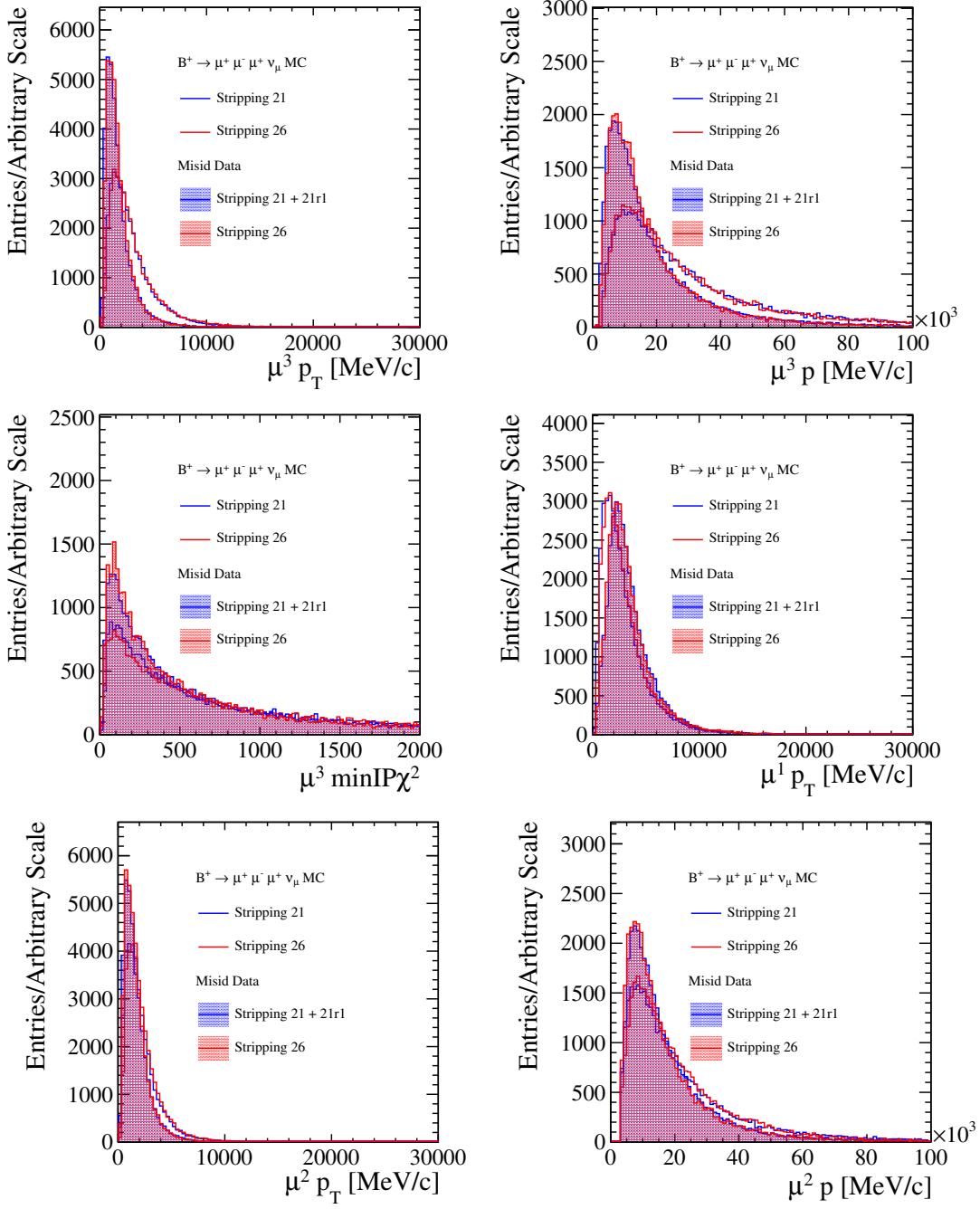
Figure 5.13: The variables with the most discriminative power for both misid Run I and 2016 BDTs. In these plots, for signal samples μ^1 and μ^3 are the muons with the charge that agrees with the mother B, so if the mother is B^+ then μ^1 and μ^3 are the positively charged muons (μ^+), and μ^2 is negatively charged muon (μ^-). For misID data samples, μ^1 is a true muon with the charge that agrees with the charge of the mother B, μ^2 is a true muon with the charge that does not agree with the charge of the mother B and finally μ^3 is the misidentified particle whose charge agrees with the charge of the mother B.
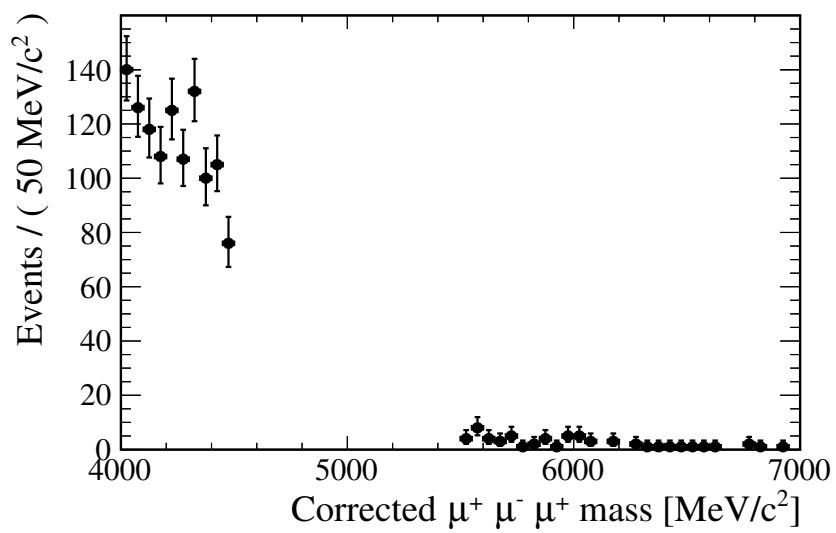
Figure 5.14: The blinded dataset after full selection.

5.9   Normalisation Channel

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

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

Secondly, all the offline signal selection cuts listed in Table 5.5 are also applied apart from a few exceptions. As this decay proceeds via the J/ψ resonance and hence the q^2 veto for J/ψ and $\Psi(2S)$, listed in Table 5.4, is not applied but rather reversed as seen in Table 5.7. As the third particle is a kaon rather than a muon, there is no explicit choice of $minq$ region.

And finally, the kaon candidates are required to have PID criteria consistent with being kaons. In addition to the preselection already imposing DLLK > 0, DLLp − DLLK < 5 is required to make a distinction with protons. To assure that the kaon is not confused with the muon, `IsMuon==0.0` is imposed. However, only kaon tracks within the geometrical muon acceptance, `InMuonAcc==1.0`, are considered.

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

Table 5.7: Original preselection of events for normalisation channel for $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ for Run I and 2016.

5.10   Fractional Corrected Mass Error (FCME) Window Split

In order to increase sensitivity, but not to decrease statistics as all the previous selection leads to a low-statistics regime, it was decided that the fitting procedure for the final fit will be in two bins of the estimated fractional corrected mass error (FCME), defined as

$$\sigma_{\{\text{lowFCME},\text{highFCME}\}} = \frac{\delta_{\text{corr}}}{M_{\text{B}_{\text{corr}}}}, \tag{5.7}$$

where δ_{corr} is the estimate of the corrected mass error. Because the corrected mass error has a clear dependence on resolution (see Figure 5.15), this split will divide the data into two bins of resolution increasing the sensitivity for observation as shown in Figure 5.16.
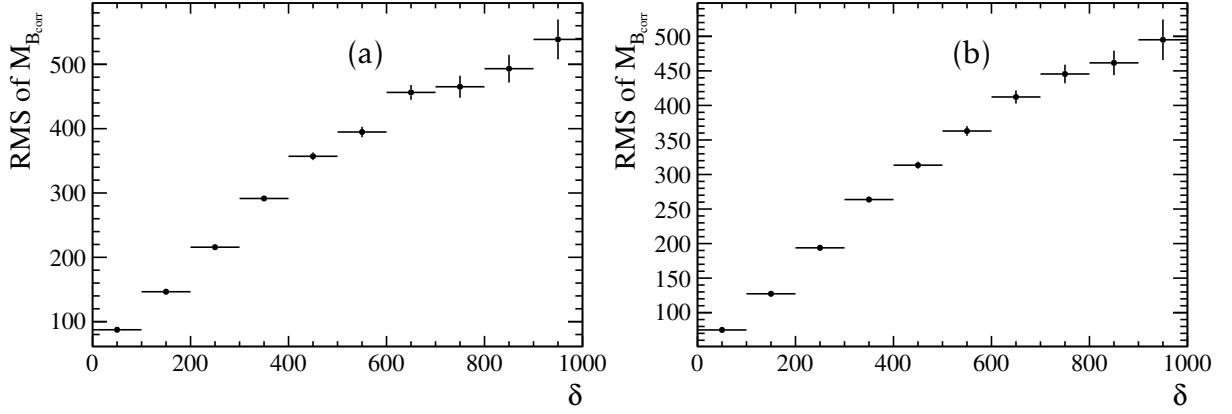
Figure 5.15: (a) The resolution of 2012 signal simulation in bins of the estimated corrected mass error δ_{corr}. (b) The resolution of 2016 signal simulation in bins of corrected mass error δ_{corr}.

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

$$\sigma_{\{\text{lowFCME},\text{highFCME}\}} = \begin{cases} \sigma_{\text{lowFCME}} & \text{if } \dfrac{\delta_{\text{corr}}}{M_{B_{\text{corr}}}} < 0.0225, \\[2ex] \sigma_{\text{highFCME}} & \text{if } \dfrac{\delta_{\text{corr}}}{M_{B_{\text{corr}}}} > 0.0225. \end{cases} \tag{5.8}$$
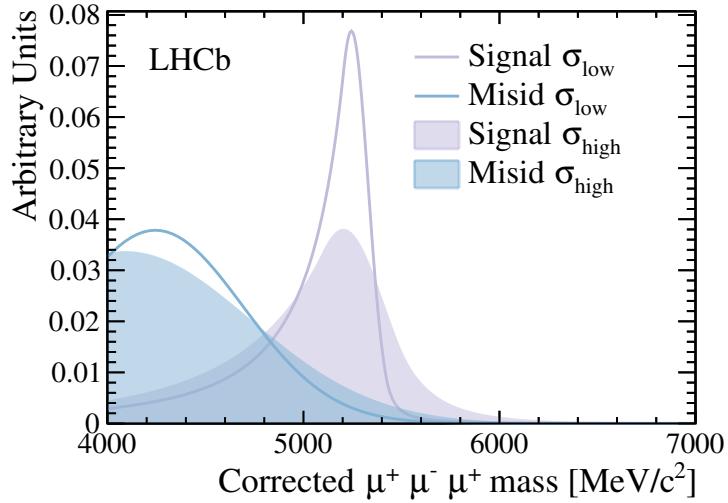
Figure 5.16: Templates for signal and misidentified background shapes in high and low fractional corrected mass uncertainty. It can be seen how a low uncertainty on the corrected mass corresponds to data with better mass resolution. The shape of the misidentification template is obtained from a control sample while the signal template is obtained from simulation. These templates are constructed after the full selection was applied.

However, in order to look at the consistency of this two bin strategy fits will also be performed with no binning in the fractional corrected mass error. These are denoted as σ_{NOFCME}.

Chapter 6

Background Studies

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

6.1 Combinatorial Background

The combinatorial background is when a random combination of tracks from different *b*-decay chains fakes the signal. The usual method at LHCb of estimating the amount and the shape of this background include extrapolation from the upper mass data sideband to the signal region. In this case, the upper mass sideband is defined as $M_{\mathrm{B_{corr}}} >$ $5500\,\mathrm{MeV}/c^2$ and the signal region is defined to be $4500\,\mathrm{MeV}/c^2 < M_{\mathrm{B_{corr}}} < 5500\,\mathrm{MeV}/c^2$. This background can be described by an exponential function in a certain range, where this range is the primary discussion of this section. Since the tight selection results in small data samples, the extrapolation from the upper mass sideband introduces a large uncertainty on the exponential constant and cannot be used to estimate the correct shape and yield of this background. What can be done, however, is to assume the exponential shape for the combinatorial component and let the exponential constant be a floating parameter in the data fit. This method for estimation of the combinatorial

134

component will be mentioned in the signal data mass fits, in subsection 7.3.3.

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

As seen in Figure 6.1(b)(c), the exponential shape is only valid for $M_{B_{corr}} > 4000$ MeV/c^2. Hence the choice of fitting region 4000 MeV/$c^2 < M_{B_{corr}} < 7000$ MeV/c^2.
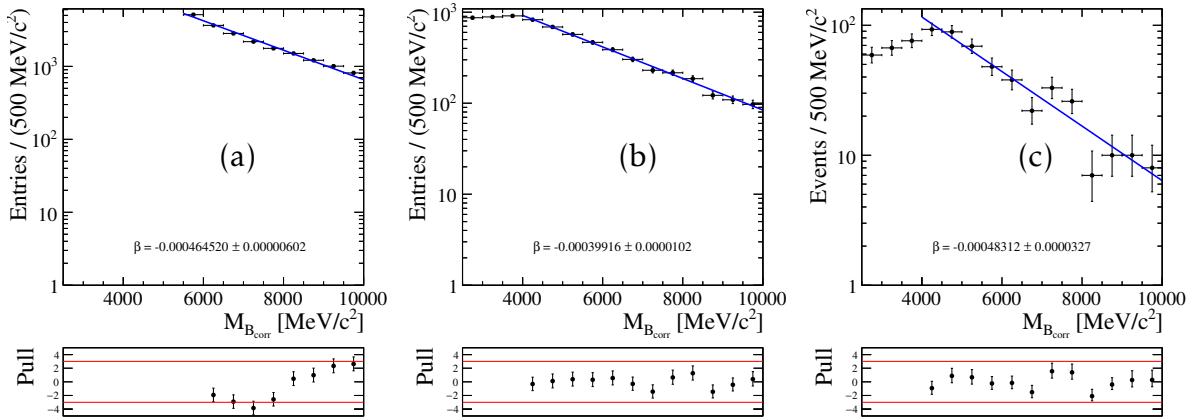
Figure 6.1: (a) Fit to upper mass side band just before application of MVA selection. (b) Fit to $\mu^+\mu^+\mu^+\nu$ same sign sample. (c) Fit to $b\bar{b}$ sample with exponential function. In (b) and (c), an exponential description is not correct below 4000 MeV/c^2.

6.2   MisID Type Background

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

As discussed in section 5.3 there are two possibilities for the charge for the misidentified background. In one case the sign of the misidentified particle agrees with the sign of the mother B, *SS misID* background. The opposite case, denoted as *OS misID* decays, arises less often as it requires decays with more additional particles. These two types of background are studied using the data-driven method described below. Finally, also double misID background was studied employing the same data-driven methods. The contribution from these events with two hadrons misidentified as muons proved insignificant.

To determine the amount and the shape of the misID background, a data sample with the same selection as for the signal sample is obtained with one marginal difference - **no PID cut** on one muon, either positive or negative. As the muon misID rate is different for pions and kaons [52], the species of the hadron, h must be determined first. The strategy for this purpose is to isolate the hadron into separate hadron PID regions, and to determine the cross-feed of one region into the other. For this, an iterative procedure as shown in Figure 6.2 is applied, ignoring the insignificant proton cross-feed. This iterative procedure splits the misidentified data sample into PID regions, where the hadron candidate is consistent with the kaon, pion and proton hypotheses. For this procedure, probabilities of identifying a given species with a given PID requirement are taken from dedicated control samples in the PIDCalib package [62] as discussed in subsection 4.3.1. The PID performance is highly dependant on the kinematic properties of the misidentified particle and hence the estimation is

performed in bins of momentum p and pseudorapidity η. At the beginning of the procedure, the number of misidentified events of a given species is assumed to be zero, and the cross-feed between regions is calculated assuming that the pion, kaon and protons regions are pure pions, kaons and protons. The procedure then corrects the distributions by taking into the account this initial cross-contamination. This procedure is repeated until the number of total misidentified particles does not change significantly from one iteration to another.
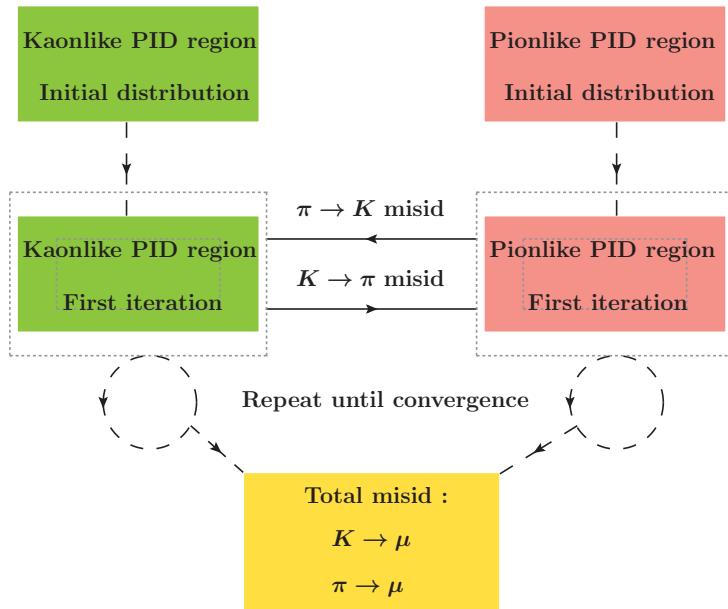
Figure 6.2: Diagram of the iterative procedure to establish contamination from decays where pions and kaons are misidentified as muons.

