



university of
 groningen

faculty of science
and engineering

Improving the sensitivity study of $B_c^+ \rightarrow \tau^+ \nu_\tau$ at the LHCb detector by rejecting the three most dangerous background modes

University of Groningen

Pietro Albanese S4103394
Bachelor's Research Project

First Examiner: dr. M. Mulder
Second Examiner: dr. L. Willmann

7 July 2022

Abstract

As part of a sensitivity study for the LHCb upgrade, the invariant mass of the pions produced in the decays $B_c^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ and $B^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ is analysed, in order to distinguish these two signal modes from the three most dangerous background modes: $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$, $B^+ \rightarrow \pi^+\pi^-\pi^+D^{0*}$ and $B^+ \rightarrow D^-(\rightarrow \pi^-\pi^+\pi^-)\pi^+\pi^+$ (normalization mode). The data were simulated using the fast simulation package RapidSim and it was found that a cut on the invariant mass of the pions at $m_{inv} = 1.8$ GeV is able to remove 99.98% of the normalization mode, 88.82% of $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$ and 86.86% of $B^+ \rightarrow \pi^+\pi^-\pi^+D^{0*}$, while maintaining 99.999% of both signal modes. Then the effect of this cut is studied by performing a MultiVariate Analysis, whose purpose is to separate signal from background by analysing multiple observables. Lastly, a likelihood fit was simulated to estimate the number of yields of B_c^+ and B^+ . It was found that the uncertainty on the number of yields of B_c^+ and B^+ decreased by 17.7% and 4.4% respectively by including the invariant mass cut.

Acknowledgements

With this section I want to thank all the people who helped me during this Bachelor's Research Project.

First of all, I want to thank my supervisor Mick Mulder for explaining all the theory behind $B_c^+ \rightarrow \tau^+ \nu_\tau$, for answering the numerous questions I had in the past two months and for giving me feedback on each step of this project.

I also want to thank Kristof de Bruyn and Maarten van Veghel for giving me suggestions during our weekly meetings on what to improve in each step of my project. I want to thank Maria Domenica Galati and Jelte Rinus de Jong for providing me with the data simulated using RapidSim, for giving me the python codes needed to run the MVA and the Likelihood fit and for answering any question I had about the data and the code.

Lastly, I want to thank Lotte de Wit, who was also working on $B_c^+ \rightarrow \tau^+ \nu_\tau$ for her thesis, for all the help and support given during this Bachelor's Research Project.

Contents

1	Introduction	4
2	Theory	6
2.1	Standard Model	6
2.1.1	Elementary particles	6
2.1.2	Lepton Flavour Universality	8
2.1.3	CKM matrix	9
2.2	$B_c^+ \rightarrow \tau^+ \nu_\tau$	10
3	LHCb Detector	12
3.1	Detector Components	12
3.2	B-Tracking Tool	14
4	Analysis Strategy	15
4.1	RapidSim	15
4.2	Signal and Background	15
5	Invariant mass	17
5.1	Invariant mass	17
5.2	Cut on invariant mass of pions	17
6	MVA	20
6.1	MultiVariate Anlaysis	20
6.2	Results of MVA training	21
7	Maximum Likelihood Fit	24
7.1	Likelihood Fit	24
7.1.1	Corrected mass	25
7.2	Results of the Likelihood Fit	25
8	Discussion	28
9	Conclusion	30

Chapter 1

Introduction

The Standard Model is a particle physics theory which describes the collection of fundamental particles that constitute the known universe and are separated into fermions and bosons. The former are comprised of leptons and quarks which have half-integer spin, while the latter have integer spin. Many experiments about particle physics are conducted at CERN (European Council for Nuclear Research) in Switzerland, where one of the most important experiments is the LHCb (Large Hadron Collider beauty) detector. As a result of recent upgrades to the VERtex LOcator (VELO), the LHCb detector will be able to detect the charge left by a direct hit of a b-meson. Consequently, more precise measurements of beauty mesons will be possible, and the detection of decay modes that have never been seen before will also be achievable. For instance, the decay mode $B_c^+ \rightarrow \tau^+ \nu_\tau$, which is the main topic of this thesis, could be observed.

In order to understand the importance of this $B_c^+ \rightarrow \tau^+ \nu_\tau$, the concept of Lepton Flavour Universality (LFU) must be explained. LFU states that in the Standard Model no force discriminates between the three different families of leptons: electron, muon and tauon. However, in the last decade, the research undertaken at CERN revealed the possibility that LFU might not be a law of physics [1]. If the data taken at LHCb will show proof of LFU violation, they would show the presence of new physics, since this result would be impossible in the Standard Model. In such case, theorists argue that a new particle, called Leptoquark which couples quarks with leptons, may exist, or that a charged Higgs boson, which has not been discovered yet, might be responsible for LFU violation.

Therefore, the decay $B_c^+ \rightarrow \tau^+ \nu_\tau$ is quite important not only because it has never been observed, but also because it could provide further evidence of LFU violation.

To estimate the sensitivity of a search of this decay with the upgraded LHCb detector, simulations are used. One of the simulation packages is called RapidSim which is a fast simulator of b-decays at LHCb and allows to simulate the decay modes of B_c^+ and B_u^+ mesons. These particles have numerous decay modes, therefore $B_c^+ \rightarrow \tau^+ \nu_\tau$ and $B_u^+ \rightarrow \tau^+ \nu_\tau$, which are the most interesting ones, are considered as signal. The main issue with measuring these decays is that the particles B_c^+ , B_u^+ and τ^+ decay quickly and the ν_τ cannot be detected. In order to detect the meson, the VELO detector is being upgraded as explained in Section 3.1, while the τ^+ decays before reaching the detector. The tauon has numerous decay modes. For instance one of the most important decays is $\tau^+ \rightarrow \mu^+ \nu_\mu \bar{\nu}_\tau$ or the equivalent decay with the positron instead of the muon. However, since the neutrinos cannot be detected, it is difficult to reconstruct the trajectory of tauons from only one particle. Therefore, the decay $\tau^+ \rightarrow \pi^+ \pi^- \pi^+ \bar{\nu}_\tau$ is considered, so the trajectories of the three pions can be used to precisely calculate where the tauon decayed. Hence the full decay chain studied in this analysis is $B_c^+ \rightarrow \tau^+ (\rightarrow \pi^+ \pi^- \pi^+ \bar{\nu}_\tau) \nu_\tau$.

Moreover, among all the other decay modes there are some which look similar to the signal mode due to the presence of three pions in the final state of the decay, therefore they could be recognized as signal and affect the measurements. Hence, these decay modes are considered as background and it is important to be able to distinguish signal from the other decay modes. **In this thesis, the three most dangerous background modes will be studied by taking into account the invariant mass of the three pions produced during the decay, in order to find a cut which separates most of the background from the signal and to propagate this selection to the final sensitivity estimate. Then, a MultiVariate Analysis will be performed to separate signal from background modes using decision trees. In the end, a Likelihood fit will be simulated to analyse the performance of the invariant mass cut.**

Chapter 2

Theory

2.1 Standard Model

The Standard Model (SM) is a particle physics theory which describes the elementary particles that form matter and the fundamental forces, such as electromagnetism, the weak force and the strong force, with the exception of gravity. Although this theory is able to provide accurate predictions of several phenomena, it is considered the low-energy part of a more comprehensive theory [2]. For instance, it cannot explain the problem of the asymmetry of matter and anti-matter in the universe and it does not include a Dark Matter particle.

2.1.1 Elementary particles

In the SM there are two groups of elementary particles as shown in Figure 2.1: fermions with half-integer spin, and bosons with integer spin. Among the first class, there are two families: leptons and quarks. There are six leptons which are divided in three generations, each one with one negatively charged and one neutral lepton. Therefore there are three charged leptons (electron, muon and tauon) and three neutral ones (electron neutrino, mu neutrino and tau neutrino). The electron is the lightest charged lepton and the tauon is the heaviest; the neutrinos were considered to be massless until recently, when it was discovered that they oscillates between generations, which is an effect that requires the neutrinos to have mass. Among the charged leptons, only the electron and muon can normally be detected directly since they have a long enough lifetime. On the contrary, the tauon has a mean life of $\tau = (290.3 \pm 0.5) \cdot 10^{-15}$ s, which is too short to reach the detector, so only its decay products can be detected. As for the neutrinos, they rarely interact with matter, so they are not tracked by the detectors.

