A search for the $B^0_{(s)} \to \mu^+\mu^-(\gamma)$ decays at LHCb

Thesis submitted in accordance with the requirements of the
University of Liverpool for the degree of
Doctor in Philosophy
by

Lauren Emma Yeomans

Sept 2021 (Compiled January 9, 2022)

Abstract

Feasibility studies are conducted to test if the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction can be measured using a partial reconstruction technique, using the $B_{(s)}^0 \to \mu^+\mu^-$ analysis with no optimisation for the radiative counterpart. An improved measurement of the decay $B_s^0 \to \mu^+\mu^-$ and searches for the decays $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ are performed at the LHCb experiment using data collected in proton-proton collisions at $\sqrt{s} = 7,\ 8$ and $13\,\text{TeV}$, corresponding to integrated luminosities of 1, 2 and $6\,\text{fb}^{-1}$, respectively. The $B_s^0 \to \mu^+\mu^-$ branching fraction is measured to be $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \left(3.09\,^{+\,0.46}_{-\,0.43}\,^{+\,0.15}_{-\,0.11}\right) \times 10^{-9}$, where the uncertainty includes both statistical and systematic contributions. No significant signal is found for $B^0 \to \mu^+\mu^-$ and the upper limit of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.6 \times 10^{-10}$ at 95% confidence level is determined. Positive results from the feasibility studies led to the first search for $B_s^0 \to \mu^+\mu^-\gamma$ decays using partial reconstruction, and an upper limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) < 2.0 \times 10^{-9}$ is set at 95% confidence level, limited to the range $m_{\mu^+\mu^-} > 4.9\,\text{GeV}/c^2$.

Additionally, the ratio between the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions is measured to be $\mathcal{R}_{\mu^+\mu^-} < 0.095$ at 95% confidence level. The results are all in agreement with the Standard Model predictions.

For you, Dad, thank you for showing me the stars.
And for Nanny, who shines the brightest of them all.

Declaration

I hereby confirm this work is my own, except where other works are referenced. This work has not previously been submitted to any institute, including this one. This thesis does not exceed the relevant word count.

Lauren Emma Yeomans

Acknowledgements

The work presented in this thesis would not have been possible without the guidance and advice of many of my LHCb colleagues. Firstly I would like to thank Francesco Dettori, who took me under his wing in the early stages of my PhD and offered invaluable support. He always went above and beyond, so thank you, Francesco. My supervisors, David Hutchcroft and Tara Shears, have always been by my side through the ups and downs of completing a PhD, and many thanks especially to David for all of your help with the writing of this thesis itself; your hours of proof-reading and note-making have been hugely appreciated!

Thanks also to my peers at Liverpool University; Matt, James, Phil, Tom, Abbie, and many more, who have kept me going, even in the most difficult times. You've kept me sane (ish) through this journey; I'll miss our AJ trips! So thanks guys, for the help and the memories. A special thank you to Matt for being so patient with my never ending coding questions. Thanks to Magda for prosecco nights and helping me with my work-life balance.

My analysis group have also been amazing, encouraging, and selfless with their time, so thank you to Flavio, Matteo, Marco, Harry, Ifan, Titus, Silvia, Maik, Mick and Jacco. Working in a team with you guys was an amazing experience.

Thanks to my Mum and Steve for your endless support, and for not getting mad at the endless amounts of printing and whining it took to get this thesis finished. Thank you for your encouragement and your belief in me. Thank you to Nanny for being my biggest cheerleader. I miss you every day. And finally, thank you Dad, for waking me up in the middle of the night when I was younger, to show me the stars with your telescope. Thank you for making me curious about the universe... You started this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Standard Model (SM) of particle physics describes the nature of matter. It covers the properties and interactions of the fundamental particles that make up our universe, explaining the observations made by scientists over several decades. There are, however, a few observations which cannot be explained by the SM as it currently stands, including:

1. The matter/antimatter asymmetry in the universe.
2. The nature of dark matter and dark energy, which combined are thought to make up over 95% of the universe.
3. The force of gravity; we do not know how to quantize it to add it in to the SM consistently.

In order to explain these shortcomings, several New Physics (NP) models have been suggested and are currently being searched for by physicists all over the world. This can be done in one of two ways; directly, via the production of potential new particles in high energy collisions, or indirectly, by discovering discrepancies between the way particles *should* behave according to the SM, and how they *actually* behave in experiment. Finding that a particle behaves differently to what is theorised may be an indication that there are new particles making a contribution to the process. Indirect searches have the benefit of not being limited to the collision energy, meaning that higher mass particles could potentially be discovered. For direct searches, the mass of any new particle is limited by the input energy.

According to the SM, the $B^0_{(s)} \to \mu^+\mu^-$ decays are extremely rare, occurring only four times in a billion for the B^0_s meson and one in ten billion for the B^0 meson. This makes them sensitive probes of NP, as the decay rate is so small and precisely predicted by theory. For this reason, these decay channels have been studied by several collaborations, with the search beginning more that 30 years ago. Before the LHC, no evidence had been found for either decay mode and the upper limits on the branching fractions were an order of magnitude above the SM predictions.

The search culminated in the recent observation of the $B_s^0 \to \mu^+\mu^-$ decay and evidence of the $B^0 \to \mu^+\mu^-$ decay reported by the LHCb and CMS collaborations [1] in 2015. Since then, more data collected by the LHC has resulted in more and more precise measurements and limits. The most recent results are discussed in Sect. 1.2.

A summary of the history of the limits and measurements of these decays up to 2015 is given in Fig. 1.1.

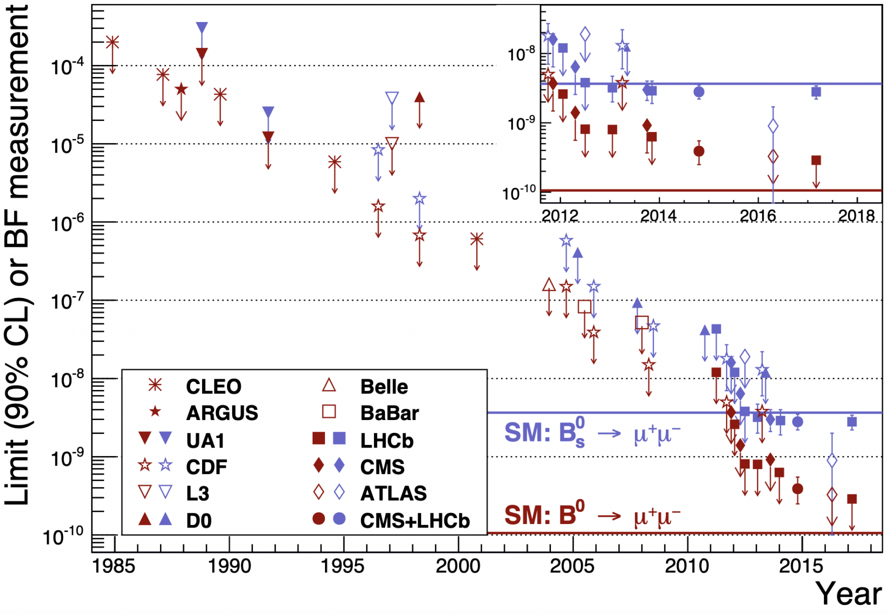


Figure 1.1: History of the limits and measurements of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays over the past 30 years, from Ref. [2]

The $B_s^0 \to \mu^+\mu^-\gamma$ decay, involving initial state radiation, has not yet been studied; there are no current limits on the branching fraction of this decay. This radiative decay offers sensitivity to a wider range of new physics due to the additional photon.

1.1 Outline of this Thesis

This thesis will present the most recent and improved measurement of the branching fraction of the $B_s^0 \to \mu^+\mu^-$ decay, as well as the search for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays performed at LHCb. The search for $B_s^0 \to \mu^+\mu^-\gamma$ is conducted only for a certain dimuon invariant mass range, being $m_{\mu^+\mu^-} > 4.9\ \text{GeV}/c^2$. The present analysis uses data collected in pp collisions at $\sqrt{s} = 7\,\text{TeV}$, $8\,\text{TeV}$ and $13\,\text{TeV}$, corresponding to integrated luminosities of $1.0\,\text{fb}^{-1}$, $2.0\,\text{fb}^{-1}$ and $5.9\,\text{fb}^{-1}$, respectively. Tools in place from the previous LHCb analysis [3] are used with several improvements. $B_s^0 \to \mu^+\mu^-\gamma$ is included as an additional observable, following the success of feasibility studies conducted on Monte Carlo. The $B_s^0 \to \mu^+\mu^-\gamma$ decay is partially reconstructed, using only the two muons in the final state and not reconstructing the photon.

In this thesis, emphasis is placed on my personal contributions to the analysis. The $B_s^0 \to \mu^+\mu^-\gamma$ feasibility studies were conducted entirely by myself, with guidance from another group member. Following these studies, my role within the main analysis group was primarily focused on the particle identification of the signal and various background contributions, as well as the calculation of other efficiencies entering into the normalisation and computation of the background yields.

1.1.1 Overview of the theory and LHCb experiment

Chapters 2 and 3 cover an introduction to both the theory behind the analysis and the physics of the detector used to collect the data. The framework of the Standard Model is explained and theoretical branching fractions are defined in Chapter 2. The rarity of $B_{(s)}^0 \to \mu^+\mu^-$ decays is also discussed along with the possible New Physics contributions. Chapter 3 outlines the experimental environment for the analysis recorded in this thesis, being the LHCb detector at CERN's Large Hadron Collider. Chapter 4 then discusses how events are reconstructed within LHCb, and how certain particles are identified.

1.1.2 $B_s^0 \to \mu^-\mu^+\gamma$ Feasibility Studies

Chapter 5 presents the studies conducted using the tools and dataset from the previous LHCb analysis round [3], to see whether the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction could be extracted as a partially reconstructed decay, with no photon reconstruction. The signal efficiencies for the $B_s^0 \to \mu^+\mu^-\gamma$ decay are calculated using Monte Carlo and compared to those of the $B_s^0 \to \mu^+\mu^-$ decay. The variables used in the selection process are compared, and selection cuts are examined at each stage, to compare signal retention and plan possible optimisations.

1.1.3 The $B_{(s)}^0 \to \mu^-\mu^+(\gamma)$ Analysis

Chapter 6 presents the full branching fraction analysis of $B_{(s)}^0 \to \mu^+\mu^-$ with the inclusion of $B_s^0 \to \mu^+\mu^-\gamma$, using the full Run 1 + Run 2 datasets collected by LHCb. The particle identification procedure is explained, including the calculations for the efficiencies of misidentified backgrounds. Background yields are estimated, and signal efficiencies entering the normalisation are calculated. The final results are presented in this chapter.

1.1.4 Conclusions and Future Outlook

The final chapter present the implications of the results of the current analysis. The future prospects for the decay channels studied are also discussed.

1.2 Current Experimental Status

The last analysis conducted by LHCb in the search for $B^0_{(s)} \to \mu^+\mu^-$ decays, prior to the analysis presented here, is detailed in Ref. [3]. An excess of $B^0_s \to \mu^+\mu^-$ decays was observed with a significance of 7.8 standard deviations, representing the first observation of this decay in a single experiment. The branching fraction was measured to be:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.0 \pm 0.6^{+0.30}_{-0.20}) \times 10^{-9}, \tag{1.2.1}$$

where the first uncertainty is statistical and the second is systematic.

No significant excess of $B^0 \to \mu^+\mu^-$ events was found, so an upper limit was placed on the branching fraction of:

$$\mathcal{B}(B^0_d \to \mu^+\mu^-) < 3.4 \times 10^{-10}, \tag{1.2.2}$$

at 95% confidence level.

The most recent results for the branching fractions of $B^0_{(s)} \to \mu^+\mu^-$ are from the combination of ATLAS, LHCb and CMS Run 1 (2011 and 2012) and partial Run 2 (2015 and 2016) datasets, described in Ref. [4]. The results are shown in Figs. 1.2 and 1.3.

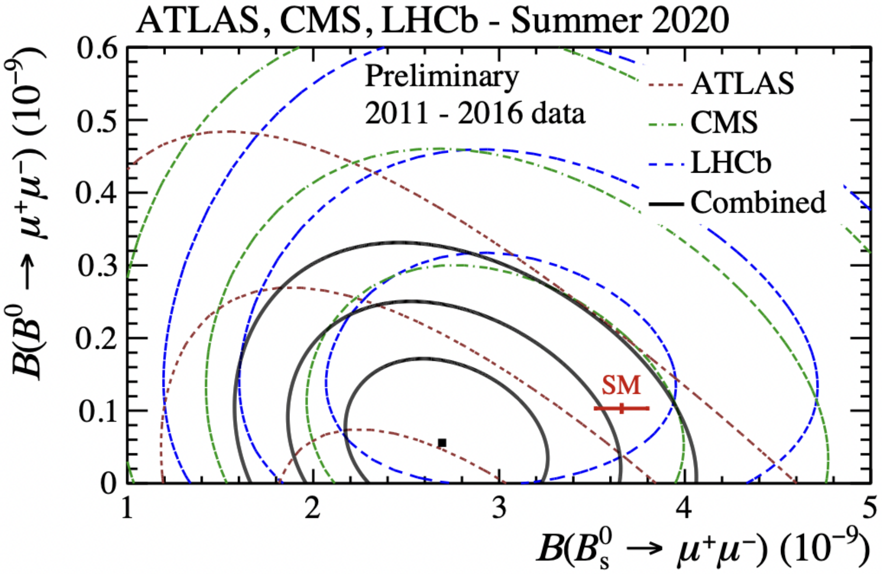


Figure 1.2: The 2D likelihood contours of the results for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays for the three experiments are shown, together with their combination. The red dashed line represents the ATLAS experiment, the green dot-dashed line the CMS experiment, the blue long-dashed line the LHCb experiment and the continuous line their combination. For each experiment and for the combination, likelihood contours correspond 1σ, 2σ and 3σ. The red point shows the SM prediction.

The branching fraction of the $B^0_s \to \mu^+\mu^-$ decay from the combined analysis is:

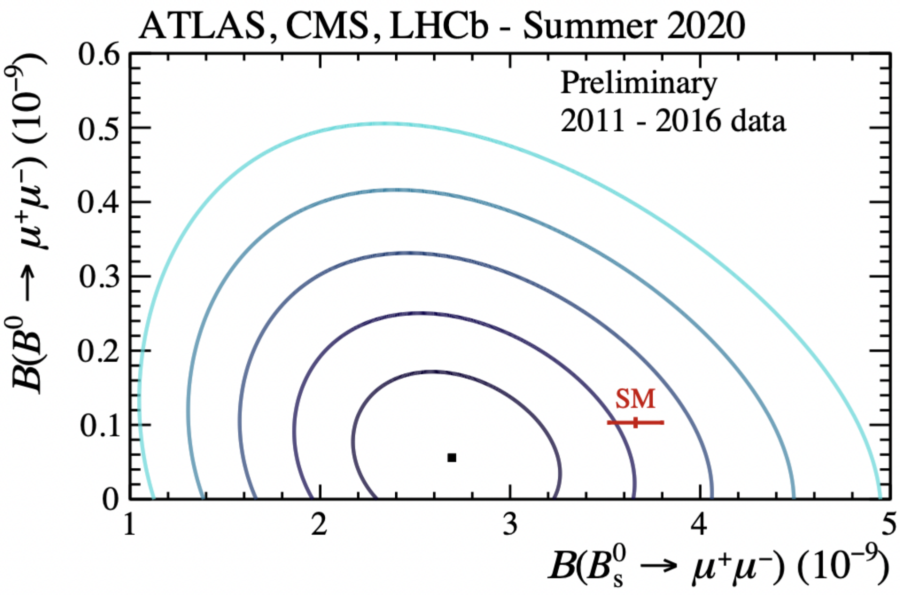


Figure 1.3: The combination of the three experiments is shown alone, with contours of different shades corresponding to 1σ, 2σ, 3σ, 4σ, and 5σ represented in order by darkest to less dark colour. The red point shows the SM prediction.

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (2.69^{+0.37}_{-0.35}) \times 10^{-9}, \tag{1.2.3}$$

and the upper limit on the $B^0 \to \mu^+\mu^-$ branching fraction is determined as

$$\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 1.6(1.9) \times 10^{-10} \tag{1.2.4}$$

at 90% (95%) confidence level. From these two results, an upper limit on the ratio of the $B_d^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ branching fractions is calculated to be 0.052 (0.060) at 90% (95%) confidence level. The measured $B_s^0 \to \mu^+\mu^-$ branching fraction is compatible with the SM within 1.1 standard deviations, the $B^0 \to \mu^+\mu^-$ branching fraction at 2.3 standard deviations, and the relative branching fraction at 2.4 standard deviations. Prior to the analysis presented in this thesis, has been no previous limit set on the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction.

The present analysis, discussed in this thesis, has recently been published by the LHCb collaboration and can be found in Refs. [5] and [6].

Chapter 2

Theory

This chapter will outline the Standard Model (SM), describing the theoretical background needed to understand the processes behind the $B^0_{(s)} \to \mu^-\mu^+$ and $B^0_s \to \mu^-\mu^+\gamma$ decays. The structure of the SM will be described, including an introduction to modern flavour physics, the CKM matrix, and Quantum Field Theory (QFT). Possible New Physics (NP) contributions are discussed at the end of the chapter, as well as the constraints on NP from current experimental results.

2.1 The Standard Model

The Standard Model is built within the framework of Quantum Field theory (QFT), and describes our observations of the fundamental particles of the universe. QFT states that each fundamental particle has a quantized field associated to it, and the particles themselves are seen as excitations of their corresponding fields.

2.1.1 The Fundamental Particles and Forces

Our universe is thought to be made up of three main components; dark matter (27%), dark energy (68%), and visible matter (5%). Dark matter is the name given to an unknown form of matter that does not interact with light, and it was theorised to explain the movement of galaxies, including their rotation curves, which do not behave as expected. It seems that there is more *stuff* distributed throughout the galaxies than can be seen, although we still don't know exactly what this matter is. Dark energy is more mysterious still, but is thought to be what causes the expansion of the universe to speed up.

The dynamics of universe can be described by four fundamental interactions: the gravitational force, the electromagnetic force, and the weak and strong nuclear forces [7]. A description of the gravitational force is not yet included within the SM framework, and it is the only one of the four forces which is thought to interact not only with visible matter, but also with dark matter.

Visible matter can be broken down into fermions and bosons. The fermions have half-integer spin, and obey Fermi-Dirac statistics, while bosons have integer spin, and obey Bose-Einstein statistics [8]. Fermions are grouped into quarks and leptons, with only the quarks interacting via the strong force. Each fermion has a corresponding ‘antifermion’, which has equal mass but opposite electric charge. The fermions and bosons that make up the SM, together with some of their properties, are shown in Fig. 2.2.

Particles with an electric charge all couple to the electromagnetic force, including the e^-, μ^- and τ^- leptons which each have a charge of -1e, and their corresponding antiparticles (each with a charge of +1e). Quarks also carry charge, with the ‘up-type’ quarks (up, charm, top) carrying a charge of $+\frac{2}{3}e$, and the ‘down-type’ quarks (down, strange, bottom) carrying a charge of $-\frac{1}{3}e$. The neutrinos (ν_e, ν_μ and ν_τ) are neutral, and hence do not interact electromagnetically.

Quarks and leptons are the building blocks of matter. Leptons can be found as free particles, whilst quarks are always observed as part of a larger composite particle. Combinations of quarks make up different hadrons, which can either be mesons (a quark antiquark pair), or baryons (particles containing three quarks). In recent years LHCb have also found evidence of more exotic combinations of quarks including the four quark state $Z(4430)^-$ [9] and five quark states $P_c^+(4380)$ and $P_c^+(4450)$ [10].

The weak force affects all twelve of the fermions (and all twelve corresponding antifermions). It can transform up type quarks into down type quarks, and can also transform charged leptons into neutrinos (and vice versa). Both of these processes occur in nuclear β decay: $n(udd) \rightarrow p(udu) + e^- + \nu_e$.

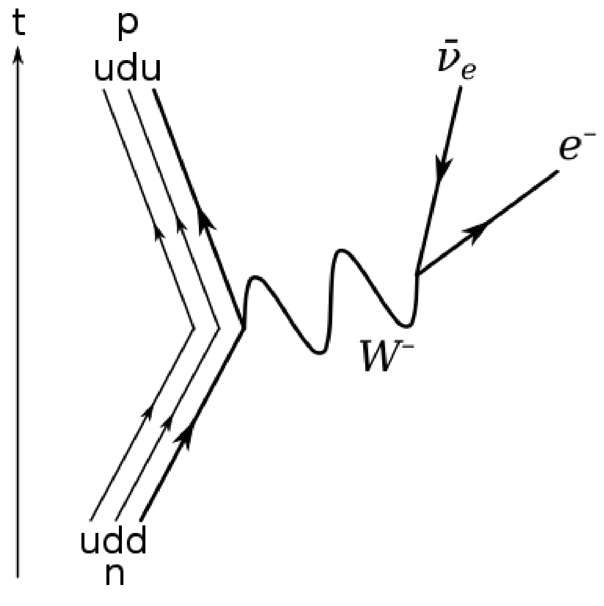
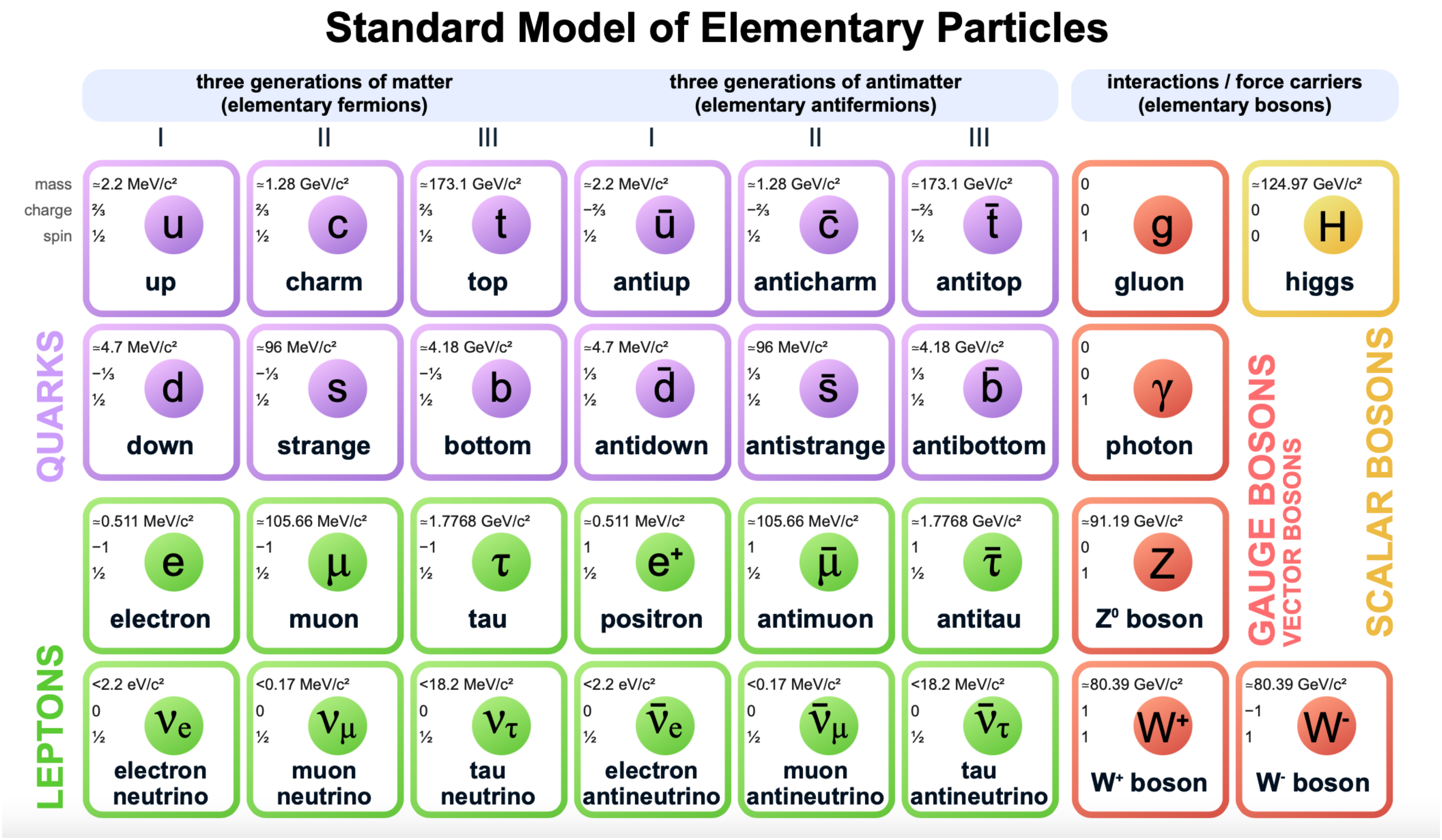


Figure 2.1: Nuclear β decay

Studies into the strength of the weak force among the different fermions has allowed categorisation of the quarks and leptons each into three pairs, also known as 'generations'. The quarks are grouped into $\binom{u}{d}$, $\binom{c}{s}$, $\binom{t}{b}$ and the leptons into $\binom{e}{\nu_e}$, $\binom{\mu}{\nu_\mu}$, $\binom{\tau}{\nu_\tau}$ [11]. The strong force is, as the name suggests, the strongest of the four fundamental forces. It holds together the quarks inside composite particles.

Each of the forces discussed are mediated via bosons, shown on the right hand side of Fig. 2.2. These bosons can be thought of as force carriers, with each force having its own boson(s) responsible for the force carrying; the gluon mediates the strong force, the photon mediates the electromagnetic force, and the $\mathrm{W}^{\pm}$ and Z^0 bosons mediate the weak force. The force carrier for the force of gravity is known as the graviton, although this particle is still hypothetical.



	I	II	III	I	II	III		
mass charge spin	≈2.2 MeV/c² ⅔ ½ u up	≈1.28 GeV/c² ⅔ ½ c charm	≈173.1 GeV/c² ⅔ ½ t top	≈2.2 MeV/c² −⅔ ½ $\bar{u}$ antiup	≈1.28 GeV/c² −⅔ ½ $\bar{c}$ anticharm	≈173.1 GeV/c² −⅔ ½ $\bar{t}$ antitop	0 0 1 g gluon	≈124.97 GeV/c² 0 0 H higgs
QUARKS	≈4.7 MeV/c² −⅓ ½ d down	≈96 MeV/c² −⅓ ½ s strange	≈4.18 GeV/c² −⅓ ½ b bottom	≈4.7 MeV/c² ⅓ ½ $\bar{d}$ antidown	≈96 MeV/c² ⅓ ½ $\bar{s}$ antistrange	≈4.18 GeV/c² ⅓ ½ $\bar{b}$ antibottom	0 0 1 γ photon	
LEPTONS	≈0.511 MeV/c² −1 ½ e electron	≈105.66 MeV/c² −1 ½ μ muon	≈1.7768 GeV/c² −1 ½ τ tau	≈0.511 MeV/c² 1 ½ e^+ positron	≈105.66 MeV/c² 1 ½ $\bar{\mu}$ antimuon	≈1.7768 GeV/c² 1 ½ $\bar{\tau}$ antitau	≈91.19 GeV/c² 0 1 Z Z^0 boson	
	<2.2 eV/c² 0 ½ ν_e electron neutrino	<0.17 MeV/c² 0 ½ ν_μ muon neutrino	<18.2 MeV/c² 0 ½ ν_τ tau neutrino	<2.2 eV/c² 0 ½ $\bar{\nu}_e$ electron antineutrino	<0.17 MeV/c² 0 ½ $\bar{\nu}_\mu$ muon antineutrino	<18.2 MeV/c² 0 ½ $\bar{\nu}_\tau$ tau antineutrino	≈80.39 GeV/c² 1 1 W^+ W^+ boson	≈80.39 GeV/c² −1 1 W^- W^- boson

Figure 2.2: A breakdown of the elementary particles of the Standard Model [12]

2.1.2 Mathematical Framework

Quantum Field Theory states that each fundamental particle (as described in Sect. 2.1.1) has a quantized field associated to it, and the particles themselves are seen as excitations of their corresponding fields. There are 12 fermion fields, although the SM does not attempt to explain the number, structure or properties of the corresponding fermions. Just as the fermions are excitations of their respective fields, the bosons are excitations of relevant gauge fields, which must be included in the SM to ensure that the model is invariant under local phase transformations of the type:

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{2.1.1}$$

Invariance under $SU(3)_C$ transformations are ensured by the mediator of the strong force, the gluon. The 'C' here stands for colour charge, with $SU(3)$ being a unitary transformation acting on a triplet of quark colour states. Being invariant under this transformation requires all hadrons that exist in nature to be colourless. This is a conclusion of Quantum Chromodynamics (QCD) which is the theory of quarks and gluons, postulating that each of the quarks have a colour charge that is either red, blue or green (and the antiquarks have anticolour charge). Since a meson is made up of a quark antiquark pair, if the quark has a green colour charge, then the antiquark must have an antigreen charge to ensure that the particle overall is colourless. Colourless combinations can be made by either a colour and its anticolour, or by all three of the colours at the same time. For example, a baryon consisting of three quarks with colour charges blue, red, and green, is also colourless overall. Quark confinement is believed to be a consequence of QCD, but is not rigorously derived from it, so for now it is included as an additional hypothesis.

Invariance under $SU(2)_L \times U(1)_Y$ transformations require the introduction of four further gauge fields. These fields give rise to the other force mediators, $\mathrm{W}^{\pm}$, Z^0 (Weak) and γ (EM). The $SU(2)_L$ transformation corresponds to the weak force, with the $SU(2)$ part being a unitary transformation acting on a doublet of quark spin states. Here it is important to introduce the concept of 'chirality', which is an intrinsic property of particles that describes how they behave in a weak interaction. Particles can either be 'left handed' or 'right handed', and due to the parity violating feature of the weak force, the $SU(2)_L$ transformation applies only to the left-handed chiral component (L). This means that only fermions with 'left-handed' chirality, or antifermions with 'right-handed' chirality interact via the weak force. Each of the fermion fields can be separated into the right and left handed components:

$$u_L = \frac{1}{2}(1-\gamma^5)u = \mathcal{P}_L u, \qquad u_R = \frac{1}{2}(1+\gamma^5)u = \mathcal{P}_R u, \tag{2.1.2}$$

where $\mathcal{P}_L$ and $\mathcal{P}_R$ are respectively the left handed and right handed chiral projection operators and γ^5 is the product of the four Dirac matrices $(i\gamma^0, \gamma^1, \gamma^2, \gamma^3)$.

The $U(1)_Y$ transformation is related to the weak hypercharge, Y, which depends on the electromagnetic charge (Q) and the third component of the weak isospin I^3_W:

$$Y = 2Q + 2I^3_W. \tag{2.1.3}$$

2.1.3 The Higgs Mechanism

The masses of the Weak force mediators, $W^{\pm}$ and Z^0, are explained only with the introduction of a new scalar field. This field is called the Higgs field, and it allows the bosons to acquire mass by breaking the electroweak symmetry in a process called Spontaneous Symmetry Breaking. The mechanism relies on the Goldstone Theorem [13], which states that a Lagrangian with continuous global symmetry (other than the vacuum symmetry) must have one massless boson associated to each generator. Spontaneous Symmetry Breaking relies on introducing an isospin $SU(2)_L$ doublet, Φ, to the SM Lagrangian comprised of charged and neutral complex fields, where:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.1.4}$$

These charged and neutral scalar fields have hypercharge Y=+1 in $SU(2)_L \times U(1)_Y$ space, and have an associated potential of:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2, \qquad \lambda > 0 \tag{2.1.5}$$

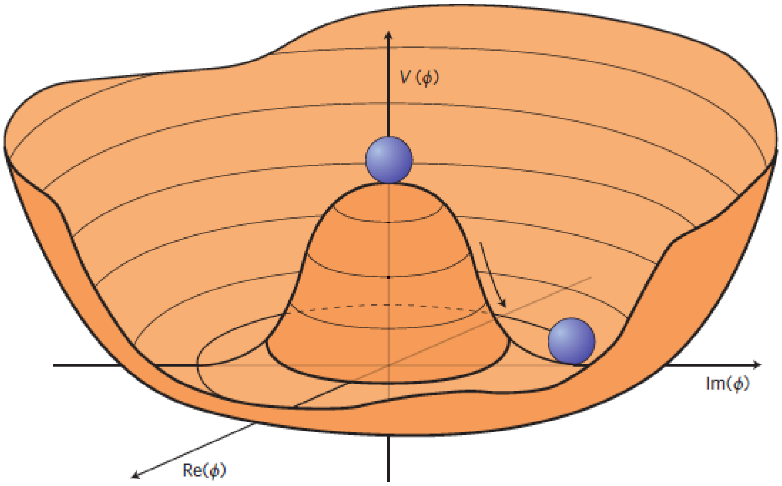


Figure 2.3: The Higgs potential for $\mu^2 < 0$ [14]

If $\mu^2 > 0$ then the potential has a minimum at $\Phi = 0$, and is symmetric. If, however, $\mu^2 < 0$, then the potential has the shape shown in Fig. 2.3. This potential has an infinite number of minima, at the points:

$$|\phi| = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}. \tag{2.1.6}$$

Where v is the vacuum expectation value which is the lowest energy configuration, being ≈ 246 GeV for the Higgs field. According to Goldstone's Theorem this potential will result in four Goldstone Bosons, one for each degree of freedom of the doublet, corresponding to the four real scalar fields. However, three of these degrees of freedom mix with the three bosons from the weak interaction ($W^{\pm}$ and Z^0) which are made massive, leaving only one single degree of freedom which becomes a new scalar particle, the Higgs boson, with $M_H = \sqrt{2\lambda v^2}$.

The addition of the Φ field means another component must be added to the electroweak Lagrangian:

$$L_{EW} = L_{Bosons} + L_{Fermions} + L_{Higgs}. \tag{2.1.7}$$

2.1.4 Fermion Masses

As well as containing information of the couplings of the electroweak gauge bosons, the Higgs Lagrangian also contains couplings to the SM fermions. The part of the electroweak Lagrangian that is responsible for the mass of the fermions is called the Yukawa Lagrangian, which describes the coupling between the Higgs field and massless quark and lepton fields. The Yukawa couplings enable the fermions to acquire masses through the Higgs mechanism. These couplings are described by Eq. 2.1.8.

$$-\mathcal{L}_{\text{Yukawa}} = Y_{ij}\bar{\psi}_{L,i}\phi\psi_{R,j} + h.c., \tag{2.1.8}$$

where $h.c.$ stands for the hermitian conjugates, $\psi_{L,i}$ is a vector of the weak isospin doublets (with one component for each of the three quark generations described in Sect. 2.1.1), ϕ is the scalar Higgs field, $\psi_{R,i}$ are the weak isospin singlets, and Y_{ij} is a complex 3×3 matrix in flavour space. $\mathcal{L}_{Yukawa}$ is $SU(2) \times U(1)$ invariant. After spontaneous symmetry breaking this Lagrangian becomes

$$-\mathcal{L}_{\text{Yukawa}} = M_{ij}^{up}\bar{q}_{L,i}^{up}q_{R,j}^{up} + M_{ij}^{down}\bar{q}_{L,i}^{down}q_{R,j}^{down} + h.c., \tag{2.1.9}$$

where the M_{ij} terms represent matrices that represent the mass terms for the quarks. These mass terms are determined experimentally, meaning that the Yukawa coupling terms themselves are also experimentally determined.

2.1.5 The CKM Matrix

The Cabibbo-Kobayashi-Maskawa (CKM) matrix is a complex 3×3 unitary matrix containing information on the strength of the flavour-changing weak interaction. It determines the probability that a particular quark will turn into a quark of another flavour.

The ‘up type’ quarks (charge $= +\frac{2}{3}e$), being u, c and t, are by convention chosen to be pure states [15], and flavour mixing is described in terms the CKM matrix, V_{CKM}, acting on the d, s and b quark states:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

Where d’, s’ and b’ are partners of u, c and t respectively within the weak isospin doublets. The quark mixing matrix, and the most recent calculations for the values of its elements [16] are given below:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.974 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix}$$

The matrix is almost diagonal, so the diagonal elements are close to one, while the off-diagonal elements are very small (but not zero). This means that the most probable transitions are between quarks of the same generation. The values of the matrix elements are determined by experiment, by measuring the rates of weak decays of the relevant quarks. For the $B_s^0 \to \mu^+\mu^-$ analysis, the relevant elements are V_{ts} and V_{tb}, which describe the strength of the coupling between the $t \leftrightarrow s$ quarks and $t \leftrightarrow b$ quarks respectively. The fact that the value of V_{ts} is so small gives rise to a suppression of the branching fractions of the $B_s^0 \to \mu^-\mu^+(\gamma)$ decays (see Sect. 2.2).

2.2 $B_{(s)}^0 \to \mu^-\mu^+(\gamma)$ in the Standard Model

Mesons are the simplest form of hadron, consisting of a quark antiquark pair. The focus of this thesis is decays of neutral B mesons. The B_s^0 meson is a neutral combination of $\bar{b}s$ quarks, with an antiparticle, $\bar{B}_s^0$, made up of a $b\bar{s}$ quark combination. For this analysis, reference to the B_s^0 meson covers both the particle itself and the antiparticle (since we do not differentiate between the two). The same applies to the B^0 meson, which is made up of either $\bar{b}d$, or $b\bar{d}$ (which is the antiparticle, $\bar{B}^0$) quark combinations. The $B_{(s)}^0$ notation refers to either of these neutral B mesons.

The $B_{(s)}^0 \to \mu^+\mu^-$ decays involve a transition between two ‘down type’ quarks of different flavours. This is known as a Flavour Changing Neutral Current (FCNC) which is forbidden in the SM at leading order. These transitions are only allowed to proceed via a quantum loop involving an ‘up type’ quark and two up-down transitions, as shown in Fig. 2.4. The fact that the decays cannot proceed at leading order causes suppression, which is one of the reasons why the $B^0_{(s)} \to \mu^+\mu^-$ decays are so rare. The loop suppression factor is calculated using the weak coupling constant, g_W as:

$$\left(\frac{g_W}{4\pi}\right)^2 \sim 10^{-3}. \tag{2.2.1}$$

The probability of any decay decreases rapidly depending on the number of electroweak vertices (from the mass of the W$^\pm$ bosons), as well as the heaviness of the virtual particles involved.

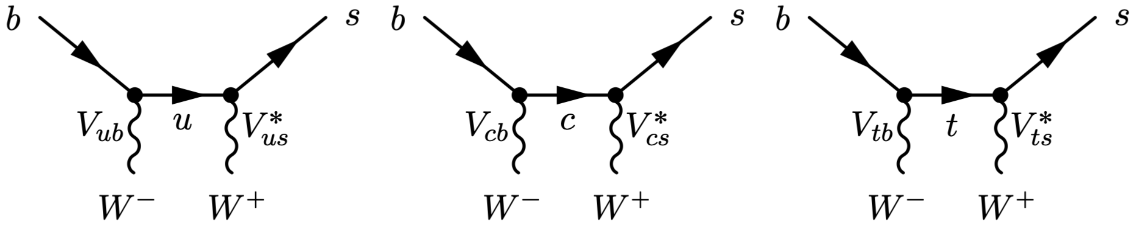


Figure 2.4: Possible $b \to s$ transitions via a loop with an ‘up-type’ quark.

In addition to this loop suppression factor, the unitary nature of the CKM matrix causes further suppression for the $b \to s$ transition:

$$V^*_{ub}V_{us} + V^*_{cb}V_{cs} + V^*_{tb}V_{ts} = 0. \tag{2.2.2}$$

The decay is only saved from being forbidden entirely due to the different quark masses, which mean that complete cancellation is avoided. If all of the up type quarks had the same mass, the total contribution would be reduced to zero. The strength of the contributions is defined by:

$$\mathcal{A}(B^0_{(s)} \to \ell^+\ell^-) \propto V_{ub}V^*_{uq}Y_0(x_u) + V_{cb}V^*_{cq}Y_0(x_c) + V_{tb}V^*_{tq}Y_0(x_t), \tag{2.2.3}$$

where q is the d or s quark depending on the type of B meson (B^0 or B^0_s respectively), $Y_0(x_i)$ a gauge-independent Inami-Lim factor [17], and x_i $(i = u, c, t)$ is m_i^2/m_W^2, where m_i is the quark mass and m_W is the mass of the $W^\pm$ boson.