Once the cross-feed between the different hadron species has been taken into account, the final step is to calculate the probability for a specific hadron to pass the stringent muon PID requirements applied in the analysis. The presence of the two real muons in the $\mu^+\mu^- hX$ background increases the probability to misidentify the hadron as a muon, mainly due to sharing of hits in the muon stations. Therefore the hadron misID probability is obtained from a dedicated control sample designed to emulate the topology of the mis-identified background, $B^0 \to J/\psi(\to \mu^+\mu^-)K^*(\to \underline{K^+\pi^-})$, as shown in subsection 4.3.4.

This process can be summarized mathematically in the following way:

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

Region	PID cuts
Proton-like	DLLp > 5, DLLp - DLLK > 5
Kaon-like	DLLK > 0, DLLp - DLLK < 5
Pion-like	DLLK < 0, DLLp < 5

Table 6.1: Species region definitions.

- ID efficiencies are obtained from PIDCalib in bins of p, η for all three regions.

- MisID efficiencies are obtained from the specific calibration sample with two other muons in the sample in bins of p, η.

- In order to account for cross-contamination between the kaon and pions species the following procedure is applied:

  - The data in each region is binned to obtain two dimensional $N(p,\eta)$ distributions. The true kinematic distributions for kaons and pions are given by

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Examples of misID distributions with unweighted, weighted by probability with no cross-feed correction, and weighted with cross-feed correction can be seen in Figure 6.3 for the *SS misID* and Figure 6.4 for the *OS misID*. These are the misID distributions before misid BDTs are applied, which minimize the contamination of this background as discussed in subsection 5.8.4.
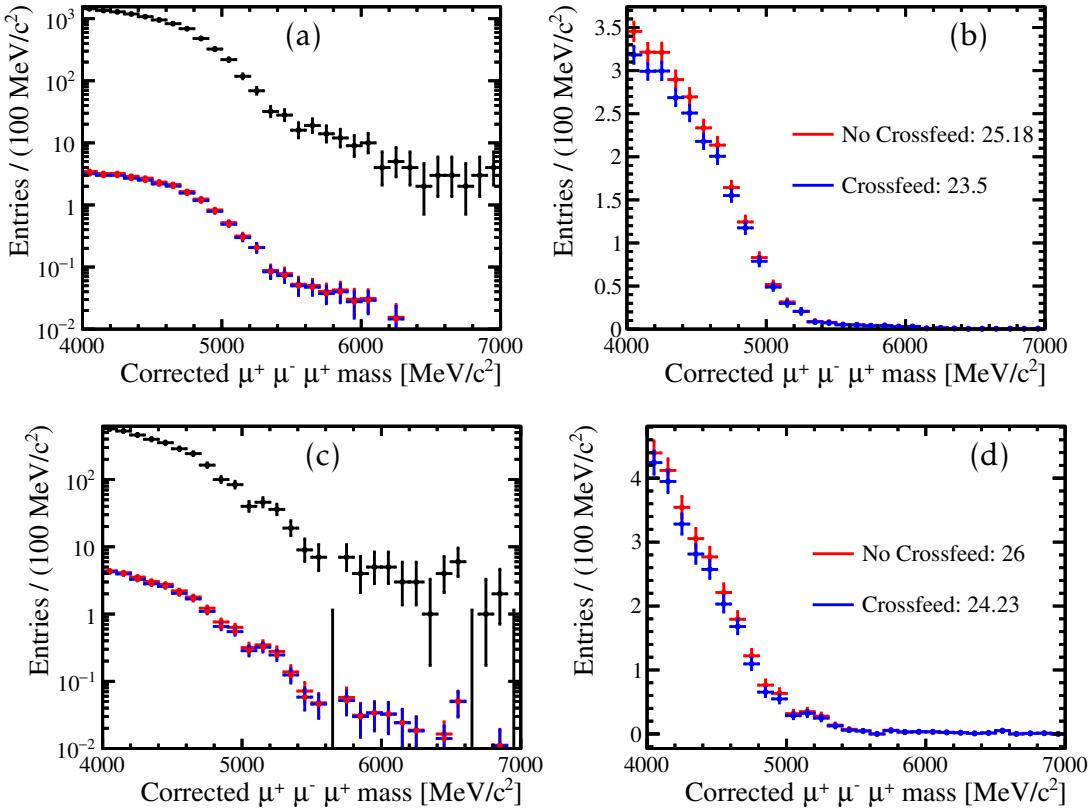
Figure 6.3: Examples of the data distributions where the misID procedure is applied to obtain yields and shapes for Run I. On the left, unweighted misID distributions (black), weighted with no cross-feed misID distributions (blue) and weighted misID distributions with cross-feed (red) for (a) kaon SS (c) pion SS . On the right, only weighted misID distributions for Run I (b) kaon SS (d) pion SS are shown together with the yield estimates. These shapes are obtained after the combinatorial BDT was applied, but before the misid BDT was applied. Total yields need to be multiplied by 100 to counteract the prescale that was applied on this data.

This parametrisation of the misID background is crucial for the final fit model described in section 7.3.3.
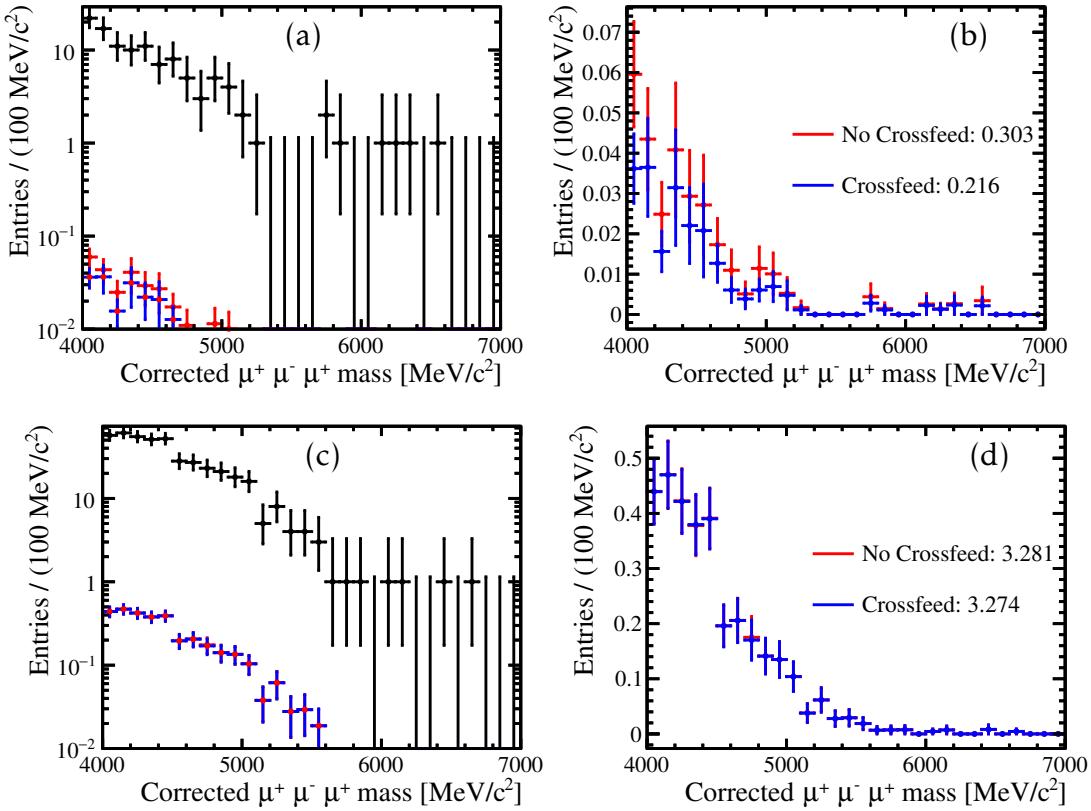
Figure 6.4: Examples of data distributions where the misID procedure is applied to obtain yields and shapes for Run I. On the left, unweighted misID distributions (black), weighted with no cross-feed misID distributions (blue) and weighted misID distributions with cross-feed (red) for (a) kaon OS (c) pion OS. On the right, only weighted misID distributions for Run I (b) kaon OS (d) pion 0S are shown together with the yield estimates. These shapes are obtained after the combinatorial BDT was applied, but before the misid BDT was applied. Total yields need to be multiplied by 100 to counteract the prescale that was applied on this data.

6.3 Partially Reconstructed Background

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

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

In order to estimate both the shape and the size of the partially reconstructed background, one of the most dangerous examples is studied: $B^+ \to (D^0 \to K^- \pi^+ \mu^+ \mu^-) \mu \nu$. The expected $\mathcal{B}(B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$ is obtained by multiplying $\mathcal{B}(D^0 \to K^+ \pi^- \mu^+ \mu^-) = (4.17 \pm 0.12 \pm 0.40) \times 10^{-6}$ [73] and $\mathcal{B}(B^+ \to Dl^+ \nu X) = (9.8 \pm 0.7) \times 10^{-2}$ [13] yielding $\mathcal{B}(B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu) \approx (4.1 \pm 0.5) \times 10^{-7}$.

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

For this reason all muon cuts from the selection apart from the PID are also applied to the pions. Relative efficiency ratios including all the efficiencies after the MVA stage are obtained, where for signal the total selection efficiency is $\varepsilon^{total}_{B^+ \to \mu^+ \mu^- \mu^+ \nu} = (2.63 \pm 0.03) \times 10^{-3}$ and for the partially reconstructed background $\varepsilon^{total}_{B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu} = (6.82 \pm 0.07) \times 10^{-4}$. Assuming the branching fractions for $\mathcal{B}(B^+ \to \mu^+ \mu^- \mu^+ \nu) = 1 \times 10^{-8}$ and for $\mathcal{B}(B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu) = (4.1 \pm 0.5) \times 10^{-7}$, the relative contamination between signal and partially reconstructed background results in Figure 6.5(a).

To check the fact that there is no dangerous shaping of the background using this particular proxy simulation sample, the full selection efficiency in bins of dipion mass is plotted. The flat efficiency shown in Figure 6.5(b) means that this selection does not have model dependence and hence it is safe to use for shape estimates for the partially reconstructed backgrounds.
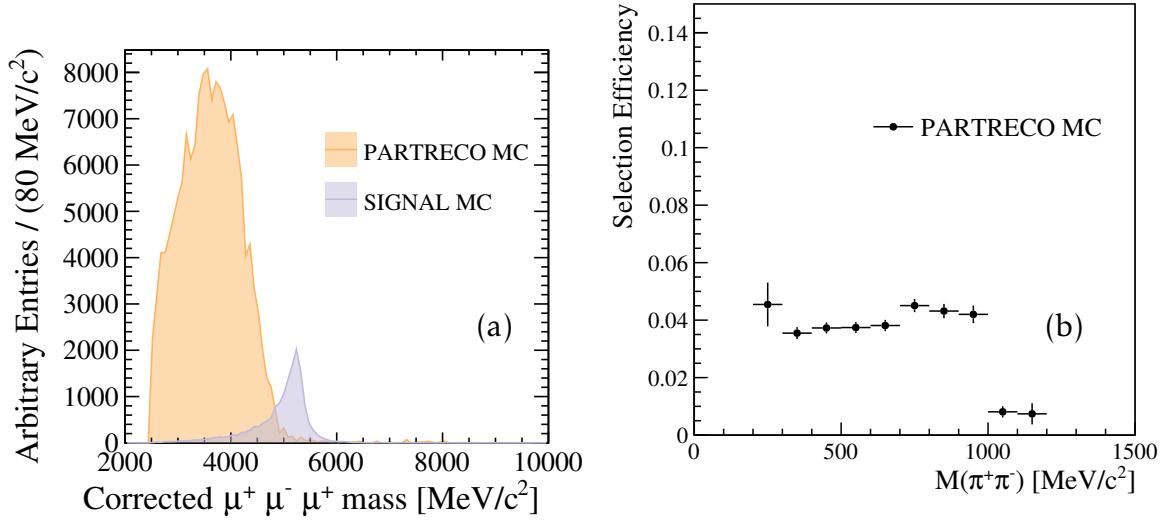
Figure 6.5: (a) Signal and partially reconstructed background distributions scaled to their expected ratio after the full MVA selection assuming the following branching fractions: $\mathcal{B}(B^+ \to \mu^+ \mu^- \mu^+ \nu) = 1 \times 10^{-8}$ and $\mathcal{B}(B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-)\mu^+ \nu) = (4.1 \pm 0.5) \times 10^{-7}$. (b) Full selection efficiency as a function of invariant mass of the proxy pions is constant.

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

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

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

In this section, the total estimate for the branching fraction of the partially re-

constructed backgrounds proceeding via η/η' from D decays is computed. The full inclusive rate $\mathcal{B}(D^0 \to \eta/\eta'X)$ is $\sim 10\%$). However, the most relevant decay chains are the ones where the mass of the missed particle(s) is small. This is because if only a light particle is missed, the shape of the corrected mass of partially reconstructed background comes closest to the signal region. Such decay chains are considered in Table 6.2.

It can be seen that the total cumulative contribution is much smaller then the one considered with $D^0 \to K^+\pi^-\mu^+\mu^-$, where $\mathcal{B}(D^0 \to K\pi^+\mu^+\mu^-) = (4.17 \pm 0.12(\text{stat}) \pm 0.40(\text{syst})) \times 10^{-6}$ [73]. No further consideration hence is necessary for this type of decay.

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

Table 6.2: Contribution to total $D^0 \to (\eta/\eta' \to \mu\mu\gamma)X$ rate made from all the decays considered above. In total, this cumulative contribution is approximately three times smaller than $D^0 \to K^+\pi^-\mu^+\mu^-$. All the branching fractions are obtained from [13].

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

The backgrounds with $\eta(')$ resonances from partially reconstructed decays that proceed via D decays were considered in subsection 6.3.1. In this section backgrounds with $\eta(')$

along with vector resonances ω/ρ coming directly from the B are estimated. The total branching fractions for these processes are listed in Table 6.3 and since they are very small this type of background is discarded and will not be considered further.

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

Table 6.3: Different and total contribution to $B^0 \to \eta(')\rho(\omega)$. All the branching fractions are obtained from [13].

6.4   Rare and Resonant $B^+ \to \pi^+/K^+\mu^-\mu^+$ Backgrounds

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

It is, however, necessary to evaluate the impact of the rare equivalent of this back-

ground, namely $B^+ \rightarrow \mu^+\mu^-\pi^+$ decays, where the π^+ is misidentified as muon. The $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\pi^+) = 1.79 \pm 0.23 \times 10^{-8}$ [13]. The contribution of this background is accounted for in section 6.2, but it is crosschecked as this particular background peaks just under the corrected mass of the B. For the same rare decay but with a kaon instead, $B^+ \rightarrow \mu^+\mu^-K^+$, the mass is expected to be shifted away from this peak because of the higher kaon mass.

In order to establish the severity of this background, the $B^+ \rightarrow \mu^+\mu^-\pi^+$ simulation for Run I and 2016 is reconstructed where the muon mass hypothesis is used for the pion track candidate. Afterwards the same selection as in the signal case is applied. The expected number of $B^+ \rightarrow \mu^+\mu^-\pi^+$ decays after the full selection in a given Run can be calculated by normalising to $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ decays. In the end 0.06 (0.03) $B^+ \rightarrow \mu^+\mu^-\pi^+$ events are expected in Run I (2016) which is negligible given that there are ~ 17 expected signal events with $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) = 1 \times 10^{-8}$. Hence no further specific action for this background is taken, however, its contribution is directly accounted for in section 6.2.

6.5 Summary

In conclusion, different backgrounds that can mimic the signal were studied. From all backgrounds considered only the combinatorial, misID and partially reconstructed backgrounds have considerable contribution after all the selections and hence need to be modelled. The exact contribution of these three backgrounds is discussed in subsection 7.3.3.

Chapter 7

Mass Fits and Efficiencies

To be able to obtain a signal branching fraction estimate, the normalisation channel of $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ is used. Both, for signal and normalisation channel, the absolute efficiencies, luminosity, the b-quark cross-section and fragmentation fractions will cancel. There are, however, efficiencies that will not cancel and will be necessary in the final limit setting procedure. In this section, methods of obtaining efficiencies of the selection for the normalisation and signal channel are described. Later, the fitting procedure is outlined.

7.1 Efficiency Ratio

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

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

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

Three methods for signal efficiency determination are required for different parts of the selection chain.

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

$$\varepsilon_{GEN,minq} \times \varepsilon_{REC,minq} = \frac{N_{in\_acc,minq}}{N_{generated,minq}} \times \frac{N_{REC,minq}}{N_{in\_acc,minq}}, \tag{7.1}$$

$$N_{in\_acc,minq} = N_{in\_acc} \times \varepsilon_{minq}. \tag{7.2}$$

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

- Method II - Divide the number of events that passed the selection by the total number of events prior to this particular selection step.

- Method III - The data-driven approach using PIDCalib package explained in section 3.9 of determining the PID efficiency is used.