The other family of fermions is comprised of quarks, which are six in total and are called: up, down, charm, strange, top and bottom. They are also separated into three generations and their mass also increases from the first to the third generation. They mainly form mesons, which are particles with a quark and an anti-quark, and baryons, which have three quarks or three anti-quarks, however tetraquark (particles with four quarks) and pentaquarks (particles with five quarks) have also been observed.

Lastly, the second class of particles is comprised of bosons, which are force mediators. Among them, there are photons, which are the mediators of the electromagnetic interaction, gluons, which are the mediators of the strong interaction, and W^\pm and Z^0 bosons which are the mediators of the weak interaction. Besides these, there is also the gravitational interaction which is thought to be mediated by the graviton, however it has never been detected. Moreover, there is the Higgs boson, which gives mass to fermions and bosons due to the interaction of these particles with the Higgs field [3].

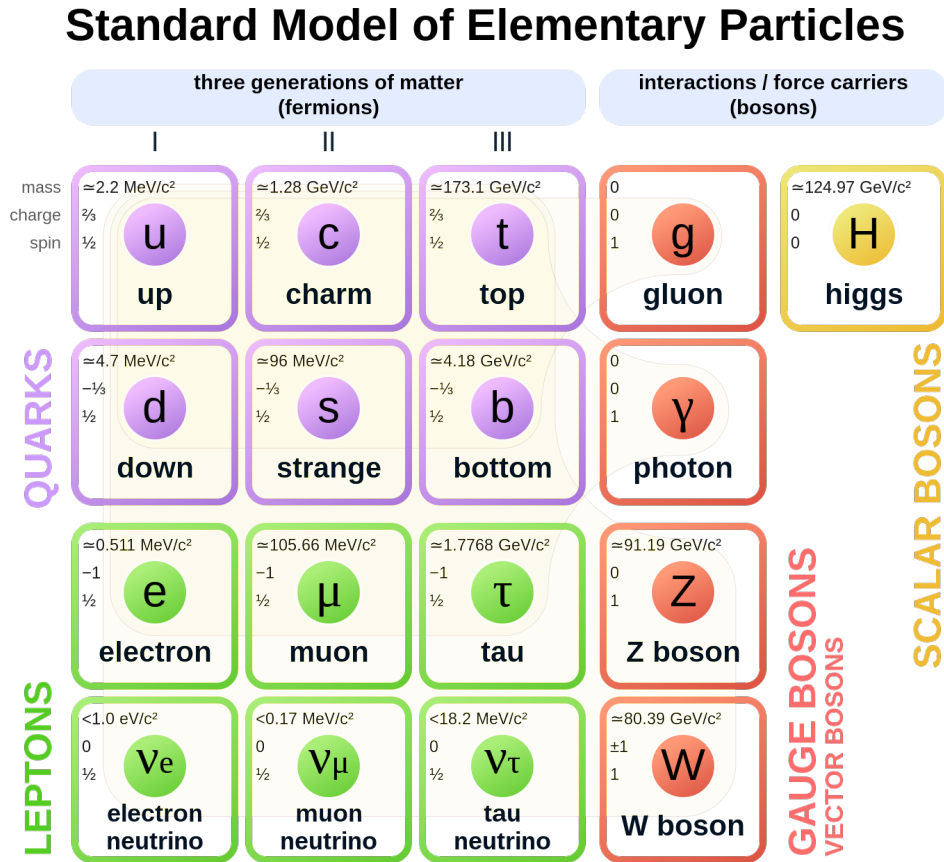


Figure 2.1: Standard Model of Particle Physics [4].

2.1.2 Lepton Flavour Universality

In the SM, even though the charged leptons have different masses and lifetimes, they couple with photons, W^\pm and Z^0 boson with the same strength regardless of their generation. This phenomenon is called Lepton Flavour Universality (LFU) [2].

This property has been tested by many experiments. For example, at the Large Electron-Positron Collider (LEP), LFU was tested in couplings between the Z boson and the charged leptons by measuring the partial width of $Z \rightarrow e^+e^-$, $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow \tau^+\tau^-$ and calculating their ratios. The obtained results were [5]:

$$\frac{\Gamma_{Z \rightarrow \mu^+\mu^-}}{\Gamma_{Z \rightarrow e^+e^-}} = 1.0009 \pm 0.0028 \quad (2.1)$$

$$\frac{\Gamma_{Z \rightarrow \tau^+\tau^-}}{\Gamma_{Z \rightarrow e^+e^-}} = 1.0019 \pm 0.0032 \quad (2.2)$$

However, recent measurements of $R(D^*)$ and $R(D)$ show a significant deviation between the SM prediction and the experimental value. The equations for these two ratios are:

$$R(D) = \frac{BR(B^0 \rightarrow D^- \tau^+ \nu_\tau)}{BR(B^0 \rightarrow D^- \mu^+ \nu_\mu)} \quad (2.3)$$

$$R(D^*) = \frac{BR(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)}{BR(B^0 \rightarrow D^{*-} \mu^+ \nu_\mu)} \quad (2.4)$$

The predicted values according to the SM are: $R(D) = 0.300 \pm 0.008$ and $R(D^*) = 0.252 \pm 0.003$, however the values calculated with the measurements conducted at BABAR, BELLE and LHCb are: $R(D) = 0.407 \pm 0.039 \pm 0.024$ and $R(D^*) = 0.304 \pm 0.013 \pm 0.007$. Therefore, the experimental values of $R(D)$ and $R(D^*)$ are 2.3σ and 3.4σ above the SM prediction respectively [1].

For this reason, there have been many theoretical studies which introduce new particles in order to explain this deviation. A possible solution proposes the introduction of a charged Higgs boson H^\pm in a 2 Higgs Doublet Model, while another suggests the existence of the Leptoquark, a particle that couples lepton and quarks as shown in Figure 2.3. Both theories could explain LFU violation, because these new particles would couple more strongly with the tauon compared to lighter leptons [7]. Therefore, LFU should be further investigated, not only to better understand this property, but also for the possibility to discover New Physics (NP) beyond the SM.

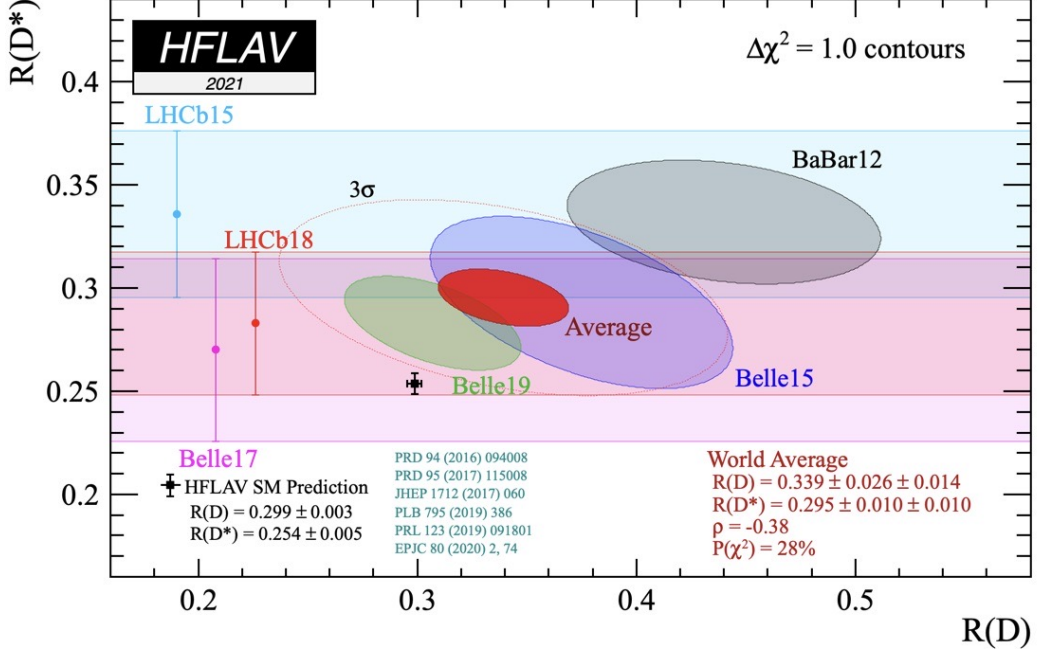


Figure 2.2: Measurements of $R(D)$ and $R(D^*)$ taken by different experiments [6].

In order to study this phenomenon, the decay $B_c^+ \rightarrow \tau^+ \nu_\tau$ will be analysed at LHCb during Run 3.