The strength of the contribution from each quark loop to the $B^0_s \to \mu^+\mu^-$ decay is therefore determined by the values of the CKM matrix and the quark masses. The strength of each contribution is calculated in Ref. [17], and is proportional to the quark mass at first order, meaning that it strongly favours contributions from heavier quarks. Due to the large mass of the top quark, this contribution is dominant and the contributions from the c and u quark loops become negligible. The cancellation effects on the up, charm, and top loops from the unitary conditions of the CKM matrix are less effective on the top because of the large mass difference. The top loop contribution is dominant because of these factors, and also because of the fact that V_{tb} is large (0.999) compared to V_{cb} and V_{ub} (0.041 and 0.004 respectively). The dominating ‘Z penguin’ loop process (shown in

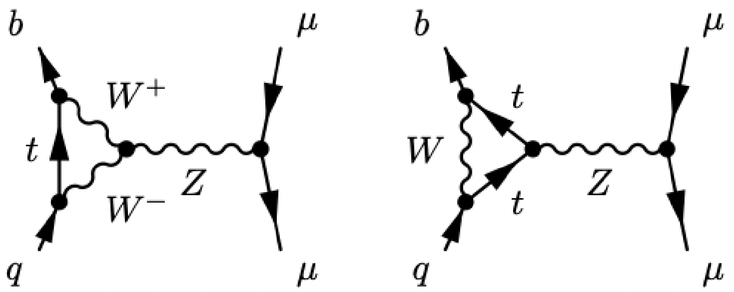


Figure 2.5: The ‘Z penguin’ contributions to the $B^0_{(s)} \to \mu^+\mu^-$ decays, where q is a strange or down quark.

Fig. 2.5, responsible for the largest contribution (75%) to the decay [18]) also involves four electroweak vertices. Furthermore, there are quark transitions from $b \to t$ and $t \to s$, so the CKM matrix transitions relating to these flavour changes (V_{tb} and V_{ts}) are important; the fact that V_{ts} has such a low strength contributes to the rareness of the $B^0_s \to \mu^+\mu^-$ decay. The $B^0_d \to \mu^+\mu^-$ decay is rarer still, since the value of V_{td} is even smaller than V_{ts}.

In addition to these suppression factors, we must also take into account helicity for $B^0_s \to \mu^+\mu^-$. The B^0_s meson is a spin-0 particle, so it is only allowed to decay to a final state that also has zero total angular momentum. In terms of helicity, this means that the μ^+ and μ^- in the final state must both be in either positive or negative helicity states. However, weak interactions produce muons with opposite chirality, meaning that the helicity of one of the muons has to be flipped. This adds a helicity suppression factor proportional to the mass of the muon squared: $(m_\mu/m_B)^2 \sim 4 \times 10^{-4}$. Because of this factor, the suppression of the $B^0_s \to e^+e^-$ decay is much greater (as $m_e < m_\mu$), and $B^0_s \to \tau^+\tau^-$ is less suppressed (as $m_\tau > m_\mu$) but is experimentally more difficult to detect.

These suppression factors mean that overall the fraction of B^0_d and B^0_s decays to $\mu^+\mu^-$ are reduced to the level of 10^{-11} and 10^{-9} respectively.

2.2.1 Effective Field Theory

The SM prediction of the transitions involved in these decays are typically calculated in an effective field theory framework, with an effective Hamiltonian ($\mathcal{H}_{eff}$) derived from the SM Lagrangian, parameterized in terms of a sum of operators $\mathcal{O}_i$ and Wilson Coefficients C_i giving

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i V^i_{CKM} C_i(\lambda)\mathcal{O}_i(\lambda). \tag{2.2.4}$$

The Wilson coefficients (C_i) provide information about short-distance effects at high energy scales, while the operators ($\mathcal{O}_i$) encompass the information about the long distance (non-pertubative) effects. The subscript of the operator $\mathcal{O}_i$ indicates the type of current, and a primed operator ($\mathcal{O}_i^{'}$) refers to the contribution from currents of quarks with opposite handedness. The Wilson coefficients are independent of the initial and final states, and are usually computed using lattice QCD. Their values are found by matching the effective field theory expression to the amplitude obtained using the full theory [19], at the scale of the mass of the W boson, $\mathcal{O}(m_W)$.

For the $B^0_{(s)} \to \mu^+\mu^-$ decays, the only non-vanishing contributions to the decay amplitudes are the axial-vector, scalar, and pseudo-scalar lepton currents, corresponding to the operators $\mathcal{O}^{(\prime)}_{10}$, $\mathcal{O}^{(\prime)}_{S}$, $\mathcal{O}^{(\prime)}_{P}$ respectively [20], defined in Eqs. 2.2.5 and 2.2.6. That said, in the SM, the contributions from $\mathcal{O}^{(\prime)}_{S}$ and $\mathcal{O}^{(\prime)}_{P}$ are suppressed by a factor of M_{B_q}/M_W [21] (where M_{B_q} is the mass of the $B^0_{(s)}$ meson) meaning that they are negligible.

$$\mathcal{O}_{10} = \frac{\alpha_{em}}{4\pi}\bar{s}\gamma^{\mu}P_L b\bar{\ell}\gamma_{\mu}\gamma_5\ell \tag{2.2.5}$$

$$\mathcal{O}_P = \frac{\alpha_{em}}{4\pi}m_b\bar{s}P_R b\bar{\ell}\gamma_5\ell, \qquad \mathcal{O}_S = \frac{\alpha_{em}}{4\pi}m_b\bar{s}P_R b\bar{\ell}\ell, \tag{2.2.6}$$

where $P_{L,R} = 1 \pm \gamma_5$. Here, the $\mathcal{O}^{\prime}_P$ and $\mathcal{O}^{\prime}_S$ are defined by replacing P_R with P_L and m_b with m_s.

The V_{CKM} factor in Eq. 2.2.4 encompasses the transition values from the CKM matrix relevant to the decay. For example, in the case of $B^0_s \to \mu^+\mu^-$, the V_{CKM} factor is the product of two CKM elements: V_{ts} and V_{tb}, so V_{CKM} in the effective Hamiltonian is replaced with $|V^*_{ts}V_{tb}|^2$.

2.2.2 Branching Fraction Computation

Using the effective Hamiltonian defined earlier in Eq. 2.2.4, the SM branching fraction for $B^0_s \to \mu^+\mu^-$ can then be calculated using Eq. 2.2.7:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-)_{SM} = \tau_{B_s}\frac{G_F^2\alpha_{em}^2}{16\pi^2}f_{B_s}^2 m_{\mu}^2 m_{B_s}\sqrt{1-\frac{4m_{\mu}^2}{m_{B_s}^2}}|V_{tb}V^*_{ts}|^2|C^{SM}_{10}|^2, \tag{2.2.7}$$

where τ_{B_s} is the lifetime of the B^0_s meson, f_{B_s} is the B^0_s meson decay constant (obtained in lattice QCD [22]), G_F is the Fermi decay constant, α_{em} is the fine structure constant, and m_{μ} and m_B are the mass of the μ and B^0_s meson respectively. Using this calculation the most recent SM prediction for the branching fraction, from Ref. [23], is:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-)_{SM} = (3.65 \pm 0.23) \times 10^{-9}. \tag{2.2.8}$$

2.2.3 From $B^0_s \to \mu^+\mu^-$ to $B^0_s \to \mu^+\mu^-\gamma$

The suppression of the $B^0_s \to \mu^+\mu^-\gamma$ decay differs slightly to that of its non-radiative counterpart. Since we now have a photon in the final state, there is no helicity suppression.

However, we now have to take into account another electroweak vertex, meaning that the overall suppression of the two decays comes out to be similar.

The SM predictions of the branching ratio for the $B^0_{s,d} \to \mu^+\mu^-\gamma$ decay from Ref. [24] are given below:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-\gamma)_{SM} = 1.89 \times 10^{-8} \tag{2.2.9}$$

$$\mathcal{B}(B^0_d \to \mu^+\mu^-\gamma)_{SM} = 1.31 \times 10^{-10} \tag{2.2.10}$$

Like the $B^0_s \to \mu^+\mu^-$ decay, $B^0_s \to \mu^+\mu^-\gamma$ is sensitive to the C_{10} Wilson coefficient, however it is additionally sensitive to C_9 and C_7 due to the additional photon, offering a new probe into physics beyond the SM.

The electromagnetic operator $\mathcal{O}_7$ and semileptonic operator $\mathcal{O}_9$ are defined as:

$$\mathcal{O}_7 = \frac{m_b}{e}\bar{q}\sigma_{\mu\nu}P_R b F^{\mu\nu} \tag{2.2.11}$$

$$\mathcal{O}_9 = \bar{q}\gamma_\mu P_L b \bar{\ell}\gamma^\mu \ell \tag{2.2.12}$$

The sensitivity to C_7, C_9 and C_{10} (as well as $C_7^{'}$, $C_9^{'}$ and $C_{10}^{'}$) make the decay extremely interesting, and particularly in the light of current data which already seems to favour NP contributions to these coefficients. The current discrepancies between experimental results and the SM that favour new values of these coefficients are summarised in Ref. [25]. Two examples are given below:

- The measurement of the branching ratio of $B^0_s \to \phi\mu^+\mu^-$ is lower than the SM prediction by more than 3σ [26],
- The ratio of the branching fractions $\mathcal{B}(B^+ \to K^+e^+e^-)$ to $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$, R_K, shows evidence for the breaking of lepton universality in beauty-quark decays, with a significance of 3.1 standard deviations [27].

In the analysis described in this thesis, only the high invariant mass region of the two muons in the $B^0_s \to \mu^+\mu^-\gamma$ decay is probed. This is due to the analysis strategy, which will be described later in Sect. 5. For this reason, there is sensitivity only to the operator $C_9^{(')}$, as sensitivity to $C_7^{(')}$ is only in the lower invariant mass range. The SM prediction for the $B^0_s \to \mu^+\mu^-\gamma$ decays is obtained from Refs. [24, 28]. It is composed of three terms: initial state radiation (ISR, photon emitted from a quark leg), final state radiation (FSR, photon emitted by a lepton leg), and their interference. Feynman diagrams of the SM processes for the ISR and FSR decays are shown in Fig. 2.6.

Figure 2.7 from Ref. [25] shows the total $B^0_s \to \mu^+\mu^-\gamma$ spectrum and the breakdown of the ISR, FSR and interference components.

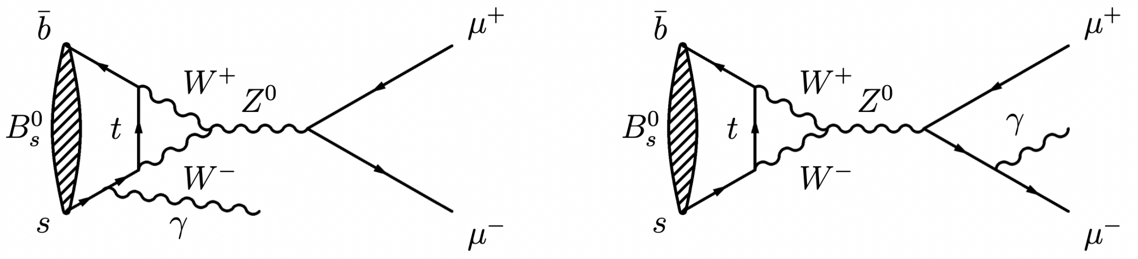


Figure 2.6: Examples of SM Feynamn diagrams mediating the $B_s^0 \to \mu^+\mu^-\gamma$ decays with contributions from ISR (left) and FSR (right).

The FSR photons are predominantly soft and their contribution is included experimentally in the $B^0_{(s)}$ mass shape (via PHOTOS [30] software) as a radiative tail[1].

The ISR process however is characterised by a larger momentum of the photon. The interference of ISR and FSR is negligible [24]. Therefore here and in the following thesis, $B_s^0 \to \mu^+\mu^-\gamma$ refers only to the ISR component. The branching fraction relative to this part is obtained by integrating the differential branching fraction of only the ISR component in the region of interest $m_{\mu^+\mu^-} \in [4.9\,\text{GeV}, M_{B_s^0}]$, and yields:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)|_{m_{\mu\mu}>4.9\,\text{GeV}} = (1.92 \pm 0.19) \times 10^{-10}. \tag{2.2.13}$$

2.3 $B^0_{(s)} \to \mu^-\mu^+(\gamma)$ and New Physics models

Within the present errors, new contributions to the Wilson coefficient of the operator $\mathcal{O}_{10}$ (defined in Eq. 2.2.5) are allowed and actually favored [25]. The sensitivity of the $B_s^0 \to \mu^+\mu^-(\gamma)$ decay to the C_{10} coefficient offers a very interesting probe into NP.

Additionally, although the SM branching ratio (Eq. 2.2.7) is dependent on only the Wilson coefficient C_{10}, NP models allow other contributions to the decay. In the SM, the pseudo-scalar and scalar operators $\mathcal{O}_P$ and $\mathcal{O}_S$ (Eq. 2.2.6), described by the coefficients 'P' and 'S' (defined in Eqs. 2.3.1 and 2.3.2), are P^{SM}= 1 and S^{SM}= 0. 'P' is normalized to the SM C_{10} value, meaning it could contain other C_P, C_{10} or C_{10}' contributions if the SM is incorrect.

$$P = \frac{C_{10}^R - C_{10}^L}{C_{10}^{SM}} + \frac{M^2_{B^0_{(s)}}}{2m_\mu}\Big(\frac{m_b}{m_b + m_{d,s}}\Big)\Big(\frac{C_P^R - C_P^L}{C_{10}^{SM}}\Big) \tag{2.3.1}$$

$$S = \sqrt{1 - \frac{4m_\mu^2}{M^2_{B^0_{(s)}}}} \cdot \frac{M^2_{B^0_{(s)}}}{2m_\mu}\Big(\frac{m_b}{m_b + m_{d,s}}\Big)\Big(\frac{C_S^R - C_S^L}{C_{10}^{SM}}\Big) \tag{2.3.2}$$

---

[1] Using Monte Carlo simulation in the PHOTOS software rather than the analytic approach outlined in Ref [29] is advantageous as detector efficiencies are taken into account.

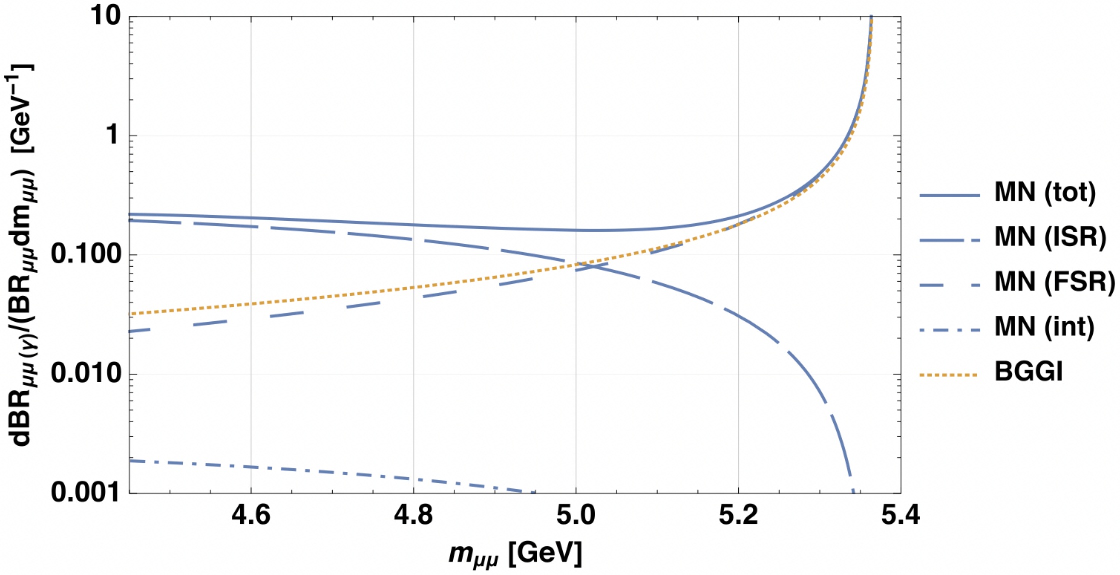


Figure 2.7: Breakdown of the $B_s^0 \to \mu^+\mu^-\gamma$ full spectrum (solid blue line) calculated in Ref. [24], denoted as MN in the legend – into its pure ISR component (long-dashed blue), FSR component (medium-dashed blue), and ISR-FSR interference (dot-dashed blue). The dotted orange line is the soft photon approximation (denoted as BGGI) from Ref. [29]. Figure taken from Ref. [25].

In NP models, S and P could have several different values, some examples of which are given below. [31]:

- S = 0: Pseudo-scalar dominated NP - For example, Constrained Minimal Flavor Violation (CMFV), the Randall-Sundrum model, and certain two Higgs Doublet Models (2HDMs).
- P = 1: Scalar dominated NP - Complementary to the above scenario of S=0 as NP effects dominated by scalar operators.
- P ± S = 1: Mixed (pseudo-)scalar NP - For example, the Minimal Super-symmetric extension of the SM (MSSM), and models involving the Minimal Flavour Violation (MFV) [32] hypothesis.

2.3.1 Implications of current experimental results

The measurements of the branching fraction of the $B_s^0 \to \mu^+\mu^-$ decay have already constrained the possible values of the Wilson Coefficients C_9 and C_{10}. Reducing the uncertainty on the measurement will further constrain these variables, and the NP models that predict larger fluctuations in these. Constraints from individual decay modes can be combined to find an overall constraint on the values. Figure 2.8 shows individual constrains

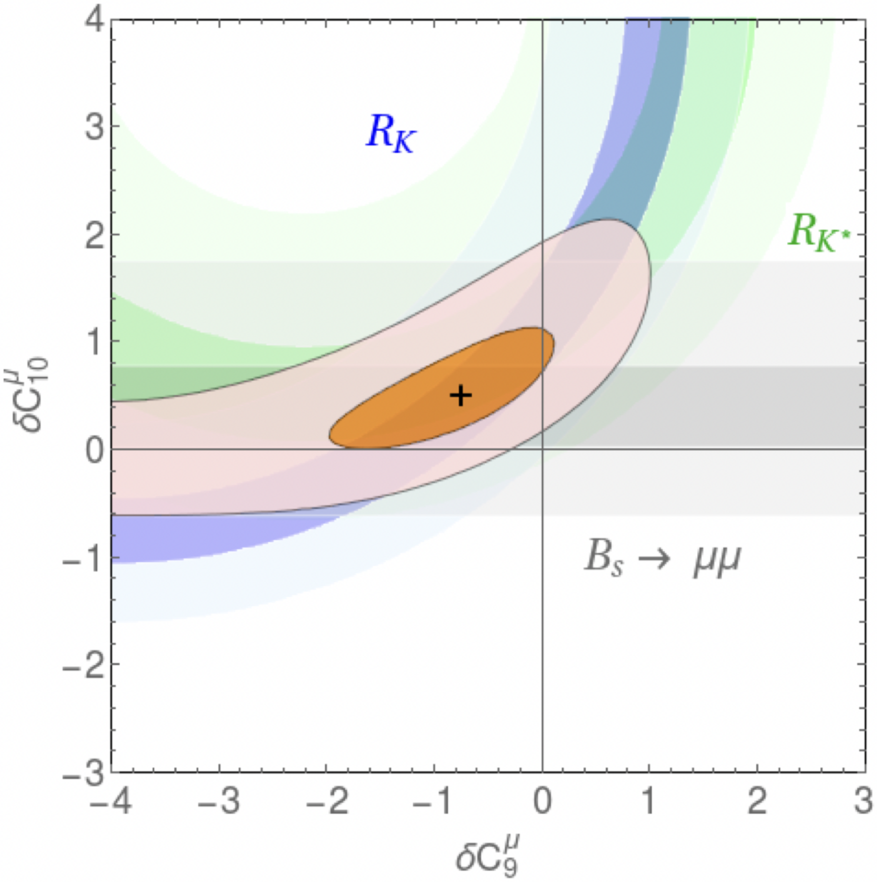


Figure 2.8: Constraint contours in the $(\delta C_9, \delta C_{10})$ plane at 1σ and 3σ from R_K (Blue), R_{K^*} (Green) and $B_s^0 \to \mu^+\mu^-$ (Grey), as well as the fit to combined $X_b \to X_s \ell^+\ell^-$ data (Orange). The cross indicates the position of the minimum. Figure taken from Ref. [33].

from $R_{K^{(*)}} = \mathcal{B}(B \to K^{(*)}\mu^+\mu^-)/\mathcal{B}(B \to K^{(*)}e^+e^-)$ and $B_s^0 \to \mu^+\mu^-$, as well as that from the fit to all three modes combined.

Measurements of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions alone also provide strong constraints on NP. Various new physics models predict fluctuations in these branching fraction measurements or the ratio between them. Figure 2.9 shows the allowed regions of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions for various new physics scenarios, including Minimal Flavour violation (MFV), the SM with a fourth generation (SM4) [34], the Randall-Sundrum model with custodial protection (RSc) [35] and four SUSY models (MSSM):

- LL - model with left-handed currents only [36],
- RVV2 - proposed by Ross, Velasco-Sevilla and Vives [37],
- AKM - proposed by Antusch, King and Malinsky [38],
- AC - proposed by Agashe and Carone [39].

The greyed out area shows what has already been ruled out experimentally.

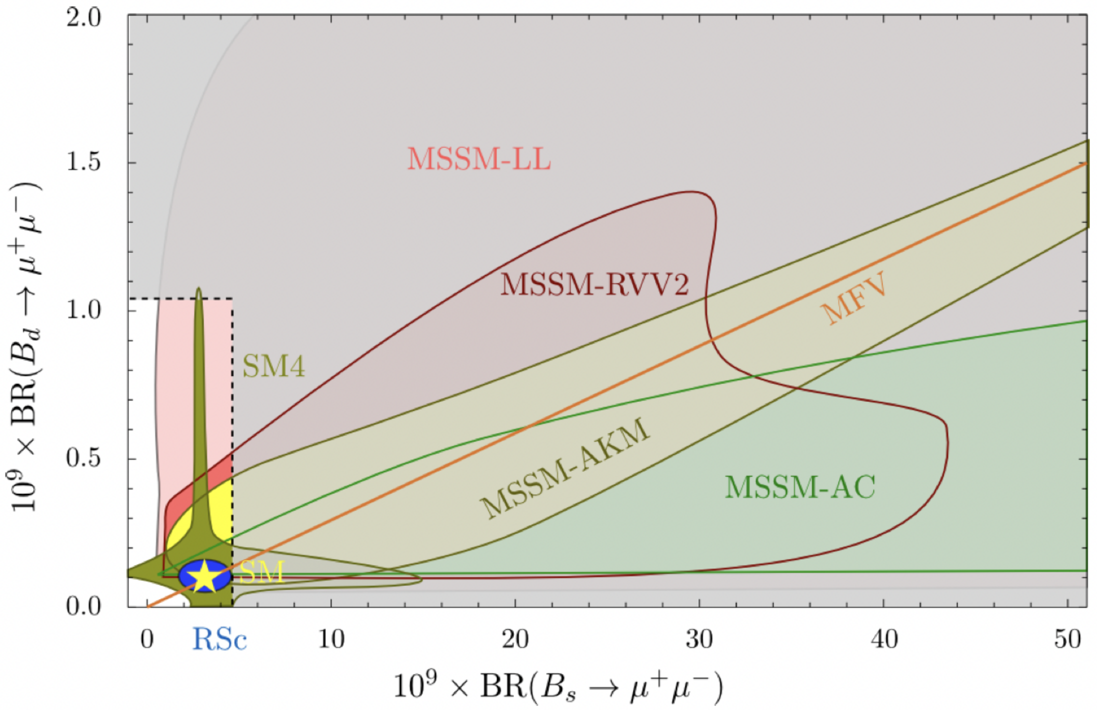


Figure 2.9: The correlation between the branching ratios of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ in MFV, the SM4, RSc and four SUSY flavour models. The SM point is marked by a star, and the grey area is ruled out experimentally. Figure taken from Ref. [40].

Chapter 3

Experimental Environment

3.1 The Large Hadron Collider

The European Organisation for Nuclear Research (CERN) is the largest laboratory in the world for particle physics research. CERN hosts several particle accelerators including the Large Hadron Collider (LHC) [41], which is a superconductive synchrotron located across the French-Swiss border near Geneva. It is installed inside a tunnel 100m underground, which was originally built to host the Large Electron-Positron collider (LEP). The primary purpose of the LHC is to study processes taking place in proton-proton (pp) collisions at high energies. Two proton beams are accelerated in opposite directions and brought together at four collision points around the ring. Surrounding these collision points are the four main experiments at the LHC:
1) ALICE [42] (A Large Ion Collider Experiment) is specifically designed to study heavy nuclei collisions.
2) ATLAS [43] (A Toroidal Lhc ApparatuS) is a multi-purpose detector for Higgs physics and NP direct searches;
3) CMS [44] (Compact Muon Solenoid) is a multi-purpose detector that complements and competes with ATLAS;
4) LHCb (Large Hadron Collider Beauty), which focuses on physics containing b quarks, and is the detector used for the analysis outlined in this thesis. LHCb will be discussed in detail in Sect. 3.2.

3.1.1 The LHC Accelerator System

The LHC represents the last element in the chain of the CERN accelerator complex, shown in Fig 3.1.

The protons for the LHC beams are provided by gaseous hydrogen. Using a strong electric field, the electrons are stripped from the hydrogen atoms leaving just the protons. These isolated protons are then accelerated to higher and higher energies using various

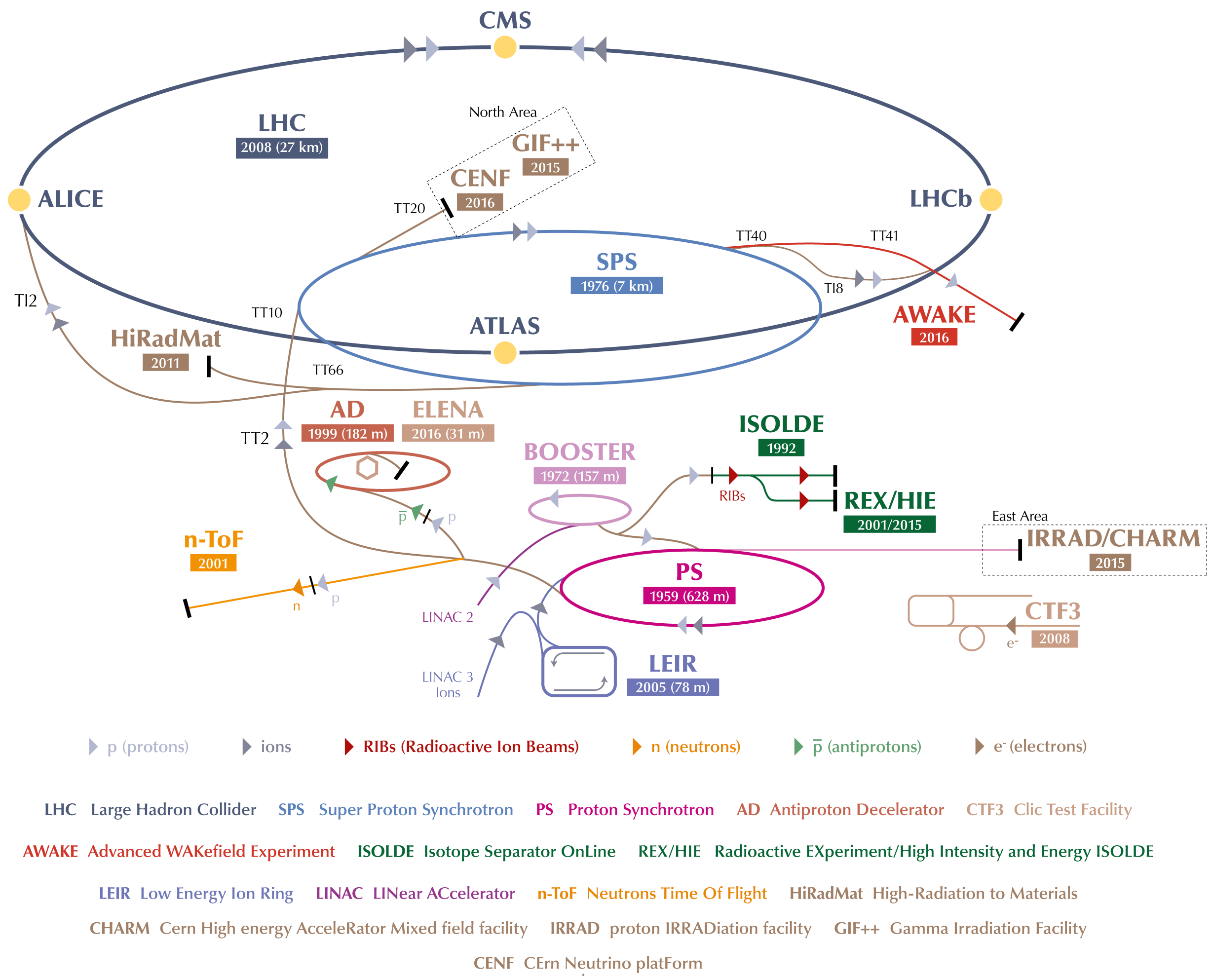


Figure 3.1: The CERN Accelerator Complex [45]

machines, the first of which is the linear accelerator LINAC2, which focuses the beam using quadrupole magnets and accelerates the protons to an energy of 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB) where it reaches an energy of 1.4 GeV before being transferred to the Proton Synchrotron (PS).

The PS accelerates the beam to 25 GeV and also forms ‘bunches’ of protons, with each bunch typically containing around 10^{11} protons and a time spacing between bunches of 25 ns. The bunches are sent to the Super Proton Synchrotron (SPS), a 7 km long accelerator ring that ramps up their energy to 450 GeV, at which time they are finally fed into the LHC.

The LHC continues to accelerate the proton bunches until they reach the required energy, using eight radio-frequency cavities located around the ring.

3.1.2 Luminosity of colliding beam experiments

The rate at which a process occurs during collisions (R) is given by the proton-proton cross section (σ_{pp}) of the process multiplied by the instantaneous luminosity ($\mathcal{L}$) of the accelerator:

$$R = \sigma_{pp} \times \mathcal{L} \tag{3.1.1}$$

The number of pp collisions (N_{pp}) can then be written as:

$$N_{pp} = \sigma_{pp} \int_{\Delta t} \mathcal{L} dt. \tag{3.1.2}$$

The luminosity is a key parameter for colliding beam experiments. The instantaneous luminosity is the number of incident particles, per centimeter squared, per second [cm$^{-2}s^{-1}$]. It includes information regarding the magnets ability to focus the beams at the interaction point, described by σ_x and σ_y, which are the transverse beam profiles in the horizontal and vertical planes respectively. The luminosity also depends on the number of protons per bunch (N_1 and N_2 are the number of particles in beam 1 and 2 respectively), and the beam crossing frequency, f (number of bunch crossings per second). In terms of these beam parameters, the instantaneous luminosity can be expressed as:

$$\mathcal{L} = f \frac{N_1 N_2}{4\pi\sigma_x\sigma_y} \tag{3.1.3}$$

Due to the large number of incident particles every second in the LHC, the instantaneous luminosity is not often used. Instead, it is more practical to use the total integrated luminosity (L), which is the time-integral of the instantaneous luminosity,

$$L = \int \mathcal{L} dt. \tag{3.1.4}$$

The integrated luminosity is measured in units of ‘barns’ with one barn being 10^{-24} cm^2. The design instantaneous luminosity[1] of the LHC is 10^{34} cm^{-2} s^{-1}. During regular operations, the beam intensity degrades over time due to the pp collisions and proton scattering on residual gas in the beam pipe. Consequently, the collision rate also decays with time. When it becomes too low, the beam is dumped on an absorber and a new fill is prepared.

A high luminosity is required for experiments searching primarily for New Physics, as the cross sections of such processes are predicted to be low. For this reason, projects like ATLAS and CMS require the luminosity to be as high as possible, as this provides higher statistics for their searches. However, a high luminosity means an increased number of collisions per bunch-crossing, μ, which is problematic for detectors designed specifically to study rare decays with high precision, such as LHCb.

---

[1] Numbers pertain mainly to ATLAS/CMS, as LHCb uses a method to specifically designed to reduce this level, discussed shortly.

The presence of multiple interactions per bunch crossing, known as pile-up, makes reconstruction for experiments such as LHCb more difficult. The ideal scenario for LHCb is to have only one collision per bunch crossing, as this allows the primary and secondary vertices of each collision to be properly determined in order to differentiate signal from background events. Because of this, the luminosity at the LHCb collision point is kept below the maximum deliverable from the LHC. This is achieved by offsetting the colliding beams slightly, in order to reduce the active collision area.

By finely tuning the transverse separation between the beams, radiation damage of the detector elements closer to the beam pipe is also reduced. The beams are progressively brought closer to each other in the transverse plane, so that the rate of collisions is almost constant even as the proton beams decay. This process is called 'luminosity levelling'.

Largely thanks to this luminosity control and excellent performance of the trigger and reconstruction system, most of the LHCb data was recorded at an instantaneous luminosity of $\sim 4 \times 10^{32}$ cm^{-2} s^{-1}, which is equivalent to $\mu \approx 1.7$, compared to the design value of $\mathcal{L} = 4 \times 10^{32}$ cm^{-2} s^{-1} and $\mu \approx 0.4$.

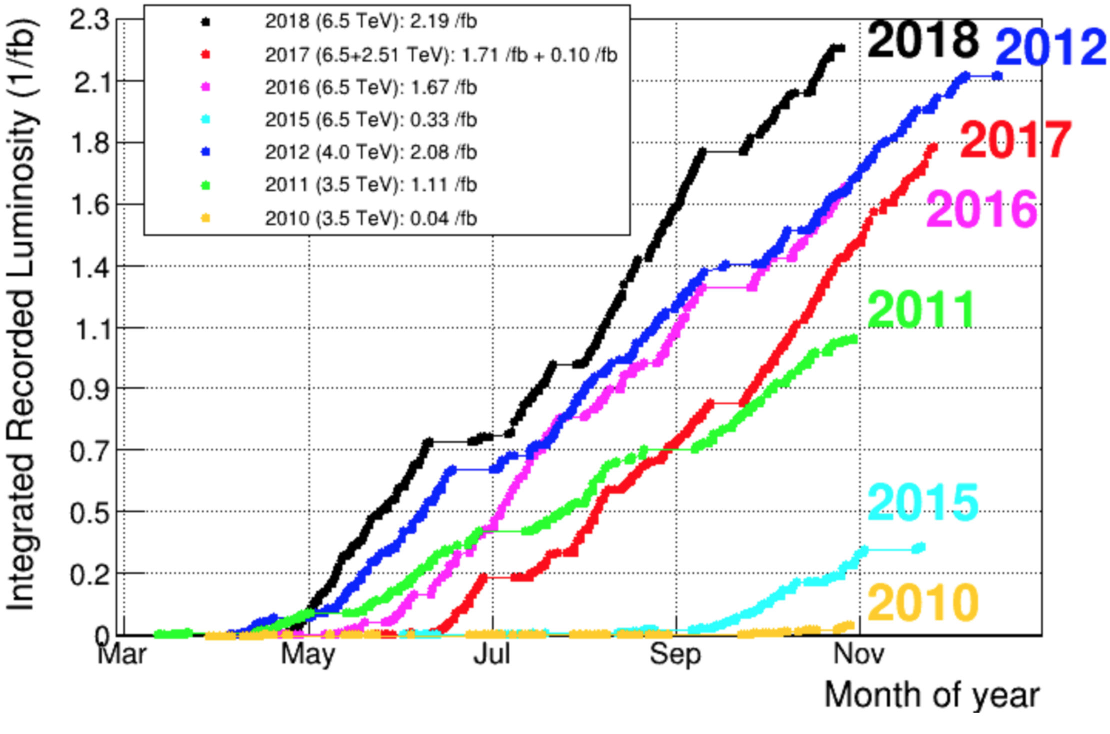


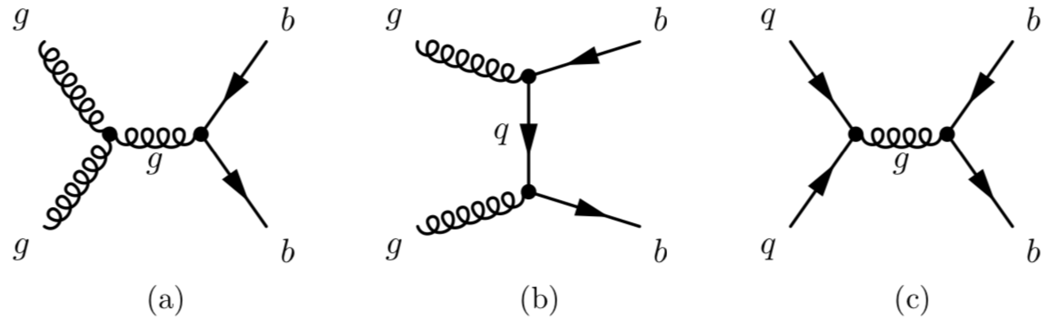
Figure 3.2: LHCb integrated recorded luminosity by data taking year.

3.1.3 Current performance

The LHC was originally commissioned to run at a centre-of-mass energy of $\sqrt{s} = 14$ TeV for proton collisions, and $\sqrt{s} = 2.76$ TeV per nucleon in the case of lead ion collisions, after a period of commissioning. Unfortunately an incident in 2008 meant that the operation had to be suspended whilst repairs were made. The decision was then made to start running the LHC at a centre-of-mass energy of $\sqrt{s} = 7$ TeV, before increasing to $\sqrt{s} = 8$ TeV in 2012 (known collectively as Run 1). Following the long-stop for upgrades in the period 2013-2014, the LHC was then restarted in 2015 at an energy of $\sqrt{s} = 13$ TeV. It ran for 4 years (Run 2) at this energy until switched off at the end of 2018 for planned upgrades of the detectors. A summary of the integrated recorded luminosity per data taking year can be seen in Fig. 3.2.

3.1.4 Production of B mesons at the LHC

The production rate of B mesons is given by the probability that a b-quark forms a particular B meson, multiplied by the $pp \to b\bar{b}$ cross section. The dominant mechanisms producing $b\bar{b}$ pairs in proton collisions are gluon fusion and quark-antiquark annihilation, as shown in Fig. 3.3. The parton distribution functions (pdf) for the initial states show that the $b\bar{b}$ pairs tend to be boosted along the direction of the beam line, due to the asymmetric incoming parton momenta. This means that the pseudorapidity (η, defined in Eq. 3.2.1) distributions for both of the produced b-quarks are relatively flat in the range $-5 < \eta < 5$, and there is a strong correlation between the b and $\bar{b}$ pseudorapidity values. This means that if one b-quark is produced in the forward (or backward) region, the other will also be produced in that same direction. The production angles of $b\bar{b}$ pairs with respect to the proton beam direction are shown in Fig. 3.4 [46].



(a) (b) (c)

Figure 3.3: The three main $b\bar{b}$ pair production processes in proton-proton collisions. Dominant contributions are from gluon-fusion (a) and (b), while the contribution from quark-antiquark annihilation (c) is small.