All of these efficiencies will have a systematic uncertainty associated with them. These systematic uncertainties are combined into one cumulative systematic which is taken into account in subsection 7.3.5 in the Signal Decay Model section.

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

$$R^{\{21,26\}}_{\mathrm{NOFCME}}(\varepsilon) = \frac{\varepsilon^s}{\varepsilon^n} = \frac{\varepsilon^s_{GEN}}{\varepsilon^n_{GEN}} \times \frac{\varepsilon^s_{REC}}{\varepsilon^n_{REC}} \times \frac{\varepsilon^s_{TRG}}{\varepsilon^n_{TRG}} \times \frac{\varepsilon^s_{OFF}}{\varepsilon^n_{OFF}} \times \frac{\varepsilon^s_{CombiBDT}}{\varepsilon^n_{CombiBDT}} \times \frac{\varepsilon^s_{MisidBDT}}{\varepsilon^n_{MisidBDT}} \times \frac{\varepsilon^s_{FR}}{\varepsilon^n_{FR}} \times \frac{\varepsilon^s_{PID}}{\varepsilon^n_{PID}},$$

$$(7.3)$$

where 21, 26 denote the stripping version for Run I and 2016. With the FCME split this efficiency ratio becomes

$$R^{\{21,26\}}_{\{\mathrm{lowFCME,highFCME}\}}(\varepsilon) = R^{\{21,26\}}_{\mathrm{NOFCME}}(\varepsilon) \times \frac{\varepsilon^{\{\mathrm{lowFCME,highFCME}\}}_s}{\varepsilon^{\{\mathrm{lowFCME,highFCME}\}}_n}. \qquad (7.4)$$

7.2 Summary of Efficiencies

The summary of individual efficiencies together with the total efficiency for σ_{NOFCME}, which is calculated for signal using the numerator of Equation 7.3 and for normalisation the denominator of Equation 7.3, is given in Table 7.2.

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

Table 7.2: Summary of individual simulation and/or data efficiencies in % for the relative efficiency between the signal and normalisation channel. Efficiency values for 2016 are TCK-weighted averaged efficiencies, see subsection 7.2.3. The errors considered are of statistical nature, computed using binomial errors.

Hence resulting values for the relative no fractional corrected mass split efficiency ratios defined in Equation 7.3 are

$$
\begin{aligned}
R_{\mathrm{NOFCME}}^{21}(\varepsilon) &= \frac{(1.58 \pm 0.02) \times 10^{-3}}{(3.97 \pm 0.01) \times 10^{-3}} = 0.398 \pm 0.005, \\
R_{\mathrm{NOFCME}}^{26}(\varepsilon) &= \frac{(1.18 \pm 0.01) \times 10^{-3}}{(3.20 \pm 0.00) \times 10^{-3}} = 0.369 \pm 0.003.
\end{aligned}
\tag{7.5}
$$

Upon obtaining the efficiencies with the fractional corrected mass split, which defined in Equation 7.4, the efficiency ratios are

$$R^{21}_{\text{lowFCME}}(\varepsilon) = \frac{(7.44 \pm 0.12) \times 10^{-4}}{(2.33 \pm 0.01) \times 10^{-3}} = 0.320 \pm 0.005,$$

$$R^{21}_{\text{highFCME}}(\varepsilon) = \frac{(8.37 \pm 0.13) \times 10^{-4}}{(1.65 \pm 0.01) \times 10^{-3}} = 0.509 \pm 0.008,$$

$$\qquad (7.6)$$

$$R^{26}_{\text{lowFCME}}(\varepsilon) = \frac{(6.51 \pm 0.05) \times 10^{-4}}{(2.15 \pm 0.00) \times 10^{-3}} = 0.303 \pm 0.002,$$

$$R^{26}_{\text{highFCME}}(\varepsilon) = \frac{(5.33 \pm 0.05) \times 10^{-4}}{(1.05 \pm 0.00) \times 10^{-3}} = 0.506 \pm 0.004.$$

As it can be noticed, different selections that were optimised for Run I and 2016 yield different overall as well as individual efficiencies. This results in small differences in sensitivity between Run I and 2016. To better understand where this difference comes from, the ratios of individual relative efficiencies as a function of stripping version are plotted in Figure 7.1. The difference can be attributed to different BDTs.
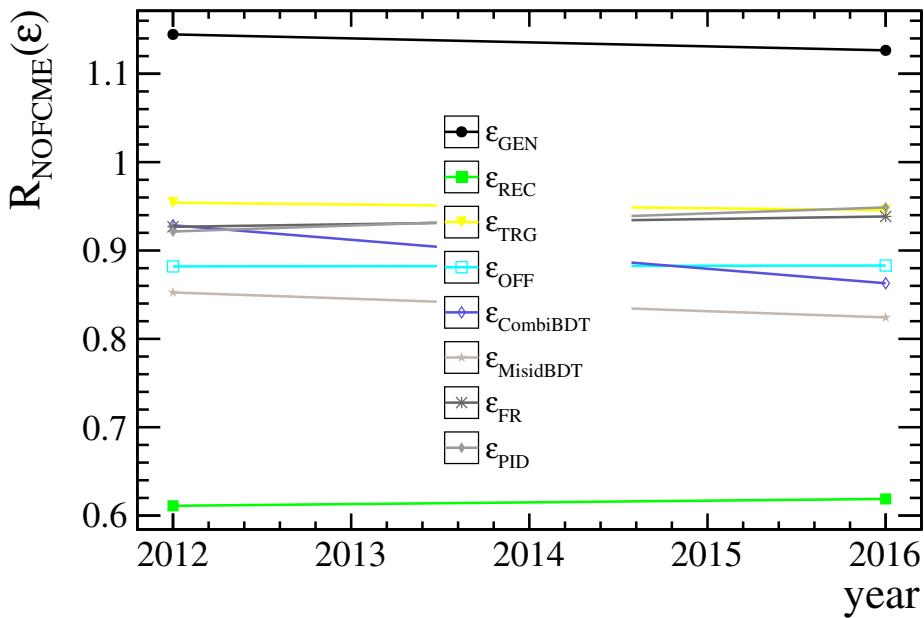
Figure 7.1: Summary of ratio of efficiencies between 2012 simulation and 2016 simulation with no FCME split. Efficiency values for 2016 are TCK-weighted averaged efficiencies.

7.2.1 Detector Acceptance Efficiency (GEN)

For charged particles, the detector acceptance efficiency describes the fraction of decays contained in the polar angle region of [10, 400] mrad. For the 2012 and 2016 simulation samples, the overall detector acceptance efficiency will be the average of two possible magnetic polarity conditions. For 2012 this will be also averaged with two different simulation versions: Pythia 6.4 [56] and Pythia 8.1 [40].

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

7.2.2 Reconstruction Efficiency (REC)

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

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

7.2.3 Trigger Efficiency (TRG)

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

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

However in 2016, the trigger thresholds have been changing often resulting in 16 different TCKs with very different p and p_T thresholds, see Table 7.3 for full details. In the second column, the luminosity proportion for 2016 is given. It can be seen that the default simulation in 2016 (corresponding to TCK decimal key 288888335) only represents around 35% of the data. For this reason, the trigger efficiencies for 2016 data have been obtained by emulation of the trigger on simulation for L0 and HLT1 level for each individual TCK, creating 16 TCK-based simulations. This trigger emulation to extract efficiencies was tested with the default trigger configuration (TCK 288888335) to validate the emulation and the correct efficiencies have been recovered. It should be noted that small differences arise from difference between *offline* and HLT1 container for PVs. This is because the PV finding-algorithm is different hence the containers store different min IPχ^2 but this has negligible effect.

In order to obtain the average efficiency for 2016, the 16 TCK efficiencies are weighted by the proportion of the integrated luminosity taken with the given TCK setting.

The full trigger efficiency for 2016 is calculated by averaging the luminosity-weighted efficiencies, as seen in Table 7.4. This TCK-dependant luminosity-weighted average efficiency is going to be given as a final efficiency for 2016 from now on for all

subsequent efficiencies unless stated otherwise.

For the HLT2 level, there were no significant changes of thresholds and hence efficiencies are obtained from full simulation regardless. The systematic effect of this assumption will be listed in subsection 7.3.5.

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

Table 7.4: Efficiencies of 2016 trigger emulation on simulation. Depending on the TCK setting, the efficiencies vary up 10% for L0 level for the signal simulation and up to 5% for normalisation. This is important as the *single event sensitivity* is sensitive to the ratio of these two efficiencies. The default configuration describes correctly 35% of the data.

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

7.2.4 Offline Selection (OFF)

In this section offline efficiencies are discussed. These include the J/ψ and $\Psi(2S)$ veto signal efficiency that were mentioned in Table 5.4, where $2946.0\,\mathrm{MeV}/c^2 < |\,\mathrm{M}(\mu^+\,\mu^-)\,| <$ $3176.0\,\mathrm{MeV}/c^2$ and $3586.0\,\mathrm{MeV}/c^2 < |\,\mathrm{M}(\mu^+\,\mu^-)\,| < 3766.0\,\mathrm{MeV}/c^2$. For the normalisation channel this is not applicable as the normalisation decay proceeds via the J/ψ resonance.

7.2.5 Combinatorial BDT and Misid BDT Efficiency

Efficiencies of the MVA selection are evaluated on simulation samples. These efficiencies are obtained using samples that passed (GEN), (REC), (TRG) and (OFF) cuts. The specific MVA for combinatorial background suppression (see subsection 5.8.3) and misID background suppression (see subsection 5.8.4) are applied to the simulation samples. As the optimisation led to different BDTs depending on the data-taking period, these are then applied parametrically to the relevant simulation samples. The efficiency results are listed in Table 7.2.

For the Misid and Combinatorial BDT selection, the normalisation $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ channel retains more signal than the $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ channel. This is due to the kaon/muon p and p_T kinematic differences as seen in Figure 7.2 and Figure 7.3, where the kaon track is generally harder than the muon track. The kaon reconstruction efficiency is worse than the muon reconstruction efficiency because about 11% of the kaons cannot be reconstructed due to hadronic interactions that occur before the last T station [46], implying that the p_T of the B meson is on average harder for the normalisation channel. As these two quantities are high in the BDT importance ranking, as mentioned in subsection 5.8.3, this makes the normalisation simulation more efficient. In the Misid BDT selection, again the kinematics of the B meson and the min IPχ^2 of the oppositely charged muon to the mother B has a better separation from background than the signal.
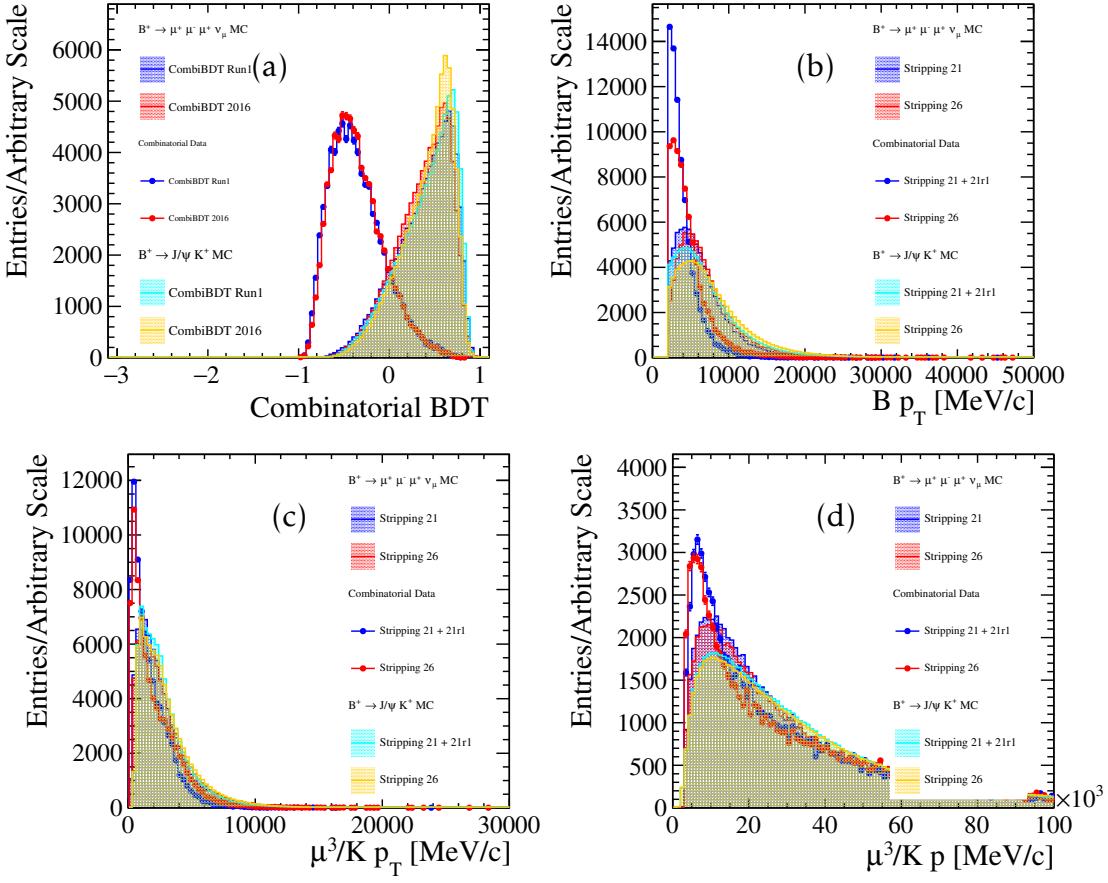
Figure 7.2: (a) Combinatorial BDT response for signal simulation and upper mass side-band as well as for the normalisation channel simulation for Stripping 21 and Stripping 26. The most discriminating variables are (b) p_T of the B meson, (c) muon/kaon p_T and (d) muon/kaon p.
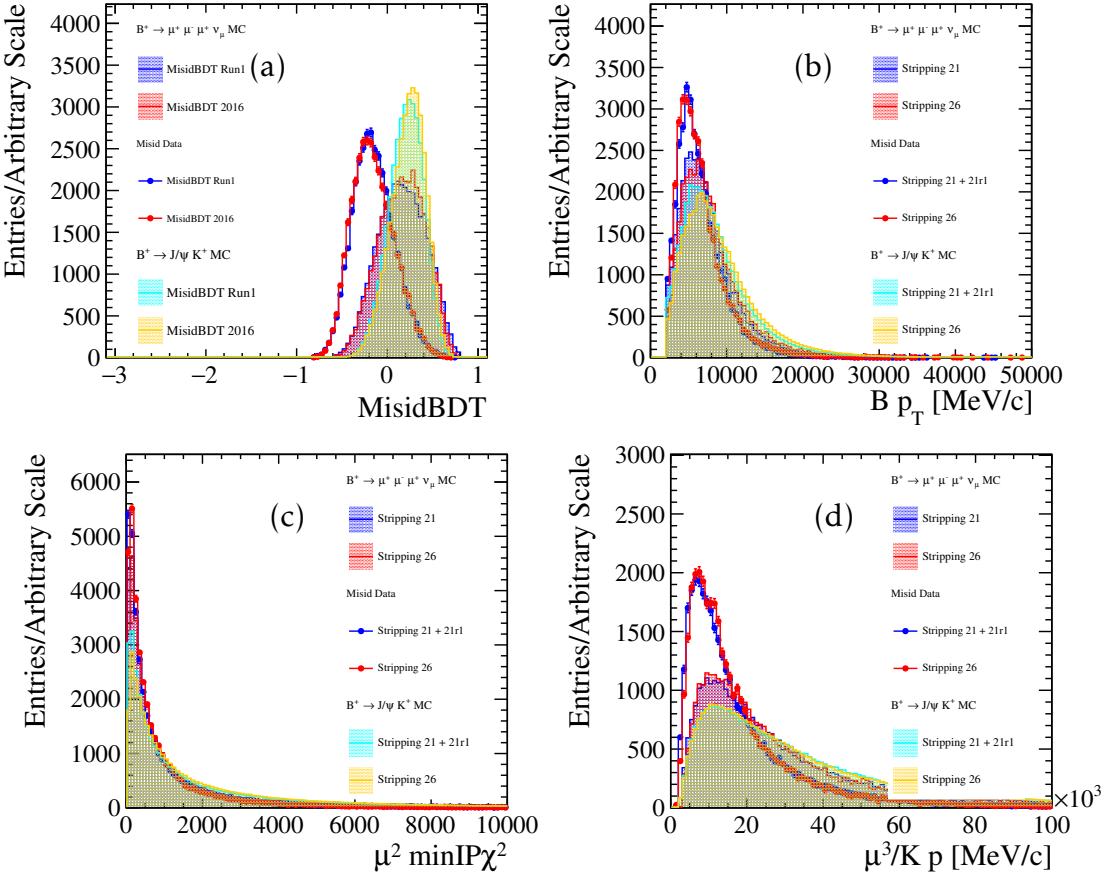
Figure 7.3: (a) Misid BDT response for signal simulation and upper mass sideband as well as for the normalisation channel simulation for Stripping 21 and Stripping 26. The most discriminating variables are (b) p_T of B, (c) muon IP χ^2 and (d) muon/kaon p. Misid BDT responses are plotted with combinatorial BDT already applied.