2.1.3 CKM matrix

In the SM, the weak interaction allows the quarks to change their flavours. The strength of this mixing between quarks is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix. Its terms are complex numbers, which allow for the introduction of CP violation in the SM. It is written as [8]:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \quad (2.5)$$

It is a 3x3 unitary matrix, which means that $\sum_i V_{ij} V_{ik}^* = \delta_{jk}$ and $\sum_j V_{ij} V_{kj}^* = \delta_{ik}$. The magnitude $|V_{ij}|$ is the amplitude of the transition from the q_i to the q_j quark, while the $|V_{ij}|^2$ is the probability of the transition [3]. Therefore, V_{cb} is the term which is involved in the decay $B_c^+ \rightarrow \tau^+ \nu_\tau$.

This matrix can also be written according to the Wolfenstein parametrisation as:

$$V_{CKM} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} \quad (2.6)$$

where $A = 0.826 \pm 0.018$, $\lambda = 0.22500 \pm 0.00067$, $\rho = 0.159 \pm 0.010$ and $\eta = 0.348 \pm 0.010$ [9].

2.2 $B_c^+ \rightarrow \tau^+ \nu_\tau$

The B_c^+ particle is a meson with an anti-bottom and a charm quark. It has many decay modes, but the one studied in this thesis is $B_c^+ \rightarrow \tau^+ \nu_\tau$, which has never been seen before and whose Feynman diagram is shown in Figure 2.2.

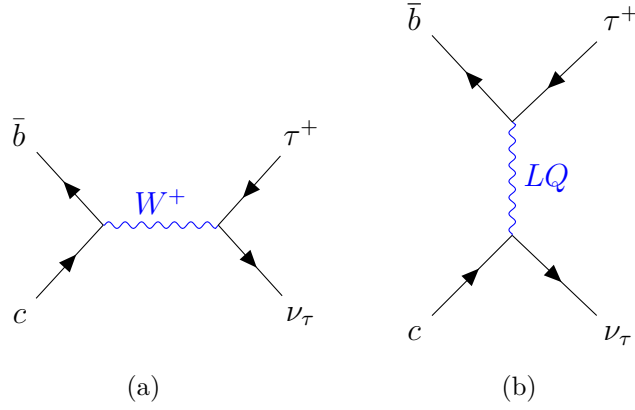


Figure 2.3: Feynman diagram of $B_c^+ \rightarrow \tau^+ \nu_\tau$: a) it shows the decay with the W^+ boson as mediator; b) it shows the decay with the Leptoquark (LQ) as mediator.

The SM prediction of the decay width of this decay is given by [7]:

$$\Gamma_{SM}(B_c^+ \rightarrow \tau^+ \nu_\tau) = \frac{G_F^2}{8\pi} |V_{cb}|^2 f_{B_c}^2 m_{B_c} m_\tau^2 \left(1 - \frac{m_\tau^2}{m_{B_c}^2}\right)^2 \quad (2.7)$$

where $G_F = 1.16638 \cdot 10^5 \text{ GeV}^2 (\hbar c)^3$ is the Fermi constant, $|V_{cb}| = (42.2 \pm 0.8) \cdot 10^{-3}$ is the CKM matrix element for when bottom and charm quarks interact with each other [9], $f_{B_c} = (0.434 \pm 0.015) \text{ GeV}$ is the decay constant¹ [10], $m_{B_c} = 6274.47 \pm 0.27 \text{ MeV}/c^2$ is the mass of the B_c^+ meson and $m_\tau = 1776.86 \pm 0.12 \text{ MeV}/c^2$ is the mass of the tauon. Moreover, given that the mean life of B_c^+ is $\tau = (0.510 \pm 0.009) \cdot 10^{-12} \text{ s}$, the predicted branching fraction can be calculated [7]:

¹The decay constant f_{B_c} is a property of Quantum ChromoDynamics (QCD) and it is determined using lattice QCD [10].

$$BR(B_c^+ \rightarrow \tau^+ \nu_\tau) = (2.36 \pm 0.19)\% \quad (2.8)$$

At LHCb, the branching fraction will be calculated with the following equation:

$$\frac{BR(B_c^+ \rightarrow \tau^+ \nu_\tau)}{BR(B^+ \rightarrow D^- \pi^+ \pi^+)} = \frac{N_{B_c^+}}{N_{B^+}} \frac{f_u}{f_c} \frac{\epsilon_{B^+}}{\epsilon_{B_c^+}} \quad (2.9)$$

where $N_{B_c^+}$ is the measured number of B_c^+ mesons, N_{B^+} is the measured number of B^+ mesons, $f_u \propto \sigma(pp \rightarrow B^+ X)$ is the production fraction of B^+ , $f_c \propto \sigma(pp \rightarrow B_c^+ X)$ is the production fraction of B_c^+ , ϵ_{B^+} is the detection efficiency of B^+ and $\epsilon_{B_c^+}$ is the detection efficiency of B_c^+ . In this sensitivity study, also a better estimate of the $N_{B_c^+}$ will be investigated.

Chapter 3

LHCb Detector

In order to study $B_c^+ \rightarrow \tau^+ \nu_\tau$, a powerful accelerator is needed, such as the Large Hadron Collider (LHC), in Geneva (Switzerland). At this accelerator, researchers study CP violation, decays of beauty and charm hadrons and matter-antimatter asymmetry. It has been taking measurements since 2009 from proton-proton collisions with a luminosity of $4 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$, which is expected to increase up to $2 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ during Run 3 [11]. The LHCb detector is a forward spectrometer, which can detect particles travelling at a pseudorapidity of $2 < \eta < 5$ [12], where $\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]$, since at high energies the hadrons are mainly produced in the forward direction. It is composed of many detectors as shown in Figure 3.1, such as the Ring Imaging Cherenkov counters (RICH), Electromagnetic Calorimeter (ECAL), Hadronic Calorimeter (HCAL), tracking systems and muon detectors.

3.1 Detector Components

The first detector is called the Vertex Locator (VELO), which takes measurements of the track coordinates near the proton-proton collisions. It was comprised of 42 modules of semicircular Silicon detectors placed at a distance of 8.2 mm from the beam, with a readout rate of 1 MHz. However, thanks to the new upgrade, the VELO will have 52 modules positioned in two L-shaped sensors, which are placed at only 5.1 mm away from the beam and have a readout rate of 40 MHz [11]. Thanks to this shorter distance, the VELO will be able to detect B mesons.

After the VELO, there is the first Ring Imaging Cherenkov counter (RICH1), which detects charged particles with momentum in the range 1-60 GeV/c, while the second one (RICH2) is placed behind the magnet and the three Scintillating Fibre (SciFi) Trackers and it is able to identify charged particles with momenta up to and past



Figure 3.1: Photo of the VELO detector at LHCb (courtesy of the LHCb collaboration).

100 GeV/c. Right next to the magnet there is the Upstream Tracker and three SciFi Trackers. The former is needed to get a first measurement of the momentum of the particles, while the latter, which is after the magnet, provide measurements of the curvature of the trajectories due to the magnetic field, and together they allow to take more accurate measurements.

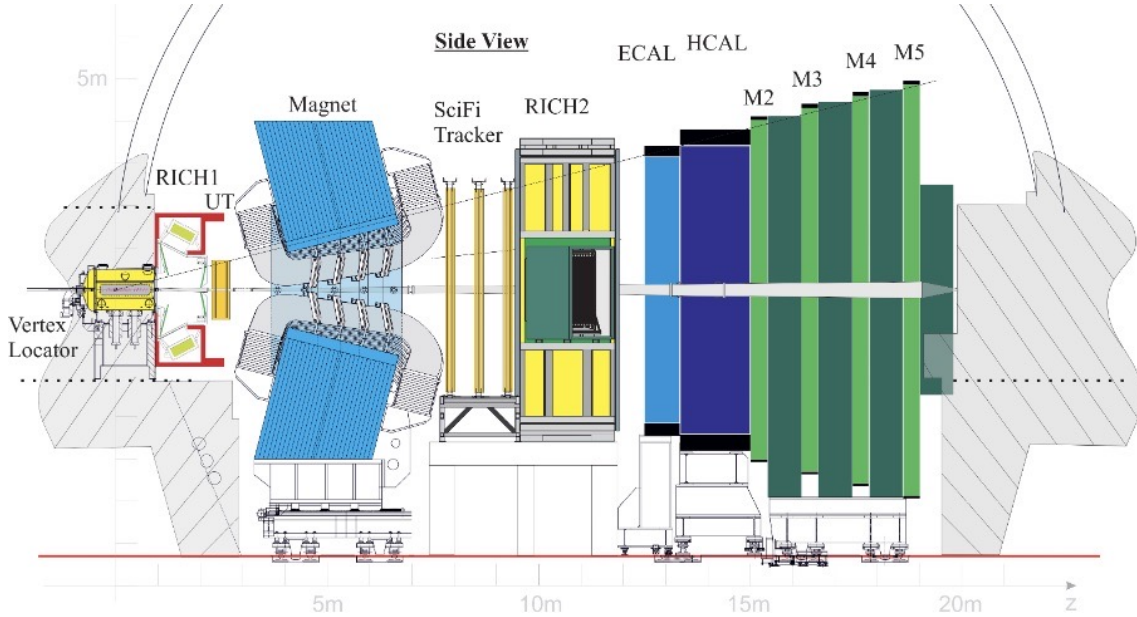


Figure 3.2: Structure of the LHCb detector [12].