3.2 The LHCb experiement

The LHCb detector is a forward spectrometer designed to study the physics of b- and c-quarks arising from proton-proton collisions at the LHC. It covers a range from 10-300 mrad in the horizontal plane and 10-250 mrad in the vertical plane. The study of the

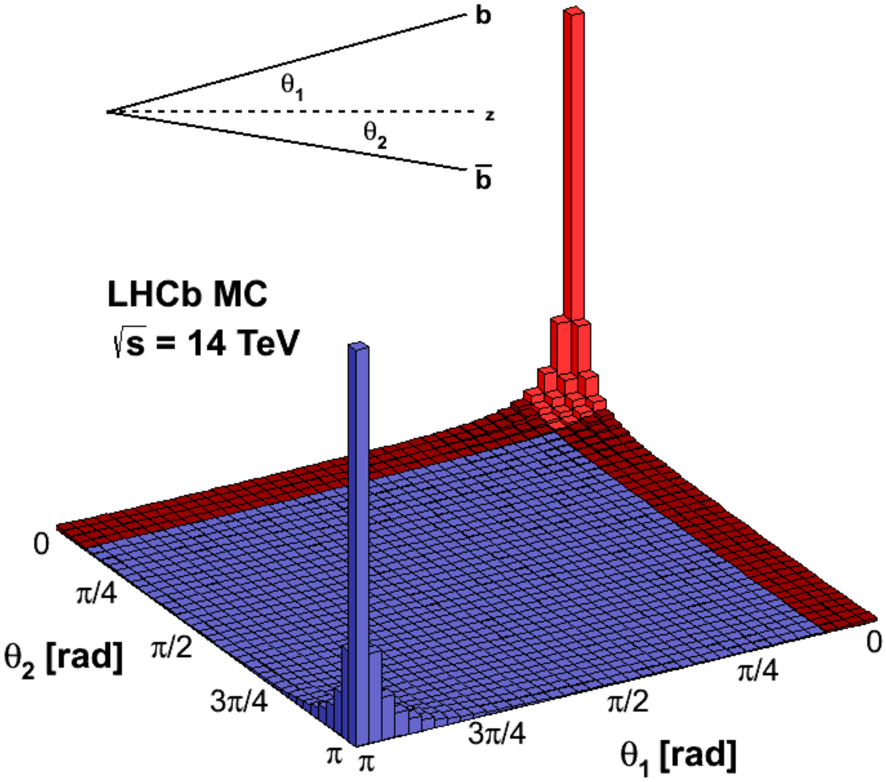


Figure 3.4: Angular distribution of simulated $b\bar{b}$ pairs at the LHC, with the LHCb detector acceptance range is shown in red. The lighter red area shows the events where both b-quarks are produced in the LHCb acceptance [47].

rare $B^0_{(d,s)} \to \mu^+\mu^-$ decays is one of the reasons for which the detector was designed. The performance and design of the LHCb detector are discussed in detail in Refs. [47,48]. This section will provide an overview of the detector components and the common variables used to describe the reconstructed particle tracks.

3.2.1 Coordinates and Variables

Particle tracks reconstructed in the LHCb detector are described using a right-handed coordinate system, with the z axis pointing along the direction of the beam line, and y being the vertical axis. The direction of the particles emerging from the collision point are described by the pseudorapidity, η:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right), \tag{3.2.1}$$

where θ is the polar angle; the angle with respect to the beam axis z. Particles emerging perpendicular to the beam will have $\eta = 0$.

The area where the two proton beams are bought together is called the 'Interaction region', with the exact proton-proton collision point being the 'Primary Vertex' (PV). The PV varies with each collision and is found by reconstructing particle tracks. If an unstable particle is produced at the PV, then it will decay at another point in the detector. The position of the secondary decay is known as the 'Secondary Vertex'. The closest distance between the reconstructed particle track and the associated PV is known as the 'Impact Parameter' (IP). Another important variable used in distinguishing signal from background events is the p_T, which is the transverse component of the particle momentum p.

3.2.2 The Layout of the LHCb detector

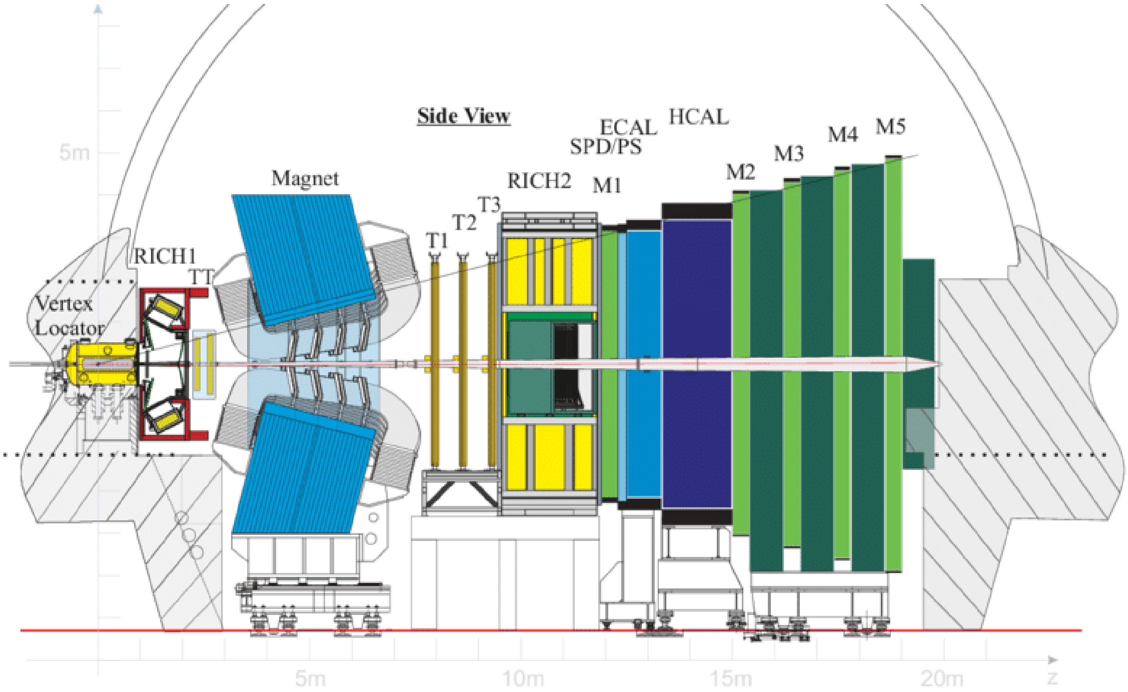


Figure 3.5: A schematic view of the LHCb detector, illustrating the positions of the various subdetectors [48].

LHCb is made up of various subdetectors adopting diverse technologies, in order to track and identify different particles with a range of energies. Figure 3.5 shows the detector elements, which can be divided into two categories based on their main purpose:

- Tracking System: Vertex Locator (VELO), Dipole Magnet, Tracker Turicensis (TT), and three tracking stations (T1-T3), which are divided into Inner Tracker (IT) and Outer Tracker (OT).

- Particle Identification (PID) system: Two Ring Imaging Cherenkov detectors (RICH1 and RICH2), Electromagnetic calorimeter (SPD+PS+ECAL), Hadronic calorimeter (HCAL) and five muon stations (M1-M5).

The tracking system allows reconstruction of charged particles passing through the magnetic field.

3.2.3 The LHCb Magnet

The LHCb magnet is a dipole, made up of two 25 tonne saddle shaped coils, inside a 1450 tonne iron frame [49]. Each of the coils is constructed from 10 'pancakes', made out of almost 3,000 m of aluminium cable. The high precision necessary for the tracking detectors demand an integrated magnetic field of 4 Tm for tracks originating near the primary interaction point. The magnetic field is aligned with the y-axis and covers the full LHCb acceptance, with a free aperture of ±300 mrad horizontally and ±250 mrad vertically. Since one of the primary goals of LHCb is to study CP violation, the polarity of the magnet can be switched, providing reduced systematic uncertainties for these asymmetry studies. The behaviour of a particle in a magnetic field provides information not only about the charge of a particle (from the direction of curvature), but also its momentum (from the radius of curvature).

3.2.4 Vertex Locator

Displaced secondary vertices are a distinctive feature of $b-$hadron decays, as the particles are long-lived meaning that they fly a significant distance before decaying. Since the VELO is the first subdetector, surrounding the interaction region, it must provide an accurate measurement of the particle coordinates in this area, which are used in the LHCb trigger to reconstruct the PV and IP defined earlier in Sect. 3.2.1. Precision measurements of the primary and displaced secondary vertices can be achieved via the accurate reconstruction of tracks in the region immediately surrounding the interaction point. Within the VELO, a series of silicon modules spaced along the beam axis are used, positioned 8mm away from the beam to detect almost all of the tracks originating from the interaction point, as shown in Fig. 3.6. The silicon technology provides good spatial resolution and a strong resistance to radiation. In order to allow fast reconstruction of the tracks in the LHCb trigger system the modules have cylindrical geometry, with each module separated into two halves, and each half equipped with 300 μm thick R and ϕ sensors. The R sensors measure the r-coordinate of the hit (the distance from the z axis) using circular strips split into four segments around the beam axis. The ϕ sensors have radial strips to measure the azimuthal coordinate. The modules on each side of the beam are staggered by 1.5cm in z, meaning that when the VELO is fully closed there is a slight overlap to ensure no acceptance holes and improve alignment.

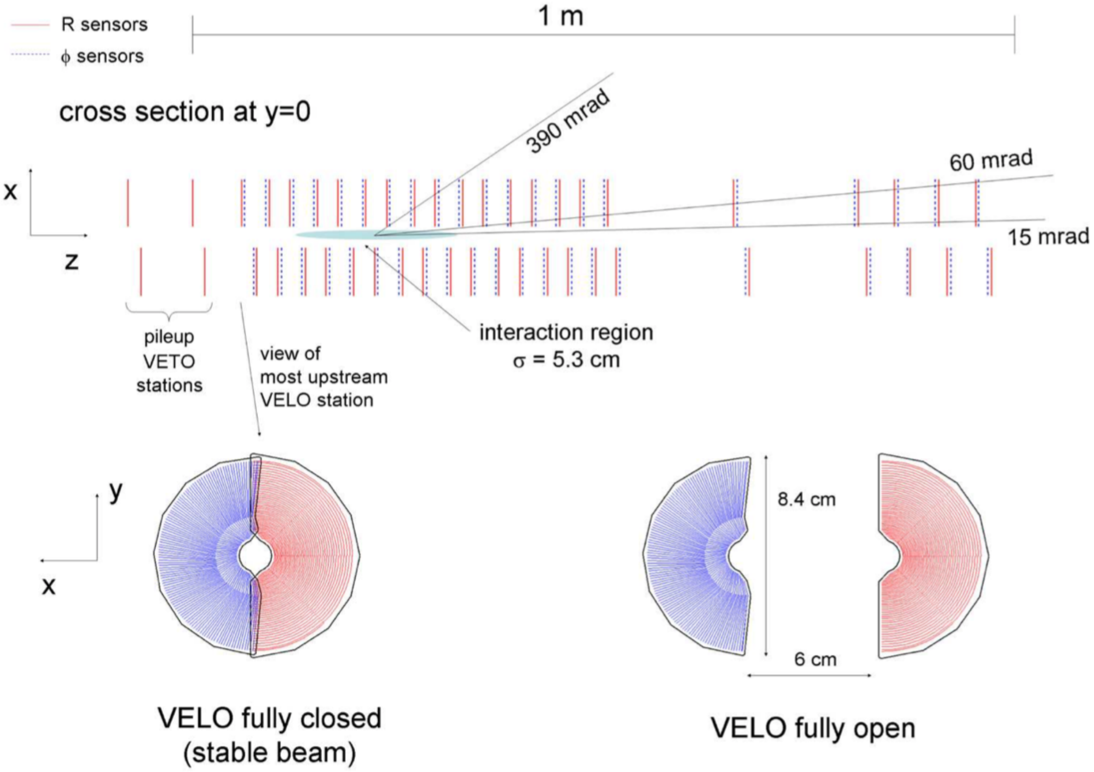


Figure 3.6: The set-up of the VELO silicon modules along the beam line (top), and the layout of the r (red) and ϕ (blue) sensors in closed (left) and open (right) positions. Indicated are the average crossing angle for minimum bias events (60 mrad), and the minimal (15 mrad) and maximal (390 mrad) angle for which at least 3 VELO stations are crossed.

There are 23 modules in total on each side of the beam, housed in an 0.5 mm thick aluminium case which separates the beam vacuum from the VELO vacuum, and also shields the detector from the radio frequency (RF) electromagnetic fields generated by the LHC beam. Two of the VELO modules are placed behind the interaction point (in the -z direction), and their purpose is to measure the number of interactions per collision, producing pile-up information for the trigger.

3.2.5 Tracking System

In addition to the VELO, there are a number of other subdetectors comprising the tracking system. These are placed either side of the magnet, and collect information about the particle trajectory of charged particles in order to determine the charge and momentum of the particle. These properties are determined by measuring the direction and magnitude of the particle's deflection in the magnetic field. Silicon microstrip technology is used in the Tracker Turnicensis (TT) and the inner tracker (IT), meaning that they are collectively called the Silicon Tracker (ST). The ST uses a strip pitch of $\sim$200 μm which provides high resolution close to the interaction region. The TT consists of four silicon sensor ladders as shown in Fig. 3.7. Each sensor is 500 μm thick, 9.64 cm by 9.44 cm in area, and carries 512 readout strips with a strip pitch of $\sim$183 μm. The second (u) and third (v) layers are rotated by stereo angles of -5° and +5° respectively, whilst the first and last layers contain vertical (x) strips. The x-u-v-x geometry and dimensions of the TT allow for better track reconstruction and coverage of the full detector acceptance. The TT stands between the VELO and the magnet, while the IT is located at the far side of the magnet at the centre of the tracking stations T1-T3.

The IT plays a key role in providing accurate momentum estimates, which are important for determination of the invariant mass of a particle and decay-time resolutions. The three IT stations consist of four boxes in a cross shape (covering a region approximately 120 cm wide and 40 cm high) around the beam pipe, configured in the same geometry as the TT.

In the IT, each sensor is 7.6 cm by 11 cm in area and has 384 readout strips, with a strip pitch of 198 μm. The modules in the IT are 310 μm thick for single sensor modules or 410 μm thick for two-sensor modules.

The outside of these stations is called the Outer Tracker (OT), which is a drift-time detector covering a large area, using straw-tube drift chambers with 5 mm cell diameter, ideal for the tracking of charged particles over a large acceptance area. The chambers are arranged in the same geometry x-u-v-x as the TT and the IT, with each module containing two layers of drift tubes. The three OT stations are shown in Fig. 3.8. The individual tubes have a diameter of 4.9 mm and are filled with a mixture of Ar (70%), CO_2 (28.5%), and O_2 (1.5%) [50], which guarantees a fast drift time (<50 ns) whilst maintaining good spacial resolution. When a charged particle passes through, the gas becomes ionised and the freed electrons drift to the anode, where an electron avalanche causes a signal current. The delay between the electron being freed and causing the signal can be used to determine the position of the ionisation within the tube, and therefore the position of the detected particle.

The tracking system provides an overall charged particle momentum resolution of $\delta p/p \sim 0.5$ for $p = 20$ GeV/c and $\delta p/p \sim 0.8$ at $p = 100$ GeV/c [47]. This leads to a B mass resolution of $\sim$25 MeV/c^2 for dimuon modes.

3.2.6 Cherenkov detectors (RICH1 and RICH2)

The Ring Imaging Cherenkov (RICH) detectors are used to distinguish particles of different mass at low (RICH1) and high (RICH2) momentum. If a particle passes through the RICH detectors at a speed faster than the speed of light in that medium of the detector (known as the ‘radiator’), then electromagnetic radiation is emitted. This radiation, known as Cherenkov radiation, is emitted in a cone shape with a polar angle θ_C (the Cherenkov

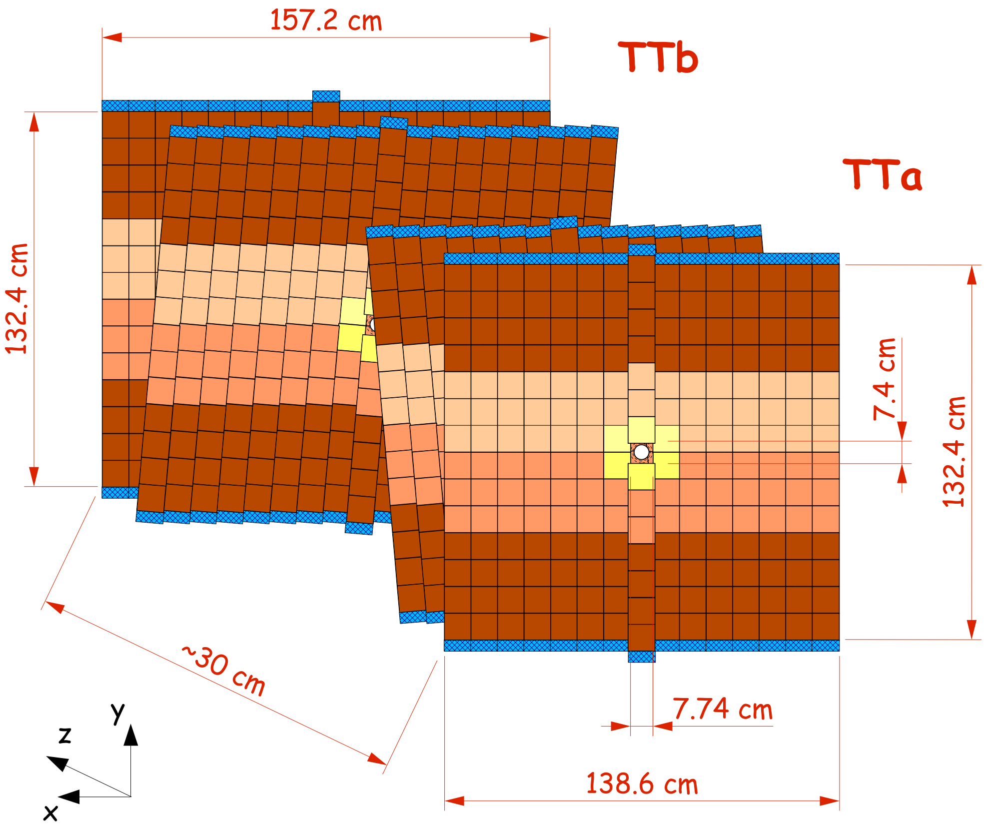


Figure 3.7: Layout of the TT detection layers, in the x-u-v-x configuration.

angle) and is related to the particle velocity by Eq. 3.2.2, where n is the refractive index of the medium and β is the ratio of the speed of the particle (v_P) to the speed of light (c): $\beta = v_P/c$.

$$\cos\theta_C = \frac{1}{n\beta} \tag{3.2.2}$$

The basic principles of the two RICH detectors are the same: in both detectors, the cone of light is reflected by a series of flat mirrors. Hybrid Photon Detectors, placed outside the detector acceptance, detect the circular projection of the reflected cones. A schematic view of the RICH detectors is shown in Fig. 3.9. The RICH1 detector uses aero-gel and C_4F_{10} radiators and covers $1-60\,\mathrm{GeV}/c$ particle momentum range, while the RICH2 detector uses a CF_4 radiator to cover the momentum range $15-150\,\mathrm{GeV}/c$. Combining the particle velocity information from the RICH detectors with the momentum information from the tracking stations, particle masses can be determined, allowing for particle identification. The RICH detectors mainly provide discrimination between charged kaons and pions, but also complement the calorimeters and muon chambers in the identification of muons and electrons [51].

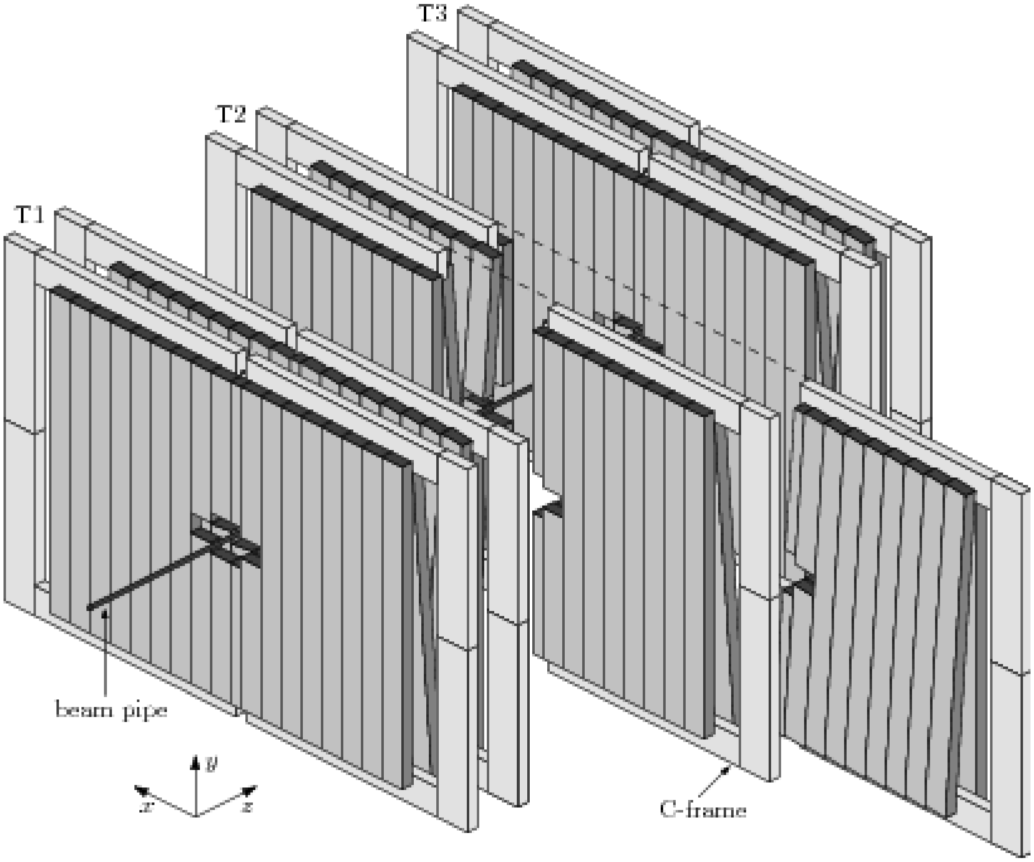


Figure 3.8: The three LHCb Outer Tracker Stations.

3.2.7 Calorimeters

The calorimeter system is designed to measure the position and energy of electrons, photons and hadrons, allowing identification of these particles. This is done by utilising a material which will stop the incoming particle within this part of the detector. The system is used to select events in the LHCb trigger discussed in Sect. 3.2.9.

The Scintillating Pad Detector (SPD) and pre-shower (PS) are positioned in front of the calorimeters, separated by a lead wall of 1.5 cm width, equal to 2.5 X_0. The lead wall causes showers that deposit significant amounts of energy for electrons but not for hadrons, thus distinguishing them from each other. This system helps with background rejection and efficiency of photon detection. The SPD/PS system identifies charged particles, and allows electrons to be separated from photons.

After the SPD/PS there are two calorimeters in LHCb; the ECAL (Electronic calorimeter) and the HCAL (Hadronic calorimeter). Both have a sandwich-like structure, with alternating layers of metal and plastic. Calorimeters make use of particle showering, which occurs when particles hit the metal plates. The secondary particles excite molecules in the plastic layers which emit ultraviolet light, the amount of which is proportional to the energy of the incoming particle. Figure 3.10 illustrates which particles leave signals in the

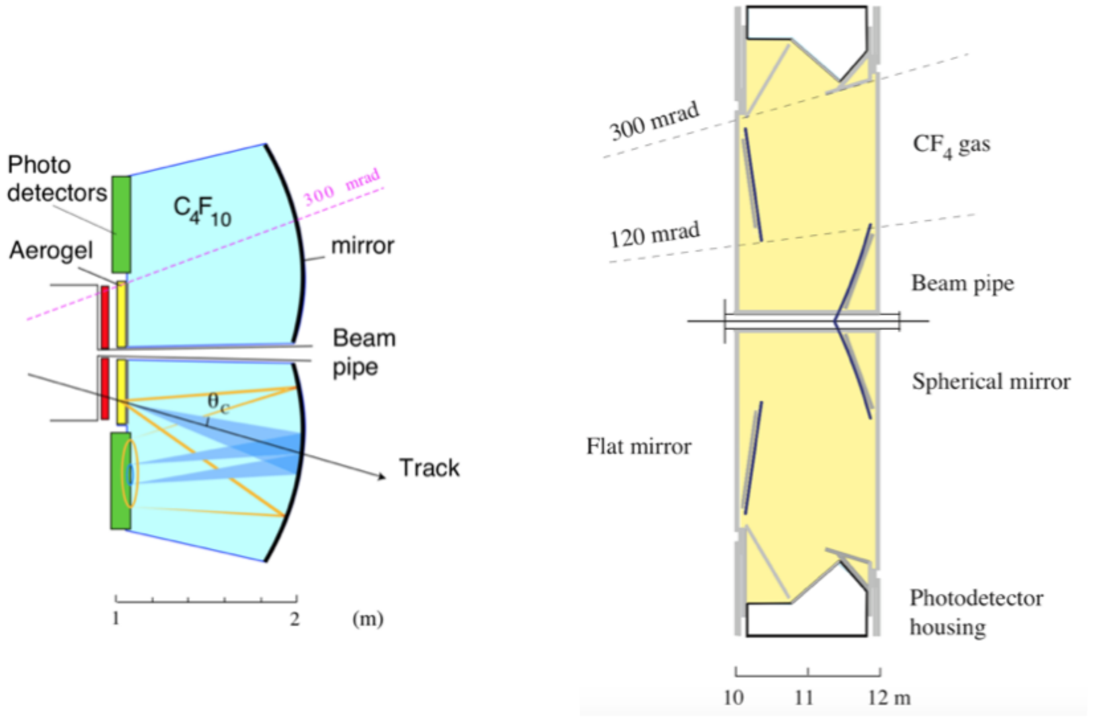


Figure 3.9: Schematics of RICH1 (left) and RICH2 (right), as seen from above from Ref. [52].

different parts of the calorimeter system.

The ECAL is made of alternating layers of 4 mm thick scintillating tiles and 2 mm thick lead plates, with 10-stage photomultipliers used to detect the light that is emitted. Since the hit density varies significantly with the distance from the beam, three different cell sizes are used; 4×4 cm, 6×6 cm and 12×12 cm in the inner, middle and outer parts of the detector respectively. In total there are 3312 modules made of 6016 individual cells, with a total detector dimension of 7.6×6.3 m. The ECAL covers a range of $25 - 300$ mrad in the horizontal plane and $25 - 250$ mrad in the vertical plane. The amount of radiation lengths contained in the ECAL is 25 X_0, which fully contains the shower of highly energetic photons. The ECAL achieves an energy resolution on electromagnetic showers of

$$\delta E/E = 10\%/\sqrt{E} \oplus 1\%. \tag{3.2.3}$$

The HCAL is positioned behind the ECAL, and uses thin iron plates between the scintillating tiles rather than lead. In this system the iron is 16 mm thick and the scintilator is 4 mm thick. As with the ECAL, the cells in the centre of the HCAL are smaller than those on the outside of the detector (13×13 cm and 26×26 cm respectively). The HCAL is a total of 5.6 hadronic interaction lengths deep, which is determined by the space available within the LHCb cavern. Hadrons may start to develop initial showering in the PS/ECAL,

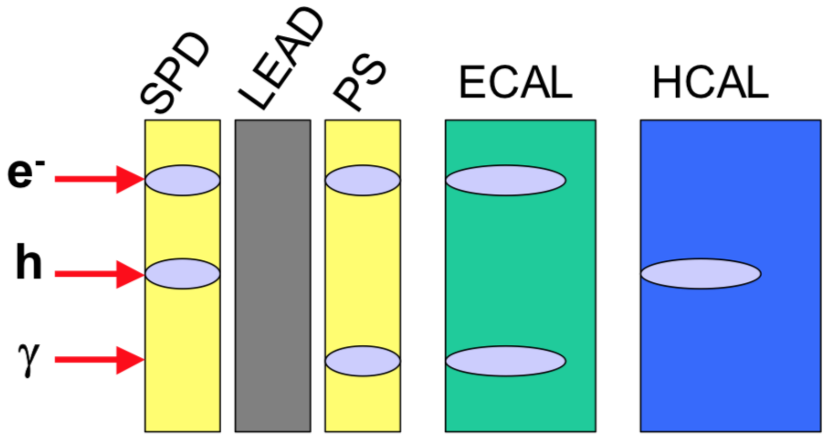


Figure 3.10: Signal deposited on the different parts of the calorimeter by an electron, a hadron, and a photon [53]. In addition to that shown, there are also hadronic showers in the PS and ECAL but the showers are much smaller.

but most of the hadrons energy will be deposited in the HCAL. The energy resolution of the HCAL on hadronic showers is

$$\delta E/E = 80\%/\sqrt{E} \oplus 10\%. \tag{3.2.4}$$

Both of the calorimeters use wavelength-shifting (WLS) fibres to transport scintillation light to a set of photomultipliers. Further details on calorimeter performance can be found in Ref. [54].

3.2.8 The Muon System (M1-M5)

Muons are minimum ionising at the energies produced in the LHC and they do not produce hadronic showers, meaning that they are not stopped by the calorimeter system. This means that they are primarily identified by a set of separate detectors, which allow fast detection of muon candidates with a high efficiency for muon identification, which is of particular significance for the offline selection of rare $B^0_{(s)} \to \mu^+\mu^-$ decays.

The muon system covers a range of 20-306 mrad and 16-256 mrad in the horizontal and vertical planes respectively and is comprised of five stations (M1-M5). The stations are placed along the beam axis as seen in Fig. 3.11, with the first chamber located before the calorimeter and the other four located downstream of the HCAL. This separation of the stations allows for muon momentum measurements to be calculated in the L0 trigger (see Sect. 3.2.9).

Any hadrons, photons and electrons that are not stopped in the calorimeters are absorbed by 80 cm thick iron plates which are positioned between the muon stations, reducing the likelihood that these particles will be misidentified as muons. Each of the stations is

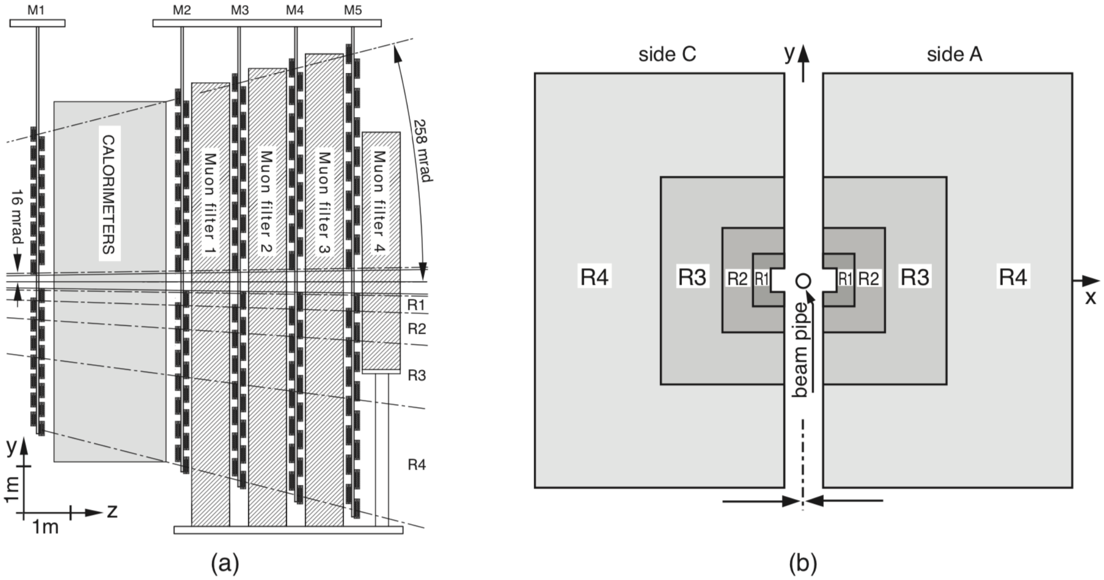


Figure 3.11: (a) Side view of the LHCb Muon Detector. (b) Muon station layout with the four regions R1–R4 [55].

divided into four regions (R1-R4), as seen in Fig. 3.11, with each region at an increased distance from the beam axis. The granularity of the detector is finest in R1, and the linear dimensions of the regions is such that the channel occupancy and particle flux are approximately the same over all four regions. All five of the muon stations, (with the exception of the inner region of M1) use multi-wire proportional chambers (MWPCs) which contain 5 mm gas gaps. The innermost part of M1 is where the occupancy is highest, so triple-GEM (gas electron multiplier) detectors are used here. The triple-GEM detectors comprise of three GEM foils separated by a gas mixture of Ar, CO_2 and CF_4. The muon chambers are positioned in such a way as to provide a high acceptance of high momentum particles coming from the interaction region. The chambers provide space-point measurements, with the detector being split into rectangular 'pads' defined in the x, y plane. The granularity of the detector, which determines the resolution of the x and y coordinates, depends on these pads. The pad size is smaller for the inner regions of the detector, and larger for the outer regions since the resolution is dominated by multiple scattering in this region anyway. The distribution of these pads is shown in Fig. 3.12. The number of pads in the y direction is the same for each of the muon stations M1-M5, while the number in the x direction follows the ratio 2:4:4:1:1 for M1-M5 respectively.

3.2.9 Trigger

The trigger is a key component of the detector system for rejecting background events and selecting signal. At LHCb, the trigger has two stages; a 'Level-0' hardware trigger (L0), and a 'High Level' software trigger (HLT). The trigger can be adapted as run conditions change,

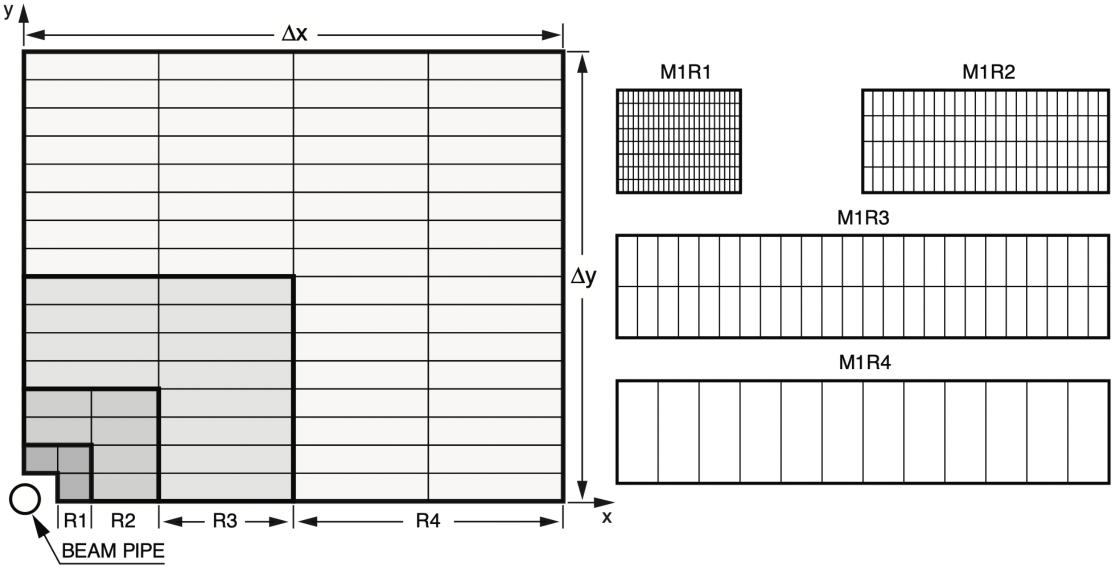


Figure 3.12: The image on the left shows a quadrant of a muon station with the regions R1-R4 highlighted in different shades. The pad layout is shown in each of R1-R4 as rectangular regions. The image on the right shows a comparison of the pad sizes for each of the regions within M1. Image from Ref. [56].

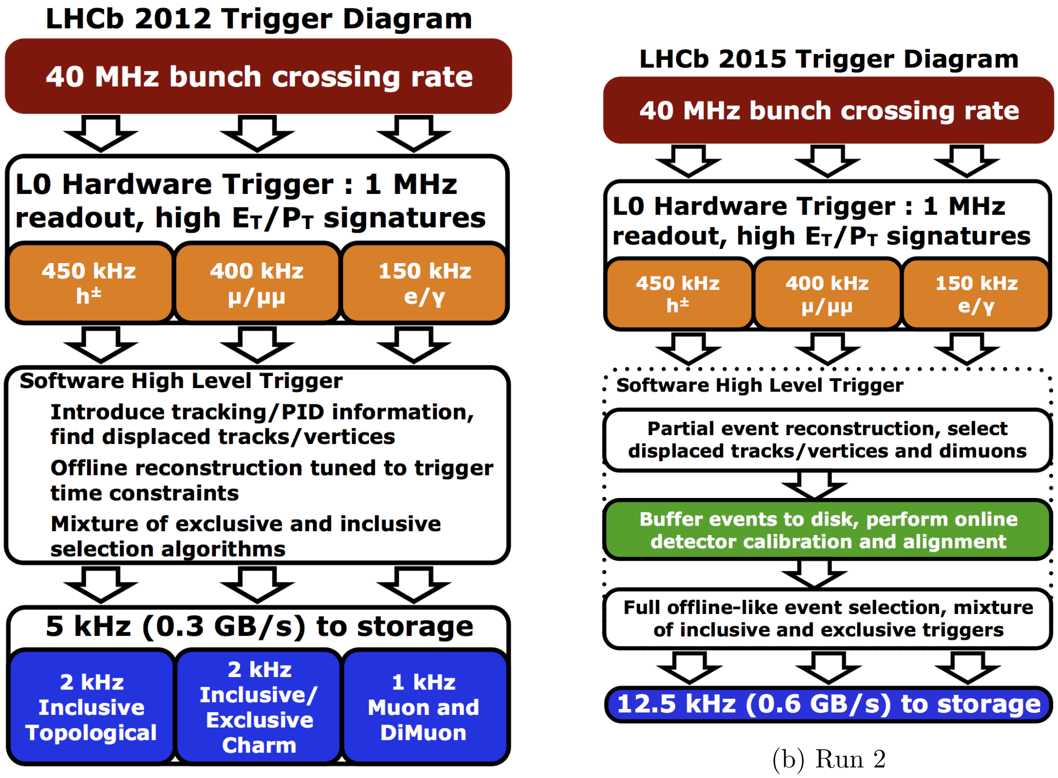
and uses a variety of techniques for event classification including the use of multivariate classifiers to identify b-hadrons. An overview of the trigger system is given in Fig 3.13 for Run 1 (a) and Run 2 (b).

Level-0

The first stage of the trigger system, L0, makes a decision about whether or not to store an event in under 4 μs and reduces the event rate from ∼13 MHz to 1 MHz, which can then be read out of the LHCb detector. In this time, the L0 trigger makes a decision for each of the three independent subtriggers; L0-Muon which processes hits from the five muon stations, L0-Calorimeter which uses information from the SPD, PS, ECAL and HCAL, and finally L0-PileUp. If any of these three decisions is positive, then the event will be passed to the next stage of the trigger, whilst those with all negative decisions will not be stored.

Events with either high p_T muons are selected by the muon trigger, with either a single muon with $p_T > 1.76\,\text{GeV}/c$ or a pair of muons with $p_{T1} \times p_{T2} > (1.6\,\text{GeV}/c)^2$. The relative resolution for the p_T estimate for muons reconstructed using exclusively information from the muon chambers is 20%. The output rate of the L0 muon triggers is about 400 kHz.

The calorimeter trigger selects events with large transverse energy (E_T) deposits in the calorimeter. For hadrons, this is events with $E_T > 3.68\,\text{GeV}$ deposited in the HCAL and for electrons and photons this is $E_T > 3\,\text{GeV}$ deposited in the ECAL. The L0-Calorimeter trigger uses this transverse energy information to build three candidates; L0Hadron (hadron



(a) Run 1

(b) Run 2

Figure 3.13: Overview of the LHCb trigger stages [57]

candidate), L0Photon (photon candidate), and L0Electron (electron candidate).

The output rates of the hadron and electromagnetic triggers are 450 kHz and 150 kHz respectively. The L0-Pileup is an additional veto which uses information from the VELO and SPD to exclude bunch-crossings containing multiple collisions, retaining those containing single collisions.

HLT1

The HLT is a software based trigger which runs C++ algorithms on a farm of multiprocessor PCs called Event Filter Farm (EFF). The aim of the HLT is to reduce the event rate down to a few kHz. The full offline reconstruction takes a total of 2 s per event, but due to restrictions of the computing (the EFF allows only 30 ms per event), the HLT must be simplified to meet the time requirements.

The first stage of the High-Level trigger (HLT1) performs a partial event reconstruction on all events accepted by L0 and is based on VELO reconstruction software. At the start of each LHC fill, the mean position of the proton-proton collisions within the VELO is determined from track reconstruction. This quantity, the PV^{mean}, remains stable within a few μm throughout the fill. The tracks in a 300 μm radius that have either a large IP or can be matched to hits in the muon chambers are extrapolated into the main tracker. The selected track segments are extended in the tracking stations, by searching for hits consistent with a high p_T track.

For events triggered by the L0 muon line, there is a fast reconstruction which extrapolates VELO tracks in a straight line to M3, defining a search window in this subdetector. The information from the VELO and M3 is then used to search for hits in the other muon stations, where a candidate is accepted if there is a hit in at least one other muon station.