7.2.6   Fitting Range Efficiency (FR)

As discussed in subsection 5.8.5 and in section 6.1, the fitting region was chosen in order to avoid modelling a drop in the low corrected mass region for the combinatorial background (exclusion below 4000 MeV/c^2) and secondly in order to not include a region where there are very few/no events (exclusion above 7000 MeV/c^2) in **corrected mass**. As seen in Table 7.2 and in Figure 7.4 the normalisation channel does not loose many candidates compared to the signal channel.
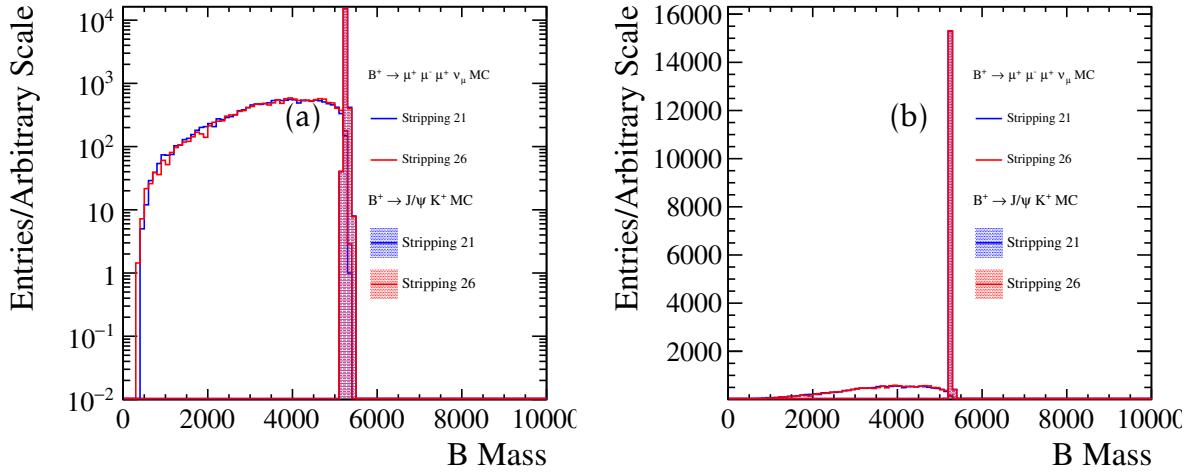
Figure 7.4: (a) Visible mass of normalisation and signal simulation. It can be seen that normalisation's previous preselection has a sharp cut around the visible *B* mass leading to much higher fitting efficiency. (b) The corresponding logarithmic version of plot (a).

7.2.7 PID Efficiencies (PID)

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

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

Species	2012 PID Simulation	2016 PID Simulation
muon	DLLmu > 0	DLLmu > 0
muon	(DLLmu - DLLK) > 0	(DLLmu - DLLK) > 0
muon	-	IsMuonTight==1.0
muon	Nshared==0	Nshared<2
muon	Probnnmu> 0.35	Probnnmu> 0.35
ε_{PID}	0.631± 0.005	0.623±0.006

Table 7.5: Signal simulation efficiency using PIDCalib efficiencies.

Species	2012 PID Simulation	2016 PID Simulation
muon	DLLmu > 0	DLLmu > 0
muon	(DLLmu - DLLK) > 0	(DLLmu - DLLK) > 0
muon	-	IsMuonTight==1.0
muon	Nshared==0	Nshared<2
muon	Probnnmu> 0.35	Probnnmu> 0.35
kaon	DLLK > 0	DLLK > 0
kaon	(DLLp - DLLK) < 5	(DLLp - DLLK) < 5
ε_{PID}	0.685± 0.001	0.656±0.001

Table 7.6: Normalisation simulation efficiency using PIDCalib efficiencies.

TCK dec	%$\mathcal{L}$	$\mathcal{L}$	HLT1TrackMuon				LOMuon	
			P_{ghost}	p_μ	$p_T(\mu)$	min IPχ^2	SPD	$p_T(\mu)$
	%	pb^{-1}		[MeV]	[MeV]			[MeV]
2016 MD 0.86 fb^{-1}								
287905280	0.8	12.7	–	6.0	0.91	10	450	14
287905283	2.1	35.0	–	6.0	0.91	10	450	23
287905284	1.5	24.8	–	6.0	0.91	10	450	27
287905285	4.7	78.4	–	6.0	0.91	10	450	31
288822793	4.4	72.1	0.2	6.0	1.1	35	450	27
288822798	1.4	22.8	0.2	6.0	1.1	35	450	27
288888329	0.4	6.9	0.2	6.0	1.1	35	450	31
288888334	2.0	31.7	0.2	6.0	1.1	35	450	31
288888335	34.7	575.3	0.2	6.0	1.1	35	450	37
2016 MU 0.80 fb^{-1}								
288495113	6.5	107.0	–	6.0	0.91	10	450	27
288626185	7.1	118.1	–	6.0	0.91	10	450	27
288691721	1.4	23.5	0.2	6.0	1.1	35	450	27
288757257	25.0	414.6	0.2	6.0	1.1	35	450	27
288888337	2.7	44.1	0.2	6.0	1.1	35	450	31
288888338	5.4	89.8	0.2	6.0	1.1	35	450	33
288888339	0.1	1.1	0.2	6.0	1.1	35	450	27
Default Simulation								
1362630159	–	–	0.2	6.0	1.1	35	450	37

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

161

7.3 Mass Fits

In this section, firstly the parametrisation of the normalisation channel is shown in subsection 7.3.1. The fit to the normalisation channel using the full fit model is described in subsection 7.3.2. This is followed by the signal fit parametrisation in subsection 7.3.3, resulting in the blinded and non-blinded data fits described in subsection 7.3.4.

7.3.1 Normalisation Channel Parametrisation

To obtain the $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ yield in Run I and 2016, an unbinned extended maximum likelihood fit to the invariant $\mu^+\mu^-K^+$ data distribution in each respective Run dataset is performed. In this section contributions to the normalisation fit model are considered.

Signal

The first component is the signal itself, which is modelled with PID-weighted simulation and can be best described by a double-sided Ipatia function, detailed in section A.1, where all the parameters apart from the mean μ^{IP} and width σ^{IP} are fixed from the signal simulations. These simulations pass through the same selection process as the corresponding $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ data, described in section 5.9, with one exception. Instead of directly cutting on PID variables, the simulations are reweighted with the relevant PID weights, because of the known mismatch between simulation and data. More on this will be covered in subsection 7.2.7.

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

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

$B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ simulation and $B^+ \to (J/\psi \to \mu^+\mu^-)\pi^+$ simulation from Stripping 21 are showed using different scales. For signal, Run I Stripping 21 sample is used and for 2016 the Stripping 26 - TCK 288888335 sample is used. For the $B^+ \to (J/\psi \to \mu^+\mu^-)\pi^+$ the Stripping 21 sample is used for both Run I and 2016.
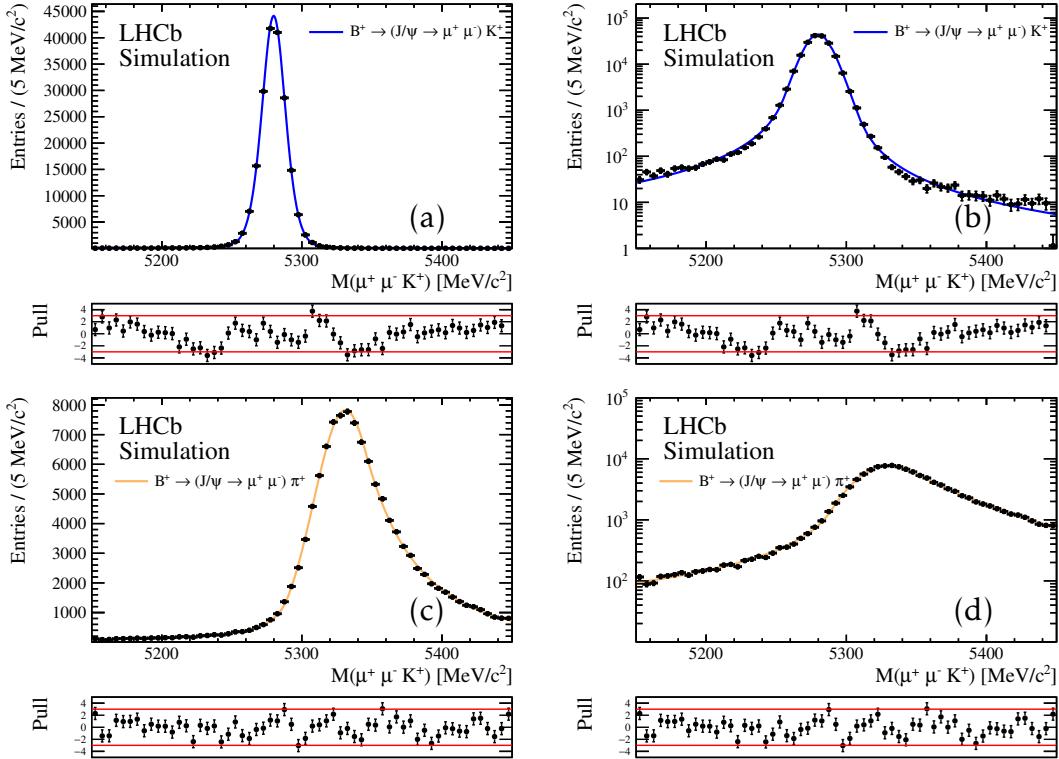
Figure 7.5: Fit to 2012 (a) $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ simulation and (c) $B^+ \to (J/\psi \to \mu^+\mu^-)\pi^+$ simulation under the kaon mass hypothesis. On right, the same plots but with logarithmic scale instead.

Combinatorial Background

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

7.3.2 Normalisation Fit

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

Fit Parameter	Status
Yields	
$N_{B^+ \to (J/\psi \to \mu^+\mu^-)K^+}$ (Signal)	Free
$N_{B^+ \to (J/\psi \to \mu^+\mu^-)\pi^+}$	Free
$N_{Combinatorial}$	Free
Signal Shape Parameters (double-sided Ipatia)	
$\mu^{IP}_{B^+ \to (J/\psi \to \mu^+\mu^-)K^+}$	Constrained from signal simulation
$\sigma^{IP}_{B^+ \to (J/\psi \to \mu^+\mu^-)K^+}$	Constrained from signal simulation
Others	Fixed from simulation
$B^+ \to (J/\psi \to \mu^+\mu^-)\pi^+$ **Shape Parameters (double-sided Crystal Ball)**	
$\mu^{CB}_{B^+ \to (J/\psi \to \mu^+\mu^-)\pi^+}$	Constrained from signal simulation
$\sigma^{CB}_{B^+ \to (J/\psi \to \mu^+\mu^-)\pi^+}$	Constrained from signal simulation
Others	Fixed from simulation
Combinatorial Shape Parameters	
exponential par.	Free

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

The $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ signal yield is extracted by performing an unbinned extended maximum likelihood fit with the full fit model to the invariant $\mu^+\mu^-K^+$ distribution in the range $5150\,\mathrm{MeV}/c^2 < M_{B^+} < 5450\,\mathrm{MeV}/c^2$. Fits to the $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ for Run I and 2016 are shown in Figure 7.6. Yields from the fit to $B^+ \to (J/\psi \to$

$\mu^+\mu^-)K^+$ are obtained and summarized in Table 7.8.
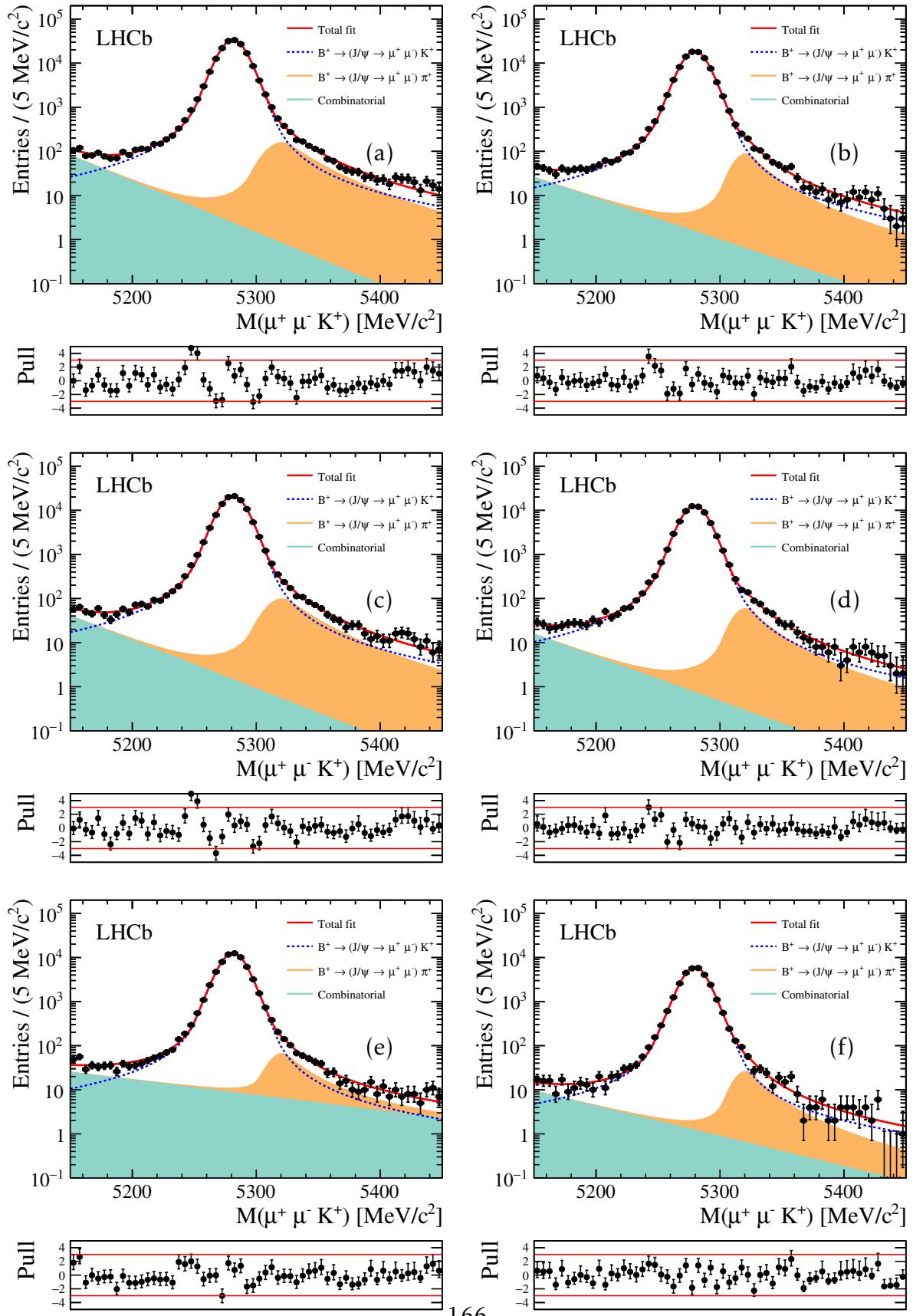
Figure 7.6: Fit results in logarithmic scale to (a) Run I (b) 2016 $\mu^+\mu^-K^+$ mass spectrum with no fractional corrected mass split, (c)(d) low FCME bin, (e)(f)high FCME bin.

Sample	Stripping	Split	Yields
$N_{B^+ \to J/\psi K^+}$	Run I	σ_{NOFCME}	173422±446
$N_{B^+ \to J/\psi K^+}$	2016	σ_{NOFCME}	94491±313
$N_{B^+ \to J/\psi K^+}$	Run I	σ_{lowFCME}	109224±337
$N_{B^+ \to J/\psi K^+}$	2016	σ_{lowFCME}	64723±259
$N_{B^+ \to J/\psi K^+}$	Run I	σ_{highFCME}	64078±257
$N_{B^+ \to J/\psi K^+}$	2016	σ_{highFCME}	29760±176

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

7.3.3   Signal Channel Parametrisation

To fit the signal data, the default fitting strategy is to use a simultaneous unbinned maximum likelihood fit to the $\mu^+\mu^-\mu^+$ corrected mass spectrum for the Run I and 2016 dataset after the full selection in two bins of fractional corrected mass error, which is denoted as the simultaneous fit. In this fit, the branching fraction for $B^+ \to \mu^+\mu^-\mu^+\nu$ is directly fitted as the parameter of interest. As a cross-check a non-simultaneous fit, where no splitting in bins of corrected mass is done, is denoted as the non-simultaneous fit. The parametrisation of all the components for the full signal model fit is described below.

Signal

The full fit to Run I and 2016 data requires the knowledge of the signal shape for the combined dataset. This is obtained from a Run I and 2016 signal simulation *cocktail*. The cocktail is created after full selection by assigning event-by-event weights, w^i, which capture the differences between the Run I and 2016 simulation, that were not

considered by the full selection.

Firstly, weights that reflect the expected difference due to the increased luminosity are computed. Defining $\mathcal{L}^{2012}, \mathcal{L}^{2016}$ to be the integrated luminosities and $\sigma^{2012}_{pp \to b\bar{b}}, \sigma^{2016}_{pp \to b\bar{b}}$ to be the cross-sections in a given year the following definitions are provided

$$n^{2012} = \mathcal{L}^{2012} \times \sigma^{2012}_{pp \to b\bar{b}}, \tag{7.7}$$

$$n^{2016} = \mathcal{L}^{2016} \times \sigma^{2016}_{pp \to b\bar{b}}. \tag{7.8}$$

The signal weights can be obtained by solving the following system of two equations

$$w^{2012} \times N^{2012} + w^{2016} \times N^{2016} = N^{2012} + N^{2016}, \tag{7.9}$$

$$\frac{w^{2012} \times N^{2012}}{w^{2016} \times N^{2016}} = \frac{n^{2012}}{n^{2016}}. \tag{7.10}$$

where N^{2012}, N^{2016} are the number of events at *generator level*. N^{2012}, N^{2016} number of events at *generator level* is obtained by dividing the number of reconstructed events N_{REC} by the reconstruction efficiency ε_{REC} (see Table 7.2). Values for these variables are summarized in Table 7.9.