Behind RICH2, there is an electromagnetic calorimeter and a hadronic calorimeter. Their purpose is to stop almost all the particles reaching these two detectors and provide measurements of their energy and position. The former is used for electrons and photons, whereas the latter for hadrons. However, the only particles that are able to go past them are muons, which are identified using muon detectors, and neutrinos, which are not tracked due to the fact that they rarely interact with other particles. Muon detectors are placed behind the calorimeters and are composed of

multiple tracking stations. The M2 and M3 stations have a high resolution and provide measurements of the transverse momentum of the particles, while stations M4 and M5 have a lower resolution and they mainly identify more penetrating muons [13].

3.2 B-Tracking Tool

There is a recent addition to the software which executes the reconstruction of the particles and their trajectories. When considering decays without intermediate particles, the meson trajectory can be deduced from the line connecting the Primary Vertex (PV), where the proton-proton collision occurred, and the Secondary Vertex (SV), where the meson decayed. However, in a decay with an intermediate particle, such as $B_c^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$, the direction of the B-meson is not exactly along the line connecting PV to SV. Therefore, the B-Tracking Tool searches for hits in the VELO near the line PV-SV, to measure the direction of the meson². The VELO has been placed closer to the beam to improve the general reconstruction of the PV and SV and also to achieve a better B-tracking [14].

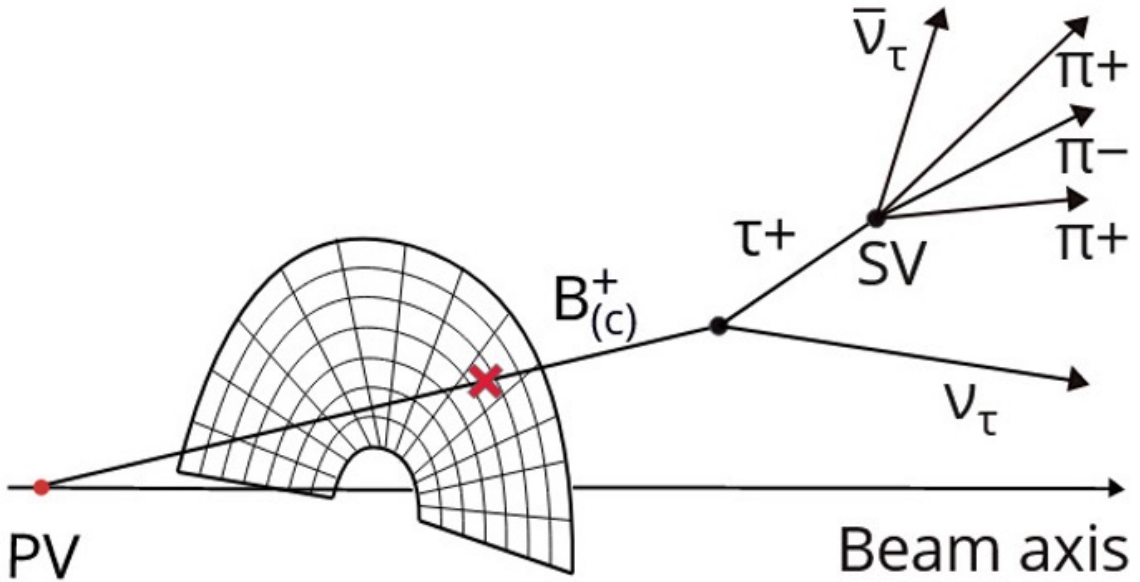


Figure 3.3: Diagram of the B-Tracking tool [14].

²It is only an approximation, since it is assumed that the hit on the VELO is near the line connecting PV to SV. In RapidSim there is no explicit hit; the simulator places the hit on the PV-SV line when this intersects the VELO detector.

Chapter 4

Analysis Strategy

4.1 RapidSim

For the analysis conducted in this thesis the data were generated using RapidSim, which is a fast simulation of decays of hadrons with beauty or charm quarks. This simulation allows to generate, in just a few seconds, millions of events which are an approximation of what a full simulation might show. For this reason, RapidSim allows researchers to conduct fast analysis on a set of data, so to understand what should be further investigated in a more advanced simulation [15].

In addition, RapidSim allows the user to specify the momentum resolution of the detectors, in order to reproduce the effect of errors during the reconstruction of the particles trajectories. In order to do so, it smears the momentum resolution of each particle based on a specific resolution function. Moreover, since every decay could happen along any direction, it is also possible to define a solid angle in which the decays can be detected to simulate the geometrical acceptance of the LHCb detector [15].

4.2 Signal and Background

In this analysis also the B_u^+ (or B^+) meson, which has an anti-bottom quark and an up quark, will be taken into account³. The meson B^+ decays similarly as B_c^+ , thus for this analysis the decays $B_c^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ and $B^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ are considered as signal modes. The major difference between these two decays is that the second one is suppressed by a factor of $\frac{|V_{cb}|}{|V_{ub}|}$, but enhanced by its production

³In theory, the bottom quark could also couple to the top quark to form B_t^+ , however the top quark decays so quickly that it doesn't form a bond with other particles.

fraction.

As mentioned in the previous section, only the three pions can be detected and are considered as the signature of the two decays explained above. However, there are other decay modes which produce three pions and are not considered as signal modes. These are considered as background modes and could affect the measurements, thus they have to be separated from the signal as best as possible. Since there is a large number of background modes, only the three most dangerous decays (according to Ref. [16]) will be considered as background in this thesis: $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$, $B^+ \rightarrow \pi^+\pi^-\pi^+D^{0*}$ and $B^+ \rightarrow D^-(\rightarrow \pi^-\pi^+\pi^-)\pi^+\pi^+$ (called normalization mode). In order to reject as many background events as possible, a cut on the invariant mass of the three pions produced in each decay is found by plotting the invariant mass on a histogram. Then a MultiVariate Analysis (MVA) is performed to further separate signal from background by using multiple observables together. Lastly, a Likelihood fit is simulated to study the performance of the cut.

Chapter 5

Invariant mass

In order to know whether or not the signal modes could be separated from the three background modes, the invariant mass of the pions is analysed. Then a cut on the invariant mass will be found.

5.1 Invariant mass

The main goal of this thesis is to understand how the three background modes can be distinguished from the signal by looking at the invariant mass of the three pions. In order to calculate it, the following formula is used:

$$q = (E_1 + E_2 + E_3, \vec{p}_1 + \vec{p}_2 + \vec{p}_3) \quad (5.1)$$

$$m^2 = q^2 = (E_1 + E_2 + E_3)^2 - (\vec{p}_1 + \vec{p}_2 + \vec{p}_3)^2 \quad (5.2)$$

where q is the four vector of the system comprised of the three pions, m is the invariant mass, E is the energy and \vec{p} is the momentum of the pions.

5.2 Cut on invariant mass of pions

First of all, the invariant mass is investigated. Figure 5.1 and 5.2 show the 1-dimensional histograms of the invariant mass of the three pions produced in each decay mode.

In figure 5.1 and 5.2, the histograms of the two signal modes are the same, because in both cases the three pions are produced by the tauon decay. As for the background, it can be seen that the two modes with the D^0 meson have a similar shape, however $B^+ \rightarrow \pi^+ \pi^- \pi^+ D^{0*}$ is shifted towards lower energies since the excited D^{0*} meson takes more energy away from the pions compared to D^0 . Lastly, the normalization

mode has a different shape compared to the other decays, because the three pions taken into account when calculating the invariant mass are produced from the decay of D^- . For this reason, there is no neutrino or D^0 meson which takes energy away from the pions, so the resulting invariant mass graph is a spike at the rest energy of the D^- meson.

Figure 5.2 is showing the same invariant mass graph, but with a logarithmic scale on the y-axis. Based on this plot, bins with a low number of events can be observed more clearly and a cut on the invariant mass is found. After analysing this graph, the cut at 1.8 GeV was chosen. This means that the events above this value are rejected since they are mostly background, while the ones below 1.8 GeV are kept because they are mostly signal events.