If a good quality track with a $p_T > 1.6\,\mathrm{GeV}/c$ ($1\,\mathrm{GeV}/c$ for muon and $0.5\,\mathrm{GeV}/c$ for dimuon tracks) can be reconstructed, then the event is accepted and processing continues in the next stage.

HLT2

Events selected by HLT1 are passed on to the second stage of the HLT (HLT2), which performs a full offline reconstruction of all tracks with $p_T > 300\,\mathrm{MeV}/c$. The trigger lines in HLT2 are more specific than HLT1. The HLT2 trigger lines are classified as either *inclusive* or *exclusive*. Exclusive lines require all final state particles for specific decay to be reconstructed, whilst inclusive lines trigger on all b-hadron decays with a displaced vertex and at least two charged particles in the final state. In general there are three types of trigger lines used in LHCb: generic beauty triggers, muonic triggers and charm triggers. The performance and efficiency of the different trigger lines depend on the particular process being studied. The trigger efficiencies for the $B^0_{(s)} \to \mu^+\mu^-$ analysis are very high; the trigger requirements will be discussed in Sect. 6.4.

3.2.10 Data Acquisition

The role of the Data Acquisition (DAQ) system is to collect non-suppressed data in the front end electronics, and assemble complete events, in order for this information to be reduced by further selection. The DAQ system comprises of Readout Units, a Readout Network, Sub-Farm controllers, and a CPU Farm [58]. The first components in the system, the Readout Units, act as a Front-end multiplexer. The modules are used as an interface to the Readout Network, which assembles all of the event fragments coming from one bunch crossing in one place. There is then a second interface, the Sub-Farm controllers, which come between this network and the processor farm which executes the high level trigger algorithms, such as the HLT2 lines mentioned in the last section. Figure 3.14 shows a schematic of the integrated LHCb Trigger and Data Acquisition systems.

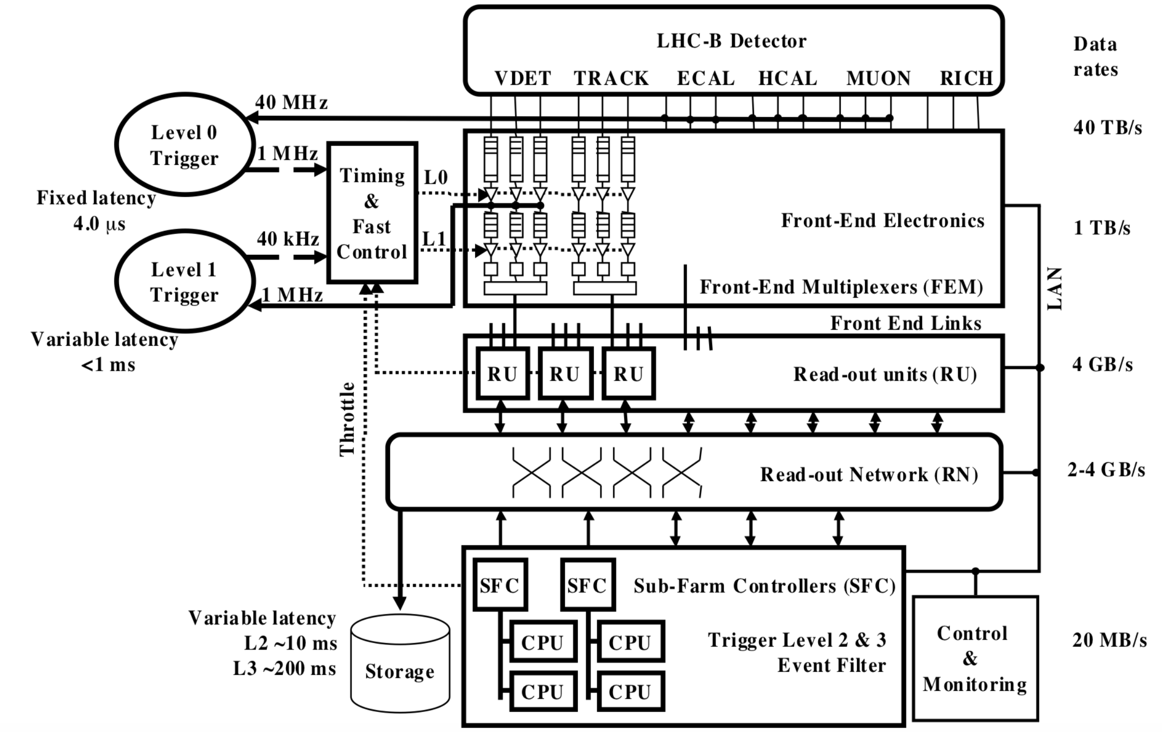


Figure 3.14: Schematic diagram of the DAQ architecture integrated with the Trigger system for the LHCb experiment [58].

Data Quality Shifts

Once the data is collected, it must be checked to ensure that it is consistent with expectation and nothing has gone wrong with any of the subdetectors during a run. The data quality monitoring is summarised here, and described in detail in Ref. [59].

The most valuable method of assessing data is visual inspection. Data quality monitoring (DQM), is an important step for high-energy physics experiments, and ensures that the experimental apparatus is functioning correctly during data taking. The checking of the data is the responsibility of research scientists, operating on a rotating shifts schedule. The scientist on shift is known as the Data Quality (DQ) Shifter, who compares the data collected for each part of the detector with existing performance histograms.

LHCb software Prompt DQ analyzes a subset of events that pass HLT2, a sample that is typically is enriched in Λ^0, D^0, D^+, D_s and J/ψ. When 93% of the HLT1 output files of a run have been processed by HLT2, the run is added to a small database where it will eventually be marked as 'Done', or 'Failed' in the case that the HLT2 lines cannot be found. When a run is marked as 'Done', a ROOT [60] output file is created and the run is added to a second data quality database (DQDB).

Through Monet [59], the DQ Shifter can check histograms from that specific run to ensure that they line up with expectation. At this point, they can either be marked as 'OK', in which case they will be used for physics analysis, or 'BAD', in which case they will not. The flagging decision is made by the DQ Shifter. If they think a decision cannot be reached, for example the statistics of the run are too small, the run is left UNCHECKED in the bookkeeping but flagged UNKNOWN in the DQDB. Figures 3.15 and 3.16 show examples of some of the performance histograms from the muon detectors, since they are of particular relevance to this analysis. Tag-and-probe muons are used to estimate the particle identification and misidentification rates later in the analysis.

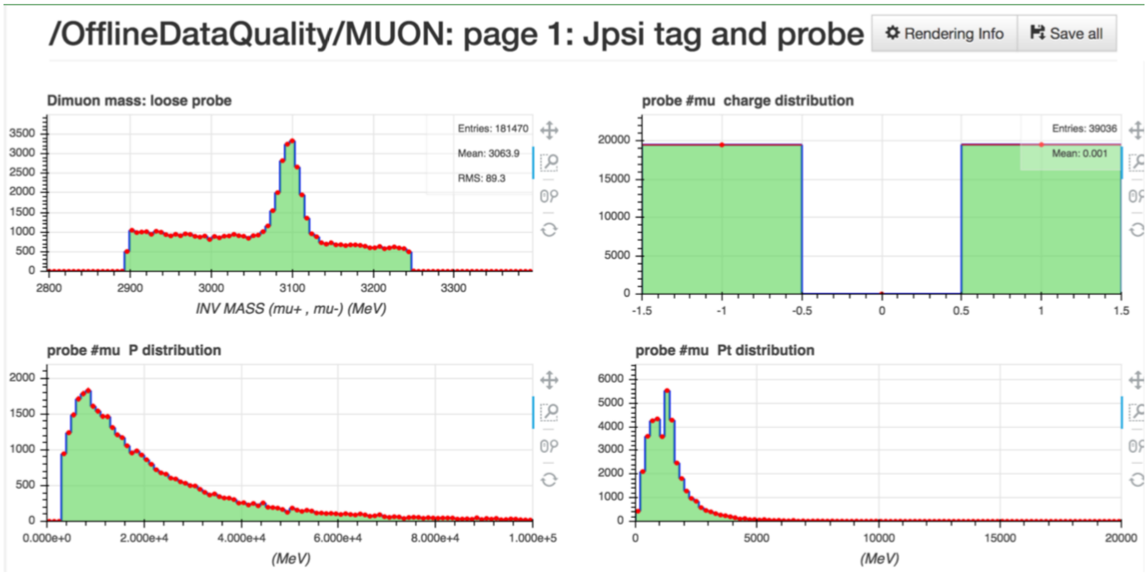


Figure 3.15: Data Quality plots for Muon detectors for probe Muon from J/ψ decays.

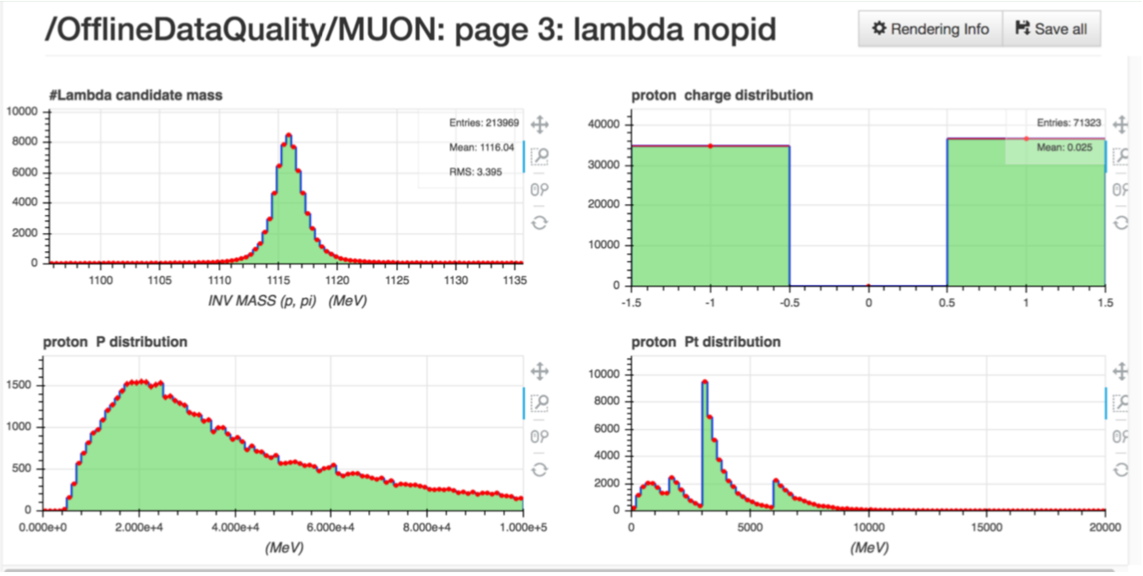


Figure 3.16: Data Quality plots for Muon detectors for Protons from Λ decays.

Chapter 4

Event Reconstruction and Particle Identification

In this chapter, the LHCb data flow will be described; how the hits in the detectors are turned into usable information for physics analysis. The general data flow is discussed first, followed by a more detailed description of how tracks, and primary and secondary vertices are reconstructed from the information provided by the subdetectors. Once the tracks are reconstructed, the particles can be identified and their properties determined. The LHCb particle identification procedure will be discussed in Sect. 4.3, with particular emphasis on muon identification due to its significance in the $B^0_{(s)} \to \mu^+\mu^-$ analysis.

4.1 LHCb Data flow

The collisions recorded by the LHCb detector pass through a specific 'data flow', which is designed to maximise the quality of the data and the data taking efficiency.

BRUNEL [61] is an application based on the GAUDI [62] software framework, that is used for the reconstruction of tracks and vertices. Once the data has been reconstructed by the BRUNEL software into useful objects, it is stored in an output 'DST' file. The data in the file is then filtered further through certain selection criteria called the *stripping*. The stripping involves applying general cuts that reduce the size of the dataset by only keeping those events that look interesting for physics analysis. This is done by the DaVinci [63] application, which writes out data either in the DST or μDST (micro-DST) format. These output files are put into groups known as streams, which contain similar selections. This saves time for the physicists analysing the data as they can look only at the selections relevant to their analysis. For the $B^0_{(s)} \to \mu^+\mu^-$ analysis, for example, the group of interest for the signal would be the dimuon stream. The DaVinci application is used to create datasets, or 'ntuples', containing the variables that are relevant for a specific analysis.

The data flow changed between Run 1 and Run 2, as the increase in energy at the

LHC provided an increase in particle production and a higher rate of interesting events. Due to limited computing resources, a method had to be introduced that would filter out the uninteresting events as soon as possible. To achieve this, the reconstruction went from being offline, to being performed in the trigger, known as online reconstruction. This became possible for Run 2 after major improvements to the online reconstruction software.

As such, in 2015, a Turbo stream was introduced which selected candidates solely in HLT2 and saved them directly to disk to be used in physics analysis. Any tracks or hits that don't form part of the triggered decay are not saved, meaning that it is not possible to re-reconstruct this stream, as the information that would be needed to do this is not saved. An overview of the LHCb data flow is shown in Fig. 4.1, with the Turbo Run 2 part highlighted in red.

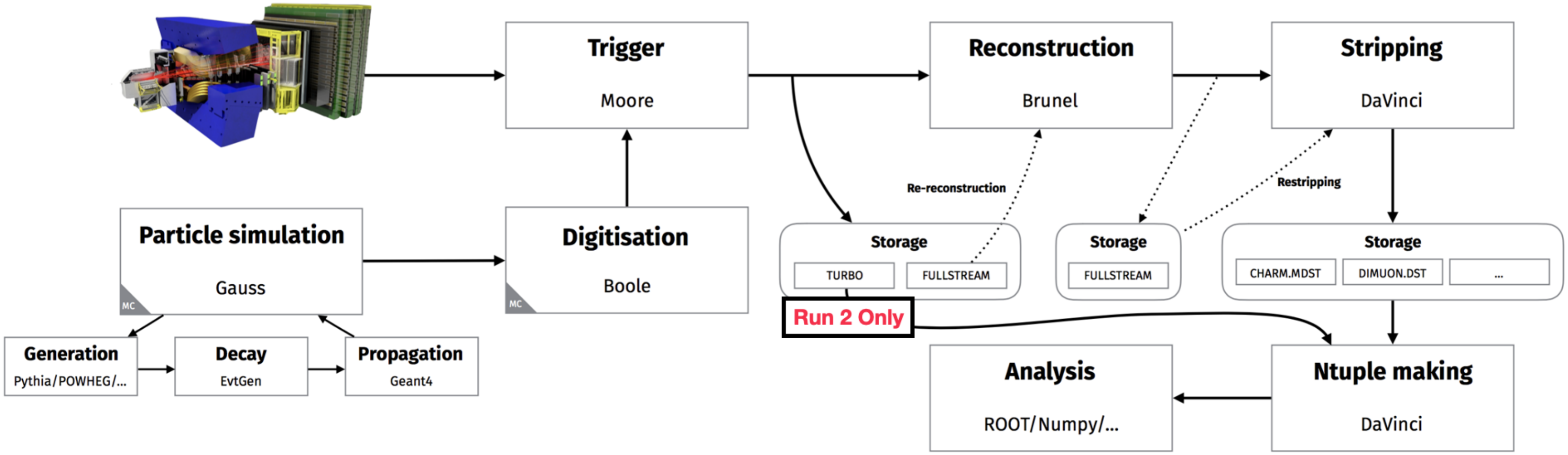


Figure 4.1: The LHCb data flow. The Turbo stream was only introduced in Run 2 (as highlighted in red).

4.2 Event Reconstruction

Only events passing the HLT1 and HLT2 triggers (described in Sect. 3.2.9) are stored for physics analysis. The information from these 'triggered' events is stored at a rate of 500MB/s, and includes the raw hit data from the subdetector cells as well as the trigger decisions (which particular trigger lines were passed for each event).

For Run 1, offline computer algorithms were used to reconstruct the particle tracks (the trajectory of a particle through the detector) from this stored information. These tracks could be extrapolated back to a vertex, which represented some source of particle production in space. For Run 2 the reconstruction was done within the trigger. Information regarding the *pp* collision position (primary vertex) is also stored with the other trigger information.

4.2.1 Track Reconstruction

The Brunel application reconstructs objects such as tracks from the raw data. The reconstruction algorithm aims to find all tracks in an event by combining hits from the VELO, the TT, the IT and the OT subdetectors. The reconstruction begins with a search for an initial track candidate in the VELO region and the T stations. A trajectory is formed for each event using a Kalman filter [64] which accounts for multiple scattering and energy loss. A track will only be reconstructed if a minimum number of hits in the relevant subdetectors is met. In LHCb, tracks are classified into five groups according to their hits in the tracking stations:

- Velo tracks: Hits only in the Vertex Locator (VELO),
- Ttrack tracks: Hits only in the T seeding stations,
- Long tracks: Hits traversing the detector, from the VELO to the T stations. They can have hits in all tracking stations,
- Upstream tracks: Hits in the VELO and the Trigger Tracker (TT),
- Downstream tracks: Hits in the TT and T seeding stations.

Figure 4.2 shows a schematic view of these five types of tracks within the LHCb tracking system.

4.2.2 Vertex reconstruction

The Primary Vertex (PV) is the point at which the proton-proton collision happens, while the Secondary Vertex (SV) is the point at which the resultant particle (for example a B meson) decays into the final state particles. Figure 4.3 provides a sketched example of the extrapolated tracks and vertices, all within the VELO subdetector. Reconstruction of

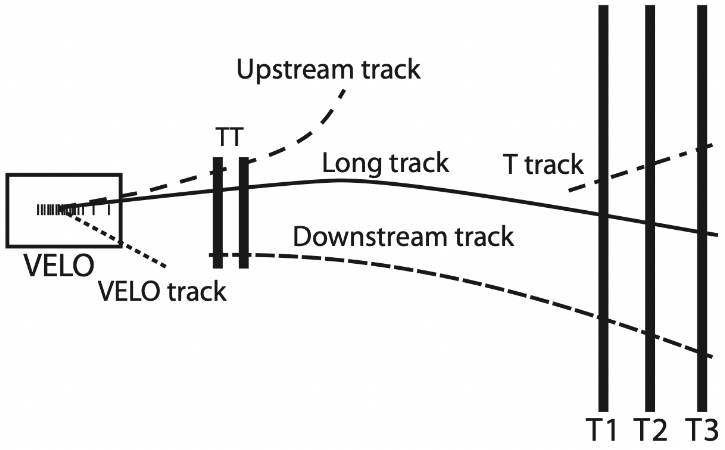


Figure 4.2: Schematic view of the track types in the LHCb tracking system: long, upstream, downstream, VELO and T tracks, from Ref. [65].

the PV must happen first, as this provides criteria for selection of the correct SV for the decay of interest. The reconstruction of the PV is performed in two steps: seeding and fitting, which will be described briefly in the next sections (see Ref. [66] for a more detailed discussion).

Primary Vertex

'Seeding' requires finding space points at which the extrapolated particle trajectories accumulate along the z axis (or in the luminous region). This provides PV candidates with a sufficient number of tracks passing close to one another. The seeding process is performed by evaluating each track individually, and determining the number of other 'close tracks' to that one. A track is determined to be close to another if the distance between the closest points of the two tracks is less than 1 mm. Space points passing these criteria, or *seeds*, are determined to be a PV candidate if at least four close tracks are found. The seeds are then sorted according to multiplicity, with the highest multiplicity candidate being reconstructed first. Using this method means that the number of SVs incorrectly reconstructed as PVs is reduced.

A PV fit is then performed for every seed. Every successfully fitted vertex is required to be spatially separated from any previously reconstructed ones. The PV position is determined using the weighted least square method. When at least 5 tracks (including at least one backward track), are assigned to the PV candidate, the true PV position is determined. The requirement of at least one backwards track reduces the probability of mistaking a SV for a PV.

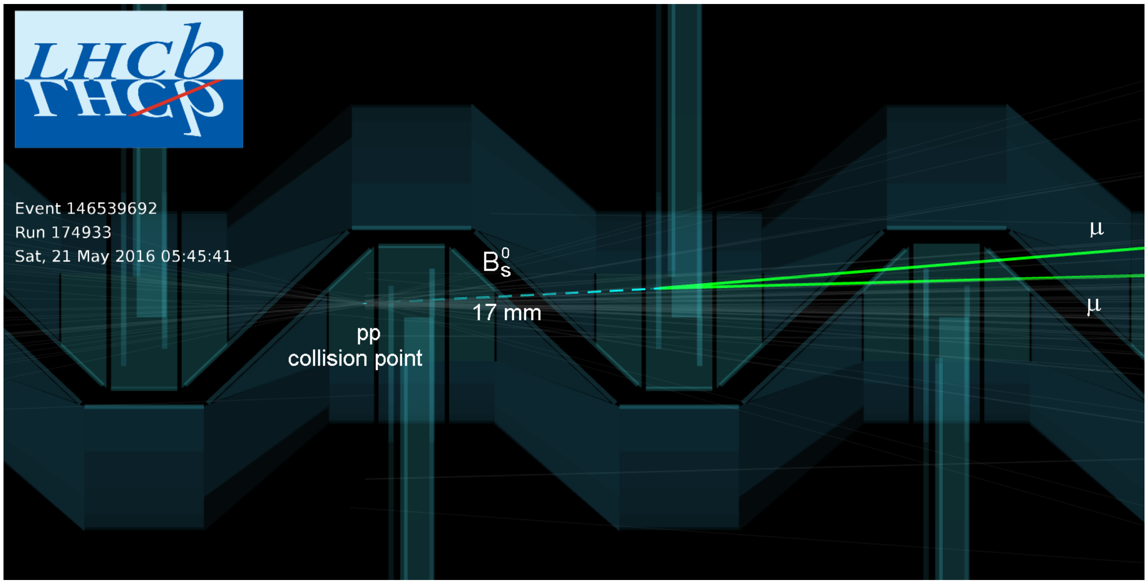


Figure 4.3: An event display showing primary and secondary vertices of a $B_s^0 \to \mu^+\mu^-$ decay from Ref. [67]. The PV is the *pp* collision point, and the SV is the point where the B meson track (dashed green) ends and the muon tracks (solid green) begin.

Secondary Vertex

Based on the selected PV, tracks within the VELO can then be selected as originating from an SV. The relevant SV is analysis dependant, as it is the vertex at which a certain particle from the PV decays into other particles, and there might be several SVs for each PV. The reconstruction technique is similar to that for the PV, but fewer tracks are required to converge at the space point, since there could be as few as two particles originating here (as is the case with the $B^0_{(s)} \to \mu^+\mu^-$ analysis). The relevant SV for this analysis is the vertex at which the B_s^0 meson decays into a μ^+ and a μ^-. For cases involving B mesons such as this, the SV is required to be separated from the PV to allow a certain flight distance of the B meson. The SV will be the point at which the extrapolated muon tracks come together. The particular selection criteria for this analysis are discussed in Sect. 5.3.

4.3 Particle Identification at LHCb

A high muon identification efficiency is crucial for the $B^0_{(s)} \to \mu^+\mu^-$ analysis and many others within LHCb. However, as well as correctly identifying the real muons, particles ‘faking’ muons must also be successfully rejected. Hadrons for example can be misidentified (misID) as muons, with pion misID providing one of the major sources of background for this analysis. Since the $B^0_{(s)} \to \mu^+\mu^-$ decays are extremely rare, hadron misID probabilities must be kept at a very low level to allow accurate discrimination of signal from backgrounds with a similar topology.

Each proton-proton collision within the LHCb detector will produce approximately 100 particles within the geometric acceptance range of $2 < \eta < 5$. In $b\bar{b}$ events, the majority of the charged particles produced are pions (∼75%), followed by kaons (∼15%), electrons (∼6%), and protons (∼4%) [68]. Muons make up less that 1% of the charged particle tracks passing through the full tracking system.

Several subdetectors provide information contributing to the identification of these particles. Charged hadron identification is achieved with two RICH detectors, which distinguish between kaons, pions and protons. The RICH system also provides some separation between leptons and hadrons, which is used to improve the overall particle identification performance. The calorimeter system (ECAL, HCAL) identifies electrons, photons and neutral pions. The muons system, as the name suggests, identifies muons. It provides information for the selection of high p_{T} muons at the trigger level and for the offline muon identification. The muon detection strategy and performance, discussed in Ref. [69], is summarised in the next section.

A schematic view of particles passing through each stage of the Particle Identification (PID) system is shown in Fig. 4.4. The PID strategies relevant for the signal and backgrounds of the $B^0_{(s)} \to \mu^+\mu^-$ analysis are presented in the following section.

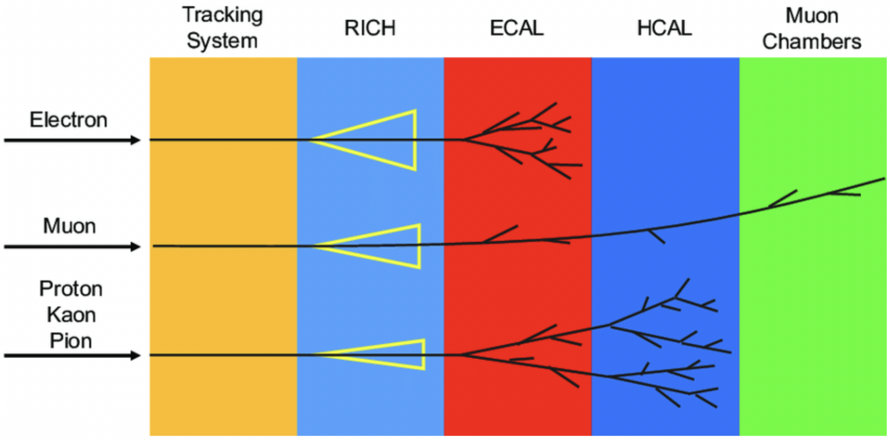


Figure 4.4: Illustration of different charged particle type responses in the LHCb subdetectors. Image from Ref. [70].

4.3.1 Muon Identification

Muon identification in LHCb relies on four key pieces of information which discriminates muons from other particles:

- **Muon mass** - The particle momentum combined with the Cherenkov radiation angle (defined in Eq. 3.2.2) measured in the RICH detectors can give an estimate of the mass of the muon candidate. This enables a basic differentiation between muons and heavier particles, such as protons and kaons.

- **Muon penetration power** - Charged particles lose energy via different means as they traverse the detector. Light particles, like the electron, mainly lose their energy via bremsstrahlung radiation (radiation of photons in material). Electrons will shower and eventually be stopped within the ECAL. However, since bremsstrahlung is inversely proportional to the mass of the particle, the heavier muons are less affected. Muons are minimum ionising particles in the momentum range $0.5 < p < 50\,\text{GeV}/c$ [71]. The amount of ionisation energy lost by a particle traversing a material ($\beta\gamma$) is defined as:
$$\beta\gamma = \frac{p}{Mc}, \tag{4.3.1}$$
where p is the particle momentum, M is the particle mass and c is the speed of light. An electron, being ∼200 times lighter than a muon, has $\beta\gamma$ around 200 times larger for the same momentum. Therefore electrons have much bigger ionisation losses in the same momentum range, and muons can therefore travel much further than electrons before being stopped. Since hadrons are also more massive, they do not lose as much energy to bremsstrahlung either. However, unlike muons they interact strongly with nuclei, meaning that for the most part they are stopped within the HCAL following hadronic showering. Muons therefore have high penetration power compared to most other particles; they can travel right to the end of the detector without much interruption.

- **Hit-pattern in the muon stations** - Using hits in the tracking detectors, track predictions within the muon stations can be extrapolated. These track predictions provide an area in the muon stations in which the muon hits should lie. This allows differentiation between muons that originate at the interaction point and those that are the decay products of particles such as kaons or pions that decay in flight. If the hits in the muon stations are in line with the predictions then this is a good indication that the muons were formed in the interaction region. Muons coming from decays in flight will generally exhibit a spread of hits larger than the true muon tracks. Tracks containing a hit that is shared with another track can be rejected as false candidates.

- **Calorimeter information** - The information from the ECAL and HCAL can be used to reject hadrons and electrons, reducing the misID rates. Real muons can be identified by a small energy loss in the calorimeters close to the extrapolated muon track, whilst hadrons and electrons deposit energy over a wider area predominately, in the HCAL and ECAL respectively.

All of this information can be used and combined to make decisions on the identification of a given candidate, using computer algorithms. LHCb uses a few different algorithms which are applied to every track to identify muons.

For an initial loose selection, an algorithm called ISMUON is used. ISMUON is a selection tool for muon candidates, which gives a binary output for each candidate deciding if it

Table 4.1: ISMUON hit requirement for low, medium and high momentum muon candidates.

Candidate momentum, p	Required muon station hits
$3 < p < 6$ GeV/c	M2 & M3
$6 < p < 10$ GeV/c	M2 & M3 & [M4 or M5]
$p > 10$ GeV/c	M2 & M3 & M4 & M5

could be a muon (1) or not (0). The ISMUON algorithm is based on the penetration of the muons through the calorimeters and iron filters of the muon stations. More specifically, it requires muon station hits in a defined 'Field Of Interest' (FOI) which is identified using the extrapolated muon candidate track. This FOI is dependant on the candidate momentum and is parameterized separately for the 4 regions (R1-R4) of stations M2-M5, in both x and y directions. The candidate momentum also defines how many muon station hits (and in which particular stations) are necessary for a candidate to receive a positive muon decision. The hit requirement for each candidate momentum range is summarised in Table 4.1. ISMUON is very effective at identifying muons, and reduces the misID probability of hadrons to the percent level.

The candidates with a positive ISMUON outcome can be further selected using the other muon identification strategies. From these differentiation methods, a candidate can be given a muon 'likelihood', being the probability that a candidate is actually a real muon. Non-muon likelihoods are also computed; the probability that a candidate is a particle other than a muon. In the $B^0_{(s)} \to \mu^+\mu^-$ analysis, it is important that protons, kaons and pions are not misidentified as muons. The muon likelihood arising from the hit-pattern in the muon stations is computed using the distribution of the average hit distance significance squared (D^2), defined as,

$$D^2 = \frac{1}{N}\sum_i \left\{ \left(\frac{x^i_{\text{closest}} - x^i_{\text{track}}}{\text{pad}^i_x} \right)^2 + \left(\frac{y^i_{\text{closest}} - y^i_{\text{track}}}{\text{pad}^i_y} \right)^2 \right\}, \tag{4.3.2}$$

where i iterates over the stations that have hits in the FOI, up to the total, N. The values x, y_{closest} refer to the coordinates of the closest hit to the track extrapolation coordinates x, y_{track} for each of the stations. $\text{pad}^i_{x,y}$ is half the pad size in the x and y planes (with the pads being predefined regions in the detector, see Sect. 3.2.8).

A narrow D^2 distribution is indicative of a real muon, while other particles which have incorrectly been passed by the ISMUON algorithm will have a wider distribution. The muon likelihood for each track is calculated by integrating the muon D^2 probability density function up to the measured D value of the candidate. The distribution is created using a sample of muons from $J/\psi \to \mu^+\mu^-$ decays.

For the non-muon likelihoods, the D^2 distribution of protons from $\Lambda^0 \to p\pi^-$ decays is used. Most protons will be stopped in the HCAL, however, there are a few reasons why protons might provide hits in the muon stations that are around the track extrapolation

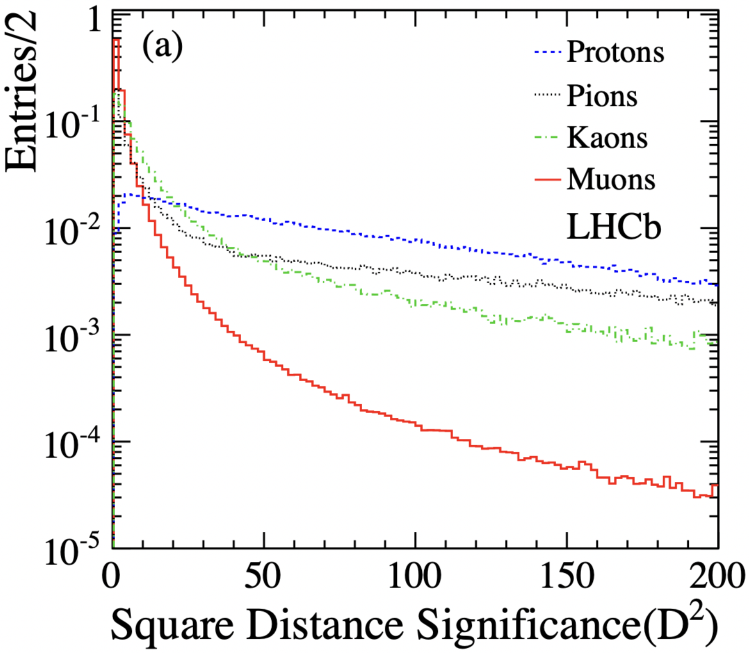


Figure 4.5: Average Distance Significance (D^2) distributions for protons, kaons, pions and muons, from Ref. [69].

region:

1. the hits are random,
2. the proton is in line with a true muon pointing in the same direction,
3. they are punch-through or sneak-through protons[1].

The D^2 distributions for protons, kaons, pions and muons are shown in Fig. 4.5.

4.3.2 Hadron Identification

Protons, kaons and charged pions are identified in the RICH detectors by an algorithm that again outputs likelihoods for each particle hypothesis. The algorithm matches the pattern of hits in the photodetectors to expected patterns of each particle [72].

In order to calculate the likelihood of a candidate to be a certain particle, each of the pixels displaying hits are matched to a track that could have produced the hit via Cherenkov radiation. The hit patterns take the form of rings, and the emission angle can

---

[1] A punch-through proton is one that passes through the calorimeters without depositing all of its energy. This could be due to very highly energetic initial particles, or simply that the proton kept most of the momentum in the interaction. Sneak through protons occur when there are fewer than normal number of proton/nucleus interactions, and a protons might ‘sneak through’ the detector by finding a gap in the calorimeters.

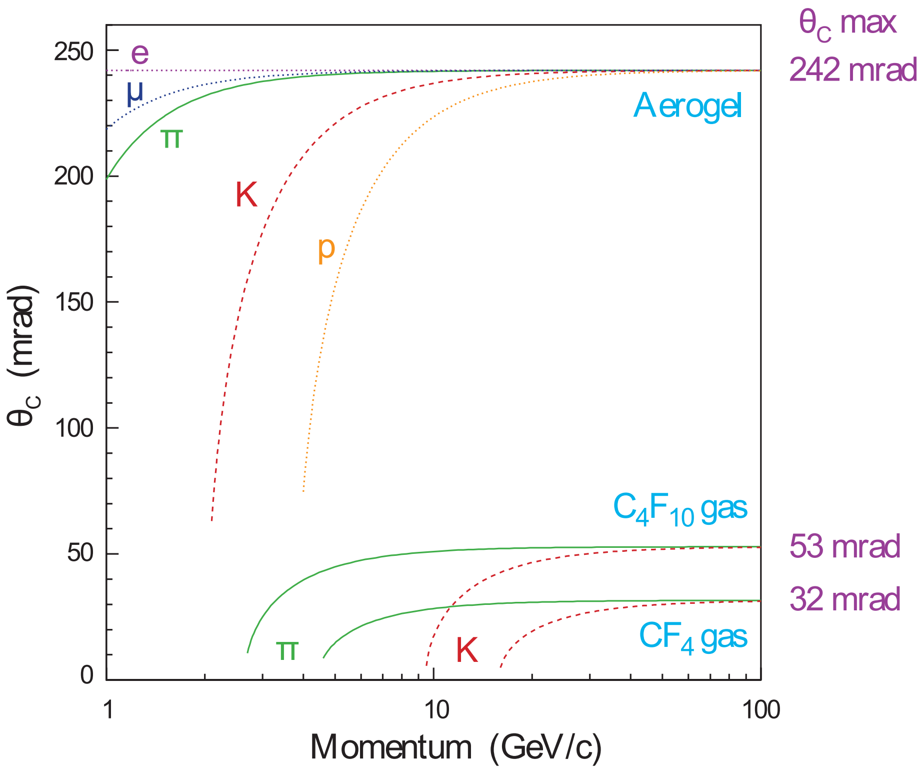


Figure 4.6: The Cherenkov angles, θ_C, of various particles as a function of momentum, for the radiators used in the LHCb RICH system from Ref. [48]

then be calculated for these pixel-track combinations. The expected emission angles, or *Cherenkov angles*, for particles in each of the three RICH radiators can be seen in Fig. 4.6.

The overall likelihood for each particle hypothesis is found by considering all of the three radiators simultaneously and performing a global pattern-recognition to associate Cherenkov photons (hits) to tracks. Figure 4.7 shows an example of hits in the RICH detector from a low occupancy event prior to Run 1. As shown, sometimes these Cherenkov rings will overlap, which makes it more difficult to associate a hit to a particular track. The particle most likely to have caused the hit is found first, then the likelihood of each particle hypothesis for that hit-track combination is determined.

The calorimeter system also aids particle identification, by discriminating between photons, electrons and neutral pions. Differentiating between charged and neutral particles is done by checking if there is a visible track corresponding to the energy deposit in the ECAL/HCAL, as tracks are only left by charged particles. Photons and neutral pions can then be recognised by the shape of the energy deposit or ‘cluster’. The classification of the particle due to the cluster shape is performed by a neural network and used to create a likelihood for each particle hypothesis.

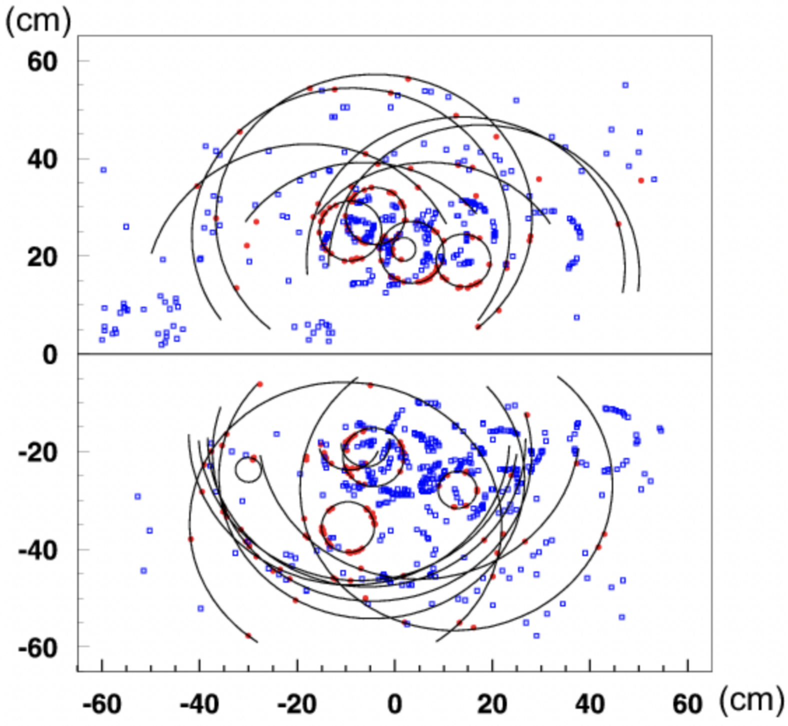


Figure 4.7: A typical simulated LHCb event in the RICH1 detector [48]. The data from the two photodetector planes are drawn in the upper and lower halves.

4.4 PID Variables

The muon station likelihoods are combined with the likelihoods from the RICH and calorimeters, to come up with a combined particle likelihood for each track. The likelihoods are calculated from the RICH using different particle mass hypotheses, and from the calorimeters using the size and spread of the energy deposits.

There are two main types of PID variables used in LHCb, obtained by computing the combined particle likelihoods in slightly different ways: the Delta Log Likelihood (DLL) variables, and the ProbNN variables. The first (DLL variables), are a sum of the logarithms of the subdetector likelihoods, relative to the pion hypothesis, while the second (ProbNN variables) are the pseudo-probability from 0 to 1 for each hypothesis returned by a Neural Network. ProbNN takes into account correlations between the subdetector likelihoods. It also adds information from the tracking system, to help reject ghost particles and particles that result from decays in flight.

The variables used and requirements placed on them is dependant on the analysis. ProbNN variables are used for the signal in $B^0_{(s)} \to \mu^+\mu^-$ analysis, since it was found they provide a high signal efficiency and stronger background rejection compared to DLL [69]. The background rejection rate for each variable is shown in Fig 4.8, as a function of signal efficiency for selected muons from simulated samples and data sidebands.