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

$$w^{2012} = \frac{N^{2012} + N^{2016}}{N^{2012} \times (1.0 + \frac{n^{2016}}{n^{2012}})} = 0.931 \tag{7.11}$$

$$w^{2016} = \frac{N^{2016} + N^{2012}}{N^{2016} \times (1.0 + \frac{n^{2012}}{n^{2016}})} = 1.073, \tag{7.12}$$

Secondly, an event-by-event weight which differs between Run I and 2016 that needs to be accounted for is the PID efficiency ε_{PID} (see Table 7.2), which depends on the kinematics of the final state particles. It is denoted as $w^{i \in \{2012, 2016\}}_{\varepsilon^i_{PID}}$.

The final weight of an event depending on Run I and 2016 is calculated a

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

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

$$w_{total}^{i \in \{2012,2016\}} = w^{i \in \{2012,2016\}} \times w_{\varepsilon_{PID}^i}^{i \in \{2012,2016\}}. \tag{7.13}$$

After obtaining a combined Run I and 2016 signal *cocktail*, a fit to this weighted simulation is done with **the shape** in the corrected mass modelled by a double-sided Crystal Ball function section A.2. The fit and its parameters can be seen in Figure 7.7. The function describing this shape is denoted as f^{sig} and hence is a function of 6 parameters for the non-simultaneous fit and 12 parameters for the simultaneous fit.
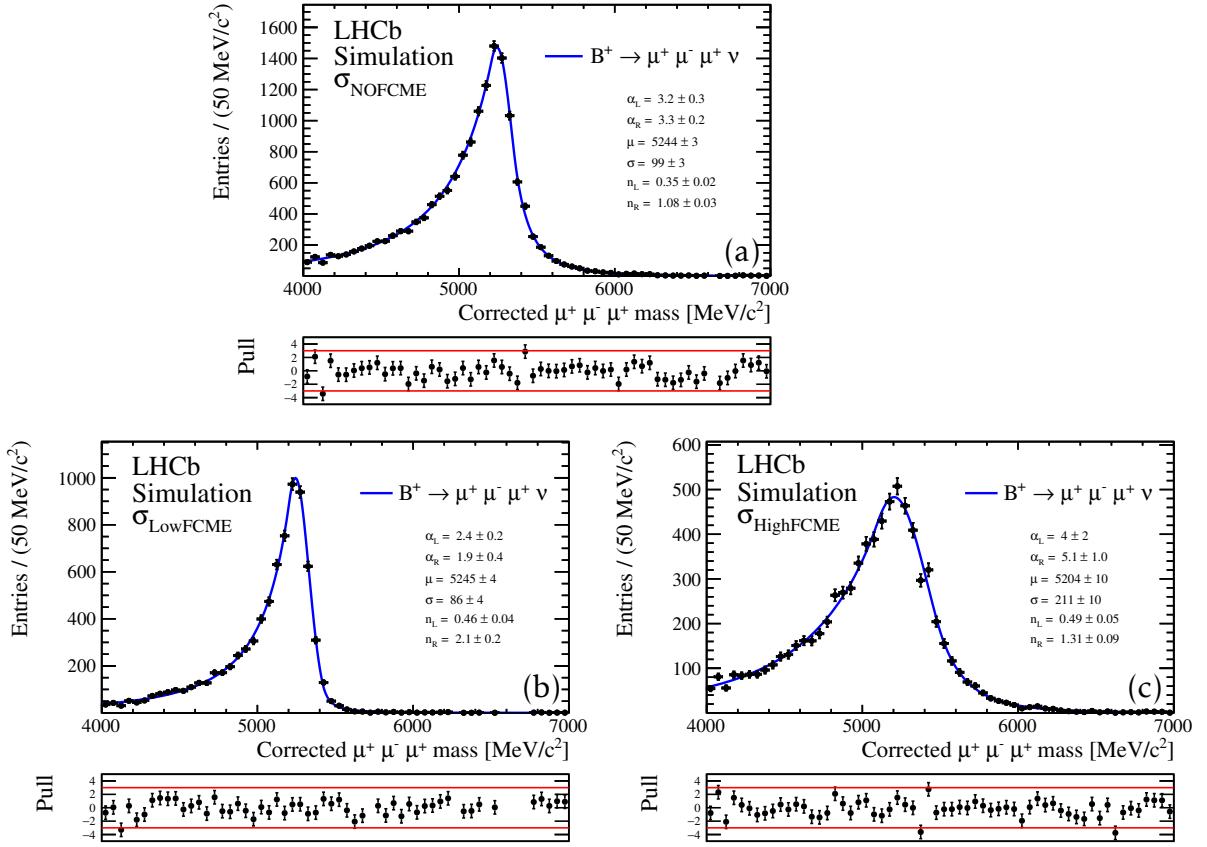
Figure 7.7: Fit to the weighted combined signal *cocktail* for (a) NO FCME (b) Low FCME and (c) High FCME split.

The signal yield, $N^{sig} = N(B^+ \to \mu^+\mu^-\mu^+\nu)$, is calculated using several parameters. More specifically, it is related to the branching fraction using the normalisation channel in the following way:

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

$$\tag{7.14}$$

170

where ε^x is the total selection efficiency for the channel x, N_x is the number of x decays, $\mathcal{B}(x)$ is the branching fraction of decay x, R_{FCME} is the relevant efficiency ratio between the two decays detailed in section 7.2 and finally α is the *single event sensitivity*, a variable which describes the sensitivity for the search. Hence, for Run I and 2016 the non-simultaneous fit N^{sig} is a function of six parameters $N^{sig}(R^{21}_{\mathrm{FCME}}, R^{26}_{\mathrm{FCME}}), N^{Run\ I}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), N^{2016}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), \mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+), \mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu))$. In the simultaneous case $R^{21}_{\mathrm{FCME}}, R^{26}_{\mathrm{FCME}}$, $N^{Run\ I}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)$ and $N^{2016}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)K^+)$ are further split into $\sigma_{\mathrm{lowFCME}}$ and $\sigma_{\mathrm{highFCME}}$ bins, resulting in 10 parameters for N^{sig} in the fit, as shown for the non-simultanenous fit parameter summary in Table 7.14. Again, the parameter of interest in the end is $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu)$ which is directly fitted for so $N(B^+ \rightarrow \mu^+\mu^-\mu^+\nu)$ is given here just to be able to translate the $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu)$ into $N(B^+ \rightarrow \mu^+\mu^-\mu^+\nu)$.

Partially Reconstructed Background

Partially reconstructed backgrounds are still non-negligible after the full selection chain. The simulation sample for the partially reconstructed background that originates through D^0 decays (section 6.3) is used for both the yield estimate and shape modelling.

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

$$\frac{N_{B^+\rightarrow(D^0\rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu}}{N_{B^+\rightarrow(J/\psi\rightarrow\mu^+\mu^-)K^+}} = \frac{\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu)}{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)} \times \frac{\varepsilon^{B^+\rightarrow(D^0\rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu}}{\varepsilon^{B^+\rightarrow(J/\psi\rightarrow\mu^+\mu^-)K^+}}$$

$$= \frac{\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu)}{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)} \times R_{\mathrm{FCME}},$$

(7.15)

where ε^x is the total selection efficiency for the channel x and R_{FCME} is hence the efficiency ratio between the two channels, N_x is the number of x decays and $\mathcal{B}(x)$ is the branching fraction of decay x. The quantity of the interest, $N_{B^+\rightarrow(D^0\rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu}$, can

therefore be calculated given knowledge of all other terms. $N_{B^+ \to (J/\psi \to \mu^+ \mu^-)K^+}$ is obtained from Table 7.8. $\mathcal{B}(B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-)\mu^+ \nu)$ is obtained by multiplying $\mathcal{B}(D^0 \to K^+ \pi^- \mu^+ \mu^-) = (4.17 \pm 0.12 \pm 0.40) \times 10^{-6}$ [73] and $\mathcal{B}(B^+ \to Dl^+ \nu X) = (9.8 \pm 0.7) \times 10^{-2}$ [13] yielding $\mathcal{B}(B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-)\mu^+ \nu) \approx (4.1 \pm 0.5) \times 10^{-7}$. The $B^+ \to (J/\psi \to \mu^+ \mu^-)K^+$ branching fraction is obtained with the same approach: multiplying $\mathcal{B}(B^+ \to J/\psi K^+) = (1.026 \pm 0.031) \times 10^{-3}$ [13] and $\mathcal{B}(J/\psi \to \mu^- \mu^+) = (5.961 \pm 0.0033) \times 10^{-2}$ [13] yielding

$$\mathcal{B}(B^+ \to (J/\psi \to \mu^+ \mu^-)K^+) = (6.12 \pm 0.19) \times 10^{-5}. \tag{7.16}$$

All the relevant total selection efficiencies are obtained from the full simulation sample and are shown in Table 7.10. Due to the usage of a proxy simulation for these partially reconstructed decays (using a pion rather than a muon in one case), as discussed in section 6.3, the trigger efficiency ε_{TRG} cannot be obtained from simulation, because the HLT2 trigger (see Table 5.3) would make a positive decision only either because of finding dimuon pair or two or three-body decays, hence the trigger ratio $\dfrac{\varepsilon_{TRG}^{B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-)\mu^+ \nu}}{\varepsilon_{TRG}^{B^+ \to (J/\psi \to \mu^+ \mu^-)K^+}}$ is assumed to be 1, which is rather a conservative estimate (overestimate) but makes sure that other partially reconstructed backgrounds are accounted for. Another efficiency that was not accounted for because of the same reason is PID efficiency, ε_{PID}. Moreover as this proxy simulation was not accessible for 2016, the same ratio of efficiencies as in Run I is used.

The summary of the expected yield is summarized in Table 7.10. For the non-simultaneous final fit there are hence 5 parameters $N^{PR}(R^{21}_{\text{FCME}}(B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-)\mu^+ \nu), N^{Run\,I}(B^+ \to (J/\psi \to \mu^+ \mu^-)K^+), N^{2016}(B^+ \to (J/\psi \to \mu^+ \mu^-)K^+), \mathcal{B}(B^+ \to (J/\psi \to \mu^+ \mu^-)K^+), \mathcal{B}(B^+ \to (D^0 \to K^+ \pi^- \mu^+ \mu^-)\mu^+ \nu))$ and for the simultaneous data fit with FCME splitting 8 parameters. The total yield expected for this type of background is very low compared to the other expected backgrounds.

The shape for partially reconstructed backgrounds, f^{PR}, is also obtained from the simulation proxy after all the selections. The shape is best described with the sum of two Crystal Ball functions, more in section A.2, with free means μ^1, μ^2 and widths

Properties	Run I	2016
$\mathcal{B}(B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu)$	$(4.1 \pm 0.5) \times 10^{-7}$	$(4.1 \pm 0.5) \times 10^{-7}$
$\mathcal{B}(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)$	$(6.12 \pm 0.19) \times 10^{-5}$	$(6.12 \pm 0.19) \times 10^{-5}$
$\varepsilon_{total}^{B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu}$	$(1.87 \pm 0.04) \times 10^{-4}$	Using 2012
$\varepsilon_{total}^{B^+ \to (J/\psi \to \mu^+\mu^-)K^+}$	$(5.80 \pm 0.01) \times 10^{-3}$	Using 2012
$R_{FCME}^{21}(B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu)$	$(3.22 \pm 0.07) \times 10^{-2}$	Using 2012

	$\sigma_{lowFCME}$	
$N_{B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu}$	19.8 ± 2.6	11.7 ± 1.5

	$\sigma_{highFCME}$	
$N_{B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu}$	17.0 ± 2.2	7.9 ± 1.0

	σ_{NOFCME}	
$N_{B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu}$	37.3 ± 4.8	20.3 ± 2.6

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

σ^1, σ^2 as seen in Figure 7.8. Because the shape of this background suggests that the majority of the contamination is below $5000\,\mathrm{MeV}/c^2$, it is one of the least dangerous backgrounds. In the non-simultaneous fit, f^{PR} is a function of 9 parameters and for the simultaneous fit 18 parameters coming from the sum of two Crystal Ball functions.
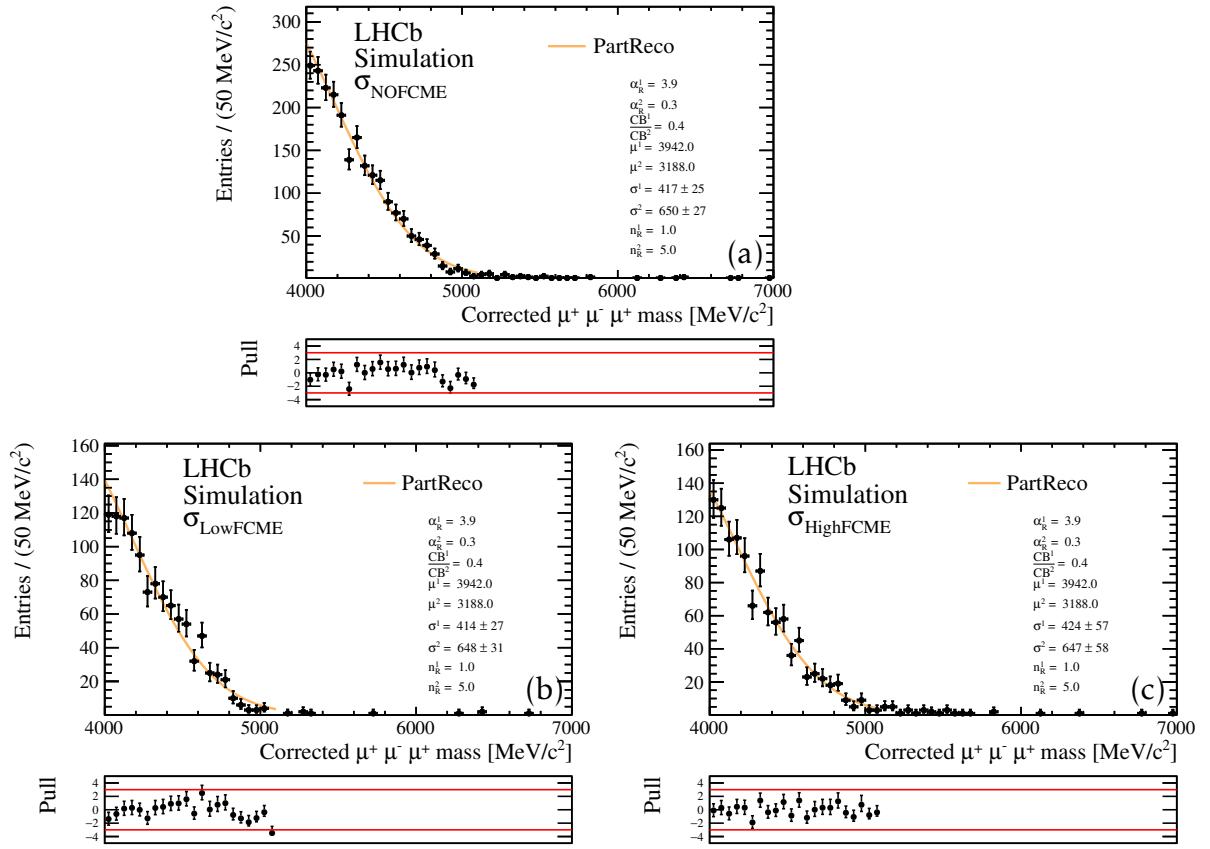
Figure 7.8: Fit to weighted combined partially reconstructed background simulation proxy for (a) NO FCME (b) Low FCME and (c) High FCME split.

MisID background

The level and the shape of misID background is determined by fitting the misID data samples obtained using the method described in section 6.2. A binned χ^2 fit is used to extract the shape and yields parameters. The reason for usage of the binned χ^2 fit is that the misID samples are little populated weighted samples and the shape and yield needs to be propagated to the final data fit while preserving the fit parameter correlations. Since there is a prescale factor of 1% at stripping level, to obtain the correct yield, the final number needs to be multiplied by 100 to counteract the prescale.

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

$\pi^+ K^-$) samples (see subsection 4.3.1) have uncertainties associated with them as shown in Figures 4.11(a), 4.12(a), 4.13(a), 4.14(a). These uncertainties are accounted for in the fit by Gaussian variation of the weights within the uncertainty in a given kinematic bin of p, η for each particle species and then folded in to the misID calculation. In this case 100 variations were used. Each variation results in a different template for the misID shape. This misID template is then subsequently binned in 15 bins of corrected mass. From each corrected mass bin, mean μ_{var} and error σ_{var} from gaussianly distributed number of misID events is obtained.