As a consequence of this cut, 99.999% of both signal modes is kept, while only 0.018% of the normalization mode remains, 11.18% of $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$ and 13.13% of $B^+ \rightarrow \pi^+\pi^-\pi^+D^{0*}$ remain.

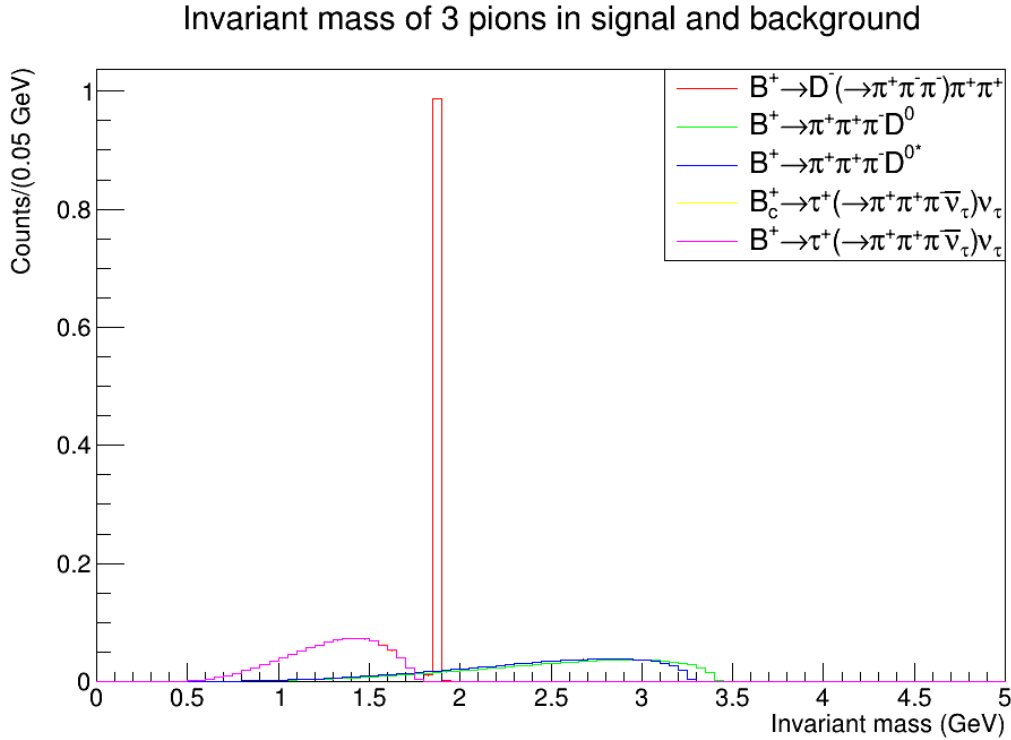


Figure 5.1: Invariant mass of three pions with a linear scale on the x-axis.

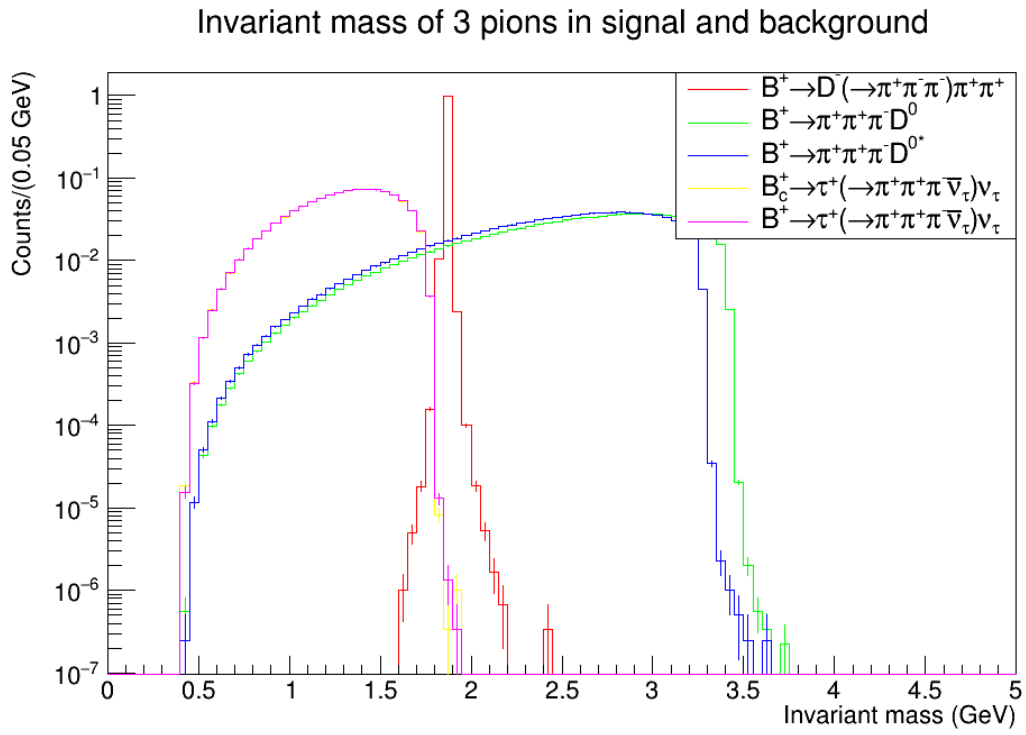


Figure 5.2: Invariant mass of three pions with a logarithmic scale on the x-axis.

Chapter 6

MVA

6.1 MultiVariate Anlaysia

After finding the cut on the invariant mass, its effect on the MultiVariate Analysis (MVA) is studied. The MVA is a statistical method which uses multiple variables to analyse a set of data. In this thesis it serves to separate signal from background modes using several observables that are simulated for each event with RapidSim, however the invariant mass is not included among these observables. For instance, the variables used in the analysis are the total momentum of the pions, transverse momentum of the pions, the opening angle, the full flight distance (distance of SV from PV), number of hits in the VELO and the impact parameter of the pions (closest distance of the trajectory of the pions from the PV).

This MVA is run using the python code written by Maarten van Veghel (and modified by Jelte Rinus de Jong and Maria Domenica Galati) and uses the Gradient Boosting Classifier (GBC) from the sklearn library. This is a machine learning algorithm which gives a prediction on whether an event is signal or background based on decision trees, which are tree-like models used to make classifications or decisions based on the features of the available data set.

Since the data were simulated with RapidSim and it was already known whether an event was signal or background, the method of supervised learning was used. This method consists of marking each event as either signal or background and then dividing the dataset in a training and testing sample. The algorithm is trained on the former group and its performance is then tested on the latter one, by checking how the algorithm recognized each event compared to how they were flagged at the beginning. In addition, each event of the testing sample receive an MVA score between

0 and 1, which tells how likely that event was recognised as signal or background (0 means it is certainly background, whereas 1 means it is certainly signal).

For the MVA, 400K events were used for both signal and background with a 50-50% division between the training and the testing sample. Only the background mode $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$ is used for the MVA, since, as Ref. [16] found, this decay is the most dangerous one and it is similar to $B^+ \rightarrow \pi^+\pi^-\pi^+D^{0*}$. Also the normalization mode was not studied with MVA, because the cut at 1.8 GeV on the invariant mass of the three pions removes 99.98% of this mode, and therefore it is considered sufficiently repeated.

6.2 Results of MVA training

After the training and testing processes, an histogram of the MVA score is created, as shown in Figure 6.1. Then, multiple cuts are made between the score 0 and 1, in order to calculate the amount of signal events on the right side of this cut, which is called True Positive Rate (TPR), and the amount of background events on the right side of the cut, which is called False Positive Rate (FPR). At the end, the FPR are plotted against the TPR in a graph called Receiver Operating Characteristic (ROC) curve. The Area Under the Curve (AUC) gives a value between 0 and 1 (0 being the worst and 1 the best score), which tells how successful the MVA was.