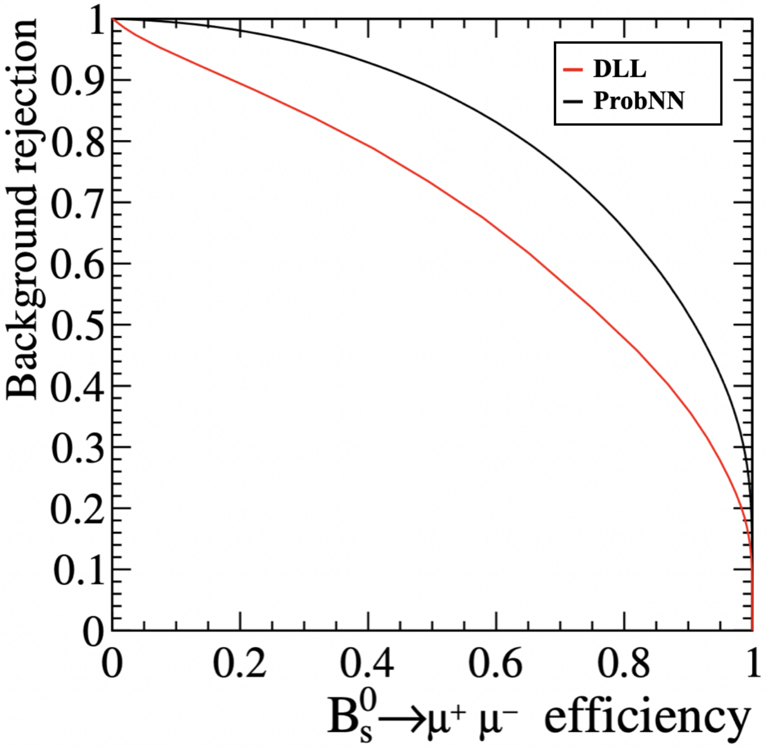


Figure 4.8: The efficiency to select muons versus rejecting combinatorial background, using the DLL variable (red) and ProbNN variable (black) for muons from $B_s^0 \to \mu^+\mu^-$ MC samples and sideband data.

The exact PID requirements for the $B_{(s)}^0 \to \mu^+\mu^-$ analysis will be discussed in more detail in Sect. 6.8.1.

Chapter 5

$B_s^0 \to \mu^+\mu^-\gamma$ feasibility studies

5.1 Introduction

The $B_s^0 \to \mu^+\mu^-\gamma$ decay offers sensitivity to a wider set of effective operators than its non-radiative counterpart $B_s^0 \to \mu^+\mu^-$, as discussed earlier in Chap. 2. That said, the direct measurement of $B_s^0 \to \mu^+\mu^-\gamma$ in hadron collider experiments is much more difficult than that of $B_s^0 \to \mu^+\mu^-$ for several reasons:

1. The detection and reconstruction efficiencies of photons are typically much smaller than that of charged tracks, such as muons.
2. The energy of the final state muons is being shared with an additional photon, meaning they have lower momentum and resulting in lower trigger and reconstruction efficiencies.
3. Including a photon in the reconstruction of the invariant mass results in a lower resolution, collecting more background candidates under the signal peak.

In order to overcome some of these difficulties, it was suggested in Ref. [25] that the $B_s^0 \to \mu^+\mu^-\gamma$ decay could be searched for in the same event sample selected for the $B_s^0 \to \mu^+\mu^-$ measurement. The method involves searching for the $B_s^0 \to \mu^+\mu^-\gamma$ decay as a contamination to the $B_s^0 \to \mu^+\mu^-$ signal peak, by not reconstructing the photon in the final state. The $B_s^0 \to \mu^+\mu^-\gamma$ measurement would be extracted as an additional observable, without any specific optimisation to the selection. However, a search for $B_s^0 \to \mu^+\mu^-\gamma$ using this method will only possible if enough of the $B_s^0 \to \mu^+\mu^-\gamma$ signal is retained during the $B_s^0 \to \mu^+\mu^-$ selection process.

This chapter outlines the feasibility studies conducted on Monte Carlo using the tools and dataset from the previous LHCb $B_{(s)}^0 \to \mu^+\mu^-$ measurement [3], to see if the $B_s^0 \to \mu^+\mu^-\gamma$ signal efficiency is close enough to that of $B_s^0 \to \mu^+\mu^-$. The signal efficiencies are compared at each stage of the selection process, and the number of events we would expect to see in our dataset is computed.

In the $B_{(s)}^0 \to \mu^+\mu^-$ analysis, the branching fraction is extracted from fits to the data as a function of the invariant mass of the two final state muons, $m_{\mu^+\mu^-}$. This dimuon invariant mass distribution will peak at the B_s^0 meson mass and B^0 meson mass for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays respectively. The invariant mass window used in the search for these decays is limited to $4.9\ \text{GeV}/c^2 < m_{\mu^+\mu^-} < 6.0\ \text{GeV}/c^2$. All of the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ signal events are expected to lie comfortably within that region, so cutting out events outside of this region will only reduce background yields.

The possibility of enlarging this mass window for the $B_s^0 \to \mu^+\mu^-\gamma$ analysis is investigated. Although all of the signal events for $B_{(s)}^0 \to \mu^+\mu^-$ should be contained within the $4.9\ \text{GeV}/c^2 < m_{\mu^+\mu^-} < 6.0\ \text{GeV}/c^2$ region, this is not the case for $B_s^0 \to \mu^+\mu^-\gamma$. Since the final state muons are now sharing their energy with a photon, which is not reconstructed, the dimuon invariant mass will *appear* to be lower for the radiative decay. The signal will therefore increase as $m_{\mu^+\mu^-}$ decreases from the B_s^0 peak. For this reason, the possibility of moving the lower boundary of $m_{\mu^+\mu^-}$ down to $4.5\,\text{GeV}/c^2$ will also be investigated here, to see how much the signal yields would increase if the analysis was adapted in this way. The shapes of the $B_{(s)}^0 \to \mu^+\mu^-$ backgrounds will be explored using MC down to this lower $m_{\mu^+\mu^-}$ boundary.

Figure 5.1 from Ref. [25] shows a sketch of the predicted $B_s^0 \to \mu^+\mu^-\gamma$ signal overlayed onto the final $B_{(s)}^0 \to \mu^+\mu^-$ fit from the previous LHCb analysis round. Also shown are two NP scenarios which could enhance or suppress the $B_s^0 \to \mu^+\mu^-\gamma$ signal. From the data it can already be seen that a large enhancement of the SM branching fraction (for example 10×SM), of $B_s^0 \to \mu^+\mu^-\gamma$ is not likely. On the other hand, a suppressed signal is possible and would be an indication of NP.

5.2 Monte Carlo Samples

This study uses Monte Carlo (MC) simulation corresponding to 2011 - 2012 (Run 1), and 2015 - 2016 (Run 2) data collected by LHCb. Creating the MC distributions for this study requires several stages:

$$\text{Generation} \to \text{Acceptance} \to \text{Tracking} \to \text{Reconstruction} \to \text{Selection} \to \text{Trigger}$$

The first stages, generation and acceptance, involve generating events in as much detail as they are created in nature; all of the possible combinations of particles resulting from a *pp* collision are generated in MC. At LHCb, this event simulation is performed by a software package called Gauss [73], which simulates *pp* collisions and the resultant particle production, as well as the decays of these particles. Gauss relies on Pythia [74] to simulate processes within the *pp* collisions and the hadronisation that follows. B meson decays are simulated by a dedicated package called EvtGen [75], which excludes particles outside of the LHCb acceptance range, meaning that the polar angle between the two muons in the

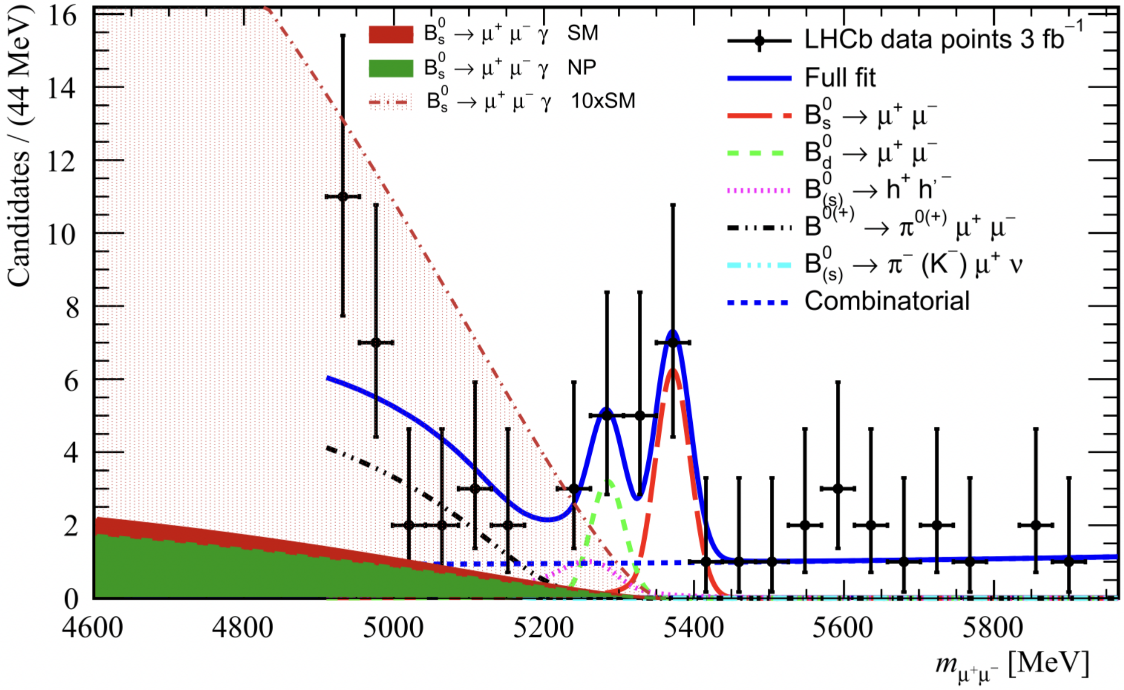


Figure 5.1: The $B_s^0 \to \mu^+\mu^-\gamma$ signal estimate (dark red) overlayed onto LHCb's dimuon invariant mass distribution from the previous measurement of $B_{(s)}^0 \to \mu^+\mu^-$ [3]. Also shown are two NP scenarios for $B_s^0 \to \mu^+\mu^-\gamma$: one where the signal is enhanced to a level of 10×SM (light red), and one where the Wilson coefficient C_9 is less than the SM value such that $\delta C_9 = -12\%\ C_9^{\mathrm{SM}}$ (dark green). The filled curves are not stacked on top of one another. Figure taken from Ref. [25].

final state is required to be between 10 and 40 mrad.

The role of MC is to provide simulations that mimic what really happens to particles after the collision. This means incorporating experimental conditions and performance for the particular detector. For this reason, once the particles are simulated, the GEANT4 [76, 77] software is used to pass these particles through the simulated detector materials and magnetic field, to give a more accurate representation of what would happen within the LHCb detector. Only the events that are within the detector acceptance range (defined earlier in Sect. 3.2) are kept.

The next stage, tracking and reconstruction, involves tracing back the created tracks in order to reconstruct a certain event. A particular selection criteria is then added, which could involve cuts on variables such as particle mass, momentum, or flight distance, for example. Finally, the trigger selection is added, which simulates the L0 and HLT triggers within the detector to ensure that the simulated events correspond to those that will actually be stored in real data events.

5.3 Signal Selection

In order for it to be feasible to add $B_s^0 \to \mu^+\mu^-\gamma$ as an additional observable to the $B_{(s)}^0 \to \mu^+\mu^-$ analysis, first the effect of the full selection on this new signal must be evaluated. The signal selection is taken from Ref. [3], and is optimised for selection of $B_{(s)}^0 \to \mu^+\mu^-$ signal. As previously discussed, no changes have been made to optimise for the radiative decay search.

In this analysis, signal events are selected using a series of selection cuts, made on variables that are likely to distinguish the signal events from the background ones. One of these variables is the output of a boosted decision tree (the BDTS) which will be discussed in more detail in Sect. 5.3.3. The basic principals of a boosted decision tree will described in the next section.

5.3.1 Boosted Decision Trees

Several multivariate analysis tools (MVAs) are used in the $B_{(s)}^0 \to \mu^+\mu^-$ analysis. These are machine learning techniques used to classify events based on the values of certain discriminating variables. The analysis uses boosted decision trees (or *BDTs*), which take into account that not all of the characteristics of 'signal' are shown by every signal event. Instead events are given a likelihood of being signal, not by simply ruling out those that fail one particular criterion (like a cut based selection does).

Decision trees start with a root node, then a series of 'yes/no' decisions are made for each event based on certain discriminating variables. Each of the criteria for the decision are based on the variable that gives the best signal/background discrimination at that point. This means that each of the input variables for any decision tree may be used a

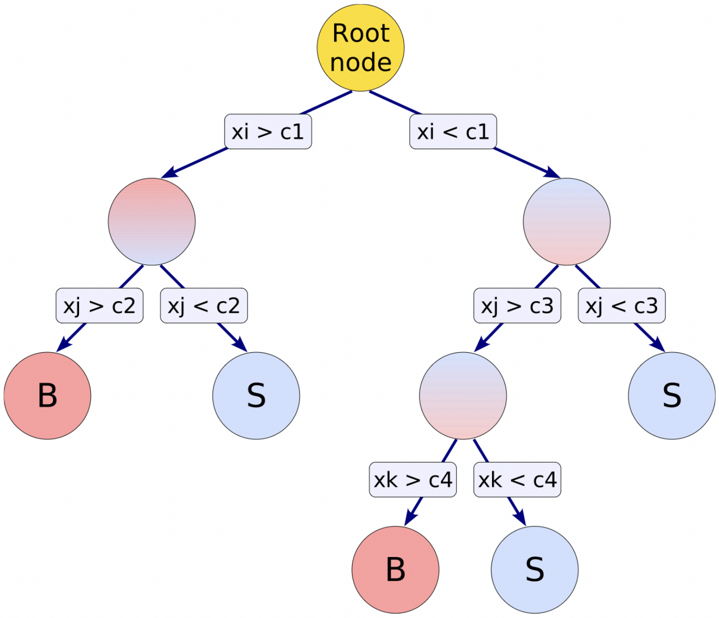


Figure 5.2: Schematic view of a decision tree from Ref. [78]. The tree begins with the root node at the top, followed by a series of binary decisions until events are classified as signal (S) or background (B) in the leaf nodes at the bottom.

number of times or not at all. At the end of the tree are the leaf nodes, which are classified as signal or background depending on how many of each event end up there. These decision trees are trained on known data, *i.e.* a signal sample and a selection of known backgrounds. This is typically done with MC simulated samples. That way, when the trees are used for unknown data samples we get a prediction of how signal-like they are. A schematic view of a decision tree is found in Fig. 5.2.

The boosting of a decision tree involves extending this concept from one tree into several, known as a *forest*. The candidates that have been mis-classified in a single tree are given a higher weighting for the training of the next tree. This collection of trees is finally grouped back together in a single classifier using the weighted average of the individual trees. Boosting generally provides better signal/background discrimination than using a single tree, and stabilises the response with respect to fluctuations in the training sample [78].

The first BDT used in this analysis is called ‘BDTS’, and is used as part of the selection process to reduce the sample size by rejecting background events, whilst keeping most of the signal. The BDTS and the variables entering it will be discussed in Sect. 5.3.3. The second BDT, simply called the ‘BDT’, is used at the final analysis stage to classify the candidates according on how likely they are to be signal or background events.

5.3.2 Selection Cuts

In this analysis several cuts are implemented in order to select the signal candidates. Specific cuts are used in the selection process to achieve certain goals. A summary of all of the selection cuts is given in Table 5.1.

Firstly, events are selected only when the two daughter tracks (the muons) have a small distance of closest approach (DOCA) and form a good vertex with one another. The tracks are also required to have a small χ^2 per degree of freedom (χ^2/ndf), which means that the tracks are a good representation of the hits in the detector. These cuts collectively ensure that the muon candidates have high reconstruction quality.

Table 5.1: The selection criteria for $B^0_{(s)} \to \mu^+\mu^-$ signal, where t is proper time, χ^2_{IP} is the significance of the Impact Parameter for each muon track (top) and for the secondary vertex fit (bottom). Definitions and explanations of other variables can be found in the text.

Cut	Applied on	Value
isMuon	$\mu^\pm$	TRUE
track χ^2/ndf		< 3
ghost prob		< 0.3
DOCA		$< 0.3\,\text{mm}$
χ^2_{IP}		> 25
p_T		$> 0.25 < 40\,\text{MeV}/c$
p		$< 500\,\text{GeV}/c$
PID$_\mu$		> 0.8
Vtx χ^2	$B^0_{(s)}$	< 9
VDS		> 15
ΔM		$\lvert M(\mu\mu) - m_{B^0_{(s)}} \rvert <$ 60 MeV/c^2
χ^2_{IP}		< 25
t		$< 9 \cdot \tau(B^0_s)$
p_{T}		$> 0.5\,\text{GeV}/c$
BDTS		> 0.05

The next group of cuts are designed to reject any non-physical candidates. Firstly, any candidates with momentum (p) or transverse momentum (p_T) outside of the LHCb acceptance range are rejected. Then a third MVA tool is used to reject 'ghost' tracks, that is, those that are not physical particle paths, but are instead artefacts created during event reconstruction. The ghost probability multi-variate operator combines 22 different parameters to compute a likelihood of a track being a ghost, including the number of subdetector hits, comparisons between measured and observed numbers of subdetector hits, and values of several other kinematic variables. From the output of this operator, any tracks that are likely to be ghosts can be rejected. Finally, any $B^0_{(s)}$ candidates with too long proper time (t) are also rejected.

A requirement on the transverse momentum of the B meson candidate rejects those that may have originated from elastic *pp* collisions, while cuts on the impact parameter significance ($IP\chi^2$) of the B meson and muons, and distance significance of the reconstructed B decay vertex (VDS) significantly reduce background muons produced from inelastic *pp* collisions. A requirement on the dimuon invariant mass means that only candidates with mass close to that of the $B^0_{(s)}$ meson are kept.

The muon identification procedure (muonID) is a key ingredient of the analysis. It consists of two steps: the requirement for the muon candidate to satisfy the ISMUON algorithm, and a more restrictive ProbNN PID selection that returns a total likelihood for a candidate to be a certain particle. The two procedures were described earlier in Sect. 4.3.1.

The PID cuts are designed to maximize the muon efficiency and reject any pions, kaons or protons that may have been misidentified. The muon selection criteria have been optimized in Ref. [79] by ensuring a strong rejection of backgrounds while preserving a high signal efficiency. The best performances for the branching fraction analysis have been obtained with the requirement:

$$\mathrm{PID}_\mu \equiv \mathrm{ProbNN}_\mu \times (1 - \mathrm{ProbNN}_p) \times (1 - \mathrm{ProbNN}_K) > 0.8, \tag{5.3.1}$$

where the $\mathrm{ProbNN}_{\mu,p,K}$ variables are outputs of the Neural Network, indicating how likely particles are to be muons, protons and kaons respectively [80]. Since pions and muons are very close in mass, and the Cherenkov rings produced in the RICH detectors are very close to one another, meaning that at most momenta the discrimination between pions and muons must be done using information from the other subdetectors, which is included in the neural network.

5.3.3 The BDTS Discriminant

The main purpose of the BDTS is to reject background that comes from combinatorial events, where two muons arise from different B decays. The cuts applied during the selection process partially reduce the number of combinatorial events, by ensuring that the two tracks originate from the same location (see next section). However, if two random muons line up in such a way as to appear to have come from the same displaced vertex, they can look like signal events. Figure 5.3 shows a sketch of two muons from different decays forming a combinatorial background event.

Further rejection of these events is therefore required and is performed using the BDTS. This variable is from an MVA tool which outputs a value for each candidate between 0 and 1, depending on how ‘signal-like’ or ‘background like’ the event appears to be. This number is used to separate signal and background.

The BDTS is built and optimized in The Toolkit for Multivariate Analysis (TMVA) [78]. Using TMVA, the BDT operators were chosen as they provided the best background rejection and signal efficiency, allowing for exclusion of candidates with low signal likelihood.

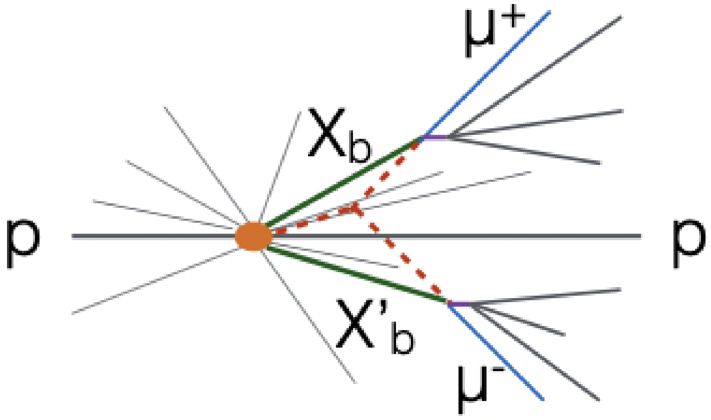


Figure 5.3: Sketch example of two muons from decays of other B mesons (X_b and X_b') forming combinatorial background. Image from Ref. [68].

The six variables used by the BDTS are:

- The B candidate impact parameter to closest PV ($B^0_{(s)}$IP),
- The B candidate impact parameter significance ($B^0_{(s)}\chi^2_{\text{IP}}$),
- The dimuon (secondary) vertex significance (Vtx χ^2),
- The minimum distance between the two daughter tracks (DOCA),
- The angle between the direction of the momentum of the B candidate and the direction defined by the secondary and the primary vertices (DIRA),
- The minimum impact parameter of each muon with respect to any primary vertex (μ min IP).

The BDTS is trained on $b\bar{b} \to \mu\mu X$ (representing combinatorial background) and $B_s^0 \to \mu^+\mu^-$ (representing signal) simulated events with the standard selection applied. The trained BDTS is applied to all signal candidates. The BDTS itself and the BDTS cut have not been optimized since Ref. [81], where a cut of BDTS > 0.05 was found to be optimal for signal retention and background rejection. Figure 5.4 shows the BDTS distributions for signal and background events, from Ref. [82].

5.4 Signal Efficiencies

There are several stages that impact the total signal efficiency in the $B^0_{(s)} \to \mu^+\mu^-$ analysis. The total detection efficiency is split according to Eq. 5.4.1.

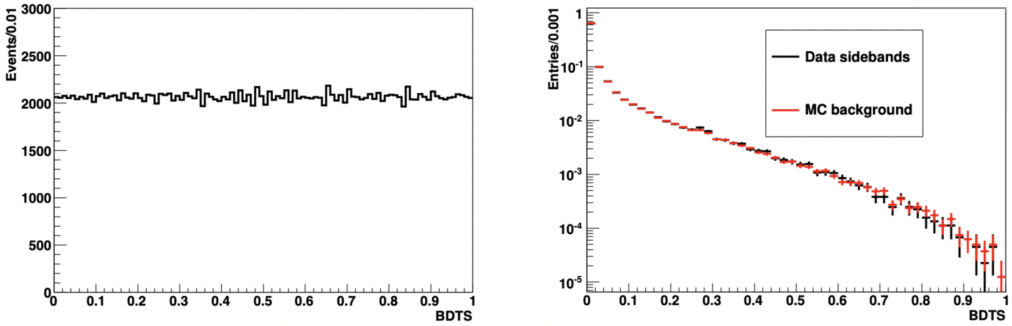


Figure 5.4: Left: BDTS distribution for $B_s^0 \to \mu^+\mu^-$ Monte Carlo events; right: BDTS distribution for $B_s^0 \to \mu^+\mu^-$ sideband events (black markers) and $b\bar{b} \to \mu\mu X$ simulated events (red markers).

$$\epsilon_{Tot} = \epsilon_{Acc} \times \epsilon_{Rec} \times \epsilon_{Sel} \times \epsilon_{Trig}, \tag{5.4.1}$$

where ϵ_{Acc}, ϵ_{Rec}, ϵ_{Sel} and ϵ_{Trig} are the efficiencies for the geometric detector acceptance, reconstruction, selection, and trigger respectively.

The detector acceptance efficiency is defined as the fraction of the decays that have all of their decay products within the geometrical detector acceptance. Not all of the candidates in the acceptance will be reconstructed, as they are affected by the magnetic field and interactions with the detector materials. The reconstruction efficiency is then the fraction of decay candidates in the detector acceptance that are successfully reconstructed. The selection efficiency is the fraction of reconstructed decay candidates that pass the selection described in Sect. 5.3. Finally, the trigger efficiency shows how many of these simulated events would actually pass the initial trigger criteria within the detector if they were real data. Each of these efficiencies need to be computed in order to compare the total signal efficiencies of $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$.

5.4.1 Acceptance and Reconstruction

The acceptance efficiencies for the two signal channels are first compared using simulated samples of 2016 data. As this efficiency includes the detector acceptance, it will naturally be less than $\sim 50\%$ due to the forward design of the LHCb detector. The decay products are actually required to be within an acceptance defined by the polar angle [10, 400] mrad, which is slightly larger than the physical LHCb detector acceptance to allow for the recovery of particles by the magnetic field. The acceptance efficiencies for each signal channel are given in Table 5.2.

In principle these should be very similar for the two channels. However, during the creation of the MC, the geometric acceptance cut of the detector that is usually applied at that level was not included for the $B_s^0 \to \mu^+\mu^-\gamma$ sample. This just means that this efficiency will be taken into account in the reconstruction stage (as a candidate outside of the acceptance can not be reconstructed). The reconstruction efficiency is the fraction of decay candidates in the detector acceptance that are successfully reconstructed. The reconstruction efficiency for $B_s^0 \to \mu^+\mu^-\gamma$ is much lower than that of $B_s^0 \to \mu^+\mu^-$ as the geometric cut is taken into account at this stage rather than at the generation stage, however the products of the two efficiencies show that the combined $\epsilon_{Rec|Acc}$ efficiency is similar for each of the decay modes, as expected. Table 5.2 shows the acceptance and reconstruction efficiencies for $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$, as well as the product of the two efficiencies. As shown, the total $\epsilon_{Rec|Acc}$ is 12.11% for $B_s^0 \to \mu^+\mu^-$ and 12.91% for $B_s^0 \to \mu^+\mu^-\gamma$.

Table 5.2: The acceptance and reconstruction efficiencies for $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$.

Channel	ϵ_{Acc}	ϵ_{Rec}	$\epsilon_{Rec\|Acc}$
$B_s^0 \to \mu^+\mu^-$	19.33%	62.64%	12.11%
$B_s^0 \to \mu^+\mu^-\gamma$	47.38%	27.24%	12.91%

5.4.2 Selection

It is important to calculate exactly how much of the $B_s^0 \to \mu^+\mu^-\gamma$ signal will pass the $B_s^0 \to \mu^+\mu^-$ event selection. So far, the product of the reconstruction and acceptance efficiencies are similar, but there are likely to be some differences between the efficiencies of the two channels at the selection stage, due to the additional photon. Figure 5.5 shows a sketch of a $B_s^0 \to \mu^+\mu^-\gamma$ decay, and how the addition of a photon in the final state might shift the dimuon extrapolation back to the PV. This in turn will affect the χ^2_{IP} of the B candidate and also the output of the BDTS, which uses the χ^2_{IP} as a discriminating variable.

Pre- and post-selection variable comparisons

An estimate of which variables involved in the selection process cause the largest drop in signal efficiency for $B_s^0 \to \mu^+\mu^-\gamma$ can be seen using plots of the candidates before and after the selection cuts. These comparisons were performed using MC, and mostly showed good agreement between $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$. However, there were some significant differences with certain variables. The variables showing a good agreement between the two signals have been included in Appendix A for reference.

Figure 5.6 shows the pre- and post-selection dimuon invariant mass distributions for the two signals. As expected, the $B_s^0 \to \mu^+\mu^-\gamma$ appears as a 'shoulder' on the left hand side of the $B_s^0 \to \mu^+\mu^-$ signal peak in both plots. It is important to note here that at this stage, no mass cut is explicitly applied at the selection stage. However there is a mass cut

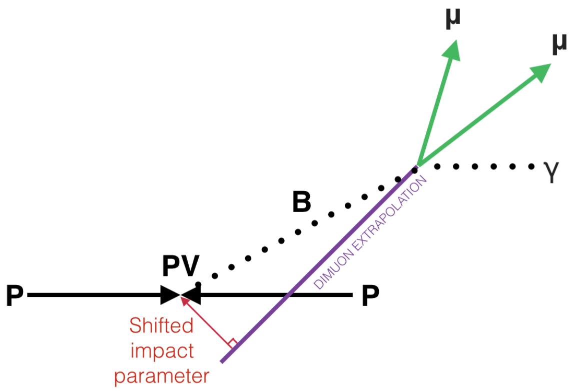


Figure 5.5: Sketch of how the dimuon extrapolation back the PV might be shifted due to the unreconstructed photon

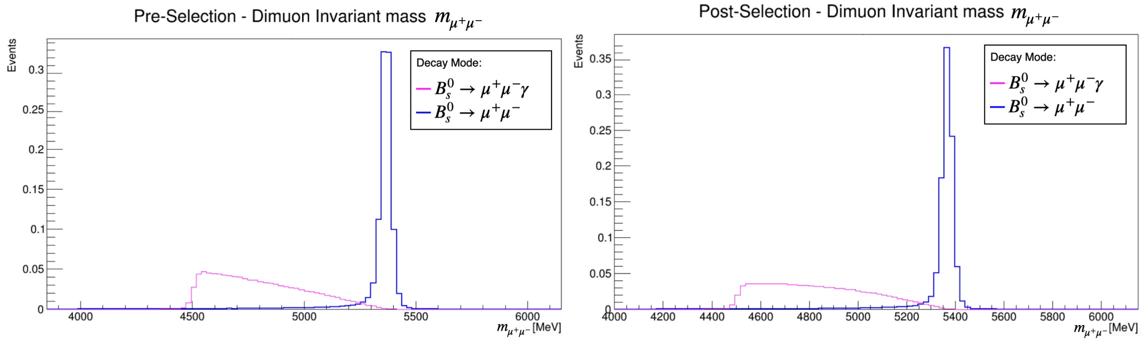
in the generation of the MC at 4.5 GeV/c^2, which is why the $B_s^0 \to \mu^+\mu^-\gamma$ signal drops away at this point.

There is not a significant difference in the shapes following the application of the selection cuts. The $\chi^2_{\rm IP}$ distributions on the other hand shown in Fig. 5.7 show that the selection cuts remove much more $B_s^0 \to \mu^+\mu^-\gamma$ signal than $B_s^0 \to \mu^+\mu^-$. The distribution for $B_s^0 \to \mu^+\mu^-\gamma$ does not fall as quickly as the $B_s^0 \to \mu^+\mu^-$ signal, meaning that the cut at IP$\chi^2 > 25$ causes a greater loss of signal for $B_s^0 \to \mu^+\mu^-\gamma$. In addition to this, the BDTS distributions for the two signals are quite different. The BDTS is *designed* to give a flat distribution between 0 and 1 for signal. However, the BDTS was not designed to include $B_s^0 \to \mu^+\mu^-\gamma$ as a signal, so the distribution for this is more 'background like' with a peak at 0 (see Fig. 5.8). Like the $\chi^2_{\rm IP}$ cut, the BDTS cut will therefore decrease the signal efficiency of $B_s^0 \to \mu^+\mu^-\gamma$ more that $B_s^0 \to \mu^+\mu^-$. The exact amount of signal efficiency loss will be discussed in the next section.

Cutflow Table for Signal Selection

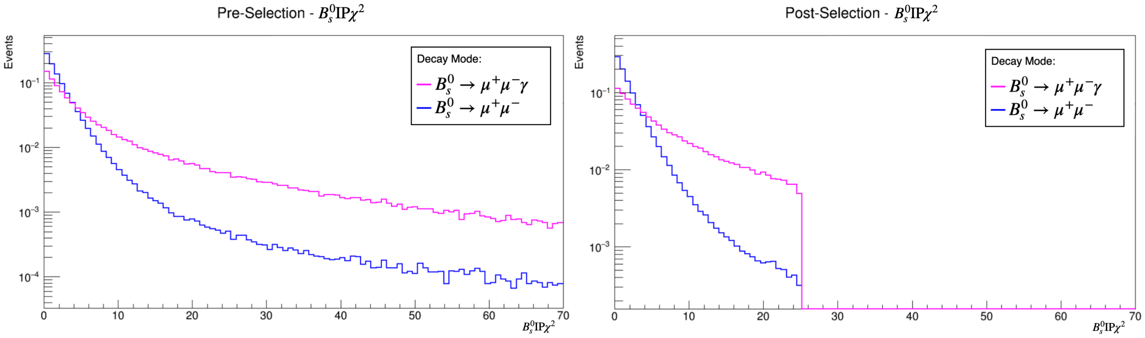
Having compared the shapes of the $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ variables from simulated samples at both pre- and post-selection stages, it is important to calculate exactly how much of the signal is passing the event selection. In order to do this, each of the specific cuts used in the $B_s^0 \to \mu^+\mu^-$ selection are investigated to ensure that the $B_s^0 \to \mu^+\mu^-\gamma$ signal efficiency remains high enough to enable an analysis using this selection. Table 5.3 lists all of the cuts applied in the selection process for $B_s^0 \to \mu^+\mu^-$. Again, the table excludes the specific mass cut for both signal as two mass cut options will be investigated. The requirement at generation of the $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ MC that $m_{\mu^+\mu^-} > 4.5\,\text{GeV}/c^2$ is implicitly included.

The table shows that the overall cumulative efficiency of the selection cuts for $B_s^0 \to \mu^+\mu^-\gamma$ is 29.61%, and for $B_s^0 \to \mu^+\mu^-$ it is 44.92%, determined using MC. Inspection of the



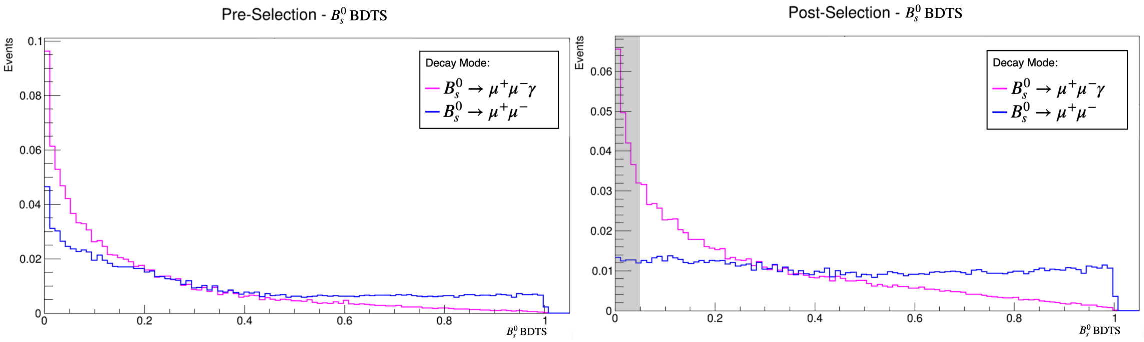
(a) (b)

Figure 5.6: Invariant mass of the two muons, $m_{\mu^+\mu^-}$ for $B_s^0 \to \mu^+\mu^-$ (blue) and $B_s^0 \to \mu^+\mu^-\gamma$ (pink), normalised for shape comparison (y-axis arb. units). The plot on the left shows the comparison before selection, and the right shows the comparison after the selection from Table 5.1.



(a) (b)

Figure 5.7: Impact parameter significance, IP χ^2, of the B candidate, for $B_s^0 \to \mu^+\mu^-$ (blue) and $B_s^0 \to \mu^+\mu^-\gamma$ (pink), normalised for shape comparison (y-axis arb. units). The plot on the left shows the comparison before selection, and the right shows the comparison after the selection from Table 5.1.



(a) (b)

Figure 5.8: The BDTS output for $B_s^0 \to \mu^+\mu^-$ (blue) and $B_s^0 \to \mu^+\mu^-\gamma$ (pink), normalised for shape comparison (y-axis arb. units). The plot on the left shows the comparison before selection, and the right shows the comparison after the selection cuts. The BDTS > 0.05 cut itself is not applied, so that the signal lost from this cut can be seen in the shaded gray region.

Table 5.3: Cutflow table, showing the cumulative efficiencies for each signal decay mode following individual selection cuts (using MC). No explicit invariant mass cut is applied at this stage. The two lines in bold show the variables with the biggest difference in efficiency between the two signals.

	$B_s^0 \to \mu^+\mu^-\gamma$		$B_s^0 \to \mu^+\mu^-$	
Variable	Eff (%)	Cumulative Eff (%)	Eff (%)	Cumulative Eff (%)
$\mu^\pm$ ISMUON =1	84.0194	84.0194	84.1533	84.1533
$\mu^\pm$ Track χ^2/ndf <3	99.6584	83.7441	99.6036	83.8345
$\mu^\pm$ ghost prob <0.3	99.5834	83.4441	99.6107	83.5656
B_s^0 DOCA <0.3	99.8011	83.3037	99.7903	83.4273
$\mu^\pm$ χ^2_{IP} >25	63.6491	54.327	62.881	53.6526
B_s^0 Vtx χ^2 <9	98.232	53.4546	98.1088	52.7305
B_s^0 χ^2_{IP} <25	**85.4531**	**41.7508**	**97.6886**	**51.4316**
B_s^0 VDS >15	68.9392	37.9626	68.2196	47.7566
B_s^0 DIRA >0	98.3907	37.9457	98.2924	47.747
B_s^0 τ <13.248 $\times 10^{-3}$	99.9419	37.9423	99.9337	47.7331
$\mu^\pm p_T$ <40 MeV/c	99.9831	37.9370	99.9711	47.7178
$\mu^\pm p_T$ >0.25 MeV/c	99.9206	37.9273	99.9232	47.7073
$\mu^\pm p$ <500 GeV/c	99.729	37.8881	99.5653	47.6285
$B_s^0 p_T$ >500	98.7995	37.4633	98.9819	47.1921
B_s^0 BDTS >0.05	**71.1036**	**29.9624**	**84.7498**	**45.4001**
$\mu^\pm$ PID$_\mu$ >0.8	82.3341	29.6092	82.8538	44.9215

cutflow table indicates that the dramatic decreases in signal efficiency for $B_s^0 \to \mu^+\mu^-\gamma$ compared to $B_s^0 \to \mu^+\mu^-$ are due to the cuts relating the variables discussed at the beginning of this section, being B_s^0 $\chi^2_{\rm IP}$, and BDTS, which are shown in bold on the table. The signal efficiency for the B_s^0 $\chi^2_{\rm IP}$ and BDTS cuts are $\sim 13\%$ and $\sim 16\%$ less for $B_s^0 \to \mu^+\mu^-\gamma$ respectively. As these signal decays are rare anyway, signal loss like this is not ideal, but was foreseen. Overall the selection efficiency for $B_s^0 \to \mu^+\mu^-\gamma$ is only $\sim 35\%$ lower than for $B_s^0 \to \mu^+\mu^-$, which is good considering the selection was only optimised for the non-radiative decay.

That said, given the difference in the shapes of the invariant mass distributions for these two signals, it is clear that applying a cut on $m_{\mu^+\mu^-}$ will have a significant impact on the $B_s^0 \to \mu^+\mu^-\gamma$ signal efficiency. As mentioned in the introduction to this chapter, two invariant mass cut options will be investigated; $m_{\mu^+\mu^-} > 4.5\,\mathrm{GeV}/c^2$ and $m_{\mu^+\mu^-} > 4.9\,\mathrm{GeV}/c^2$. Leaving the mass cut at $m_{\mu^+\mu^-} > 4.9\,\mathrm{GeV}/c^2$ would be much more straight forward for simply including $B_s^0 \to \mu^+\mu^-\gamma$ as an additional observable in the $B_{(s)}^0 \to \mu^+\mu^-$ analysis, since this is the cut already in place. This means that the background contributions have already been estimated for this mass region. Extending the lower mass boundary down to $4.5\,\mathrm{GeV}/c^2$ would include much more $B_s^0 \to \mu^+\mu^-\gamma$ signal, but also much more background.