The total uncertainty due to the weight w, σ_{tot}, for a given bin of corrected mass is calculated using $\sqrt{\sigma_{var}^2 + \sqrt{\sum w_i^2}^2}$, where $\sigma_{par} = \sqrt{\sum w_i^2}$ is the associated error per bin and σ_{var} is the standard deviation obtained from the variation of misID weights ((see subsection 4.3.1) in Figures 4.11(a), 4.12(a), 4.13(a), 4.14(a)). Finally, the binned χ^2 fit is made to the misID samples with the total uncertainty. **The number of misID events**, N^{MisID}, for different species-regions after all selections are seen in Table 7.11. Hence for the non-simultaneous fit this just adds 1 parameter and for the simultaneous fit there will be two parameters describing the total misID yield.

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

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

between parameters, which are hence a set if multidimensional gaussian constraints in the signal data fits. This means that all uncertainties due to misID will be directly accounted for in the signal data fits.

Sample	Region	PID	weight	misID count
Run I *SS misID*	Kaon-region	Run I PID	crossfeedweight	198
Run I *SS misID*	Pion-region		crossfeedweight	103
Run I *SS misID*	Proton-region		no-crossfeedweight	6
Run I *OS misID*	Kaon-region		no-crossfeedweight	3
Run I *OS misID*	Pion-region		no-crossfeedweight	42
Run I *OS misID*	Proton-region		no-crossfeedweight	1
2016 *SS misID*	Kaon-region	2016 PID	crossfeedweight	136
2016 *SS misID*	Pion-region		crossfeedweight	76
2016 *SS misID*	Proton-region		no-crossfeedweight	0
2016 *OS misID*	Kaon-region		no-crossfeedweight	8
2016 *OS misID*	Pion-region		no-crossfeedweight	45
2016 *OS misID*	Proton-region		no-crossfeedweight	1
Sum				619

Table 7.11: The final misID template is constructed by summing the contribution from Run I and 2016 kaon, pion and proton-like regions for both *SS* and *OS misID* contributions.
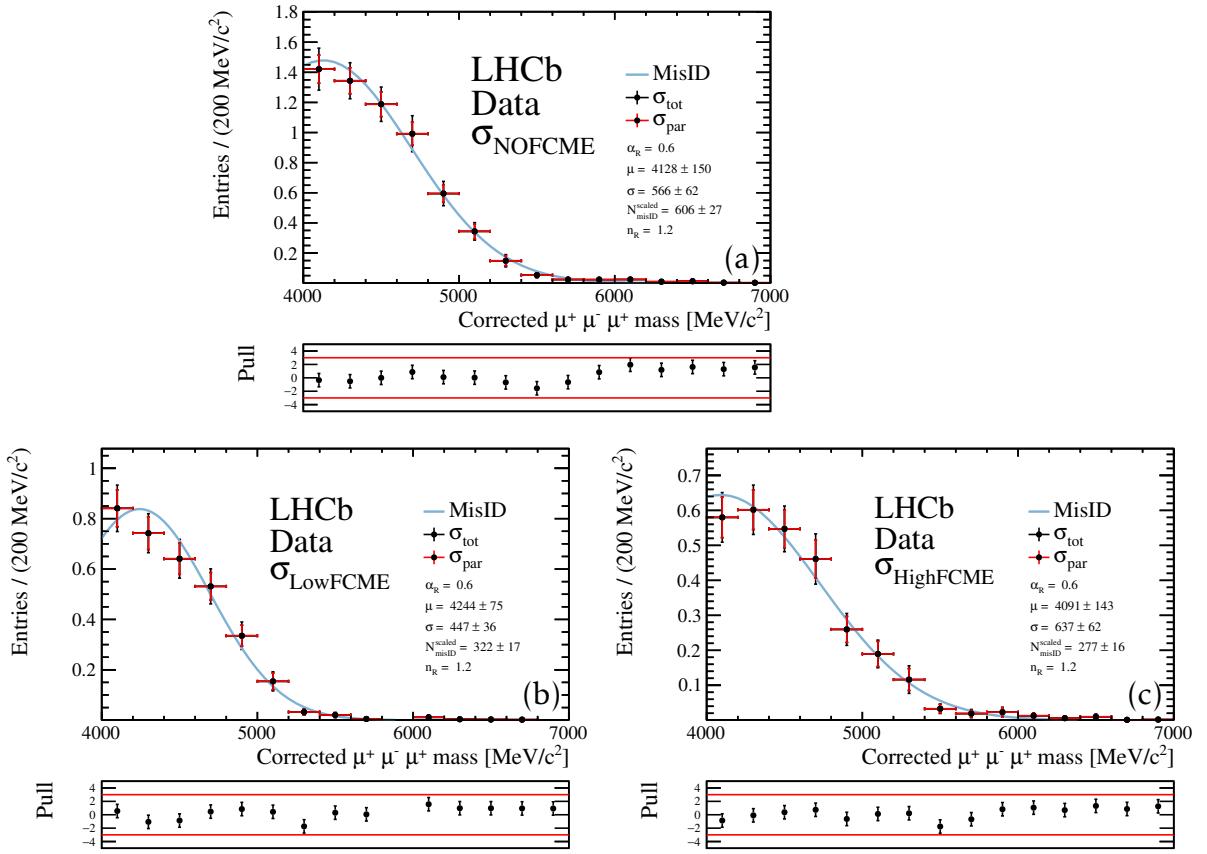
Figure 7.9: Binned χ^2 fit to the misID templates with no FCME split (b) Low FCME (c) High FCME. In high FCME case, the distribution of misID pollutes the signal window more than in the low FCME case. Both the full weight error σ_{tot} and the partial weight error σ_{par} can be seen.

Combinatorial Background

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

one (two) parameter(s) for the non-simultaneous (simultaneous) fit.

Altogether all the parameters for the non-simultanenous fit are summarized in Table 7.14.

7.3.4 Signal Data Fits

Signal fit models

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

$$f^{NS}(x, y^i) = N^{sig} \times f^{sig} + N^{MisID} \times f^{MisID} + N^{PR} \times f^{PR} + N^{Combi} \times f^{Combi}, \qquad (7.17)$$

$$f^S(x, y^k) = \sum_{j \in \sigma} (N_j^{sig} \times f_j^{sig} + N_j^{MisID} \times f_j^{MisID} + N_j^{PartReco} \times f_j^{PartReco} + N_j^{Combi} \times f_j^{Combi}),$$
$$(7.18)$$

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

Signal fit model hypotheses

As the strength of the signal is not a priori known, there are two types of hypotheses that are established: one where there is the presence of signal, also known as the signal+background hypothesis which takes exactly the form of Equation 7.17 and then a background only hypothesis, where $\mathcal{B}(B^+ \to \mu^+ \mu^- \mu^+ \nu) = 0 \to N^{sig} = 0$.

Signal data types

There are two types of signal datasets as mentioned in section 5.1. Firstly, there is the **blinded signal data** to which the simultaneous and non-simultaneous fits are performed in order to evaluate the expected sensitivity. Since these datasets are blinded only the mass regions $4000\,\mathrm{MeV}/c^2 < M_{B_{corr}} < 4500\,\mathrm{MeV}/c^2$ and $5500\,\mathrm{MeV}/c^2 < M_{B_{corr}} < 7000\,\mathrm{MeV}/c^2$ are used in the fits. These fits are shown in subsection 7.3.6. Secondly fits to the **full signal data** are performed and these are used in order to set the limit on the $\mathcal{B}(B^+ \to \mu^+ \mu^- \mu^+ \nu)$. In these fits the full mass spectrum is used.

7.3.5 Signal Fit Systematics

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

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

Table 7.12: Summary of systematic uncertainties.

Signal Decay Model

The largest systematic uncertainty arises due to the choice of decay model for the signal channel. The nominal signal model creates a photon pole, increasing the branching fraction in the low invariant dimuon mass region. The associated systematic uncertainty is estimated by replacing this decay model with one determined purely by phase-space and by comparing the selection efficiencies, resulting in a 4.6% total systematic uncertainty on the efficiency ratio. This cumulatively incorporates systematic uncertainties on individual efficiencies ε_{GEN}, ε_{REC}, ε_{TRG}, ε_{OFF}, ε_{BDTs} and ε_{FR}.

Trigger

The second most important systematic effect is due to incorrect modelling of the trigger in the simulation. This can be split into two effects: the efficiency difference between

the simulation and data and the incorrect emulation of HLT2 in the simulation in 2016 as only one TCK was considered for this trigger selection. This uncertainty is assigned by comparing the difference between the trigger efficiency of $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ decays in simulation and data using the TISTOS method.

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

$$\varepsilon_{TOS} = \frac{N_{TIS\&TOS}}{N_{TIS}}. \tag{7.19}$$

In Equation 7.19 $N_{TIS\&TOS}$ is the number of TIS and TOS events passing the trigger requirement and N_{TIS} is the number of TIS events passing the trigger requirement.

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

Kaon Interaction Probability

One difference between the signal and the normalisation channels that can have an effect on the efficiencies is that the kaon in the decay $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ can interact with the detector at a probability proportional to the amount of material traversed. The uncertainty on this amount of material leads to a 2% systematic uncertainty, derived following the procedure outlined in Ref. [46].

Kinematic Reweighting

Another source of systematic uncertainty is caused by differences in the B^+ production kinematics in the simulation. Corrected variables are momentum, p, transverse momentum p_T and vertex χ^2 of the B. The kinematic weights in a given kinematic bin are

calculated using normalised histograms of $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ data, and normalized $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ simulation in a following way

$$w_{(bin)} = \frac{w_{B^+ \to (J/\psi \to \mu^+\mu^-)K^+ data, bin}}{w_{B^+ \to (J/\psi \to \mu^+\mu^-)K^+ simulation, bin}}. \tag{7.20}$$

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

Signal Fit Bias and Coverage

In this section, the systematic uncertainty due to signal bias is evaluated. The pull is calculated using pseudo-experiments where the data is generated for a signal branching fraction $\mathcal{B} = 1.0 \times 10^{-8}$ corresponding to ≈ 17 signal events for Run I and 2016 data. These pseudo-datasets are then re-fitted with floating $\mathcal{B}$ and the corresponding number of fitted signal events is obtained. The pull is defined as the difference between number of fitted signal events and the number of events that the pseudo-experiment was created with, divided by the error on the number of signal events given by the fit:

$$\frac{N_{sig}^{orig} - N_{sig}^{fit}}{\sigma^{fit}}. \tag{7.21}$$

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

For this study 10000 pseudo-experiments were created testing the bias of both the extended non-simultaneous fit and extended simultaneous fit. The pull distributions for the non-simultaneous fit can be seen in Figure 7.10.
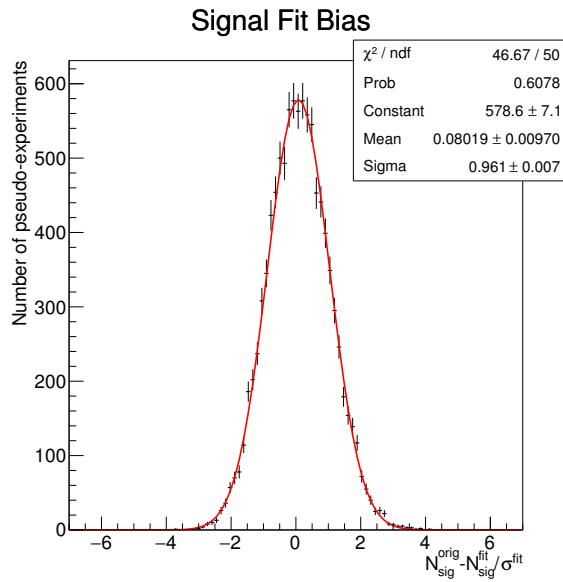
Figure 7.10: Non-simultaneous fit pulls from 10000 pseudo-experiments.

As it can be seen, for the non-simultaneous fit the pull distribution follows Gaussian distribution with the value of the standard deviation of 0.961 ± 0.007 which is reasonably good but not perfect. But since the statistical error is of the order of 100% this is not so significant. The bias shows 8% preference for lower signal yields, and hence this will be added as a systematic uncertainty. The same process was repeated for the simultaneous fit and results can be seen in Table 7.13.

Fit	$\mathcal{B}$	Statistical Error	Distribution Quality	Bias
Non-simultaneous	1.0×10^{-8}	$\approx 100\%$	4%	8%
Simultaneous	1.0×10^{-8}	$\approx 100\%$	6%	1%

Table 7.13: Signal bias estimate from 10 000 pseudo-experiments for both the simultaneous and non-simultaneous fit.

Others

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

7.3.6 Blinded Data Fits

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

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

The maximum unbinned likelihood fit to blinded signal data after all selection is shown in Figure 7.11 for all categories of FCME. As the signal region is blinded in this case the fit model uses the background-only hypothesis. The total number of expected background events, N_b, can be then obtained by integrating the total PDF in the signal region.
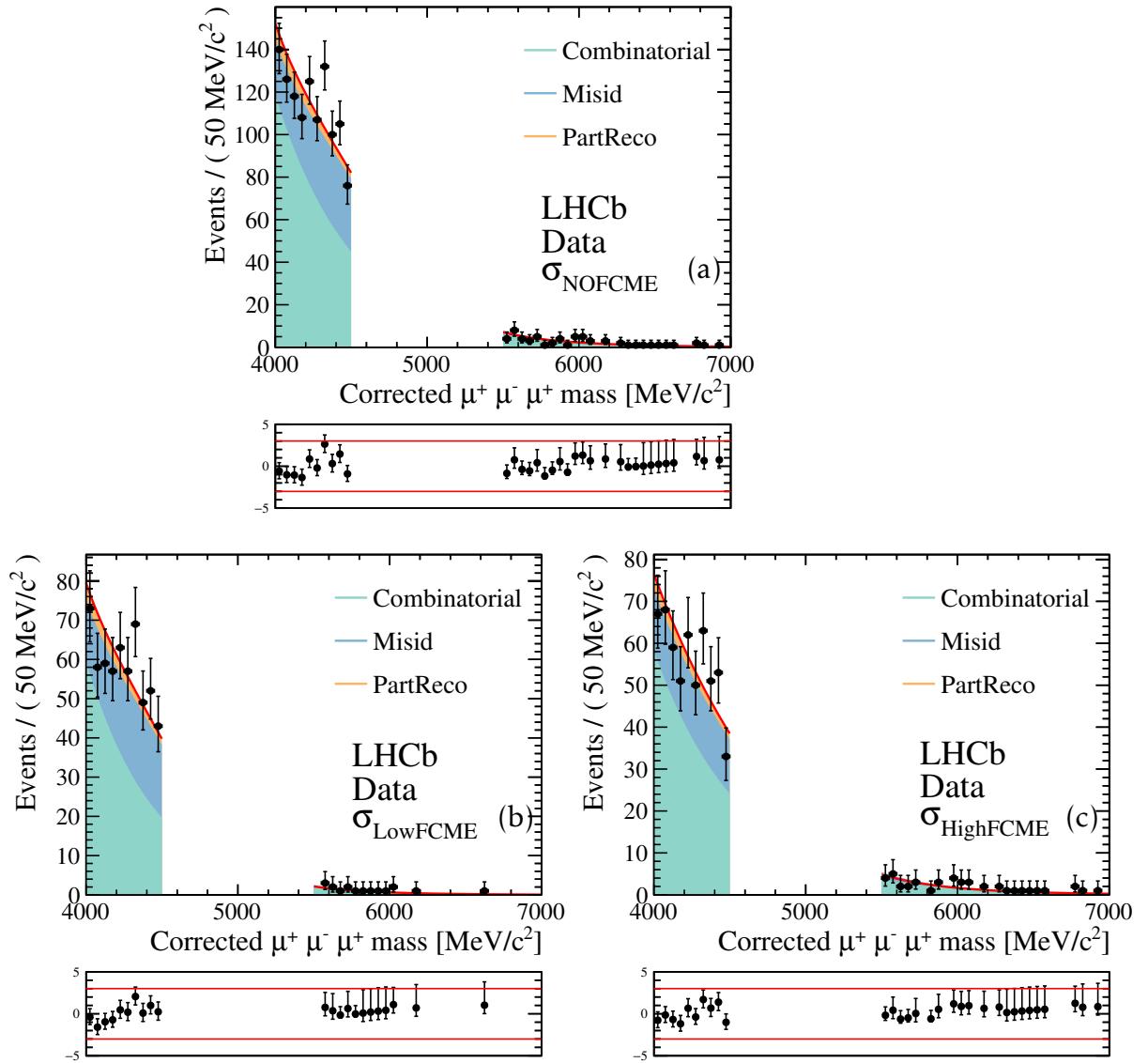
Figure 7.11: (a) Unbinned maximum likelihood fit to the blinded data in one bin of FCME. Simultaneous unbinned maximum likelihood fit to blinded data after full selection chain in two bins of FCME, with (b) fit to σ_{lowFCME} bin, (c) σ_{highFCME} bin.