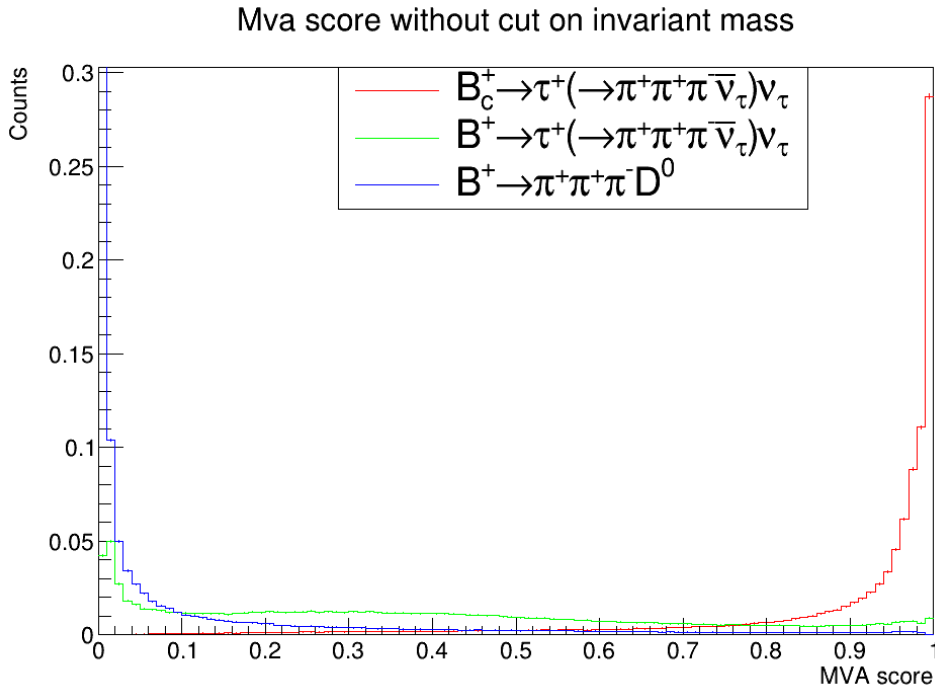


Figure 6.1: Histogram of the MVA score without the invariant mass cut.

Figure 6.1 and 6.2 are the graphs obtained from the MVA and show the FPR against the TPR with a linear and a logarithmic scale on the x-axis. It can be noticed that the AUC score decreases by 0.014 when the cut on the invariant mass is included, nevertheless it must be noted that, due to the cut, in figure 6.2 only the background events with invariant mass of pions below 1.8GeV were considered, after rejecting 88.82% of this background. Concerning the plots with the logarithmic x-axis, another noticeable effect of the cut is that the overall shape of the graphs look similar, but shifted towards lower values of true positive rate for the same false positive rate. For instance, the point with a value of $2 \cdot 10^{-2}$ as FPR corresponds to 0.8 as TPR in the MVA without the cut and 0.7 as TPR in the MVA with the cut. However, it must be remembered that the number of background events decreased to 11.18% by applying the filter on the invariant mass. Therefore, even though the AUC score is lower with the filter, the combined effect of the cut and MVA is better than only training an MVA.

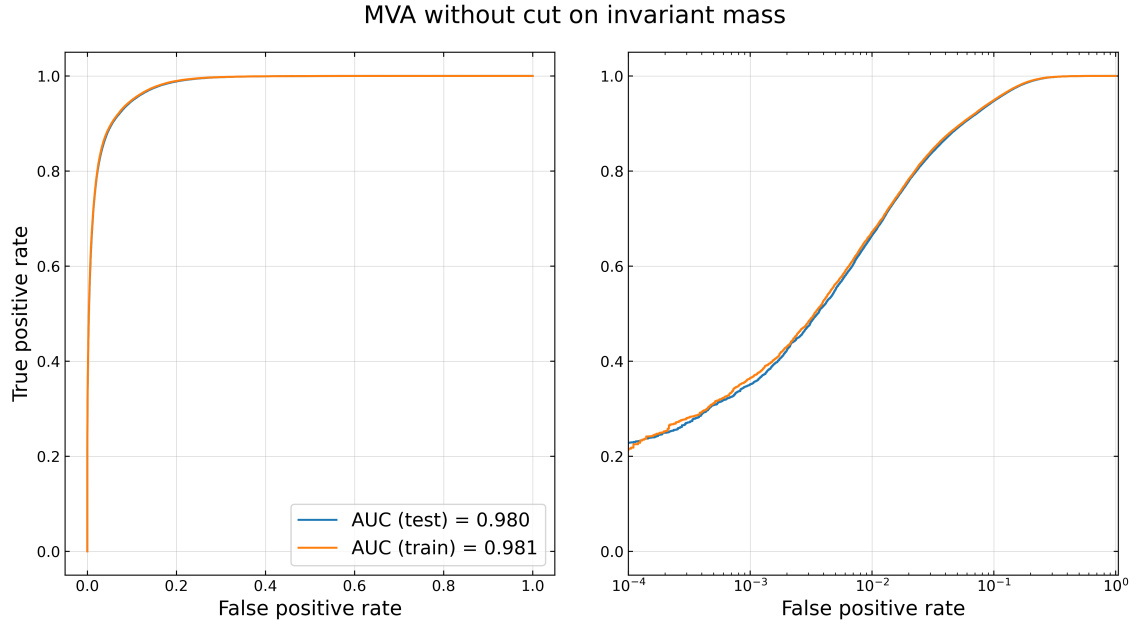


Figure 6.2: ROC curve obtained from the MVA without the cut on the invariant mass of the pions and with $B_c^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ as signal and $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$ as background.

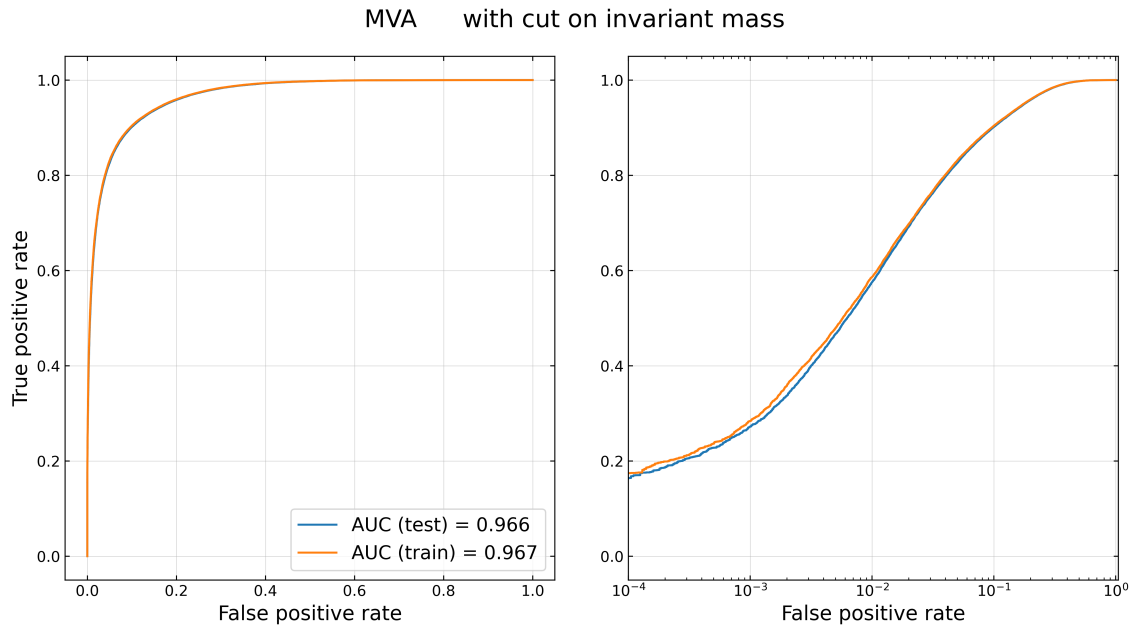


Figure 6.3: ROC curve obtained from the MVA with the cut on the invariant mass of the pions and with $B_c^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ as signal and $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$ as background.