Since the $B_s^0 \to \mu^+\mu^-\gamma$ MC being used in this study is generated for candidates with a dimuon invariant mass greater than $4.5\,\mathrm{GeV}/c^2$, adding a cut of $m_{\mu^+\mu^-} > 4.9\,\mathrm{GeV}/c^2$ at the selection stage would dramatically decrease the efficiency, since all of the events in the region $4.5\,\mathrm{GeV}/c^2 < m_{\mu^+\mu^-} < 4.9\,\mathrm{GeV}/c^2$ will be rejected. To ensure a like for like efficiency comparison between the two mass window options, the higher mass cut needs to be included in the generation stage (so that events in the $4.5\,\mathrm{GeV}/c^2 < m_{\mu^+\mu^-} < 4.9\,\mathrm{GeV}/c^2$ range are not considered from the offset) rather than being added as a selection cut.

The total selection efficiencies have been calculated including each of the $m_{\mu^+\mu^-}$ cuts at generation stage, and will be discussed in Sect. 5.4.5. These selection efficiencies will then be used in the normalisation and computation of the event sensitivity (see Sect. 5.5).

5.4.3 Trigger

No specific trigger requirements are imposed on the $B_{(s)}^0 \to \mu^+\mu^-$ data in order to maximise the available statistics. This means that events passing any of the trigger lines are used for analysis. For this study full trigger calibration was not completed, therefore to obtain a trigger efficiency for data for $B_s^0 \to \mu^+\mu^-\gamma$ the ratio of $\epsilon_{\rm Trig,Data}$ / $\epsilon_{\rm Trig,MC}$ from $B_s^0 \to \mu^+\mu^-$ is used. This gives a more accurate representation of the trigger efficiency than simply taking the number from simulation. This means that the trigger efficiency obtained in real data[1] for $B_s^0 \to \mu^+\mu^-$ is compared to the trigger efficiency obtained in MC, providing a conversion factor between the two and allowing an estimate to be made of the trigger

---

[1] The trigger efficiency for $B_s^0 \to \mu^+\mu^-$ data is calculated using the `TISTOS` method [83], which will be explained in detail in Sect. 6.11

efficiency for real data for the radiative decay mode:

$$\epsilon_{\text{TrigData}}^{B_s^0\to\mu\mu\gamma} = \frac{\epsilon_{\text{TrigData}}^{B_s^0\to\mu\mu}}{\epsilon_{\text{TrigMC}}^{B_s^0\to\mu\mu}} \times \epsilon_{\text{TrigMC}}^{B_s^0\to\mu\mu\gamma} \tag{5.4.2}$$

This data correction for the trigger efficiencies is done using values from data/MC, and is computed for both of the mass regions of interest; $m_{\mu^+\mu^-} > 4.5\,\text{GeV}/c^2$ and $m_{\mu^+\mu^-} > 4.9\,\text{GeV}/c^2$.

$$m_{\mu^+\mu^-} > 4.9\,\text{GeV}/c^2 : \epsilon_{\text{TrigData}}^{B_s^0\to\mu\mu\gamma} = \frac{0.974}{0.972} \times 0.960 = 96.2\% \tag{5.4.3}$$

$$m_{\mu^+\mu^-} > 4.5\,\text{GeV}/c^2 : \epsilon_{\text{TrigData}}^{B_s^0\to\mu\mu\gamma} = \frac{0.974}{0.972} \times 0.957 = 95.9\% \tag{5.4.4}$$

These calculations show that the trigger efficiency stays high for both of the mass options, with a ratio $B_s^0 \to \mu^+\mu^-\gamma$ to $B_s^0 \to \mu^+\mu^-$ of $\sim 99\%$ (on top of the full selection).

5.4.4 Total Signal Efficiencies without Invariant Mass Cuts

Table 5.4 provides a summary of the efficiencies at each selection stage for both decay modes. The total efficiency is obtained by finding the product of all of these. From the table the total ratio of efficiencies for $B_s^0 \to \mu^+\mu^-\gamma$ to $B_s^0 \to \mu^+\mu^-$ (using MC) is calculated at around 70%. This is without any explicit dimuon invariant mass cut.

Table 5.4: Table of total efficiencies for each signal channel

Decay Mode	ϵ_{Gen}	ϵ_{Rec}	ϵ_{Sel}	ϵ_{Trig}	ϵ_{Tot}
$B_s^0 \to \mu^+\mu^-\gamma$	47.38%	27.24%	29.39%	96.20%	3.65%
$B_s^0 \to \mu^+\mu^-$	19.33%	62.64%	44.92%	97.40%	5.30%

5.4.5 Exploring the two Invariant Mass windows

As mentioned earlier, to investigate the different mass window options for $B_s^0 \to \mu^+\mu^-\gamma$, any invariant mass region cuts should be made in the generation stage. Because these efficiency calculations are made using simulation rather than real data, if events are simulated down to $4.5\,\text{GeV}/c^2$ and a higher mass cut of $4.9\,\text{GeV}/c^2$ is imposed, the selection efficiency would be extremely low and would not represent the efficiency in real data. For this reason the efficiencies entering the normalisation are taken from efficiencies calculated with any mass cuts required at the generation stage. Figure 5.9 shows a histogram of events as a function of dimuon invariant mass for each of the stages of selection for the mass window extended down to $m_{\mu^+\mu^-} > 4.5\,\text{GeV}/c^2$. Figure 5.10 shows the same plot but with $m_{\mu^+\mu^-} > 4.9\,\text{GeV}/c^2$. Both of the plots are created with MC.

From visual inspection of both sets of histograms it is clear that the gradient of the yellow line (Generated events) is greater than the gradient of the red line (Reconstructed

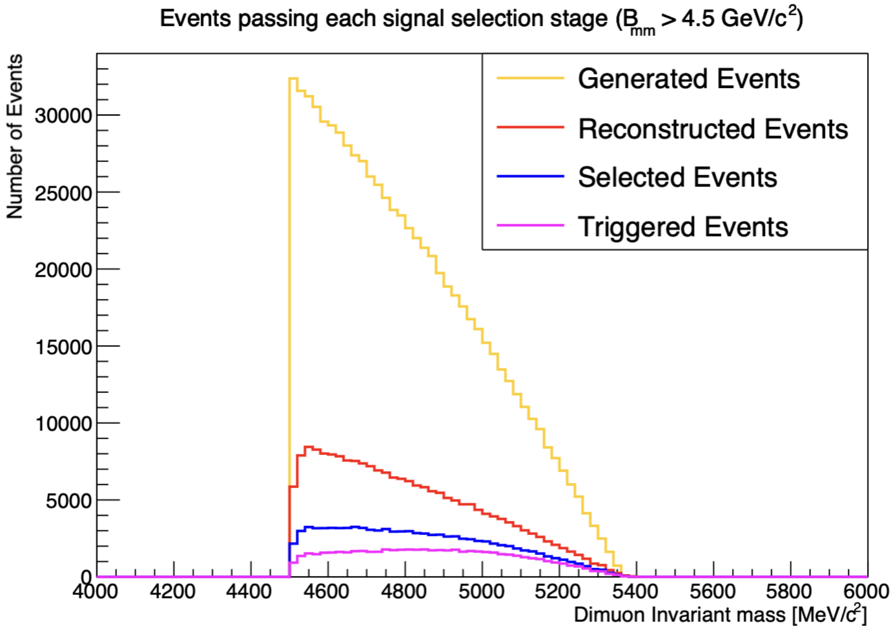


Figure 5.9: Number of events passing each of the signal selection stages for $B_s^0 \to \mu^+\mu^-\gamma$, with a cut of $m_{\mu^+\mu^-} > 4.5$ GeV/c^2 applied at generation. Plot created using MC.

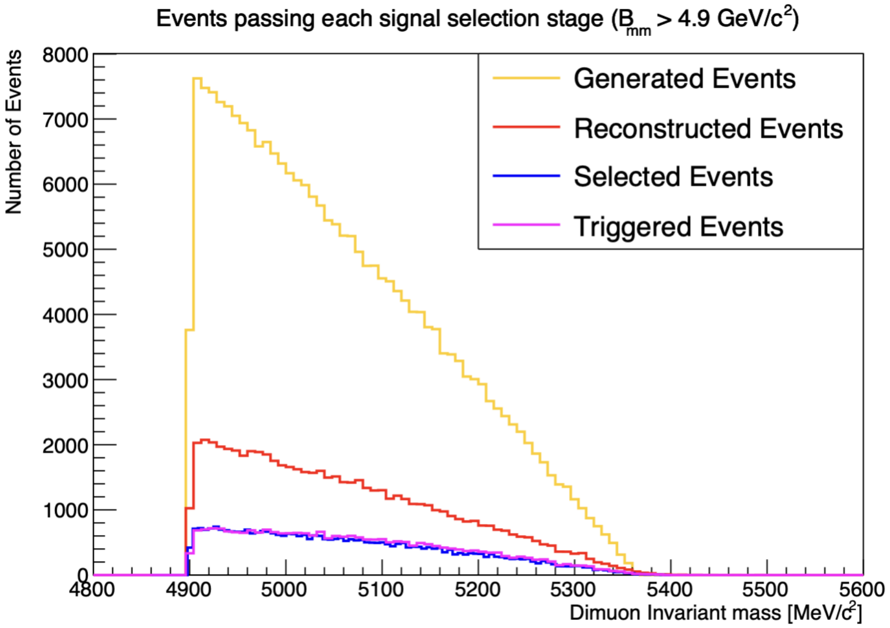


Figure 5.10: Number of events passing each of the signal selection stages for $B_s^0 \to \mu^+\mu^-\gamma$, with a cut of $m_{\mu^+\mu^-} > 4.9$ GeV/c^2 applied at generation. Plot created using MC.

events). This means that as the mass window is lowered, the reconstruction efficiency is reduced. The gradient difference between the red and blue lines (reconstructed and selected respectively) follows the same pattern, so the selection efficiency also decreases as the invariant mass window extended to the lower boundary. This decrease in efficiency and the effect on the number of detectable particles expected will be investigated in the next section.

The total signal efficiency, ϵ_{Tot} for each mass window scenario is found by dividing the Selected histogram (blue) by the Generated (yellow), then multiplying by the data corrected trigger efficiencies found in Eqs. 5.4.3 and 5.4.4. The calculations are performed for $B_s^0 \to \mu^+\mu^-\gamma$ MC for each year studied in the previous analysis round, for both mass window options. The results are presented in Table 5.5.

Table 5.5: Table of total signal efficiencies for $B_s^0 \to \mu^+\mu^-\gamma$ by MC year, for both mass window options. Efficiency differences between years is due different collision energies, as well as the use of different simulation software versions (which are continually being updates and improved) used to create the MC.

Lower $m_{\mu^+\mu^-}$ Cut	MC Year	Total Signal Efficiency
4.5 GeV/c^2	2011	3.112%
	2012	2.986%
	2015	3.294%
	2016	3.652%
4.9 GeV/c^2	2011	4.463%
	2012	4.314%
	2015	4.753%
	2016	5.213%

5.5 Normalisation and Single event sensitivity

From the efficiency calculations discussed in the previous section, the number of events that we should expect to see in our LHCb data can be computed. This number of expected events, $N_{B_s^0\to\mu^+\mu^-\gamma}$, can be described using the following relationship:

$$N_{B_s^0\to\mu^+\mu^-\gamma} = R \times f_s \times \mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) \times \epsilon_{B_s^0\to\mu^+\mu^-\gamma}, \tag{5.5.1}$$

where R is the reaction rate, f_s is the probability of a b quark forming a B_s^0 hadron, $\mathcal{B}$ is the branching ratio of the decay and $\epsilon_{B_s^0\to\mu^+\mu^-\gamma}$ is the total efficiency of the decay. The reaction rate, R is defined as:

$$R = L^{\text{LHCb}} \times \sigma_{pp\to b\bar{b}} \times 2, \tag{5.5.2}$$

where L^{LHCb} is the total integrated luminosity collected by LHCb and $\sigma_{pp\to b\bar{b}}$ is the $b\bar{b}$

production cross section. The factor of two is due to the fact that either of the b quarks can form the B meson.

The issue with calculating the number of expected events using Eqs. 5.5.1 and 5.5.2 is that there are high uncertainties involved in cross section measurement. This limits the precision of the number of events expected. Since the decays in this analysis are extremely rare, it is important to measure the rates with high precision, meaning that another method must be used. The precision problem can be alleviated by instead considering another B meson decay channel, and normalising the signal branching fraction with respect to this channel. Using a ratio of signal to normalisation channels means that many factors cancel out, and we can compute the rates with higher precision.

The normalisation channels used in the $B^0_{(s)} \to \mu^+\mu^-$ analysis are $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$. These channels are chosen as they have similar final states to $B^0_{(s)} \to \mu^+\mu^-$, with common trigger, reconstruction, and selection procedures. They are also selected due to their relative abundance and well measured branching fractions. $B^+ \to J/\psi K^+$ decays (where $J/\psi \to \mu^+\mu^-$) are often triggered by the same muon trigger lines as the signal, resulting in a very similar trigger efficiency, whilst $B^0 \to K^+\pi^-$ has two particles in the final state and will have reconstruction and selection efficiencies very similar to the signal decays. Since both channels offer their individual advantages, a weighted average of the two normalisation channels is used.

Using the efficiencies calculated in Sect. 5.3, the number of expected events can be calculated using the following relationship to a normalisation channel, using the 'single event sensitivity' factor, α_s:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-\gamma) = \underbrace{\frac{\mathcal{B}_{norm}}{N_{norm}} \times \frac{\epsilon_{norm}}{\epsilon_{sig}} \times \frac{f_{norm}}{f_s}}_{\alpha_s} \times N_{B^0_s \to \mu^+\mu^-\gamma}, \tag{5.5.3}$$

where $\mathcal{B}_{norm}$ is the branching fraction, N_{norm} is the number of expected events, ϵ_{norm} is the efficiency and f_{norm} is the hadronisation factor of the normalisation channel. The 'single event sensitivity', α_s, incorporates these factors and relates the branching fraction, $\mathcal{B}(B^0_s \to \mu^+\mu^-\gamma)$ to the number of expected events, $N_{B^0_s \to \mu^+\mu^-\gamma}$.

The branching ratios for the normalisation channels, $\mathcal{B}_{norm}$, taken from the particle data group [16], are:

$$\begin{aligned} \mathcal{B}(B^+ \to J/\psi K^+) &= (1.010 \pm 0.028) \times 10^{-3}, \\ \mathcal{B}(J/\psi \to \mu^+\mu^-) &= (5.961 \pm 0.033) \times 10^{-2}, \\ \mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+) &= (6.021 \pm 0.170) \times 10^{-5}, \end{aligned} \tag{5.5.4}$$

and

$$\mathcal{B}(B^0 \to K^+\pi^-) = (1.96 \pm 0.05) \times 10^{-5}. \tag{5.5.5}$$

The normalisation channel yields, N_{norm}, are measured on the data sample that is collected in the same period as the signal sample (or signal MC relating to that period). For this feasibility study, the yields from Ref. [3] are used, the values of which are summarised in Table 5.6 for reference. It is assumed that $f_u = f_d$, and the value of f_s/f_d was taken to be the most up to date measurement at the time of the study[2], being 0.259 ± 0.015 [85].

Table 5.6: Normalisation Channel yields used for $B_s^0 \to \mu^+\mu^-\gamma$ study, from Ref. [3].

Normalisation Channel	Year	Aprox. Yield
$B^+ \to J/\psi K^+$	2011	347000
	2012	774000
	2015	167000
	2016	684000
$B^0 \to K^+\pi^-$	2011	7000
	2012	18000
	2015	9000
	2016	28000

The theoretical branching fractions for $B_s^0 \to \mu^+\mu^-\gamma$ for the two regions of interest, calculated from Ref. [24] are given below.

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)|_{m_{\mu\mu>4.9\,\mathrm{GeV}/c^2}} = 1.91 \times 10^{-10},$$

and

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)|_{m_{\mu\mu>4.5\,\mathrm{GeV}/c^2}} = 3.70 \times 10^{-10}.$$

Using this information, α_s can be calculated for each of the mass options. For the mass window used in $B_s^0 \to \mu^+\mu^-$ analysis, of $m_{\mu^+\mu^-} > 4.9$ GeV/c^2, the single event sensitivities for each year of MC simulation are given in Table 5.7, with the corresponding integrated luminosity.

Taking into account all of the data used in the previous $B_{(s)}^0 \to \mu^+\mu^-$ analysis (up to the last quarter of 2016), the combined total event sensitivity for $m_{\mu^+\mu^-} >$ 4.9 GeV/c^2 is:

$$\alpha_s^{tot}|m_{\mu^+\mu^-} > 4.9\,\mathrm{GeV}/c^2 = (6.79 \pm 0.43) \times 10^{-11}. \tag{5.5.6}$$

---

[2] The value of f_s/f_d has since been updated by the LHCb collaboration in Ref. [84], with a value of 0.2539 ± 0.0079 (for Run 2).

Table 5.7: The single event sensitivity, α_s, for each of the data taking years, with the corresponding integrated luminosity, for $m_{\mu^+\mu^-} >$ 4.9 GeV/c^2.

Data Taking Year	Single event sensitivity (α_s)	Luminosity (fb^{-1})
2011	$(4.35 \pm 0.31) \times 10^{-10}$	1
2012	$(1.99 \pm 0.14) \times 10^{-10}$	1.9
2015	$(6.85 \pm 0.54) \times 10^{-10}$	0.3
2016	$(1.70 \pm 0.12) \times 10^{-10}$	1.1

Table 5.8: The single event sensitivity, α_s, for each of the data taking years, with the corresponding integrated luminosity, for $m_{\mu^+\mu^-} >$ 4.5 GeV/c^2.

Data Taking Year	Single event sensitivity (α_s)	Luminosity (fb^{-1})
2011	$(6.25 \pm 0.45) \times 10^{-10}$	1
2012	$(2.86 \pm 0.20) \times 10^{-10}$	1.9
2015	$(9.85 \pm 0.77) \times 10^{-10}$	0.3
2016	$(2.45 \pm 0.17) \times 10^{-10}$	1.1

Enlarging the mass region by extending the low mass boundary down to $m_{\mu^+\mu^-} >$ 4.5 GeV/c^2 provides the single event sensitivities per year found in Table 5.8, and a combined total sensitivity of:

$$\alpha_s^{tot}|m_{\mu^+\mu^-} > 4.5\,\mathrm{GeV}/c^2 = (9.75 \pm 0.63) \times 10^{-11}. \tag{5.5.7}$$

Comparing the two values of α_s, it is evident that the single event sensitivity is larger for the lower mass window as expected. Using the α_s values and the branching fraction predictions, the number of expected events can be calculated using Eq. 5.5.3:

$$N_{B_s^0 \to \mu^+\mu^-\gamma_{|m_{\mu^+\mu^-} >\ 4.9\,\mathrm{GeV}/c^2}} = 2.83, \tag{5.5.8}$$

and,

$$N_{B_s^0 \to \mu^+\mu^-\gamma_{|m_{\mu^+\mu^-} >\ 4.5\,\mathrm{GeV}/c^2}} = 3.83. \tag{5.5.9}$$

This translates to ∼1 SM event per fb^{-1} for m$_{\mu^+\mu^-} >$ 4.9 GeV/c^2, and ∼ 1.35 SM events per fb^{-1} for m$_{\mu^+\mu^-} >$ 4.5 GeV/c^2 at $\sqrt{s}$ =13 TeV. This means that extending the mass window down to 4.5 GeV/c^2 provides around 35% more expected events than keeping it at 4.9 GeV/c^2.

5.6 Backgrounds

In the last section it was shown that extending the mass window down to 4.5 GeV/c^2 gives us sensitivity to ∼35% more $B_s^0 \to \mu^+\mu^-\gamma$ signal events than leaving the boundary at the existing $B_s^0 \to \mu^+\mu^-$ lower limit of 4.9 GeV/c^2. However, lowering the mass window would only be beneficial if the backgrounds don't increase considerably in the region 4.5 GeV/c^2 $< m_{\mu^+\mu^-} <$ 4.9 GeV/c^2, since the increase in signal sensitivity is not overwhelming. A quick look at the MC distributions for some of the $B_s^0 \to \mu^+\mu^-$ backgrounds shows that they grow almost exponentially going down in mass (Figs. 5.11, 5.12).

Enlarging the mass window would only be viable with careful study of these backgrounds, and any additional backgrounds that might enter in the lower mass region that are not taken into account in the $B_{(s)}^0 \to \mu^+\mu^-$ analysis. As well as this, all of the MC samples would need to be recreated with a lower mass boundary in order to obtain reliable

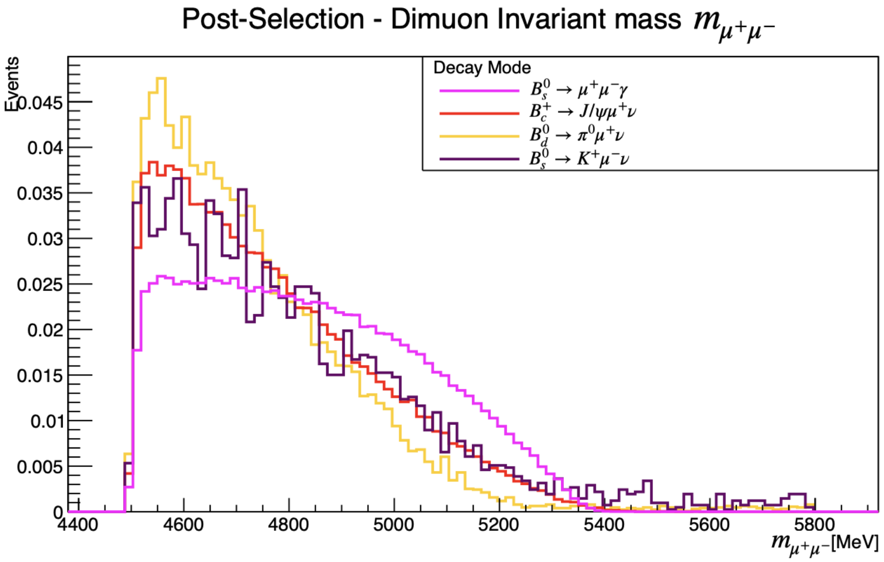


Figure 5.11: The post-selection MC distributions for $B_s^0 \to \mu^+\mu^-\gamma$ and three backgrounds from the $B_{(s)}^0 \to \mu^+\mu^-$ analysis; $B_c^+ \to J/\psi\mu^+\nu_\mu$, $B^0 \to \pi^-\mu^+\nu_\mu$ and $B_s^0 \to K^-\mu^+\nu_\mu$.

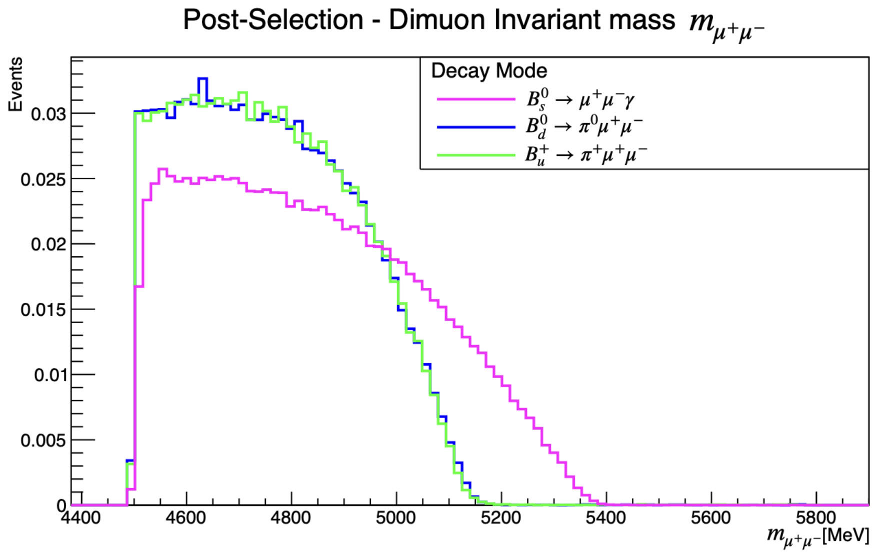


Figure 5.12: The post-selection MC distributions for $B_s^0 \to \mu^+\mu^-\gamma$ and two backgrounds from the $B_{(s)}^0 \to \mu^+\mu^-$ analysis; $B^0 \to \pi^0\mu^+\mu^-$ and $B^+ \to \pi^+\mu^+\mu^-$.

efficiencies for these backgrounds and avoid any boundary effects[3] at the $4.5\,\mathrm{GeV}/c^2$ point.

5.7 Conclusions

In conclusion, the partially reconstructed $B_s^0 \to \mu^+\mu^-\gamma$ analysis has been studied with Monte Carlo signal samples with 2011, 2012, 2015 and 2016 data taking conditions. The total signal efficiency has been found to be only 30% worse with respect to $B_s^0 \to \mu^+\mu^-$ when no explicit mass cut is taken into account, besides the one in generation of $m_{\mu^+\mu^-} > 4.5\,\mathrm{GeV}/c^2$. The variables exhibiting the largest discrepancies in their distributions are the invariant mass, χ^2_{IP} and the BDTS output, as expected.

The total signal efficiencies were computed for two lower mass boundary options: $m_{\mu^+\mu^-} > 4.5\,\mathrm{GeV}/c^2$ and $m_{\mu^+\mu^-} > 4.9\,\mathrm{GeV}/c^2$. Using these efficiencies and the theoretical branching fractions, the number of expected SM events per fb^{-1} at $\sqrt{s}$ =13 TeV was calculated to be 1.35 and 1 for the lower and higher cut options respectively. Taking into account the lower centre-of-mass energies for Run 1, 1 event per fb^{-1} at $\sqrt{s}$ =13 TeV translates to 2.8 signal events over the full dataset used for this study. Since the increase in signal sensitivity does not dramatically increase with an extended mass window, and the backgrounds grow significantly in that region, it was decided that $B_s^0 \to \mu^+\mu^-\gamma$ should be added to the existing framework for the next $B_{(s)}^0 \to \mu^+\mu^-$ analysis, without alteration to the mass window. That said, this would still be an interesting analysis to be performed in future. Following the success of the feasibility studies, $B_s^0 \to \mu^+\mu^-\gamma$ has been included as an additional observable to the analysis using the full Run 1 + Run 2 datasets, which will be described in the next chapter. From these studies it has been calculated that there will be an expected 7.5 SM $B_s^0 \to \mu^+\mu^-\gamma$ events in this full (Run 1 + Run 2) dataset using the existing invariant mass window of $4.9\,\mathrm{GeV}/c^2 < m_{\mu^+\mu^-} < 6\,\mathrm{GeV}/c^2$.

---

[3] Boundary effects are an artefact of binned histograms, where if a value is exactly on the boundary it will be assigned automatically into the higher or lower bin depending on the tools used. This can cause fluctuations at the boundary point.

Chapter 6

$B^0_{(s)} \to \mu^+\mu^-(\gamma)$ Analysis

6.1 Introduction

The $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays are flavour-changing neutral current (FCNC) processes, which means that they are highly suppressed in the SM, as discussed earlier in Sect. 2. The loop and helicity suppression of the decays means that they are extremely rare. The purely leptonic final states of these decays mean that their branching fractions can be predicted with very small uncertainties [23, 86–88]. Since the decays proceed via loop-diagrams, they can be significantly affected by the presence of non-SM particles (for example, an additional Higgs boson), entering at tree or loop level. Some examples of possible tree-level FCNC contributions to the $B^0_{(s)} \to \mu^+\mu^-$ decays are shown in Fig. 6.1. Any new heavy particles, or new processes, can significantly enhance or reduce the branching fraction values, meaning that the decays are very sensitive probes of physics beyond the SM (BSM).

That said, the results so far do not show significant deviations from the SM; the $B^0_s \to \mu^+\mu^-$ branching fraction measurements by LHCb, CMS and ATLAS are all consistent the SM prediction (all results are presented and have been combined in Ref. [4]). However, all of these measured values *do* lie below the SM value, with their average being around 2

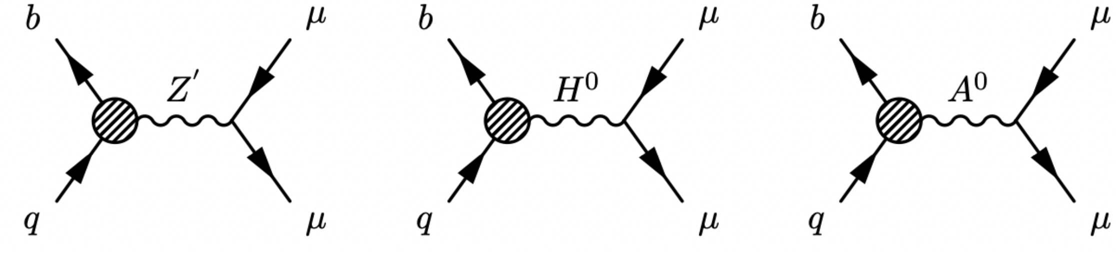


Figure 6.1: Examples of possible new physics contributions to the $B^0_{(s)} \to \mu^+\mu^-$ decays from Z' (left), and the two-Higgs doublet model with heavy scalar (centre) and pseudoscalar (right) bosons.

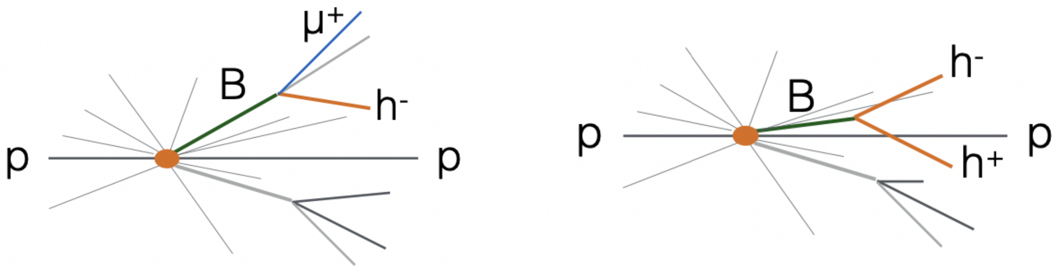
standard deviations away from the standard model prediction [4,89]. LHCb was the first of these experiments to announce results from the full Run 2 dataset [90], which are from the analysis described in this chapter. With the Run 2 results from CMS and ATLAS, it will be interesting to see how the agreement between the experiments, and indeed the combination of results evolve. The $B^0 \to \mu^+\mu^-$ decay also provides an interesting independent probe of new physics, and evidence of this decay by a single experiment has not yet been obtained. The ratio of the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ branching fractions has also been measured and will be presented in this chapter.

As discussed in Chapter 2, the physics of $b \to q\ell^+\ell^-$ transitions can be described using an effective-field theory approach. In this approach, physics beyond the standard model can be described as shifts to the Wilson coefficients of the operators of the effective Hamiltonian. The $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ channels are sensitive to the scalar and pseudoscalar operators $\mathcal{O}^{(\prime)}_{S,P}$, and to the operators $\mathcal{O}^{(\prime)}_{10}$ (as defined in Eqs. 2.2.5 and 2.2.6). The $B^0_s \to \mu^+\mu^-\gamma$ decay offers complementary sensitivity to the same operators, but also to $\mathcal{O}_9$ and to the electromagnetic-dipole operator $\mathcal{O}_7$, as well as to their chirality-flipped counterparts. That said, the sensitivity to $\mathcal{O}_7$ occurs only for values of the dilepton invariant mass squared close to zero [24, 91–95]. As discussed in the previous chapter, the invariant mass region to be investigated in this analysis is $4.9\,\text{GeV}/c^2 < m_{\mu^+\mu^-} < 6\,\text{GeV}/c^2$, which is not relevant for the $\mathcal{O}_7$ operator. This radiative decay can be searched for at LHCb by fully reconstructing the three decay products, but only when the photons are of high enough energy. For soft photons (lower energy), the partially reconstructed method described in Ref. [25], and in the previous chapter, can be used. This method involves reconstructing only the two muons of the decay, and selecting them using the $B^0_s \to \mu^+\mu^-$ selection criteria. The $B^0_s \to \mu^+\mu^-\gamma$ signal will therefore appear as a shoulder on the left hand side of the $B^0_s \to \mu^+\mu^-$ invariant mass peak, in a high $m_{\mu^+\mu^-}$ search complementary to the fully reconstructed analysis. This method is performed for the first time in the analysis discussed in this chapter. While there has been a search for $B^0 \to \mu^+\mu^-\gamma$ performed by the BaBar experiment [96], no search search for $B^0_s \to \mu^+\mu^-\gamma$ has yet been performed. When referring to $B^0_s \to \mu^+\mu^-\gamma$, it should be understood that this refers only to the initial state radiation component of the decay (see Sect. 2.2.3 for more details).

This chapter therefore presents a search for three decays; $B^0 \to \mu^+\mu^-$, $B^0_s \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$, with an analysis conducted using the full LHCb dataset to date.

6.2 Analysis Strategy

The aim of the work presented in this chapter is the measurement of the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ branching fractions and the search for the $B^0_s \to \mu^+\mu^-\gamma$ decay. Included in the same analysis is the measurement of the $B^0_s \to \mu^+\mu^-$ effective lifetime, although this will not be discussed in this thesis. Since the analysis is not optimised for the $B^0_s \to \mu^+\mu^-\gamma$



(a) Single misidentified background

(b) Double misidentified background

Figure 6.2: Different types of misidentified backgrounds for the $B^0_{(s)} \to \mu^+\mu^-$ decays [68]

search, the strategy will focus on the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays for which the analysis was designed.

The search uses the full Run 1 and Run 2 data sets from LHCb, which represents a dataset around 2.4 times larger than the size of that used in the previous analysis round presented in Ref. [3].

The strategy involves searching for the dimuon decay product using a loose event pre-selection with a tight muon requirement applied to the muon candidate tracks. The signal selection strategy has already been discussed in the previous chapter, where the effect of the signal selections on the $B^0_s \to \mu^+\mu^-\gamma$ channel was investigated. Further details will be discussed in Sect. 6.5. The selection is used to obtain a 'clean' sample, containing most of the $B^0_{(s)} \to \mu^+\mu^-$ decay candidates, while suppressing the background contributions by many orders of magnitude.

The main backgrounds that remain following the muon selection are those with a similar dimuon signature, primarily combinatorial events, or those with particles misidentified as muons, including the $B^0_{(s)} \to h^+h^{(\prime)-}$ peaking backgrounds (where $h, h' = K, \pi$). The backgrounds could have either one muon and a single misidentified hadron, or two misidentified hadrons. Figure 6.2 shows a sketch of these two types of contribution. The combinatorial backgrounds affect both the B^0_s and B^0 mass regions, while the misidentified backgrounds mainly affect only the B^0 mass region. Other backgrounds from partially reconstructed 3-body B decays populate the left mass sideband, so are of particular significance for the $B^0_s \to \mu^+\mu^-\gamma$ decay which also appears in that region.

The rejection power of the $B^0_{(s)} \to h^+h^{(\prime)-}$ peaking background is strongly dependent on the particle identification (PID) selection. In the Run 1 analysis [97] performed at LHCb, the emphasis was put on keeping a high efficiency for the $B^0_s \to \mu^+\mu^-$ signal. This translated into a loose PID selection, so as to retain as much of the signal as possible. However, this did mean that the sensitivity to the $B^0 \to \mu^+\mu^-$ signal was reduced. In the last analysis the PID cuts were re-optimised to maximize the sensitivity to the $B^0 \to \mu^+\mu^-$ channel, resulting in a large reduction of the purely hadronic background with only a small reduction in signal. The muon PID selection has not changed in this updated analysis and will be described in Sect. 6.8.1.

The evaluation of the pion ($\pi \to \mu$) and kaon ($K \to \mu$) misidentification (herein referred to as MisID) rates have, however, been changed since the last analysis round. Since pions and muons have a small mass difference, distinguishing between the two is a challenging process. As mentioned in the previous chapter, the Cerenkov rings in the RICH detector are very close and sometimes overlap in the relevant momentum range, making them hard to identify. Added to this, the bend in the track caused by the charge of the muons is sometimes not obvious due to the lack of tracking detectors between the TT and T stations in the detector. For these reasons, pions can be easily misidentified as muons. To accurately estimate the background contributions for this analysis, the percentage of misidentified particles must be carefully estimated and taken into account.

The signal and background PID efficiencies are usually calculated using a dedicated software package called PIDCalib [98], which provides access to the calibration samples of electrons, muons, pions, kaons and protons. However, it was found that using the PIDCalib software did not take into account the large pion and kaon track momentum degradation when the hadron decays to a muon (known as decays in flight) [99–101]. This effect was responsible for a bias of ∼40% of the MisID rates given by PIDCalib. For this reason, a dedicated procedure has been developed for this analysis to extract the MisID rates from $D^{*\pm} \to D^0(\to KK\pi)\pi^\pm$ control samples that are reconstructed without PID from dedicated trigger channels. With this method, the decays in flight can be taken into consideration. More details on this strategy will be given in Sect. 6.9.1. The $B^0_{(s)} \to h^+h^{(\prime)-}$ background yields are then determined from decays selected in data without the muon requirement, via a convolution of the p and p_{T} distributions of the hadrons with the $\pi \to \mu$ and $K \to \mu$ MisID probability maps.

In addition to the purely hadronic $B^0_{(s)} \to h^+h^{(\prime)-}$ backgrounds, other relevant specific decays of b hadrons, referred to as *exclusive* backgrounds, also contribute. The exclusive backgrounds mainly polluting the left sideband include:

- $\Lambda^0_b \to p\mu^-\nu_\mu$,
- $B^0 \to \pi^-\mu^+\nu_\mu$,
- $B^0_s \to K^-\mu^+\nu_\mu$,
- $B^+ \to \pi^+\mu^+\mu^-$,
- $B^0 \to \pi^0\mu^+\mu^-$ and
- $B^+_c \to J/\psi\mu^+\nu_\mu$.

These backgrounds arise when one of the hadrons in the decay is misidentified as a muon (see Fig. 6.2a), or in the case of $B^0 \to \pi^0\mu^+\mu^-$ and $B^+ \to \pi^+\mu^+\mu^-$, when the two real muons form a good vertex and pass the signal selection criteria. These backgrounds must be carefully estimated so as not to introduce bias in the determination of the combinatorial background from the mass fit. In order to calculate the yields of these backgrounds, PIDCalib or the revised $\pi \to \mu$ and $K \to \mu$ MisID method are used to calculate the PID efficiencies from data, which are then used with the other efficiencies calculated with MC and the measured (or predicted) branching fractions. This information is used to normalise the backgrounds relative to $B^+ \to J/\psi K^+$ in order to obtain yields, which will be computed in Sect. 6.10. In contrast to the method used in Ref. [3], the $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ components that were previously merged together have this time been split and estimated separately.

The majority of the background impacting the $B^0_{(s)} \to \mu^+\mu^-$ (and $B^0_s \to \mu^+\mu^-\gamma$) mass regions is the combinatorial, that is, two muons coming from two different b hadron decays. In the dataset used for this analysis, around 3 billion of this type of event are expected, making it essential that this background is suppressed as much as possible. The rejection of this combinatorial background is mainly done by the BDT (introduced in the previous chapter), which is used at the final stages of the analysis. The BDT is a multivariate classifier that is used to divide the selected sample into more or less signal-like subsamples. The BDT uses seven discriminating variables associated to the candidate topology and track isolation, including the impact parameter resolution and muon isolation criteria, which are able to suppress the muon tracks from primary vertices (rather than SVs), or from b and c semileptonic decays. The combinatorial backgrounds often have extra tracks close to the muon tracks, while real $B^0_{(s)} \to \mu^+\mu^-$ decays do not, meaning that the track isolation variables give the BDT most of its discriminating power.

The criteria for the BDT, described in more detail in Sect. 6.6, were optimised in the previous analysis and have not been updated for this round. The dimuon invariant mass and the BDT are the most discriminating variables (that is, the best at separating the signal from the background), and are therefore used in the final maximum likelihood fit. For reference, Fig. 6.3 shows the final fit from the previous analysis round, for signal with a BDT value greater than 0.5 (the most ‘signal-like’ decays).

Once the fit shapes for the signal and background, and the background yields have been computed, a final two dimensional maximum likelihood fit is performed to search for the $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$ decays. The mass shape of the $B^0_{(s)} \to \mu^+\mu^-$ decays are described by a Crystal Ball function [102] containing four parameters: the mean, resolution, and two tail parameters.

The branching fractions are measured relative to other b hadron decays namely $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $B^0 \to K^+\pi^-$ which complement each other thanks to their different final states (as described in the previous chapter). The normalisation enables the yields obtained from the fit to be converted into branching fractions using the updated value of the hadronisation fraction f_s/f_d at 13 TeV, computed recently by LHCb [84].