185

Fit Parameter		Status	Constraint	Obtained in
N^{sig}	$\mathcal{B}(B^+ \to \mu^+\mu^-\mu^+\nu)$	Free	-	This Fit
	$\mathcal{B}(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)$	Cnstr.	gaus.	Equation 7.16
	$R^{21}_{\text{FCME}}(B^+ \to \mu^+\mu^-\mu^+\nu)$	Cnstr.	gaus.	Equation 7.5
	$R^{26}_{\text{FCME}}(B^+ \to \mu^+\mu^-\mu^+\nu)$	Cnstr.	gaus.	Equation 7.5
	$N(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)^{RunI}_{\text{FCME}}$	Cnstr.	gaus.	Table 7.8
	$N(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)^{2016}_{\text{FCME}}$	Cnstr.	gaus.	Table 7.8
N^{PR}	$\mathcal{B}(B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu)$	Cnstr.	gaus.	Table 7.10
	$\mathcal{B}(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)$	Cnstr.	gaus.	Equation 7.16
	$R^{21}_{\text{FCME}}(B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu)$	Cnstr.	gaus.	Table 7.10
	$N(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)^{RunI}_{\text{FCME}}$	Cnstr.	gaus.	Table 7.8
	$N(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)^{2016}_{\text{FCME}}$	Cnstr.	gaus.	Table 7.8
N^{Misid}		Cnstr.	mvg_gaus.	Figure 7.9
N^{Combi}		Free	-	This fit
MisID Shape Parameters (Crystal Ball function)				
μ_{misid}		Cnstr.	mvg_gaus.	Figure 7.9
σ_{misid}		Cnstr.	mvg_gaus.	Figure 7.9
Others		Fixed		Figure 7.9
PartReco Shape Parameters (sum of two Crystal Ball functions)				
All		Fixed		Figure 7.8
Signal Shape Parameters (Double-sided Crystal Ball function)				
All		Fixed		Figure 7.7
Combinatorial Shape Parameters (exponential function)				
β		Free	-	This fit

Table 7.14: For all constrained variables the range is set to be within $\pm 5\sigma$. Cnstr. stands for constrained variables, gaus. for *gaussian* constraint and mvg_gaus. *multivariate gaussian* constraint.

7.3.7 Expected Sensitivity

The CL_s method [74] is used to produce an expected upper limits from the blinded data fits. This is possible thanks to the fact that this method is based on generation of pseudo-datasets with different signal hypotheses. Therefore, it only requires the resulting PDFs of the blinded fits (simultaneous and non-simultaneous) with which pseudo-datasets can be produced.

Systematic uncertainties directly influence the relative efficiencies between signal and normalisation channel. They can therefore be added as gaussian constraints on the relevant efficiency ratio. The expected exclusion limits are then computed and summarized in Table 7.15. Partial systematics only include systematics related to branching fractions, whereas full systematics include all the effects from Table 7.12. In the simultaneous fit these effects are assumed to be 100% correlated between the two bins of fractional corrected mass, but uncorrelated between themselves. As expected, the limit weakens with the addition of systematics. On the other hand there is an improvement in expected limit by 20% obtained with the use of the simultaneous fit compared to the non-simultaneous fit with the full systematics included. Addition of 2016 data also improved the expected sensitivity of the search.

Confidence Interval	Data	Systematics	Fit Type	Value
Expected 95% CL	Run I	Partial	Simultaneous	3.3×10^{-8}
Expected 95% CL	Run I	Full	Simultaneous	3.5×10^{-8}
Expected 95% CL	Run I and 2016	Partial	Non-simultaneous	3.0×10^{-8}
Expected 95% CL	Run I and 2016	Full	Non-simultaneous	3.3×10^{-8}
Expected 95% CL	Run I and 2016	Partial	Simultaneous	2.5×10^{-8}
Expected 95% CL	Run I and 2016	Full	Simultaneous	2.8×10^{-8}

Table 7.15: Resulting expected exclusion limits with both non-simultaneous and simultaneous fits. CL stands for confidence level.

7.3.8 Unblinded Data Fits

After the full analysis strategy was reviewed the signal datasets were unblinded, observing no significant signal. The simultaneous unbinned extended maximum likelihood fit to all the data was performed. As expected with no excess signal, the background-only hypothesis fit described the data very well, which can be seen in Figure 7.12.
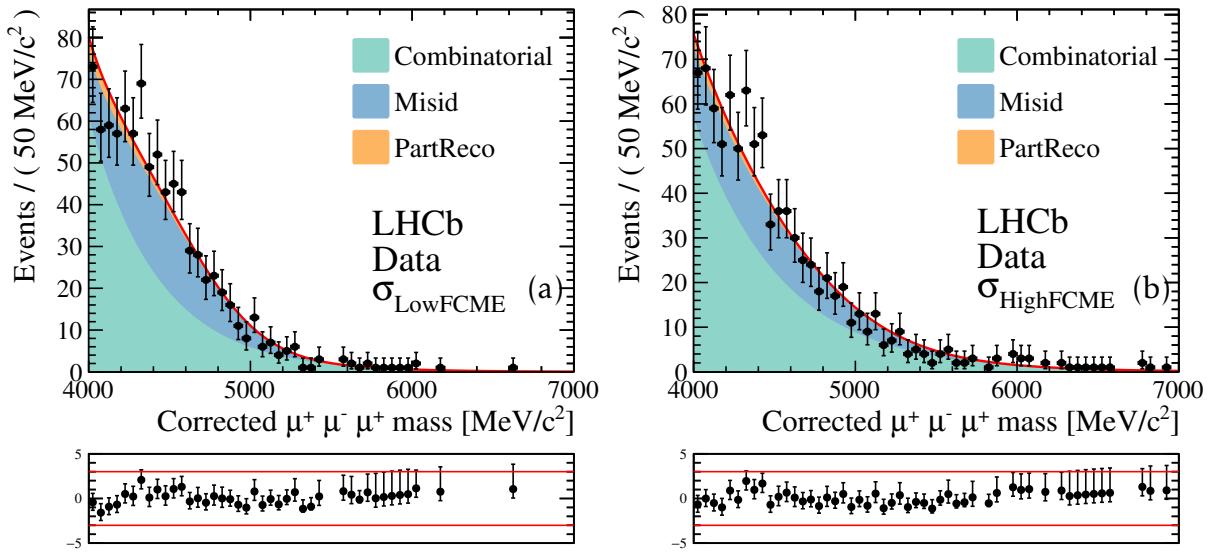
Figure 7.12: Simultaneous unbinned extended maximum likelihood fit to unblinded data after full selection chain in two bins of FCME and with $\mathcal{B} = 0$, with (b) fit to σ_{lowFCME} bin, (c) σ_{highFCME} bin.

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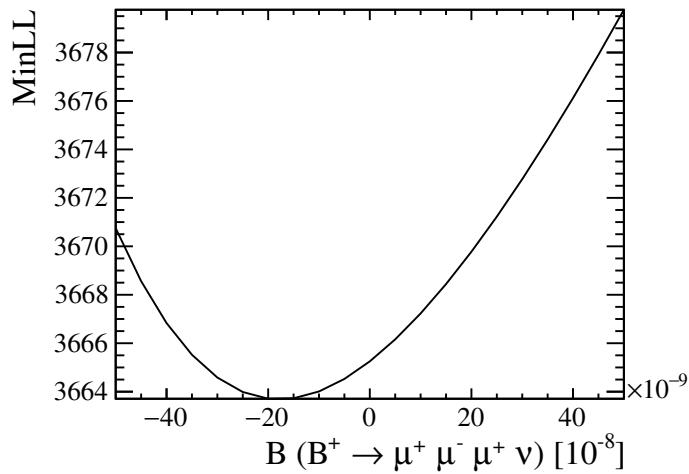
Figure 7.13: Minimized -log likelihood value at different $\mathcal{B}$.

The simultaneous unbinned extended maximum likelihood fit to all the data with signal+background hypothesis is given in Figure 7.14, converging at value of $\mathcal{B} = -1.8 \times 10^{-8}$. This hence represents a negative yield fluctuation. The full list of all free and constrained parameter values obtained in this fit is shown in Table 7.16. This fit result then can be translated into yields in all the fitting range and specifically the blinded region by integrating the resulting PDF in a relevant region as shown in Table 7.17.
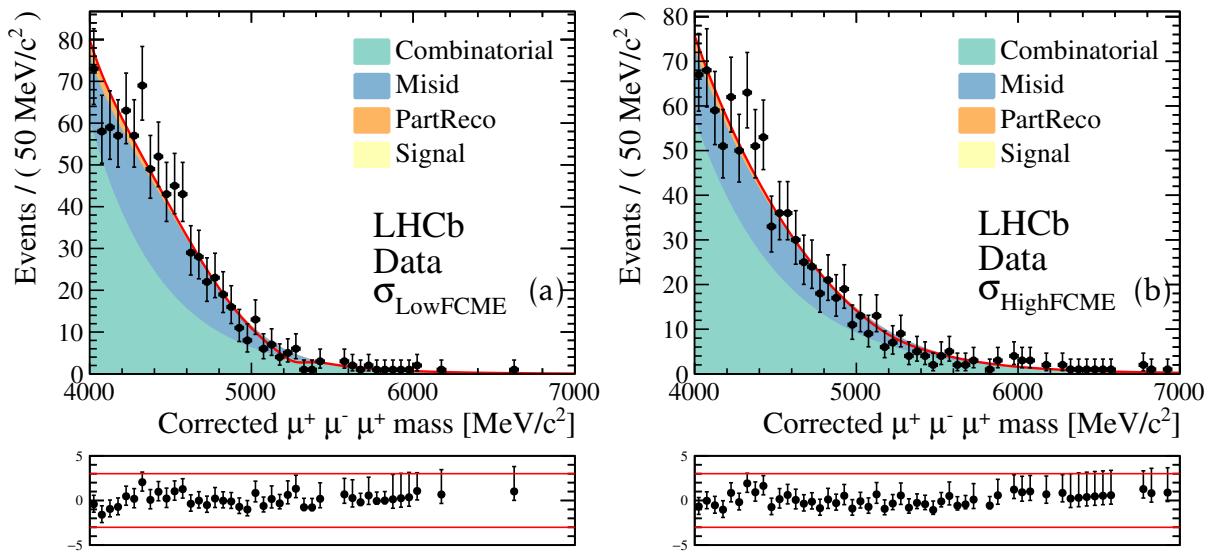
Figure 7.14: Simultaneous unbinned maximum likelihood fit to unblinded data after full selection chain in two bins of FCME, with (b) fit to σ_{lowFCME} bin, (c) σ_{highFCME} bin. The signal component is visible as negative fluctuation.

| Fit Parameter	Constraint	Fitted Value	Pull
$\mathcal{B}(B^+ \to \mu^+\mu^-\mu^+\nu)$	free	$-1.8 \pm 0.9 \times 10^{-8}$	-
$\mathcal{B}(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)$	$(6.12 \pm 0.188) \times 10^{-5}$	$(6.11 \pm 0.185) \times 10^{-5}$	0.1σ
$\mathcal{B}(B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu)$	$(4.09 \pm 0.5) \times 10^{-7}$	$(4.12 \pm 0.5) \times 10^{-7}$	-0.1σ
$\mu_{misID_{\text{highFCME}}}$	4090 ± 143	4100 ± 123	-0.1σ
$\mu_{misID_{\text{lowFCME}}}$	4240 ± 75	4330 ± 56	-1.6σ
$\sigma_{misID_{\text{highFCME}}}$	637 ± 62	619 ± 56	0.3σ
$\sigma_{misID_{\text{lowFCME}}}$	447 ± 35.9	401 ± 29.9	0.5σ
$R^{26}_{\text{highFCME}}(B^+ \to \mu^+\mu^-\mu^+\nu)$	1.98 ± 0.0189	1.97 ± 0.0788	0.1σ
$R^{26}_{\text{lowFCME}}(B^+ \to \mu^+\mu^-\mu^+\nu)$	3.30 ± 0.0	3.31 ± 0.1	-0.1σ
$R^{21}_{\text{highFCME}}(B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu)$	0.0397 ± 0.00118	0.0397 ± 0.00116	0.0σ
$R^{21}_{\text{lowFCME}}(B^+ \to (D^0 \to K^+\pi^-\mu^+\mu^-)\mu^+\nu)$	0.0271 ± 0.000809	0.0271 ± 0.000794	0.0σ
$R^{21}_{\text{highFCME}}(B^+ \to \mu^+\mu^-\mu^+\nu)$	1.97 ± 0.0305	1.96 ± 0.131	0.1σ
$R^{21}_{\text{lowFCME}}(B^+ \to \mu^+\mu^-\mu^+\nu)$	3.13 ± 0.0498	3.15 ± 0.233	-0.1σ
$N(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)^{2016}_{\text{highFCME}}$	29800 ± 176	29800 ± 173	0.0σ
$N(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)^{2016}_{\text{lowFCME}}$	64700 ± 259	64700 ± 254	0.0σ
$N(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)^{Run\,I}_{\text{highFCME}}$	64100 ± 257	64100 ± 253	0.0σ
$N(B^+ \to (J/\psi \to \mu^+\mu^-)K^+)^{Run\,I}_{\text{lowFCME}}$	109000 ± 337	109000 ± 331	0.0σ
$N^{scaled}_{misID_{\text{highFCME}}}$	277 ± 15.5	278 ± 15.2	-0.1σ
$N^{scaled}_{misID_{\text{lowFCME}}}$	322 ± 16.9	331 ± 16.3	-0.6σ
β_{highFCME}	free	$-0.00183 \pm 9.19 \times 10^{-5}$	-
β_{lowFCME}	free	-0.00226 ± 0.000129	-
$N_{combi_{\text{highFCME}}}$	free	620 ± 34.3	-
$N_{combi_{\text{lowFCME}}}$	free	531 ± 35.4	-

Table 7.16: Fit results for all floating (free and constrained) variables in the unblinded data fit. The $\mathcal{B}(B^+ \to \mu^+\mu^-\mu^+\nu) = -1.8 \times 10^{-8}$. Variables R^S_K are the efficiency ratios obtained by normalising the decays to $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ decays where S stands for stripping and K for the FCME. $N^{scaled}_{misID_K}$ is the number of misID events, N_{combi_K} is the number of combinatorial events, β_K is the exponential constant, μ_{misID_K} and σ_{misID_K} are the mean and the σ of the CB function.

191

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

Table 7.17: Resulting yields for different components from the corrected mass fit with $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = -1.8 \times 10^{-8}$.

Chapter 8

Result and Conclusion

8.1 Limit Setting

As no significant signal is observed, a limit for $\mathcal{B}(B^+ \to \mu^+\mu^-\mu^+\nu)$ is set. This is achieved by using the CL_s method described in subsection 7.3.7 and the CL_s p values are shown in Figure 8.1 together with both expected and observed curves. In order to incorporate the systematics uncertainties in the limit setting they are added as 1D Gaussian constraints on the relevant efficiency ratios when calculating the limit. They are assumed to be 100% correlated between the bins of fractional corrected mass error but uncorrelated between the different effects. This gives the following limits summarised in Table 8.1, setting the limit $\mathcal{B}(B^+ \to \mu^+\mu^-\mu^+\nu) < 1.1(1.4) \times 10^{-8}$ at 90%(95%) confidence level. As it can be seen, the observed limit is better than expected limit resulting from a downward fluctuation of around 1σ compared to the expected sensitivity.
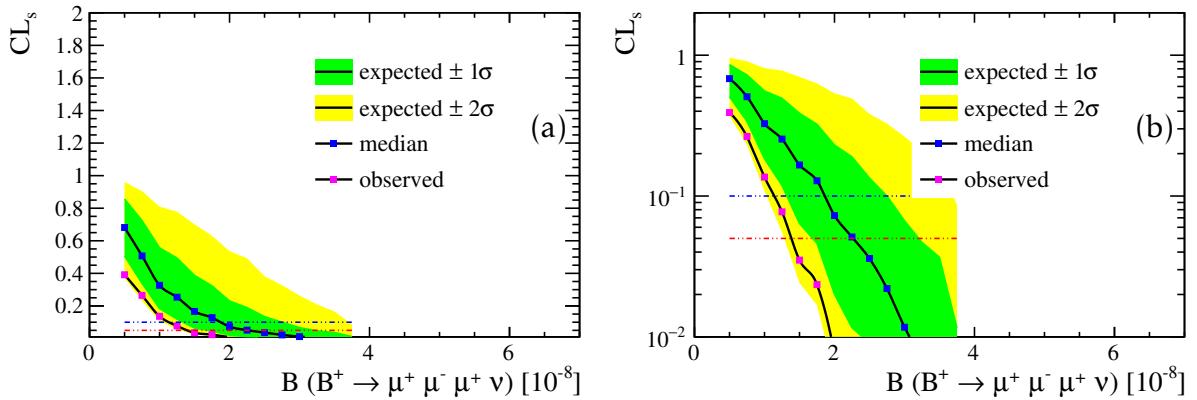
Figure 8.1: Expected and observed 90% (blue horizontal line) 95% (red horizontal line) CL exclusion limits for full Run I and 2016 simultaneous data fit accounting for all systematics with (a) normal and (b) logarithmic y-axis.

Expected/Observed	Confidence Interval	Value
Expected	90% CL	1.9×10^{-8}
Observed	90% CL	1.1×10^{-8}
Expected	95% CL	2.3×10^{-8}
Observed	95% CL	1.4×10^{-8}

Table 8.1: Resulting exclusion limits with simultaneous fit. The differences in the expected limits compared to those shown in Table 7.15 are due to the increase in data stastistics as well as different fit parameter values.