Chapter 7

Maximum Likelihood Fit

7.1 Likelihood Fit

After the MultiVariate Analysis, the events simulated by RapidSim and the MVA score of each event were used to simulate a fit. First, the theoretical number of yields for signal and background was calculated using the equation:

$$n = \mathcal{L} \cdot \sigma_{pp} \cdot f_{c/u} \cdot BR \cdot \epsilon_{VELO} \cdot (1 - \epsilon_{isol}) \quad (7.1)$$

where $L = 10 \text{ fb}^{-1}$ is the luminosity, σ_{pp} is the cross section of a proton-proton collision, $f_{c/u}$ is the production fraction of B_c^+ or B^+ , BR is the branching fraction of the decay for which the number of yields is being calculated, ϵ_{VELO} is the efficiency of the B-Tracking, and ϵ_{isol} is the reconstruction efficiency for isolation⁴. Then, by creating three 2D histograms (one for B_c^+ , one for B^+ and one for background) with corrected mass (m_{corr}) on the x-axis and MVA score on the y-axis, three 2D Probability Density Functions (PDF) were determined. Afterwards, the Maximum Likelihood method was used which allows to estimate the parameters of a probability distribution by maximizing a likelihood function, as shown in Ref. [16]. A likelihood fit was found with this method for the PDF of the total number of yields using the three PDFs. The formula of the total PDF is given by equation 7.1 of Ref. [16]:

$$\mathcal{P} = f_{B_c^+} \mathcal{P}_{B_c^+} + f_{B^+} \mathcal{P}_{B^+} + (1 - f_{B_c^+} - f_{B^+}) \mathcal{P}_{BG} \quad (7.2)$$

where $\mathcal{P}_{B_c^+}$, \mathcal{P}_{B^+} and \mathcal{P}_{BG} are the three 2D PDFs, $f_{B_c^+}$ is the fraction of B_c^+ and f_{B^+} is the fraction of B^+ compared to the total number of yields. Since the total number of events is given by $N_{tot} = n_{B_c^+} + n_{B^+} + n_{BG}$, the fractions are: $f_{B_c^+} = \frac{n_{B_c^+}}{N_{tot}}$ and $f_{B^+} = \frac{n_{B^+}}{N_{tot}}$. In order to perform the fit, $f_{B_c^+}$ and f_{B^+} were considered as free

⁴ ϵ_{VELO} and ϵ_{isol} were studied and determined by Jelte Rinus de Jong in another sensitivity study (see [16]).

parameters, while the fraction of background events is not a free parameter since it can be written as $f_{BG} = 1 - f_{B_c^+} + f_{B^+}$ [16].

7.1.1 Corrected mass

Before showing the results of the likelihood fit, the corrected mass should be explained, since it is used on the x-axis of the fit. As mentioned before, in $B_c^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ only the three pions can be detected, thus it is only possible to calculate an approximation of the mass of the B_c^+ meson, which is called the corrected mass (m_{corr}). It is given by the following equation [14]:

$$m_{corr} = \sqrt{m_{\pi\pi\pi}^2 + |p_\perp|^2} + |p_\perp| \quad (7.3)$$

7.2 Results of the Likelihood Fit

Also for the likelihood fit, only the background $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$ was considered, as it is the most worrisome decay. In Figure 7.1 it can be observed that at $m_{corr} \approx 4 \text{ GeV}/c^2$ the background reaches 10^6 events which is approximately two and three orders of magnitude greater than $B^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ and $B_c^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ respectively. However, figure 7.2 shows that the amount of background decreases by almost a factor of 10, by adding the invariant mass cut. When looking at the background mode, it can be noticed that both in Figure 7.1 and 7.2 the histogram peaks at around 4.2-4.4 GeV, nevertheless there is a difference on both sides of the peak. On the left hand side, it can be observed that there is a steeper rise in number of events in Figure 7.1 than in 7.2, while on the right hand side there is a sharper decreases in number of events in Figure 7.2 than 7.1.

The Maximum Likelihood method allows also to calculate the error on the best values of the free parameters. For the fit which does not include the cut on the invariant mass, the error on $f_{B_c^+}$ is $\frac{\sigma_{f_{B_c^+}}}{f_{B_c^+}} = 3.95\%$ and the error on f_{B^+} is $\frac{\sigma_{f_{B^+}}}{f_{B^+}} = 1.36\%$, whereas, when the cut is included, the error on $f_{B_c^+}$ is $\frac{\sigma_{f_{B_c^+}}}{f_{B_c^+}} = 3.25\%$ and the error on f_{B^+} is $\frac{\sigma_{f_{B^+}}}{f_{B^+}} = 1.30\%$. Thus, the systematic uncertainty for the fraction of B_c^+ decreased by 17.7%, and for the fraction of B^+ by 4.4%. Other than the error on the fractions, the code also returns the number of yields for B_c^+ and B^+ , which are equal to $N_{B_c^+} = 3358 \pm 133$ and $N_{B^+} = 38815 \pm 526$ without the cut on the invariant mass and $N_{B_c^+} = 3359 \pm 110$ and $N_{B^+} = 39119 \pm 509$ with the cut.

Another feature to consider is the empty bins present at higher values of m_{corr} . This issue is due to the fact that the number of events, that were used to generate the data of the background, were not enough to have an extended tail in the distribution at higher values of m_{corr} , but only a few bins contained some events while others were empty. Therefore, when the data were generated, the empty bins remained empty, while a significant number of events were created for the others.

Fit without cut on invariant mass

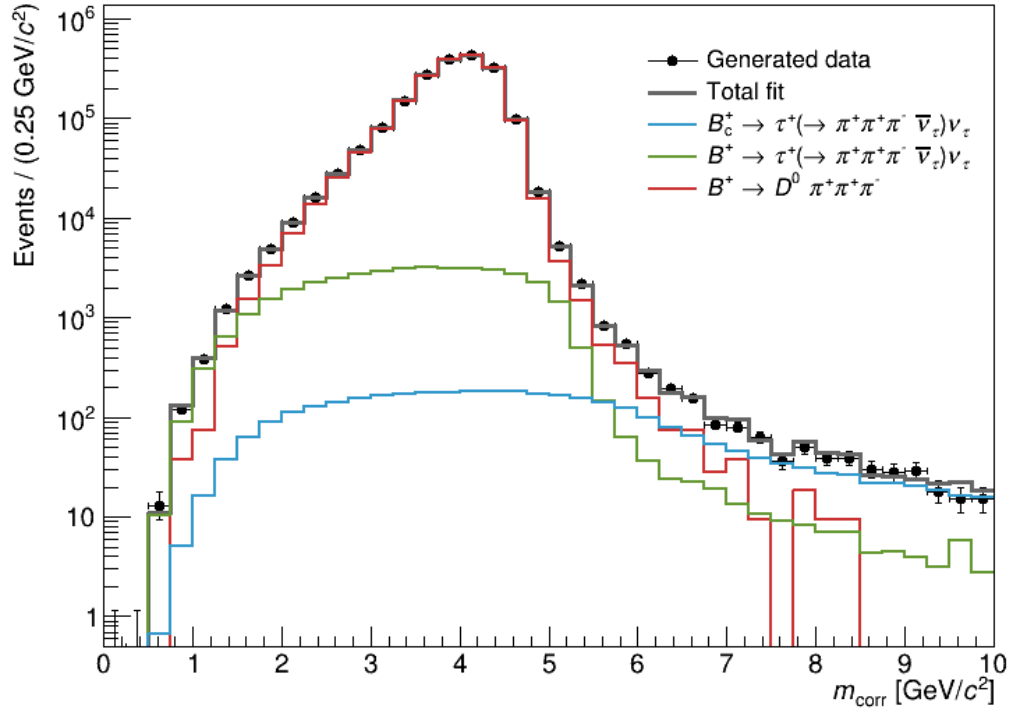


Figure 7.1: PDF projection and likelihood fit without the cut on the invariant mass.

Fit with cut on invariant mass

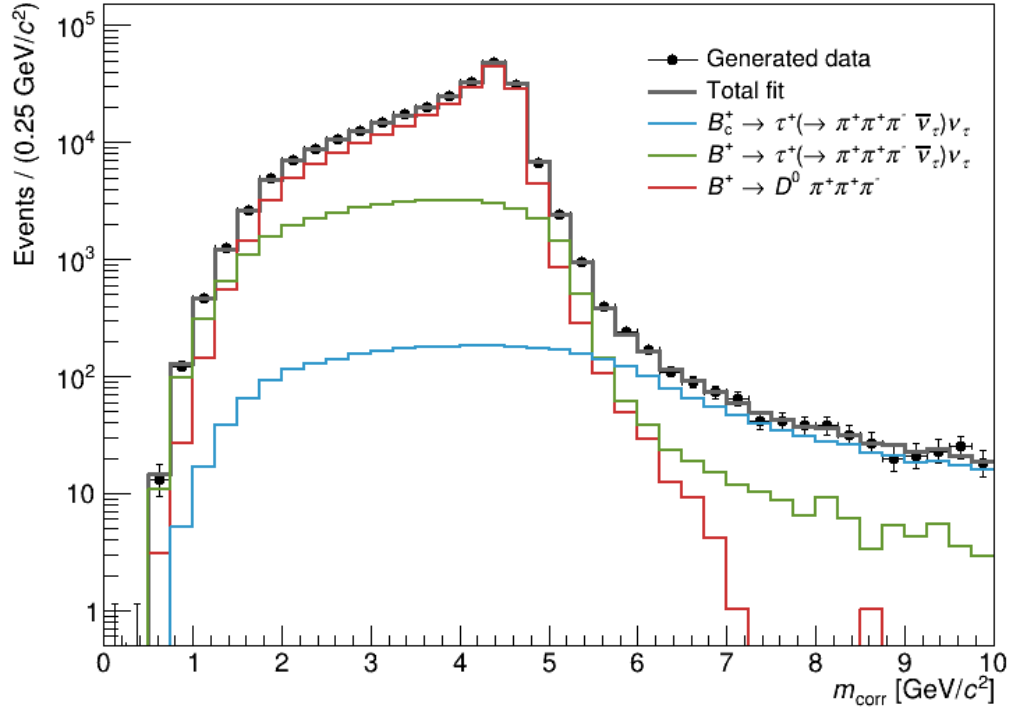


Figure 7.2: PDF projection and likelihood fit with the cut on the invariant mass.