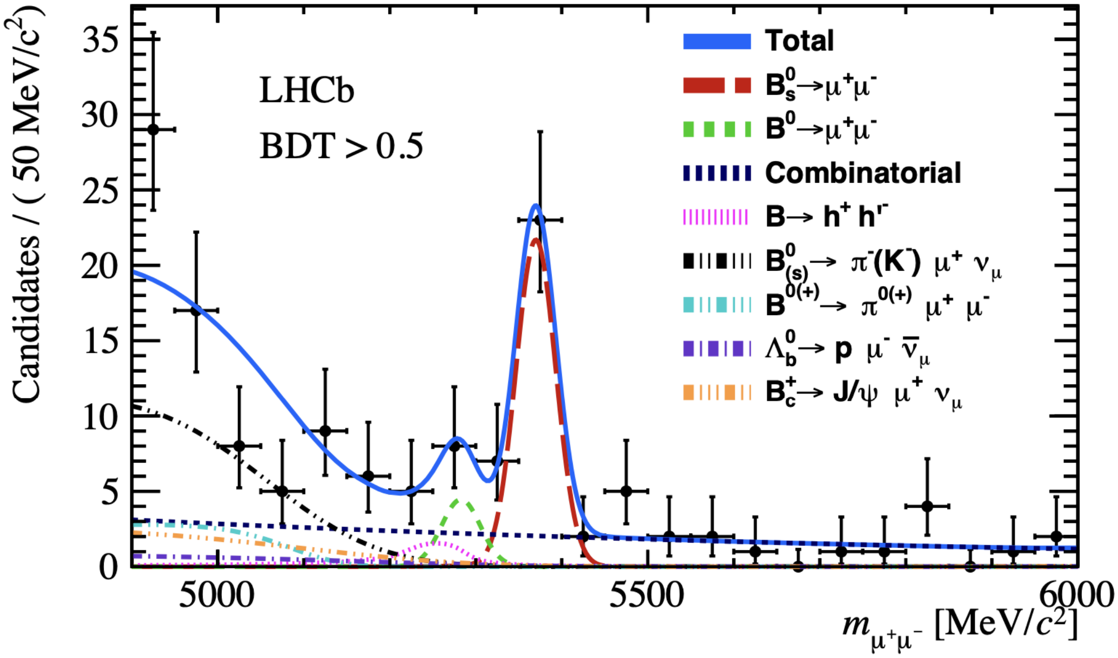


Figure 6.3: Mass distribution of the selected $B^0_{(s)} \to \mu^+\mu^-$ candidates (black dots) with BDT > 0.5, from the last analysis round. The result of the fit (blue) is overlaid and the different components detailed.

6.3 Data and Monte Carlo Samples

6.3.1 Data Samples

This analysis is performed using pp collision data collected by the LHCb experiment in 2011 and 2012 at a centre of mass energy ($\sqrt{s}$) of 7 TeV and 8 TeV respectively (known collectively as Run 1), and 2015, 2016, 2017 and 2018 at $\sqrt{s} = 13$ TeV (known collectively as Run 2). Table 6.1 lists the integrated luminosities collected each year from the dimuon stream, which is used for the signal and normalisation modes.

Table 6.1: Integrated luminosities of the data samples used in this analysis. For 2016, 2017 and 2018 no exact luminosity can be quoted since the luminosity tool is not yet calibrated for those years.

Year	$\sqrt{s}$ [TeV]	Integrated luminosity (pb^{-1}) Dimuon stream
2011	7	962.7
2012	8	1973.2
2015	13	285.3
2016	13	~ 1650
2017	13	~ 1710
2018	13	~ 2200

6.3.2 Monte Carlo Samples

Monte Carlo (MC) simulation is used for the computation of background yields and signal efficiencies which enter into the normalisation. The simulation samples used are summarised in Table 6.2. All channels are simulated with conditions (mentioned above) of each of the data taking years, being 2011, 2012, 2015, 2016, 2017 and 2018.

Table 6.2: Summary of Monte Carlo samples used in this analysis.

		Number of events	
Purpose	Channel	Run 1	Run 2
Calculation of Signal Efficiencies	$B^0_s \to \mu^+\mu^-$	2 M	6 M
	$B^0 \to \mu^+\mu^-$	2.5 M	6 M
	$B^0_s \to \mu^+\mu^-\gamma$	1.5 M	9 M
Calculation of Normalisation Efficiencies	$B^0 \to K^+\pi^-$	9.3 M	14 M
	$B^+ \to J/\psi K^+$	5.8 M	26 M
Calculation of Background Efficiencies	$B^0_s \to J/\psi\phi$	1.3 M	18 M
	$B^0_s \to K^+K^-$	8.5 M	14 M
	$B^0_s \to K^+\pi^-$	7 M	14 M
	$B^0 \to \pi^+\pi^-$	8.5 M	14 M
	$B^0 \to \pi^-\mu^+\nu_\mu$	6 M	19 M
	$B^0_s \to K^-\mu^+\nu_\mu$	6 M	9 M
	$\Lambda^0_b \to p\mu^-\nu_\mu$	2 M	14 M
	$B^0 \to \pi^0\mu^+\mu^-$	2 M	7 M
	$B^+ \to \pi^+\mu^+\mu^-$	2 M	7 M
	$B^+_c \to J/\psi\mu^+\nu_\mu$	2 M	7 M

6.4 Trigger

Events passing the trigger can be divided into three categories:

- TOS (Triggered On the Signal) - Meaning that the signal is able to trigger the event, *i.e.* an event is triggered by a dimuon trigger, where both of the muons are from the signal,
- TIS (Triggered Independently of the Signal) - Meaning that the rest of the event is able to pass the trigger, *i.e.* an event is triggered by a dimuon trigger, where both of the muons causing the trigger are not used in the reconstruction of the $B^0_{(s)}$ decay,
- TOB (Triggered On Both) - Meaning that both the signal and the rest of the event are needed to trigger, *i.e.* an event is triggered by a dimuon trigger, but one of the muons comes from the signal and one comes from the rest of the event.

Since the decays in this analysis are so rare, no specific trigger requirement is imposed in order to retain as much signal is possible. This means that all of the events that pass the trigger, regardless of the specific line, are kept. Despite this, there will still be some signal events that are not triggered at all, meaning that the trigger efficiency must be evaluated in order to correctly normalise the signal yields. This is done using a data-driven trigger efficiency estimation using events that are triggered on both TIS and TOS (known as the `TISTOS` method), described in detail in Ref. [103]. The trigger efficiencies for the signals are presented in Sect. 6.11.

6.5 Signal Selection

The rareness of the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays makes it challenging to distinguish them from backgrounds, even though the decay signatures are relatively straightforward. The signal selection procedure must accurately get rid of as much background as possible, whilst retaining almost all of the signal. The selection begins with cuts to reduce the sample size, but maintain a very high signal efficiency. The dataset is then passed through two multivariate classifiers; one used to further reduce the sample size and one used in the final classification. This section will describe each step of the candidate selection process in greater detail.

6.5.1 Stripping

The triggered events pass through a general offline selection called the stripping (see Sect. 4.1), where cuts are applied to reduce the size of the dataset while keeping the signal efficiency as high as possible. The stripping cuts help reject background events such as non-physical backgrounds (with cuts on the track χ^2/ndf, the DOCA[1] between the two tracks and the χ^2 of their vertex), quasi-elastic pp collisions (with cut on minimum p_{T} of the B meson candidate) and prompt muons produced in inelastic collisions (with cuts on the significance of the impact parameter for the muon candidates, as well as flight distance significance of the secondary vertex). The three stripping lines relevant for the $B^0_{(s)} \to \mu^+\mu^-$ analysis are $B^0_s \to \mu^+\mu^-$, $B^0_{(s)} \to h^+h^{(\prime)-}$ and $B^+ \to J/\psi K^+$, whose selections are kept as similar as possible. In order to ensure that the $B^0_{(s)} \to h^+h^{(\prime)-}$ selection is similar to $B^0_s \to \mu^+\mu^-$, the hadrons are required to be within the muon acceptance. The stripping cuts, and the channels they are applied to are summarised in Table 6.3. Some of the cuts are different for Run 1 and Run 2 as the reconstruction improved in that time meaning the cuts could be loosened slightly.

For a detailed description of the variables themselves and the purpose of each of the selection cuts, see Sect. 5.4.2, where all of these cuts were used to find the overall selection efficiency for the $B^0_s \to \mu^+\mu^-\gamma$ signal.

---

[1] Distance of Closest Approach between the two muon daughter tracks.

Table 6.3: Stripping selections for $B^0_{(s)} \to \mu^+\mu^-$, $B^0_{(s)} \to h^+h^{(\prime)-}$ and $B^+ \to J/\psi K^+$ channels; DOCA is the distance of closest approach between the two tracks, VDS is the secondary vertex flight distance significance. Cut values in parenthesis for track χ^2/ndf and ghost probability show softer cuts used starting from 2015 data taking. Tighter cut values in squared parentheses for χ^2_{IP} and VDS are used starting from 2016 data taking.

Variable	applied to	Requirement $B^0_{(s)} \to \mu^+\mu^-$ & $B^0_{(s)} \to h^+h^{(\prime)-}$	Requirement $B^+ \to J/\psi K^+$
track χ^2/ndf	μ/h	< 3 (< 4)	< 3 (< 4)
ghost probability		< 0.3 (< 0.4)	< 0.3 (< 0.4)
DOCA		$< 0.3\,\text{mm}$	$< 0.3\,\text{mm}$
χ^2_{IP}		> 9 [> 25]	> 9 [> 25] (μ: > 25)
p_{T}		$> 250\,\text{MeV}/c$	$> 250\,\text{MeV}/c$
isMuon	μ only	true	true
vertex χ^2	$\mu\mu/hh$	< 9	< 9
VDS		> 11 [> 15]	> 11 [> 15]
$\lvert m(\mu\mu) - m_{J/\psi}\rvert$			$< 100\,\text{MeV}/c^2$
$\lvert m(\mu\mu) - m_B\rvert$		$< 1200\,\text{MeV}/c^2$	
$\lvert m(hh) - m_B\rvert$		$< 500\,\text{MeV}/c^2$	
χ^2_{IP}	$B^0_{(s)}$	< 25	< 25
t		$< 9 \cdot \tau(B^0_s)$	
p_{T}		$> 350\,\text{MeV}/c$ (hh)	
$\lvert m(J/\psi K) - m_B\rvert$			$< 500\,\text{MeV}/c^2$
vertex χ^2			< 45

6.5.2 Offline Cuts

Following the stripping, fiducial or ‘offline’ cuts are applied to further reduce the size of the samples by getting rid of any non-physical candidates in the data sample; those that are artefacts of the reconstruction. These further cuts are summarised in Table 6.4.

Table 6.4: The offline selection cuts, applied after the stripping. The BDTS discriminant was described in the previous chapter.

Variable	applied to	Requirement $B^0_{(s)} \to \mu^+\mu^-$ & $B^0_{(s)} \to h^+h^{(\prime)-}$	applied to	Requirement $B^+ \to J/\psi K^+$
p	μ	< 500 GeV/c	μ/h	
p_{T}		> 250 MeV/c and < 40 GeV/c		
vertex χ^2	$\mu\mu$/hh	< 9	$\mu\mu$	< 9
VDS		> 11 (> 15)		> 11 (> 15)
ΔM		$\lvert M(hh, \mu\mu) - m_B \rvert <$ 500 MeV/c^2		$\lvert M(\mu\mu) - m_{J/\psi} \rvert <$ 100 MeV/c^2
p_{T}	$B^0_{(s)}$	> 500 MeV/c	B^+	
BDTS		> 0.05		> 0.05
ΔM				$\lvert M(J/\psi K) - m_B \rvert <$ 100 MeV/c^2

Keeping the cuts as consistent as possible between the signal and normalisation channels means that uncertainties are minimised. For the $B^+ \to J/\psi K^+$ mode, the χ^2 of the secondary vertex is substituted by the χ^2 of the J/ψ vertex, the flight distance is computed between the J/ψ vertex and the primary vertex, and the DOCA is computed between the two muons from the J/ψ decay. This means that the distributions of all of the variables are similar for the signal and normalisation channels resulting in similar efficiencies.

The BDTS was described in Sect. 5.3.3, and the cut on the BDTS output of > 0.05 retains $\sim$93% of the signal, while rejecting $\sim$70% of the background.

The events retained after the stripping and fiductial cuts detailed in Tables 6.3 and 6.4 are then subject to a J/ψ veto and particle identification cuts.

The J/ψ veto is included in the signal selection specifically to reject background from $B_c^+ \to J/\psi\mu^+\nu_\mu$ decays, where the J/ψ decays to two muons. If one of the muons from a signal candidate event can be combined with a non-signal muon that has a reconstructed invariant mass close to the J/ψ mass, then it is likely that the candidate came from this background and should therefore be rejected. For this reason, signal candidates where either muon forms a dimuon candidate with a mass within 30 MeV/c^2 of the J/ψ mass are rejected.

A particle identification cut on muon candidates further reduces the contribution from the exclusive backgrounds. Accurate particle identification is imperative for estimation of the background yields, as we need to know how often other particles look like signal by ‘faking’ muons. The particle identification process for the signal and backgrounds processes will be explained in more detail in Sect. 6.8.

Following the PID selection, the signal has been selected as much as possible, and any remaining combinatorial background cannot be reduced by means of regular cuts without loosing a significant amount of signal sensitivity. However, the most 'signal like' events can be categorised using a second boosted decision tree (the BDT), which will be described in the following section.

6.6 Signal Classification

In the previous chapter (specifically Sect. 5.3.3) the first of the boosted decision trees used in this analysis was described; the BDTS. This was used for signal selection and rejection of combinatorial background. After all of the other cuts have been implemented, a second BDT, this time used for classification rather than selection, is used. This BDT (simply called the BDT), uses the full event information to separate the signal and combinatorial background in the final stage of the analysis. The BDT has a similar response for $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h^{(\prime)-}$ decays, the latter being carefully estimated using PID techniques.

In the last section, *track isolation* was briefly introduced as an important means of signal/background discrimination. An isolated track is one that is not associated with any other tracks in an event. Muon isolation is one of the most powerful tools to distinguish the signal from the combinatorial background, since the muons from $B^0_{(s)} \to \mu^+\mu^-$ decays are expected to be relatively isolated from the other tracks of the event. The background decays on the other hand, are likely to have one or more daughter particles produced close to the muon(s).

Non-isolated tracks are defined as those tracks that originate from the same decay tree as the muon candidate, as opposed to the isolated tracks, which do not share a b quark (or c quark) ancestor with the muon. Two isolation variables are used in the BDT, which will be described in the next section.

6.6.1 Long Track and VELO Isolation

Long tracks are tracks which are reconstructed using information from all components of LHCb's tracking system. The long track isolation is a quantity which is sensitive to how close other long tracks are to the muon candidate (excluding the companion muon). If a long track has others close to it, likely to belong to the same decay tree, it is probable that is originated from a semileptonic B decay. Isolated and non-isolated long tracks are separated using a boosted decision tree (iso_{LT}) trained on simulated samples. Non-isolated tracks are taken from simulated samples of $b\bar{b} \to \mu^+\mu^- X$ candidates (where X indicates additional decay products), while isolated tracks are from simulated $B^0_{(s)} \to \mu^+\mu^-$ decays. The list of input variables were optimized in the previous analysis based on Run 1, 2015 and 2016 data. Details of the optimization procedure can be found in Refs. [79] and [104]. The discriminating variables used are:

1. The minimum of the square root of the χ^2_{IP} of the track with respect to any primary vertex (PV) of the event
2. The distance between track-muon-vertex and PV
3. The distance between track-muon-vertex and $B^0_{(s)} \to \mu^+\mu^-$ vertex
4. The DOCA between track and muon
5. The angle between track and muon
6. A variable defined as

   $$f_c = \frac{|\vec{p}_\mu + \vec{p}_{\mathrm{trk}}| \sin(\alpha^{\mu+\mathrm{trk,PV}})}{|\vec{p}_\mu + \vec{p}_{\mathrm{trk}}| \sin(\alpha^{\mu+\mathrm{trk,PV}}) + p_{\mathrm{T},\mu} + p_{\mathrm{T,trk}}} \tag{6.6.1}$$

   where $\alpha^{\mu+\mathrm{trk,PV}}$ is the angle between the sum of the muon and track momenta $(\vec{p}_\mu, \vec{p}_{\mathrm{trk}})$, and the direction defined by the PV and track-muon-vertex [105]. $f_c \to 0$ if the track-muon system originates from the PV
7. The absolute difference between the azimuthal angles of track and muon
8. The absolute difference between the pseudorapidities of track and muon
9. The transverse momentum of the track

This iso_{LT} returns an isolation score value for a pair of tracks: the track whose isolation it is computed for and the muon it is computed against. This is then used in the classification BDT as a distinguishing variable between signal and background.

The track segments composed of hits in the VELO are not used to form the long tracks. Instead, an additional isolation for VELO tracks is used, following the same procedure as with the iso_{LT}, and using the first six input variables listed above.

From these two isolation BDTs, each muon track is assigned a set of isolation scores. The larger the score, the less isolated the muon track is from others in the event.

6.6.2 The Classification BDT

The isolations described in the last section are used as inputs, along with five other variables, for a final multivariate classifier, the BDT. The purpose of this BDT is to separate the $B^0_{(s)} \to \mu^+\mu^-$ signals from the combinatorial background. This BDT is trained using the same samples described in the previous section, that is, $b\bar{b} \to \mu^+\mu^- X$ and $B^0_{(s)} \to \mu^+\mu^-$ simulated events. The seven variables used as inputs to the BDT are:

- Long track isolation,
- VELO track isolation,
- The vertex χ^2 of the B candidate,
- The impact parameter significance of the B candidate with respect to the PV,
- The angle between the B direction and the vector joining the primary and secondary vertices,
- $\sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ and $\Delta\eta$ are the azimuthal angle and pseudorapidity differences between the two muons,
- The smallest value among the muon impact parameter significance of the two muons with respect to the primary vertex associated to the $B^0_{(s)} \to \mu^+\mu^-$ candidate.

The last three variables in the list were also used in the previous selection BDTS (see description in Sect. 5.3.3). After the BDT has been trained, it is flattened for signal and divided into bins of increasing sensitivity (with the combinatorial background peaking at 0).

A few different options for binning schemes were investigated, and the optimum was found to be the scheme where the most sensitive BDT bins have a few events in the right mass sideband (so that the fit gives unbiased branching fraction estimates), but not too many (in order not to lose statistical sensitivity). Based on these criteria, the optimal binning of the BDT, was found to be;

$$[0.00, 0.25, 0.40, 0.50, 0.60, 0.70, 1.00].$$

The final maximum likelihood fit it performed individually for these BDT bins, meaning that the fraction of signal candidates per bin must be calculated. This process is called the BDT calibration and is performed using $B^0_{(s)} \to h^+h^{(\prime)-}$ data, where each h is either a pion or kaon.

6.7 Background Processes

In this section backgrounds of concern for the $B^0_{(s)} \to \mu^+\mu^-$ analysis are described. As discussed in Sect. 6.2, there are several different decays contributing to this category:

- $B^0_{(s)} \to h^+h^{(\prime)-}$ $(h, h' = K, \pi)$, with both kaon and pion misidentified as muons, which mainly pollute the B^0 signal region;
- $B^0 \to \pi^-\mu^+\nu_\mu$, with the pion misidentified as a muon, which pollutes the left mass sideband only;

- $B^0_s \to K^-\mu^+\nu_\mu$, with the kaon misidentified as a muon, which pollutes the left mass sideband only;

- $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$, with two true muons in the final state, which pollute the left sideband only;

- $\Lambda^0_b \to p\mu^-\nu_\mu$, with the proton misidentified as a muon, which mainly pollutes the left mass sideband but has a tail extending to the B^0 and B^0_s mass regions;

- $B^+_c \to J/\psi\mu^+\nu_\mu$, with two true muons in the final state, which mainly pollutes the left mass sideband but has a tail extending almost to the full mass region.

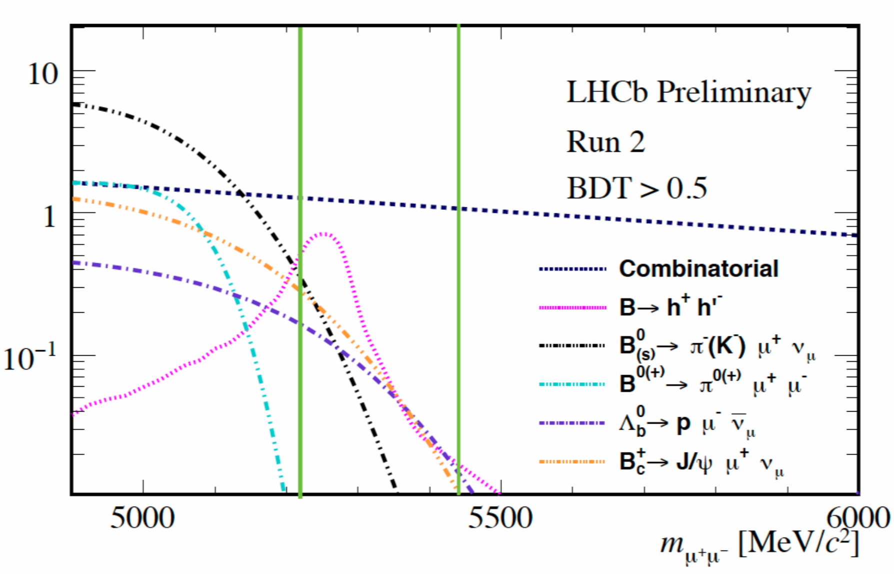


Figure 6.4: Sketch of the relevant background contributions, from the previous analysis round.

All of these backgrounds are sketched in Fig. 6.4. These backgrounds appear as signal when one or more of the decay products are misidentified as muons. For that reason, the pion, kaon, and proton to muon MisID rates have to be carefully evaluated. In the next section the identification efficiencies for muons will be calculated, as well as the MisID efficiencies for these backgrounds. Since the fit to measure the signal branching ratios is performed in bins of BDT, all of the above background sources also have to be estimated individually for each BDT bin.

6.8 Particle Identification

The PID performance for all decays, both signals and backgrounds, over all relevant data-taking years, has been evaluated from data control samples. For the signal decays, being $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$, the particle identification performance relates to the efficiency of correctly identifying the two decay products as muons. The muon identification efficiencies will be calculated in Sect. 6.8.1.

For the exclusive backgrounds listed in the previous section, it is necessary to compute how often any other particles types, which could be decay products of the backgrounds, are misidentified as muons. From the relevant backgrounds, the MisID efficiencies of interest are $\pi \to \mu$, K$\to \mu$ and p$\to \mu$. The MisID efficiencies of these will be discussed in Sect. 6.9.

6.8.1 Muon Identification

Correctly identifying the two daughter muons from the signal decays is an integral part of this analysis. The muon identification procedure consists of two steps: the requirement for the muon candidate to satisfy the isMuon algorithm, and a more restrictive PID selection using the ProbNN variables, which are based on a combination of RICH, CALO and MUON detector information. A detailed explanation of the muon ID variables was given earlier in Sect. 4.4. The muon selection criteria were optimised in Ref. [79], ensuring a high signal efficiency and a strong rejection of backgrounds. The best performance for the $B^0_{(s)} \to \mu^+\mu^-$ analysis was obtained with

$$\mathrm{PID}_{\mu,4} \equiv \mathrm{ProbNN}_\mu \times (1 - \mathrm{ProbNN}_p) \times (1 - \mathrm{ProbNN}_K) > 0.4 \tag{6.8.1}$$

for Run 1 and 2015, and

$$\mathrm{PID}_{\mu,8} \equiv \mathrm{ProbNN}_\mu \times (1 - \mathrm{ProbNN}_p) \times (1 - \mathrm{ProbNN}_K) > 0.8 \tag{6.8.2}$$

for the rest of Run 2.

Here, $\mathrm{ProbNN}_{\mu,K,p}$ are the likelihoods of the candidate being a muon, kaon or proton respectively. The ProbNN_K term is very effective in rejecting the $B^0_{(s)} \to h^+h^{(\prime)-}$ backgrounds, while ProbNN_p has been introduced specifically to reject $\Lambda^0_b \to p\mu^-\nu_\mu$.

The ProbNN variables are likelihood outputs of neural networks, which were updated and improved over time, meaning that a tighter cut could be used and was found to be optimal for most of Run 2. Compared to $\mathrm{PID}_{\mu,4}$ in Run 1, the $\mathrm{PID}_{\mu,8}$ requirement in Run 2 has the same or lower $\pi, K, p \to \mu$ probabilities (and therefore better background rejection) with similar muon efficiency.

The efficiency for muons to pass the above selection is evaluated with the PIDCalib package, which provides calibration samples for the distributions of the signal mode, and thus allows the efficiency of the above selections to be accurately computed. For muons, samples are provided from $J/\psi \to \mu^+\mu^-$ decays from b modes. The muons from this sample are selected as probes by a tag-and-probe method to ensure no bias arises from their selection.

The tag-and-probe method uses well known resonances, in this case J/ψ. One of the resultant muons from the decay is labeled as the 'tag' muon, and the other as the 'probe'. The former is a well identified triggered muon under tight selection criteria, and latter are muon candidates that have tracks compatible with the J/ψ resonance. The probe muon can either pass or fail the criteria for which the efficiency is to be measured.

To ensure that no bias arises from the trigger requirements, the probe muon candidate is also required to satisfy the condition that the track has independently triggered the event at both the hardware and software level. Further details of the trigger unbias can be found in Refs. [69] and [106].

The kinematic distribution of the probe muon in the control sample is different from the one belonging to $B^0_{(s)} \to \mu^+\mu^-$. For this reason, the single muon efficiency is evaluated from PIDCalib in 11 bins in momentum p, 4 bins in transverse momentum p_{T}, and 6 bins in nTracks (where nTracks is the number of long tracks in a candidate event), which can then be convoluted with the proper signal MC spectrum. The bin boundaries for each of the variables are chosen be:

$$p : [0, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60]\,\mathrm{GeV},$$

$$p_{\mathrm{T}} : [0.8, 1.7, 3, 5, 40]\,\mathrm{GeV},$$

$$\mathrm{nTracks} : [0, 65, 90, 115, 140, 180, 10000].$$

Efficiency maps from the calibration samples are created separately for each magnet polarity (MagUp and MagDown), and then averaged. These efficiencies are shown in Fig. 6.5 for Run 1 and Fig. 6.6 for Run 2, with each of the years shown separately as well as the combined efficiencies for the Run. These plots have been made as a 2D projection (of p, p_T) over all nTracks bins, for ease of visualisation. For the combined Run efficiencies, an average of the efficiency for all years is performed, weighted by the luminosity of that year.

Using the 3D maps, the PID selection efficiencies for our MC samples can be calculated by performing convolutions of the maps to the $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$ MC. This process involves matching the values of the p, p_{T} and nTracks from the muons in the MC event to the corresponding bins from the calibration samples, and finding the associated efficiency. For each of the events in the MC signal samples, the efficiency from PIDCalib is applied to both muons in this way, so that the output of the convolution gives the overall signal efficiency under the muon identification selection. These efficiencies are computed given the geometrical acceptance of the candidate tracks in the muon detector.

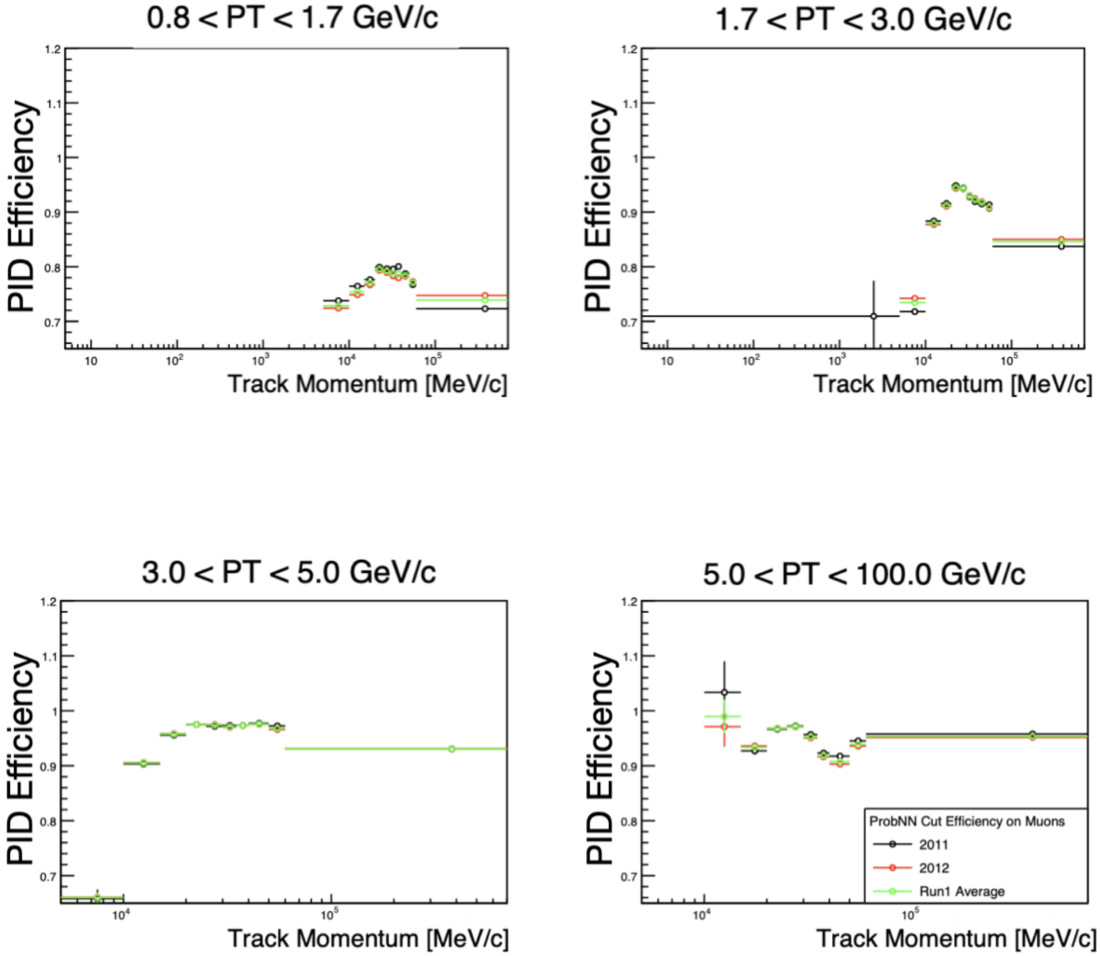


Figure 6.5: Muon efficiency vs momentum for the four p_{T} bins of the probe track, evaluated from PIDCalib for Run 1 calibration data, split by year. MagUp and MagDown performance tables have been averaged. Binned efficiencies greater than 1 can appear as a result of the sWeight procedure used in efficiency evaluation. This technique is explained in Sect. 6.8.2.

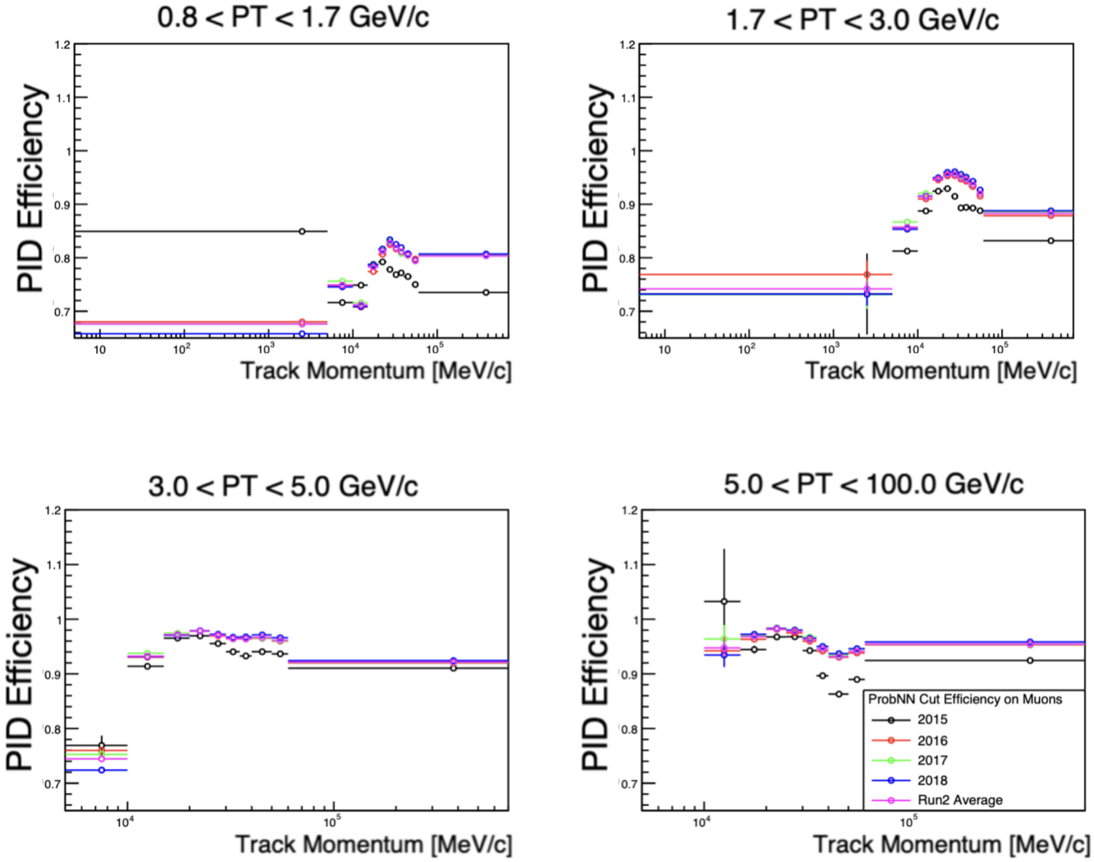


Figure 6.6: Muon efficiency vs momentum for the four p_{T} bins of the probe track, evaluated from PIDCalib Run 2 calibration data, split by year. MagUp and MagDown performance tables have been averaged. Binned efficiencies greater than 1 can appear as a result of the sWeight procedure used in efficiency evaluation. This technique is explained in Sect. 6.8.2.

For reference, the MC distributions of the three signals using the same binning scheme as the calibration sample can be found in Appendix B. The total PID efficiency for each of our signals can then be found by finding the average over all of the events in the MC sample.

The resulting efficiencies for Run 1 and Run 2, split by BDT bin, are listed in Table 6.5 for $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$, and $B^0_s \to \mu^+\mu^-\gamma$.

Table 6.5: The ISMUON $\times$PID$_{\mu,4}$ efficiency on $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$ for Run 1 and 2015, and ISMUON $\times$PID$_{\mu,8}$ for 2016, 2017 and 2018 data, as evaluated via convolution of the single muon efficiencies from PIDCalib with the MC signal spectra. Efficiencies are shown over all BDT bins and each bin individually. The uncertainty shown is statistical only, the systematic uncertainties are not yet included.

		Efficiency	
Channel	BDT Bin	Run 1	Run 2
$B^0_s \to \mu^+\mu^-$	ALL	81.12% ± 0.031%	84.20% ± 0.012%
	1	77.70% ± 0.031%	81.16% ± 0.011%
	2	78.67% ± 0.030%	82.45% ± 0.011%
	3	79.98% ± 0.030%	83.72% ± 0.011%
	4	81.15% ± 0.031%	84.49% ± 0.012%
	5	82.28% ± 0.032%	85.20% ± 0.012%
	6	85.51% ± 0.037%	87.66% ± 0.014%
$B^0 \to \mu^+\mu^-$	ALL	80.80% ± 0.031%	83.97% ± 0.011%
	1	77.10% ± 0.031%	80.75% ± 0.011%
	2	78.24% ± 0.030%	82.28% ± 0.011%
	3	79.75% ± 0.030%	83.66% ± 0.011%
	4	80.78% ± 0.031%	84.41% ± 0.012%
	5	82.06% ± 0.032%	85.15% ± 0.012%
	6	85.32% ± 0.037%	87.58% ± 0.014%
$B^0_s \to \mu^+\mu^-\gamma$	ALL	80.38% ± 0.030%	83.77% ± 0.011%
	1	76.51% ± 0.031%	80.47% ± 0.011%
	2	78.72% ± 0.030%	82.64% ± 0.011%
	3	80.19% ± 0.031%	83.90% ± 0.011%
	4	81.55% ± 0.031%	85.14% ± 0.012%
	5	83.14% ± 0.033%	85.97% ± 0.013%
	6	85.98% ± 0.037%	88.17% ± 0.015%

These muon identification efficiencies extracted from data will be used in Sect. 6.11 to evaluate data/MC correction factors when computing the ratio of efficiencies between the signal and normalization channels.

6.8.2 Muon ID Systematics

Systematic uncertainties are computed to ensure that there are no biases introduced by the methods chosen in the analysis. Since the final branching fraction measurement is measured relative to a normalisation mode, each of the systematics below is calculated as a ratio of the signal to $B^+ \to J/\psi K^+$. Each time a PID cut is applied to the signal muons, the isMuon cut is applied to the $B^+ \to J/\psi K^+$ sample under the same conditions, and the total systematic is taken as the differences in the ratios from the two approaches.

Binning scheme systematic

In order to ensure that the results presented in Table 6.5 are not biased by the choice of binning variables or boundaries used in the calibration maps, a systematic uncertainty relating to the binning scheme is computed.

This is achieved by varying the binning of the calibration samples and recomputing the efficiencies of our MC spectra, then evaluating the difference between the two sets of values. For this systematic binning scheme, a new binning variable, the pseudorapidity (η) is used, rather than the transverse momentum. In addition, the bin boundaries of the other two variables (p and nTracks) are randomly shifted to ensure that the results are not biased by the initial choice of boundaries.

The systematic binning scheme is chosen to be:

$$p : [0, 5.5, 11.68, 17.22, 23.84, 33.06, 39.70, 49.18, 55.88, 200]\,\text{GeV},$$

$$\eta : [0, 2.304, 2.736, 3.282, 3.714, 6],$$

$$\text{nTracks} : [0, 80, 100, 135, 140, 190, 10000].$$

PIDCalib efficiency maps were created with this new binning scheme, and convolutions to the MC signal spectra have been performed[2]. The total PID efficiencies for the signal channels using this new binning scheme are shown in Table 6.6. The same process is followed for the normalisation channel $B^+ \to J/\psi K^+$, with only the isMuon cut applied.

Table 6.6: Muon Efficiencies for $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$ (Run 2) using an alternative binning scheme, used for computation of the systematic uncertainty.

Channel	Efficiency	Stat. uncertainty
$B^0_s \to \mu^+\mu^-$	84.25%	0.091%
$B^0 \to \mu^+\mu^-$	83.656%	0.098%
$B^0_s \to \mu^+\mu^-\gamma$	83.97%	0.089%

The systematic uncertainty, δs, due to binning scheme variation is calculated as the value that restores 1σ between the efficiencies calculated with the two different binning schemes, using Eq. 6.8.3. Entering this calculation are the ratios of the efficiencies of the signal to $B^+ \to J/\psi K^+$ for each of the binning schemes.

$$\frac{|a-b|}{\sqrt{\delta a^2 + \delta b^2 + \delta s^2}} = 1, \tag{6.8.3}$$

---

[2] This systematic is computed only with Run 2 Data/MC, although it will be applied to all years.

where a is the ratio of efficiencies from the nominal scheme, b is the ratio of efficiencies from the systematic scheme, δa and δb are the respective statistical uncertainties [71]. The results for each of the signal channels in summarised in Table 6.7.

Table 6.7: The muon ID systematics applied to each signal channel due to binning scheme variations.

Channel	Systematic Uncertainty
$B^0_s \to \mu^+\mu^-$	0.67%
$B^0 \to \mu^+\mu^-$	0.79%
$B^0_s \to \mu^+\mu^-\gamma$	1.11%

Track Independence systematic

A second systematic uncertainty has been calculated to account for the correlation of the PID efficiency between each of the muon tracks in the event.

The PIDCalib software assumes that the total PID efficiency for a decay is given by the product of the individual track PID efficiencies. In the case of all three signals in this analysis, this is the product of the two muon PID efficiencies. However, for a given decay, the signal track kinematics *will* be correlated to some extent. For example, the rings within the RICH detector from the two tracks could overlap, resulting in a dependence of the track PID performance [107], which could violate the track independence assumption.