8.2 Conclusion

In conclusion most of this thesis was dedicated to the search for the decay $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$. This search was performed using $4.7\,\text{fb}^{-1}$ of proton-proton collision data collected by the LHCb experiment. No significant signal was observed for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ and an
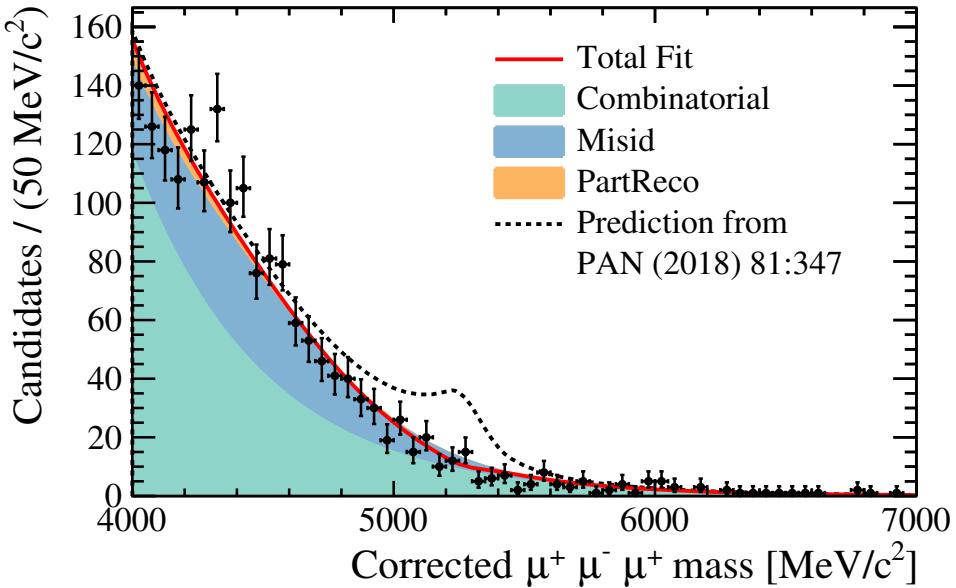
Figure 8.2: Corrected mass distribution of all selected $B^+ \to \mu^+\mu^-\mu^+\nu$ candidates with the fit overlaid. Samples with low and high corrected mass uncertainty are fitted as individual samples but are merged in the figure. As the signal yield is negative the total fit in red shows a downward fluctuation. The dashed line represents the fit result if the signal had the branching fraction predicted in [1].

upper limit of $< 1.4 \times 10^{-8}$ at 95% confidence level was set for the branching fraction, where the minimum of the two $\mu^+\mu^-$ mass combinations is below $980 \, \text{MeV}/c^2$. This limit disagrees with a recent theoretical calculation based on the vector dominance model [1], where the $\mathcal{B}(B^+ \to \mu^+\mu^-\mu^+\nu) = 1.3 \times 10^{-7}$. In order to visualize the fit with this branching fraction hypothesis, the fit is plotted in Figure 8.2. It can be seen that this signal would have been clearly visible if it was there. Under the assumption that the decay is dominated by intermediate vector mesons, the limit for the full kinematic region stays the same.

This thesis also presented PID work that concentrated on the correlation induced by having more than one muon in a final state.

8.3 Outlook

Despite the fact that no signal was observed, a stringent limit on the $\mathcal{B}(B^+ \to \mu^+\mu^-\mu^+\nu)$ was set with direct impact on the prediction given by [1]. As mentioned in section 2.10, the naive estimate of $\mathcal{B}(B^+ \to \mu^+\mu^-\mu^+\nu) = 1.0 \times 10^{-8}$ is therefore not too far from the limit set. At the moment of the writing of this thesis LHCb collected $8\,\mathrm{fb}^{-1}$ of data, which means that the dataset for the analysis doubled. So assuming that the dataset doubles with the same ratio between the background and signal the limit would improve by a factor of $1/\sqrt{2}$. This would therefore allow one to reach the naive branching fraction estimate.

Bibliography

[1] A. V. Danilina and N. V. Nikitin, *Four-Leptonic Decays of Charged and Neutral B Mesons within the Standard Model*, Phys. Atom. Nucl. **81** (2018), no. 3 347, [Yad. Fiz.81,no.3,331(2018)].

[2] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*, Physics textbook, Wiley, New York, NY, 2008.

[3] S. F. Novaes, *Standard model: An Introduction*, in *Particles and fields. Proceedings, 10th Jorge Andre Swieca Summer School, Sao Paulo, Brazil, February 6-12, 1999*, pp. 5–102, 1999. `arXiv:hep-ph/0001283`.

[4] A. Pich, *The Standard Model of Electroweak Interactions*, in *High-energy Physics. Proceedings, 18th European School (ESHEP 2010): Raseborg, Finland, June 20 - July 3, 2010*, pp. 1–50, 2012. `arXiv:1201.0537`.

[5] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321, [,157(1964)].

[7] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585, [,162(1964)].

[9] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[10] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[11] T. W. B. Kibble, *Symmetry breaking in nonAbelian gauge theories*, Phys. Rev. **155** (1967) 1554, [,165(1967)].

[12] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367.

[13] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys. **C40** (2016), no. 10 100001.

[14] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531, [,648(1963)].

[15] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[16] L.-L. Chau and W.-Y. Keung, *Comments on the Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **53** (1984) 1802.

[17] Y. H. Ahn, H.-Y. Cheng, and S. Oh, *Wolfenstein parametrization at higher order: Seeming discrepancies and their resolution*, Physics Letters B **703** (2011), no. 5 571 .

[18] C. Jarlskog, *Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Violation*, Phys. Rev. Lett. **55** (1985) 1039.

[19] UTfit, M. Bona *et al.*, *Results as of summer 2016, updated results and plots available at: http:utfit.org/UTfit/WebHome*, JHEP **10** (2006) 081, `arXiv:hep-ph/0606167`.

[20] CKMfitter Group, J. Charles *et al.*, *Results as of up to ICHEP 2016, updated results and plots available at: http://ckmfitter.in2p3.fr*, Eur. Phys. J. **C41** (2005), no. 1 1, `arXiv:hep-ph/0406184`.

[21] W. J. Marciano and A. Sirlin, *Electroweak Radiative Corrections to tau Decay*, Phys. Rev. Lett. **61** (1988) 1815.

[22] W.-S. Hou, *Enhanced charged Higgs boson effects in $B^- \to \tau \bar{\nu}, \mu \bar{\nu}$ and $b \to \tau \bar{\nu} + X$*, Phys. Rev. **D48** (1993) 2342.

[23] A. G. Akeroyd and S. Recksiegel, *The Effect of $H^\pm on B^\pm \to \tau^\pm \nu(\tau)$ and $B^\pm \to \mu^\pm$ muon neutrino*, J. Phys. **G29** (2003) 2311, arXiv:hep-ph/0306037.

[24] B. A. Dobrescu and A. S. Kronfeld, *Accumulating evidence for nonstandard leptonic decays of D_s mesons*, Phys. Rev. Lett. **100** (2008) 241802, arXiv:0803.0512.

[25] A. Arbey, F. Mahmoudi, O. Stal, and T. Stefaniak, *Status of the Charged Higgs Boson in Two Higgs Doublet Models*, Eur. Phys. J. **C78** (2018), no. 3 182, arXiv:1706.07414.

[26] Belle collaboration, A. Sibidanov *et al.*, *Search for $B^- \to \mu^- \bar{\nu}_\mu$ Decays at the Belle Experiment*, arXiv:1712.04123.

[27] Belle collaboration, I. Adachi *et al.*, *Evidence for $B^- \to \tau^- \bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle*, Phys. Rev. Lett. **110** (2013), no. 13 131801, arXiv:1208.4678.

[28] Belle collaboration, B. Kronenbitter *et al.*, *Measurement of the branching fraction of $B^+ \tau^+ \nu_\tau$ decays with the semileptonic tagging method*, Phys. Rev. **D92** (2015), no. 5 051102, arXiv:1503.05613.

[29] BaBar collaboration, J. P. Lees *et al.*, *Evidence of $B^+ \to \tau^+ \nu$ decays with hadronic B tags*, Phys. Rev. **D88** (2013), no. 3 031102, arXiv:1207.0698.

[30] BaBar collaboration, B. Aubert *et al.*, *A Search for $B^+ \to \ell^+ \nu_\ell$ Recoiling Against $B^- \to D^0 \ell^- \bar{\nu} X$*, Phys. Rev. **D81** (2010) 051101, arXiv:0912.2453.

[31] G. Burdman, J. T. Goldman, and D. Wyler, *Radiative leptonic decays of heavy mesons*, Phys. Rev. **D51** (1995) 111, arXiv:hep-ph/9405425.

[32] M. Beneke and J. Rohrwild, *B meson distribution amplitude from $B \to \gamma l \nu$*, Eur. Phys. J. **C71** (2011) 1818, arXiv:1110.3228.

[33] Y.-M. Wang, *Subleading power corrections in radiative leptonic B decay*, Nucl. Part. Phys. Proc. **285-286** (2017) 75, arXiv:1609.09813.

[34] Belle collaboration, A. Heller *et al.*, *Search for B^{++} decays with hadronic tagging using the full Belle data sample*, Phys. Rev. **D91** (2015), no. 11 112009, arXiv:1504.05831.

[35] 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.

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

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

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

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

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

[41] 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.*

[42] 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.

[43] 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.

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

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

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

[47] 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.

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

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

[50] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118** (2017), no. 19 191801, arXiv:1703.05747.

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

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

[53] 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.

201

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

[55] LHCb collaboration, I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

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

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

[58] 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.

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

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

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

[62] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.

[63] K. Dungs, *Optimisation of the software-based muon identification at the LHCb experiment*, Master's thesis, Tech. U., Dortmund (main), 2015-09-18.

[64] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

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

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

[67] U. Egede, P. Owen, and W. Sutcliffe, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, Tech. Rep. LHCb-ANA-2014-048, CERN-LHCb-ANA-2014-048, Apr, 2014.

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

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

[70] 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.

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

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

[73] LHCb collaboration, R. Aaij *et al.*, *First observation of the decay $D^0 \to K^- \pi^+ \mu^+ \mu^-$ in the ρ^0-ω region of the dimuon mass spectrum*, Phys. Lett. **B757** (2016) 558, arXiv:1510.08367.

[74] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. **G28** (2002) 2693.

Appendices

Appendix A

Fitting functions

A.1   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, [65]. Hence it has the same number of parameters and is usually denoted as $I(m, \mu_{IP}, \sigma_{IP}, \lambda, \zeta, \beta, a_1, n_1, a_2, n_2)$.

A.2   Crystal Ball Function

Crystal Ball (CB) function [66] 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

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

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 $f(x; \alpha, n, \overline{\mu}, \sigma)$:

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

A.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.

Appendix B

Correlation Matrices

B.1   Correlation Matrices for backgrounds in BDTs
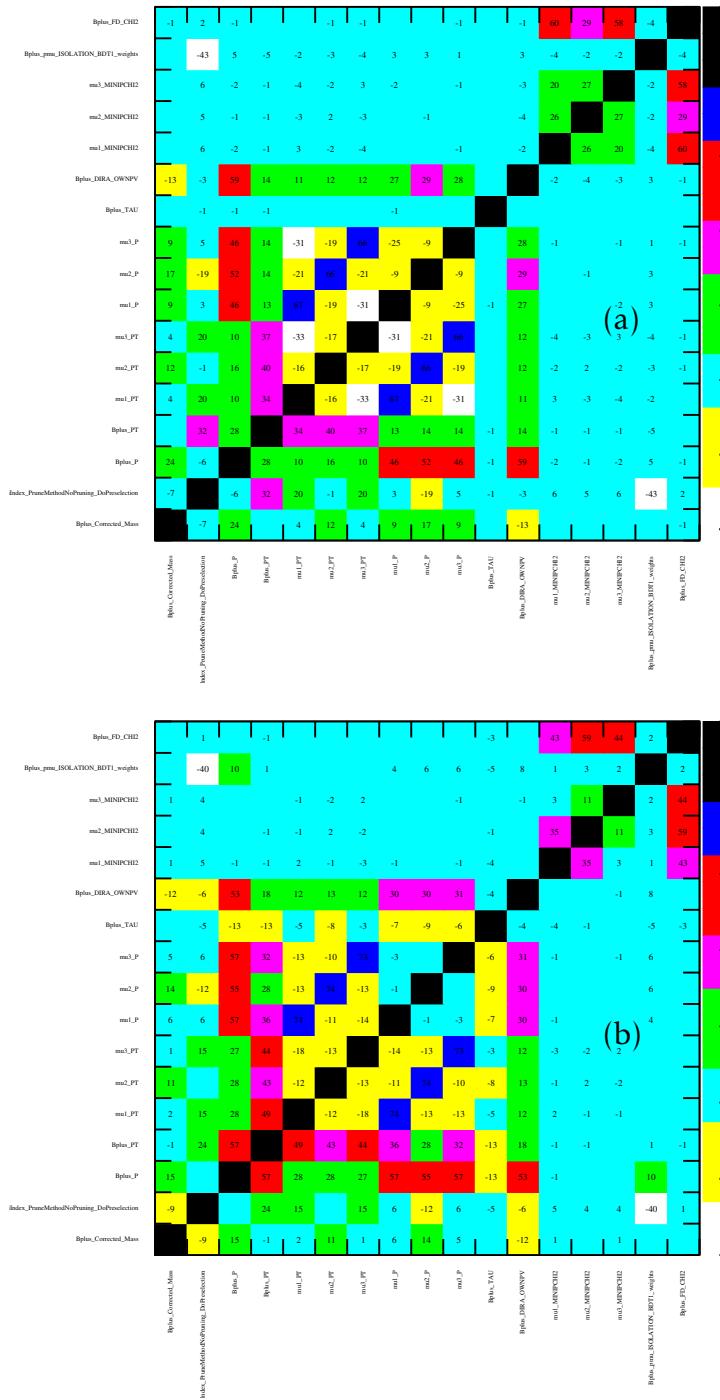
Figure B.1: Correlation matrix for all input variables, corrected mass as well resulting BDT variable for both (a) Run I Combinatorial BDT (b) 2016 Combinatorial BDT for background sample.
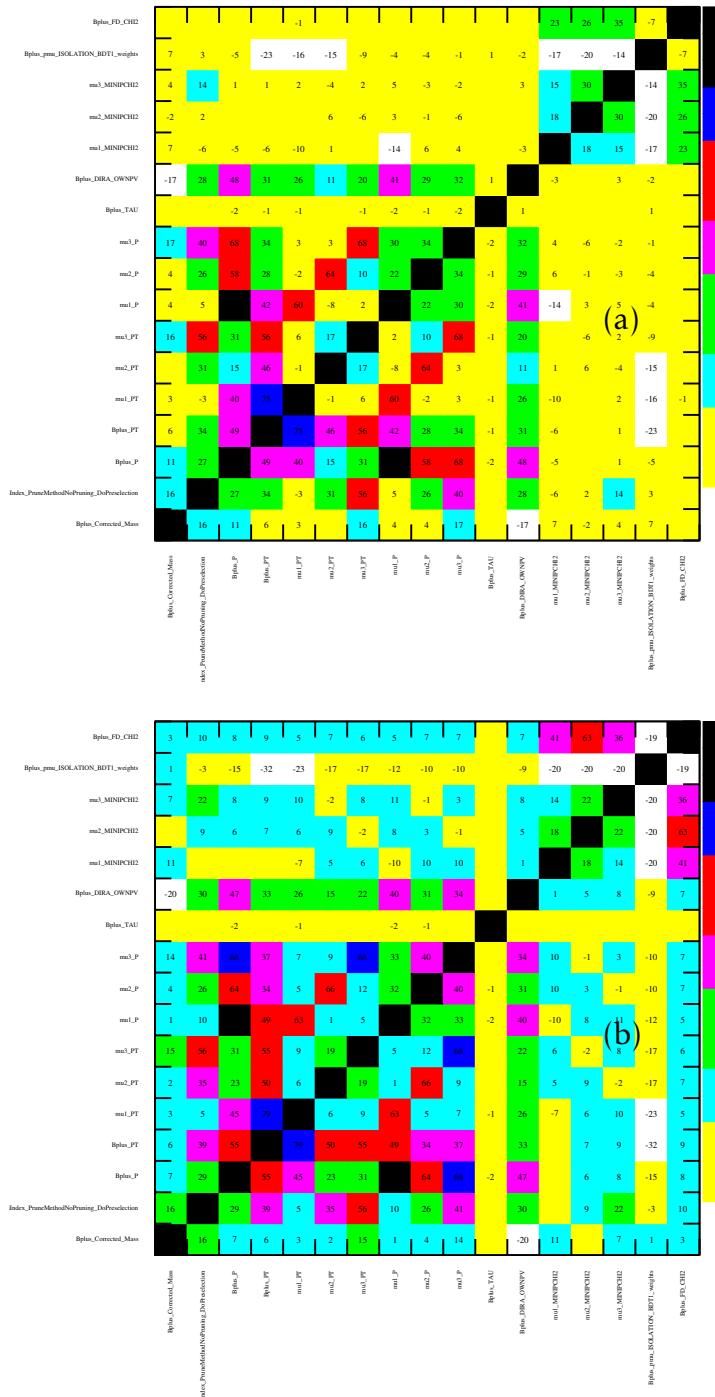
Figure B.2: Correlation matrix for all input variables, corrected mass as well resulting BDT variable for both (a) Run I Misid BDT (b) 2016 Misid BDT for background sample.

Appendix C

License

Figure C.1: License to re-use Figure 3.3 in chapter 3.