Chapter 8

Discussion

First of all, it must be said that the cut was found only by looking at the Figure 5.1 and 5.2, which might not be the optimal method. Nevertheless, even if another technique was applied, which optimizes the amount of background removed and signal kept by the cut, it might not affect the results obtained in this thesis significantly, due to the fact that between three million and nine million events were used for both plots and also because the cut keeps all the signal events.

The step of this sensitivity study which could have been improved is the MVA. At first, an insufficient number of events was used, thus a considerable overtraining issue was visible by observing that the AUC score for the training sample was higher than that for the testing sample. This problem was easily solved by simulating substantially more data. Nevertheless, even though approximately 3.78M events were available for the background mode, these would decrease to 400K after applying the filter on the invariant mass. For this reason, the maximum number of events for the training and testing sample was 200K. This amount of data was enough to reduce the overtraining issue, as it can be seen in Chapter 6 where the AUC scores differ only by 0.001.

The number of events used for the testing sample of the MVA was important, because the distribution of these events was then used to generate more data for the likelihood fit. It is clear that in Figure 7.1 and 7.2 there are empty bins above $m_{corr} = 7\text{GeV}/c^2$, due to the absence of events in the testing sample at higher values of corrected mass. This problem could have been avoided by simulating more data, so that more than 200K events for both the training and testing sample could have been used for the MVA. However, the main issue that arises from handling huge amounts of data is the time required to run the simulation, which generates

them, and the time required for the MVA, which has to analyze all the data and give a score to each event.

Moreover, even though the decay $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$ is the most dangerous background as shown in [16], the same process could have been performed for the other two background modes. Nonetheless, the study on $B^+ \rightarrow \pi^+\pi^-\pi^+D^{0*}$ would have been similar, since the histogram of the invariant mass of the pions is almost equivalent to the one for $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$. Whereas for the normalization mode the MVA and likelihood fit have not been performed since, as it was shown in Chapter 5, the cut on the invariant mass removes $\sim 99.98\%$ of events. This means that for every 10^6 events, only 200 remain, hence one billion events should have been simulated with RapidSim in order to have a comparable analysis with the MVA and the Likelihood fit as the one performed with $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$.

Chapter 9

Conclusion

In this sensitivity study, the two decays $B_c^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ and $B^+ \rightarrow \tau^+(\rightarrow \pi^+\pi^-\pi^+\bar{\nu}_\tau)\nu_\tau$ were considered as signal, while $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$, $B^+ \rightarrow \pi^+\pi^-\pi^+D^{0*}$ and $B^+ \rightarrow D^-(\rightarrow \pi^-\pi^+\pi^-)\pi^+\pi^+$ (called normalization mode) as background, since these decays are the most dangerous ones.

The signal and background modes were simulated with RapidSim. Then, a plot of the invariant mass of the three pions produced in these decays was analyzed and a cut at 1.8 GeV was chosen in order to reject all the events above this value, since they were considered to be mostly background. This cut allowed to remove 88.82% of $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$, 86.87% of $B^+ \rightarrow \pi^+\pi^-\pi^+D^{0*}$, and 99.98% of the normalization mode, while keeping 99.999% of both signal modes.

Afterwards, an MVA was performed using the machine learning method called Gradient Boosting Classifier for both the situation with and without the cut on the invariant mass. During this analysis, an MVA score was given to each event of the signal and background mode (only $B^+ \rightarrow \pi^+\pi^-\pi^+D^0$ was studied with the MVA). Lastly, a likelihood fit was performed by including and excluding the cut in order to analyse the performance of the cut and the MVA. As it can be observed from the plot shown in Chapter 7, the cut manages to reduce the amount of background by approximately a factor of 10 over the entire range of m_{corr} . Moreover, the number of yields of B_c^+ and B^+ were estimated and their systematic uncertainty decreased by 17.7% and 4.4% respectively by introducing the invariant mass cut.

Therefore the cut on the invariant mass could be useful to reject all the events of the normalization mode and most of the events of the other two background modes, and also to reduce the uncertainty on the estimated number of yields of B_c^+ and B^+ .

Bibliography

- [1] A. Akeroyd and C.-H. Chen, “Constraint on the branching ratio of $B_c^- \rightarrow \tau \bar{\nu}$ from LEP1 data and consequences for $R(D^*)$ anomaly,” *Physical Review D*, vol. 96, no. 7, oct 2017. [Online]. Available: <https://doi.org/10.1103/PhysRevD.96.075011>
- [2] S. Bifani, S. Descotes-Genon, A. R. Vidal, and M.-H. Schune, “Review of lepton universality tests in B decays,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 46, no. 2, p. 023001, dec 2018. [Online]. Available: <https://doi.org/10.1088/1361-6471/aaf5de>
- [3] D. Gerstel, “Test of Lepton Flavour Universality using the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ decays at LHCb,” Ph.D. dissertation, Marseille, CPT, Aix Marseille Université, 2020.
- [4] C. MissMJ, “Standard model of particle physics,” sep 2019, [Accessed 2-July-2022]. [Online]. Available: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg
- [5] The ALEPH Collaboration *et al.*, “Precision electroweak measurements on the Z resonance,” *Physics Reports*, vol. 427, no. 5-6, pp. 257–454, may 2006. [Online]. Available: <https://doi.org/10.1016/j.physrep.2005.12.006>
- [6] Y. S. Amhis *et al.*, “Averages of b -hadron, c -hadron, and τ -lepton properties as of 2018,” *Eur. Phys. J.*, vol. C81, p. 226, 2021, updated results and plots available at <https://hflav.web.cern.ch/>.
- [7] T. Zheng, J. Xu, L. Cao, D. Yu, W. Wang, S. Prell, Y.-K. E. Cheung, and M. Ruan, “Analysis of $B_c \rightarrow \tau \nu_\tau$ at CEPC,” *Chinese Physics C*, vol. 45, no. 2, p. 023001, feb 2021. [Online]. Available: <https://doi.org/10.1088/1674-1137/202102023001>

- [8] T. Gershon, “Overview of the Cabibbo–Kobayashi–Maskawa matrix,” *Pramana*, vol. 79, no. 5, pp. 1091–1108, nov 2012. [Online]. Available: <https://doi.org/10.1007%2Fs12043-012-0418-y>
- [9] R. Workman *et al.*, “Review of Particle Physics,” to be published (2022).
- [10] B. Colquhoun, C. Davies, J. Kettle, J. Koponen, A. Lytle, R. Dowdall, and G. Lepage, “B-meson decay constants: A more complete picture from full lattice QCD,” *Physical Review D*, vol. 91, no. 11, jun 2015. [Online]. Available: <https://doi.org/10.1103%2Fphysrevd.91.114509>
- [11] D. Dutta, “The LHCb VELO upgrade,” sep 2019, p. 022.
- [12] “LHCb Tracker Upgrade Technical Design Report,” feb 2014.
- [13] The LHCb Collaboration *et al.*, “The LHCb Detector at the LHC,” *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 005–S08 005, aug 2008. [Online]. Available: <https://doi.org/10.1088/1748-0221/3/08/s08005>
- [14] J. Rol, “b-Meson Tracking at LHCb: a Feasibility Study,” Master’s thesis, mar 2022.
- [15] G. Cowan, D. Craik, and M. Needham, “Rapidsim: An application for the fast simulation of heavy-quark hadron decays,” *Computer Physics Communications*, vol. 214, pp. 239–246, may 2017. [Online]. Available: <https://doi.org/10.1016%2Fj.cpc.2017.01.029>
- [16] J. R. de Jong, “Feasibility study of the branching fraction measurements of $B_c^+ \rightarrow \tau^+ \nu_\tau$ and $B^+ \rightarrow \tau^+ \nu_\tau$ at LHCb,” Master’s thesis, jun 2022.