For this reason, a systematic uncertainty must be assigned to allow for this effect. This systematic can be calculated using MC[3], by comparing the outcome of applying the PID cut to each of the tracks individually and finding the product of these efficiencies (ϵ_{tot}), against the efficiency of applying the cut to both tracks simultaneously (ϵ'_{tot}):

$$\epsilon_{\text{tot}} = \frac{N(\text{passcut1}) \times N(\text{passcut2})}{N(\text{tot})^2}, \qquad \epsilon'_{\text{tot}} = \frac{N(\text{passallcuts})}{N(\text{tot})}, \tag{6.8.4}$$

As with the previous systematic, the numbers entering the calculation are the ratios of the signal to the normalisation modes. For example, the numbers entering the calculation for the systematic for the $B^0_s \to \mu^+\mu^-$ channel are given in Table 6.8.

The same method is adopted for all three signal channels, and the resultant systematics are presented in Table 6.9.

sWeight systematic

Using sPlot [108] is a way to reconstruct features of a mixture of components based on known properties of distributions. The method allows the projection of a single contribu-

---

[3] Although MC is not necessarily correct for the absolute efficiencies, the difference in efficiencies in MC is a reliable measure of the systematic, as the data/MC differences cancel in the ratio.

Table 6.8: Numbers entering the calculation of the track independence systematic for Muon PID.

	Eff of simultaneous track selection	Eff of separate track selection
$B^+ \to J/\psi K^+$, isMuon	94.61%	94.61%
$B^0_s \to \mu^+\mu^-$, $\text{PID}_{\mu,8}$	88.87%	88.91%
Ratio	93.94%	93.98%

Table 6.9: Systematic for Muon PID relating to track dependence.

Channel	Track independence systematic
$B^0_s \to \mu^+\mu^-$	0.036%
$B^0 \to \mu^+\mu^-$	0.029%
$B^0_s \to \mu^+\mu^-\gamma$	0.052%

tion of a dataset using certain discriminating variables, to disentangle the contributions. Using an sPlot gives information about probabilities, which allows for an estimation the number of signal and background events within each bin. In this analysis, the reconstructed B meson mass from the $B^+ \to J/\psi K^+$ decay mode is used to describe the signal and background distributions with a fit. The fit is used to create weights (sWeights) that project out the signal distribution by subtracting the background contributions from control variables, which are independent of the discriminating variable (in this case the reconstructed invariant mass).

PIDCalib uses this sWeight method, which involves performing a global fit to calibration samples over all kinematics. However there is a slight discrepancy between the sum of the sWeights in a bin and a fit performed only in a specific bin. This is caused by the dependence of the invariant mass fit on kinematic variables such as momentum and pseudorapidity of the probed track [98]. Studies have indicated an absolute per-track systematic uncertainty at the level of 0.1% [109], which must also included in our signal efficiency uncertainties.

The use of the sWeight method in PIDCalib means that an efficiency for a given bin can be over 100% (as seen in Figs. 6.5 and 6.6, for example). This happens if the PID efficiency is already high (as it is for muons) and the events in the bin are given high weighting. Lower weightings assigned to other bins even out the efficiency over the full spectrum.

6.9 Pion, Kaon and Proton misidentification

For the background yields, the chances of one or two daughter particles being misidentified as muons must also be calculated, by measuring the probability of the decay products passing the muon selection criteria. For the backgrounds that are relevant to our search region, the MisID rates need to be calculated for $\pi \to \mu$, $K \to \mu$ and p$\to \mu$.

Some of the backgrounds, namely the $B^0_{(s)} \to h^+h^{(')-}$, involve double misIDs, while some ($\Lambda^0_b \to p\mu^-\nu_\mu$, for example), require only one misID'd particle. Evaluations of MisID probabilities are calculated under the same muon ID criteria defined in Eqs. 6.8.1 and 6.8.2.

6.9.1 Kaons, Pions and the problem with PIDCalib

The PIDCalib software, used for the muon identification efficiencies, was used in the previous analysis round to evaluate the kaon and pion MisID rates.

For the pion and kaon MisIDs, PIDCalib uses sample from the decay $D^{*+} \to D^0\pi^+_s$ with $D^0 \to K^-\pi^+$ (the inclusion of charge conjugated processes is implied throughout the following sections). Measurements of MisID rates were carried out as a function of the track momentum and transverse momentum, using these data control samples.

However, it has since been determined that there is a bias in the estimates of these rates (first reported in Ref. [99]), as the software assumes no correlation between the mass distribution of the calibration samples and muon PID requirement.

The distribution for the variable of interest, in this case PID$_{\mu,8}$, is obtained by applying the sPlot technique with the fit to the distribution of $m(D^0)$ and $\Delta m = m(D^{*+}) - m(D^0)$. The underlying assumption of the sPlot technique is that the variable of interest is independent of the variables used to separate signal and backgrounds, *i.e.* $m(D^0)$ and Δm.

This assumption is not correct in this case; it does not take into account that pions and kaons can decay-in-flight into muons, and this is actually how most of the MisID originates (especially at low momenta). For the decays-in-flight, the reconstructed momentum can be quite different from the original hadron momentum, which in turn has a significant effect on the $m(D^0)$ distribution. The momentum resolution of the reconstructed track is degraded depending on the distance that the hadron has travelled before decaying and the amount of energy inherited by the muon. The mass distribution of the D^0 candidate broadens as a consequence, meaning that the efficiency of selecting $D^0 \to K^-\pi^+$ within the given mass window decreases. This effect needs to be taken into account for correct estimation of the misidentified hadron yields.

For this reason, the MisID rates for pions and kaons are determined using a different technique. The D^0 yield is measured from a fit to the D^0 reconstructed mass, m($K^-\pi^+$), and the mass difference of the D^{*+} and D^0, $\Delta m = m(D^0\pi^+_s)$-$m(K^-\pi^+)$, both with and without the PID selection being applied. The distributions include decays-in-flight, which are estimated using simulation and shown as a tail on the left side of the mass peak. Figure 6.7 shows the $D^0 \to K^-\pi^+$ distribution from simulation, with and without the muon PID requirement on the pion.

The inclusion of the tail in the signal mass shape overcomes the bias in PIDCalib.

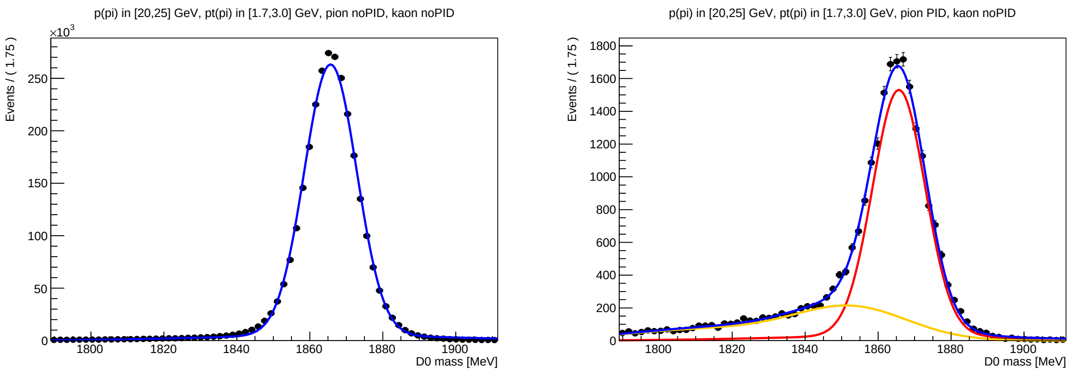


Figure 6.7: Mass distribution of reconstructed $D^0 \to K^-\pi^+$ candidates from simulation, with $p(\pi^+)$ in $[20, 35]\,\text{GeV}$ and $p_{\text{T}}(\pi^+)$ in $[1.7, 3.0]\,\text{GeV}$. Left: No PID requirements are applied to the tracks. Right: The pion track passes the $\text{PID}_{\mu,8}$ selection. The shape of the red component is fixed from the left fit, while the yellow component is the sum of two Gaussians and is determined from the right fit.

However, the fit is performed in a mass range centered around the nominal D^0 mass: $m_{\pi K} \in [1825, 1910]\,\text{MeV}/c^2$, meaning that a portion of this tail, which must be counted as signal, is still missing. For this reason that MisID probability is also corrected for the fraction of $D^0 \to K^-\pi^+$ decays where the reconstructed mass falls outside of the allowed mass window of the D^0. This correction is evaluated separately for pions and kaons by counting the events falling outside the 2D fit mass range, using a sample created with Particle Gun (pGun)[4]. Figure 6.8 illustrates this effect on misidentified pions and kaons from a 2016 pGun sample of $D^0 \to K^-\pi^+$ decays, where no mass cut is applied.

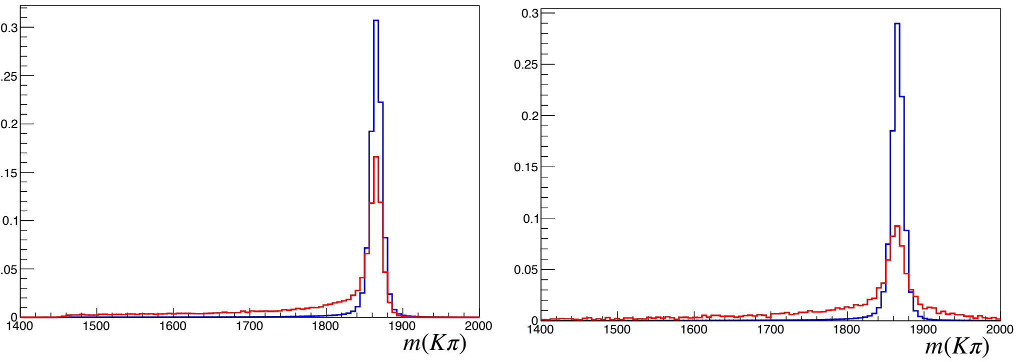


Figure 6.8: $K\pi$ invariant mass from $D^0 \to K^-\pi^+$ decays in pGun 2016 MC when no PID is applied (blue) and when $\text{PID}_{\mu,8}$ is applied (red) to the pion (left) or the kaon (right).

The BDT binning scheme is the same as used for the PID efficiencies (as defined in Sect. 6.8.1), but the last two p_{T} bins have been merged to avoid sparsely populated regions.

---

[4] Particle Gun is a fast simulation tool used in LHCb for the creation of MC in a limited timeframe. It is much faster than the normal MC creation, as instead of generating a full event, a single particle is created with kinematics that can be configured to follow various different distributions. The particle is then decayed into the desired final state using EvtGen.

These two corrections result in an increase in the MisID rates by about +50% for pions and +70% for kaons, compared to the evaluation given by PIDCalib alone.

The misID efficiency is computed by applying the pion, kaon MisID maps, comprehensive of the "out-of-range" corrections, to each $B^0_{(s)} \to h^+h^{(\prime)-}$ MC mode. The computation of the $B^0_{(s)} \to h^+h^{(\prime)-}$ yields, using these efficiencies, is discussed in Sect. 6.10.

6.9.2 Proton Misidentification

The proton misidentification rate needs to be calculated for the $\Lambda^0_b \to p\mu^-\nu_\mu$ background, where the proton is misidentified as a muon. The $p \to \mu$ probability under the muon selection is evaluated from data using $\Lambda \to p\pi$ decays as well as protons from Λ_c decays. The PIDCalib software gives unbiased results for protons and can hence be used without any corrections.

Candidates are required to be produced in the muon acceptance, which reduces the available statistics slightly. Maps for the efficiency for protons passing the muon selection criteria are created in the same bins of p and p_{T} as used for the muon efficiency for Run 2. The number of tracks (nTracks) is not used as an additional binning variable as this would result in too fine binning for the calibration sample size. Since the calibration samples are smaller for Run 1, the number of momentum bins for the maps is reduced. The proton MisID binning for Run 1 is:

$$p : [0, 20, 40, 60, 500]\,\mathrm{GeV},$$

$$p_{\mathrm{T}} : [0.5, 1.7, 40]\,\mathrm{GeV},$$

The proton MisID in the binned phase space is shown in Figs. 6.9 and 6.10, for Run 1 and Run 2 data respectively. The PID efficiency for the $\Lambda^0_b \to p\mu^-\nu_\mu$ background is found by performing a convolution of the MC to these efficiency maps. The results can be found in Table 6.10.

Table 6.10: $\Lambda^0_b \to p\mu^-\nu_\mu$ PID Efficiencies for Run 1 and Run 2, per BDT bin.

	PID Efficiency (%)	
BDT Bin	Run 1	Run 2
All	$5.5\times10^{-3} \pm 0.9\times10^{-3}$	$7.4\times10^{-3} \pm 0.3\times10^{-3}$
1	$4.6\times10^{-3} \pm 0.9\times10^{-3}$	$4.3\times10^{-3} \pm 0.3\times10^{-3}$
2	$5.1\times10^{-3} \pm 0.9\times10^{-3}$	$6.3\times10^{-3} \pm 0.3\times10^{-3}$
3	$5.6\times10^{-3} \pm 0.9\times10^{-3}$	$7.7\times10^{-3} \pm 0.3\times10^{-3}$
4	$6.2\times10^{-3} \pm 0.9\times10^{-3}$	$9\times10^{-3} \pm 0.3\times10^{-3}$
5	$6.5\times10^{-3} \pm 0.9\times10^{-3}$	$0.0105 \pm 0.3\times10^{-3}$
6	$6.9\times10^{-3} \pm 0.9\times10^{-3}$,	$0.0119 \pm 0.3\times10^{-3}$

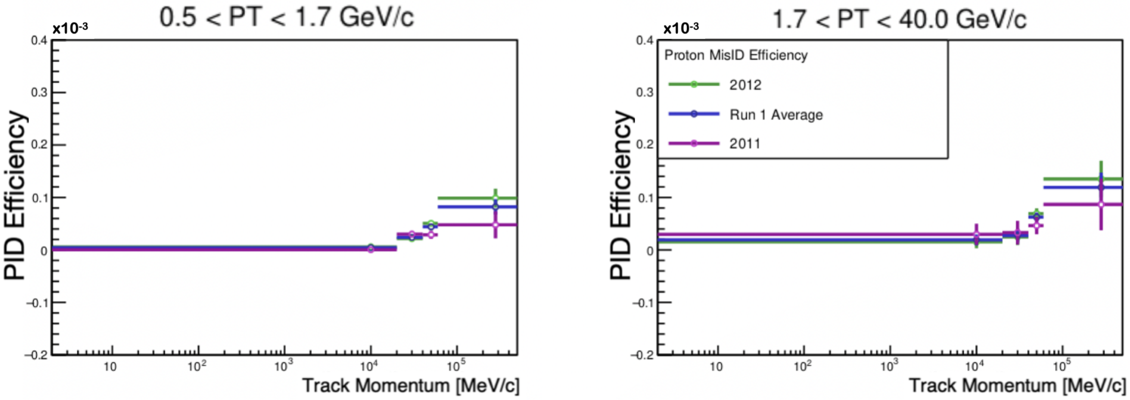


Figure 6.9: Proton misidentification versus momentum for two p_{T} bins of the probe track, evaluated from PIDCalib for Run 1 calibration data. Efficiencies are shown for 2011, 2012 and Run 1 average. The panels refer to p_{T} bins (0.5,1.7) on the left and (1.7,40) GeV on the right.

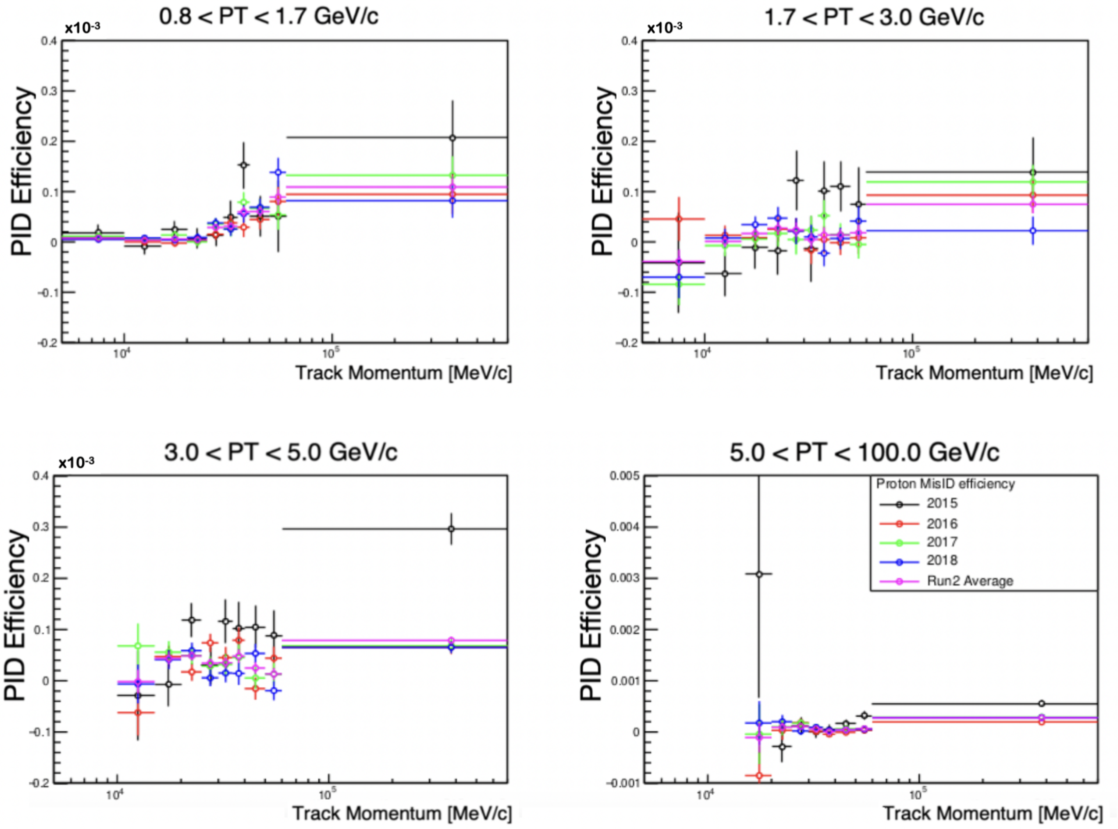


Figure 6.10: Proton misidentification versus momentum for four p_{T} bins of the probe track, evaluated from PIDCalib for Run 2 calibration data. The panels refer to p_{T} bins (0.8,1.7), (1.7,3), (5,100), and (3,5) GeV clockwise from top left.

Proton MisID Systematics

As with the Muon ID, a systematic uncertainty relating to the choice of binning scheme is computed by varying the p, p_{T} binning schemes of the calibration maps, within PIDCalib. The convolution is then performed to the MC for $\Lambda_b^0 \to p\mu^-\nu_\mu$, and the resultant efficiencies are compared. For reference the nominal binning schemes used are:

$$p : [0, 20, 40, 60, 500]\,\mathrm{GeV},$$

$$p_{\mathrm{T}} : [0, 1.7, 40]\,\mathrm{GeV},$$

for Run 1 and,

$$p : [0, 10, 15, 20, 25, 30, 35, 40, 50, 60, 700]\,\mathrm{GeV},$$

$$p_{\mathrm{T}} : [0.8, 1.7, 3, 5, 100]\,\mathrm{GeV},$$

for Run 2.

Systematic binning schemes are then chosen by varying the number of bins, and the bin boundaries. The systematic scheme used for Run 1 is different to that for Run 2 (as with the nominal scheme), given the limited statistics available in the calibration samples. The systematic binning schemes chosen are:

$$p : [2, 15, 35, 55, 500]\,\mathrm{GeV},$$

$$p_{\mathrm{T}} : [0, 2.5, 40]\,\mathrm{GeV},$$

for Run 1, and

$$p : [2, 15, 35, 55, 500]\,\mathrm{GeV},$$

$$p_{\mathrm{T}} : [0, 1.6, 3, 40]\,\mathrm{GeV},$$

for Run 2.

The systematic uncertainty is computed by finding the value that restores 1σ of the total integrated (over all BDT bins) efficiency value between the two binning schemes (nominal and systematic schemes), using Eq. 6.8.3. A summary of these values and the computed systematic are given in Table 6.11.

The overall uncertainties for the $\Lambda_b^0 \to p\mu^-\nu_\mu$ background are therefore taken to be 0.9×10^{-5} (stat) $\pm\ 0.6 \times 10^{-5}$ (syst) for Run 1, and 0.3×10^{-5} (stat) $\pm\ 1 \times 10^{-5}$ (syst) for Run 2.

Systematic studies have also been carried on in order to assess the effect of the multiplicity in the evaluation of the misidentification. A negligible effect has been found in this case.

Table 6.11: $\Lambda_b^0 \to p\mu^-\nu_\mu$ PID efficiencies computed with the nominal and systematic binning schemes for protons, with resultant systematic uncertainty

Λ_b PID efficiency (%)	Run 1	Run 2
Nominal scheme	$(5.5 \pm 0.9) \cdot 10^{-3}$	$(7.4 \pm 0.3) \cdot 10^{-3}$
Systematic scheme	$(6.1 \pm 1.1) \cdot 10^{-3}$	$(6.4 \pm 0.3) \cdot 10^{-3}$
Systematic uncertainty (%)	$0.6 \cdot 10^{-3}$	$1 \cdot 10^{-3}$

6.10 Background Yields

In this section, the yields for the backgrounds relevant to this analysis will be computed; those decays with muon candidates within the mass region of interest (4.9 to 6.0 GeV/c^2). The number of expected candidates is estimated by normalising each background decay to the $B^+ \to J/\psi K^+$ channel;

$$N_{\text{exp}} = \underbrace{\frac{f_x}{f_u} \cdot \frac{N_{B^+\to J/\psi K^+}}{\mathcal{B}(B^+ \to J/\psi K^+) \cdot \epsilon^{\text{tot}}_{B^+\to J/\psi K^+}}}_{\beta_x} \cdot \mathcal{B}_x \cdot \epsilon_x^{\text{tot}} \tag{6.10.1}$$

where N is the number of expected events and x indicates one of the background channels. β_x is used in the next section as shorthand for the braced part of the equation. The branching fraction of $B^+ \to J/\psi K^+$, where the J/ψ decays to two muons, is $\mathcal{B}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+) = (6.021 \pm 0.170) \times 10^{-5}$ (see Eq. 5.5.4).

The yields for the normalisation channel, $N_{B^+\to J/\psi K^+}$, are computed for each data taking year via a maximum likelihood fit on the reconstructed B meson mass distribution, following a selection. The selection is similar to that for the $B^0_{(s)} \to \mu^+\mu^-$ decays, with the exception that there is an additional track (the kaon) and the decay is not required to pass the ProbNN PID selections. The yields for each year are given in Table 6.12.

Table 6.12: Yields for the $B^+ \to J/\psi K^+$ normalisation.

Year	$N_{B^+\to J/\psi K^+}$
2011	$(3.479 \pm 0.008) \times 10^5$
2012	$(7.780 \pm 0.012) \times 10^5$
2015	$(1.676 \pm 0.005) \times 10^5$
2016	$(1.0369 \pm 0.0015) \times 10^6$
2017	$(1.0820 \pm 0.0014) \times 10^6$
2018	$(1.3208 \pm 0.0015) \times 10^6$

The total efficiency, ϵ_x^{tot} for a given channel includes the acceptance, reconstruction and selection, PID, and trigger efficiencies. All of these efficiencies are estimated from simulation, with the exception of the PID which uses the data-driven estimate described previously in Sect. 6.8. The trigger efficiency is evaluated from simulation after the full selection is required. The background expectations are computed in bins of BDT.

6.10.1 Semileptonic backgrounds

In this section estimates of exclusive background decays with one or two real muons in the final state are given. The decays $B^0 \to \pi^-\mu^+\nu_\mu$, $B^0_s \to K^-\mu^+\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ all pollute the lower mass sideband only, meaning that the are only present on the left hand side of the $B^0_s \to \mu^+\mu^-$ signal peak in the invariant mass distribution. $\Lambda^0_b \to p\mu^-\nu_\mu$ and $B^+_c \to J/\psi\mu^+\nu_\mu$ on the other hand span the whole mass region.

$B^0 \to \pi^-\mu^+\nu_\mu$

The semileptonic decay $B^0 \to \pi^-\mu^+\nu_\mu$ has a branching fraction [71] of:

$$\mathcal{B}(B^0 \to \pi^-\mu^+\nu_\mu) = (1.50 \pm 0.06) \cdot 10^{-4}, \tag{6.10.2}$$

and can contribute to the backgrounds of the signal decays when the pion is misidentified as a muon. The invariant mass of the dimuon candidate is shifted to the left due to the missing neutrino, and for the same reason, the BDT is shifted to lower values. Therefore, the invariant mass distribution of the decay has to be carefully estimated in bins of BDT.

MC samples are used to study this background, equally split in the two magnetic polarities. The samples were produced with a cut at generator level requiring the $\pi\mu$ invariant mass to satisfy $m(\pi\mu) > 4500\,\mathrm{MeV}/c^2$.

The samples were produced with the Isgur-Wise form factor model (ISGW2) [110,111], which is actually not in good agreement with recent data. Figure 6.11 shows the comparison of the invariant mass spectrum from the Isgur-Wise model at generator level, compared to the spectrum generated according to recent data and lattice results [112]. From these comparisons, per-event-weights are computed to scale the MC events. This scaling ensures that the MC events can accurately represent what happens in real data.

The efficiency of the PID selection on the muon track is computed by convoluting the muon efficiency maps from PIDCalib, described in Sect. 6.8.1, with spectrum of the muon in simulated $B^0 \to \pi^-\mu^+\nu_\mu$ decays.

The PID selection efficiency of the pion is computed using the following equation:

$$\varepsilon_{\text{pion}} = \frac{\text{map}_{\pi,\text{data}}(p, p_{\text{T}}) \otimes f_{\pi\mu\nu\,\text{MC}}(p, p_{\text{T}})}{\text{map}_{\pi,\text{pgun}}(p, p_{\text{T}}) \otimes f_{\pi\mu\nu\,\text{MC}}(p, p_{\text{T}})} \times \varepsilon_{\text{PID},\pi\mu\nu\,\text{MC}} \tag{6.10.3}$$

where $\text{map}_{\pi,\text{data}}$ and $\text{map}_{\pi,\text{pgun}}$ are the $\pi \to \mu$ efficiency maps, computed with data (from PIDCalib) and MC from pGun.

$f_{\pi\mu\nu\,\text{MC}}(p, p_{\text{T}})$ is the p, p_{T} distribution of simulated $B^0 \to \pi^-\mu^+\nu_\mu$ decays after the full selection is applied (without trigger). $\varepsilon_{\text{PID},\pi\mu\nu\,\text{MC}}$ is the overall efficiency of the muon selection requirement on the pion in simulated $B^0 \to \pi^-\mu^+\nu_\mu$ events, after the full selection

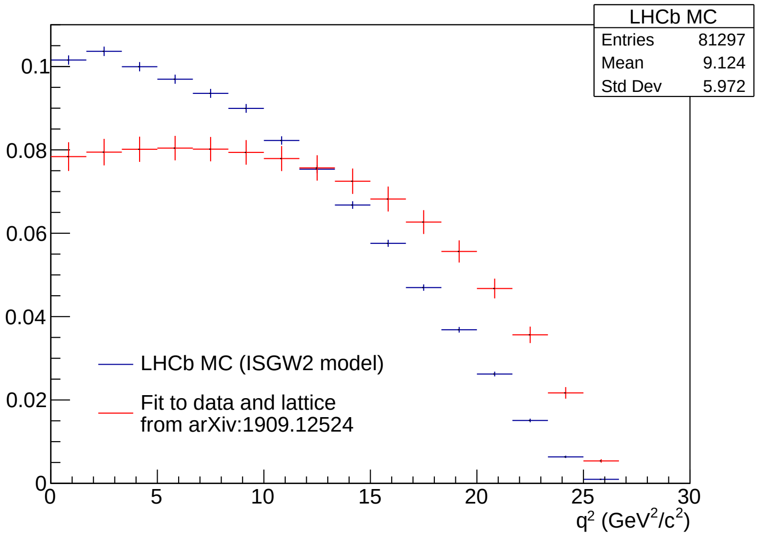


Figure 6.11: Distribution of pion-muon invariant mass (q^2) in $B^0 \to \pi^-\mu^+\nu_\mu$ decays using the ISGW2 model (blue) and using a fit to data and recent lattice results (red). Both histograms are normalized to the same area.

is applied (without trigger). The numerator of the equation is therefore the standard convolution of the MC distribution for $B^0 \to \pi^-\mu^+\nu_\mu$ events with the data, and the denominator is the convolution of the MC distribution for $B^0 \to \pi^-\mu^+\nu_\mu$ events with the pGun distributions. It can be seen as the misidentification efficiency of our $B^0 \to \pi^-\mu^+\nu_\mu$ MC, scaled by the ratio of the efficiency map in data over the efficiency map from pGun MC. Neither of the maps include the out-of-range corrections, which must be evaluated separately.

The out-of-range correction depends on the specific decay channel, and where the decay peaks within the mass window of interest. For the $B^0_{(s)} \to h^+h^{(\prime)-}$ decays the out-of-range correction is very significant as the signal peaks at $5280\,\mathrm{MeV}/c^2$, which is $\sim 400\,\mathrm{MeV}/c^2$ away from the $4900\,\mathrm{MeV}/c^2$ boundary. $B^0 \to \pi^-\mu^+\nu_\mu$ decays on the other hand peak at a lower mass, and most of the tracks from decaying-in-flight pions have a reconstructed momentum significantly smaller than the pion momentum. For that reason, the reconstructed dimuon mass is likely to take values below $4900\,\mathrm{MeV}/c^2$, and therefore the out-of-range correction is small for this channel.

The full breakdown of the numbers used to evaluate the background yield for $B^0 \to \pi^-\mu^+\nu_\mu$, including the BDT integrated PID efficiencies for each run, are given in Table 6.13. Using Eq. 6.10.1, the yield for this background has been calculated using these efficiencies, and is included in the final column of Table 6.13.

The quoted errors on the numbers of expected events include the uncertainties on the branching fraction, the normalisation with $B^+ \to J/\psi K^+$, and on the total efficiencies from MC and data control samples; this applies to all of the background estimates given in the rest of this section.

Table 6.13: Estimated $B^0 \to \pi^-\mu^+\nu_\mu$ efficiencies and yields for the Run 1 and Run 2 samples. The trigger efficiencies are evaluated on the combined sample of $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ decays to increase statistics and after checking that the pion and kaon kinematics are similar.

	β_u $[10^{11}]$	ϵ_{gen} $[10^{-3}]$	ϵ_{sel} $[10^{-2}]$	ϵ_{pid} $[10^{-3}]$	ϵ_{trig} $[10^{-1}]$	N_{exp}
Run 1	7.25 ± 0.19	6.745 ± 0.025	2.35 ± 0.10	2.91 ± 0.13	9.10 ± 0.09	46 ± 4
Run 2	21.67 ± 0.46	7.001 ± 0.012	2.60 ± 0.11	3.41 ± 0.10	9.56 ± 0.04	193 ± 13
$\mathcal{B}(B^0 \to \pi^-\mu^+\nu_\mu) = (1.50 \pm 0.06) \cdot 10^{-4}$						

The expected numbers of events in bins of BDT from the same MC-driven estimates are listed in Table 6.14, and the invariant mass distributions in bins of BDT, fitted on MC, are shown in Fig. 6.12.

Table 6.14: Estimated $B^0 \to \pi^-\mu^+\nu_\mu$ yields per BDT bin for Run 1 and Run 2.

BDT range	Run 1	Run 2
All	46 ± 4	193 ± 13
0.0-0.25	14.7 ± 1.4	63 ± 5
0.25-0.4	7.4 ± 0.9	34.4 ± 2.8
0.4-0.5	5.6 ± 0.7	22.3 ± 2.0
0.5-0.6	3.9 ± 0.6	18.9 ± 1.8
0.6-0.7	4.8 ± 0.7	18.1 ± 1.8
0.7-1.0	9.1 ± 1.1	36.0 ± 3.1

$B^0_s \to K^-\mu^+\nu_\mu$

The semileptonic decay $B^0_s \to K^-\mu^+\nu_\mu$ can represent a significant peaking background for the analysis when the kaon is misidentified as muon. The branching fraction of this decay has recently been measured by LHCb [113] to be;

$$\mathcal{B}(B^0_s \to K^-\mu^+\nu_\mu) = (1.06 \pm 0.10) \cdot 10^{-4}. \tag{6.10.4}$$

Despite having a similar branching fraction to the $B^0 \to \pi^-\mu^+\nu_\mu$ background already discussed, the contribution from $B^0_s \to K^-\mu^+\nu_\mu$ decays is expected to be smaller due to the larger mass shift, and the lower fragmentation fraction of the B^0_s compared to B^+.

In order to study this background, MC samples are split evenly between the two magnet polarities. The samples were produced with a cut at generator level requiring the $K\mu$ invariant mass to satisfy $m(K\mu) > 4500\,\text{MeV}/c^2$. Similarly to the $B^0 \to \pi^-\mu^+\nu_\mu$ case, the MC is reweighted to better represent real data.

The PID efficiency for the $B^0_s \to K^-\mu^+\nu_\mu$ decay is computed using the same procedure described for $B^0 \to \pi^-\mu^+\nu_\mu$ decays.

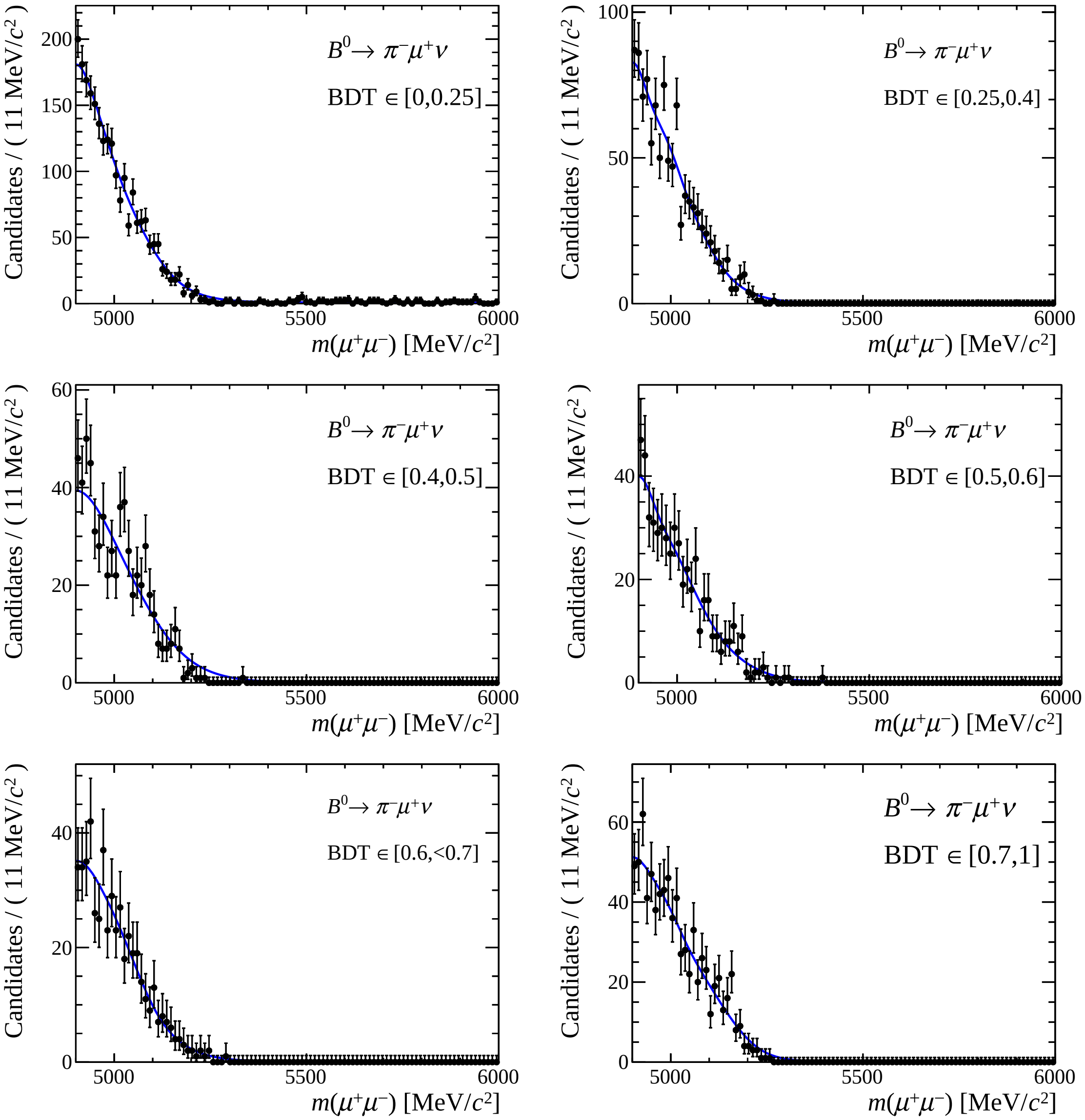


Figure 6.12: Invariant mass distribution of the $B^0 \to \pi^-\mu^+\nu_\mu$ channel and shape fitted with one-dimensional kernel density estimators using Gaussian kernels, from the sum of all simulation. The distributions are shown separately for BDT bins 1 to 6 (from left to right, from top to bottom).

Unlike the $B^0 \to \pi^-\mu^+\nu_\mu$ case however, the out-of-range contribution for $B^0_s \to K^-\mu^+\nu_\mu$ decays is significant more significant due to the position of the mass peak. The full breakdown of the numbers for the MC-driven estimate integrated over the entire BDT range is listed in Table 6.15, including the computed yields for each Run.

Table 6.15: Estimated $B^0_s \to K^-\mu^+\nu_\mu$ efficiencies and yields for the Run 1 and Run 2 samples. Note, that the selection efficiency ϵ_{sel} also contains the effect of the MC reweighting. The trigger efficiencies are evaluated on the combined sample of $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^0_s \to K^-\mu^+\nu_\mu$ decays to increase statistics and after checking that the pion and kaon kinematics are similar.

	β_s [10^{11}]	ϵ_{gen} [10^{-3}]	ϵ_{sel} [10^{-2}]	ϵ_{pid} [10^{-3}]	ϵ_{trig} [10^{-1}]	N_{exp}
Run 1	1.86 ± 0.15	9.14 ± 0.04	2.5 ± 0.4	2.72 ± 0.17	9.10 ± 0.09	11.0 ± 2.4
Run 2	5.5 ± 0.4	9.475 ± 0.019	2.7 ± 0.5	1.73 ± 0.04	9.56 ± 0.04	25 ± 5
$\mathcal{B}(B^0_s \to K^-\mu^+\nu_\mu) = (1.06 \pm 0.10) \cdot 10^{-4}$						

The expected numbers of events in bins of BDT from the same MC-driven estimates are listed in Table 6.16. The invariant mass distributions in bins of BDT are fitted on MC, and are shown in Fig. 6.13.

Table 6.16: Expected $B^0_s \to K^-\mu^+\nu_\mu$ yields per BDT bin for Run 1 and Run 2.

BDT range	Run 1	Run 2
All	11.0 ± 2.4	25 ± 5
0.0-0.25	2.0 ± 0.5	3.7 ± 1.0
0.25-0.4	1.30 ± 0.33	3.1 ± 0.8
0.4-0.5	0.73 ± 0.23	2.5 ± 0.7
0.5-0.6	1.6 ± 0.4	3.6 ± 0.9
0.6-0.7	1.40 ± 0.35	3.7 ± 0.9
0.7-1.0	4.0 ± 0.9	8.6 ± 1.9

$B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$

The decays $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ can mimic the $B^0_{(s)} \to \mu^+\mu^-$ decay if the two final state muons form a good vertex. While the dimuon invariant mass of these backgrounds will not reach the signal region (due to the additional pion in the final state), it could affect the left sideband. The decay $B^+ \to \pi^+\mu^+\mu^-$ has a branching fraction [71] of

$$\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-) = (1.75 \pm 0.22) \cdot 10^{-8}. \tag{6.10.5}$$

The decay $B^0 \to \pi^0\mu^+\mu^-$ has not been yet been observed, but a theoretical estimate [114] of its rate relative to $B^+ \to \pi^+\mu^+\mu^-$ can be used:

$$\frac{\mathcal{B}(B^0 \to \pi^0\mu^+\mu^-)}{\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-)} = 0.47^{+0.22}_{-0.18} \tag{6.10.6}$$

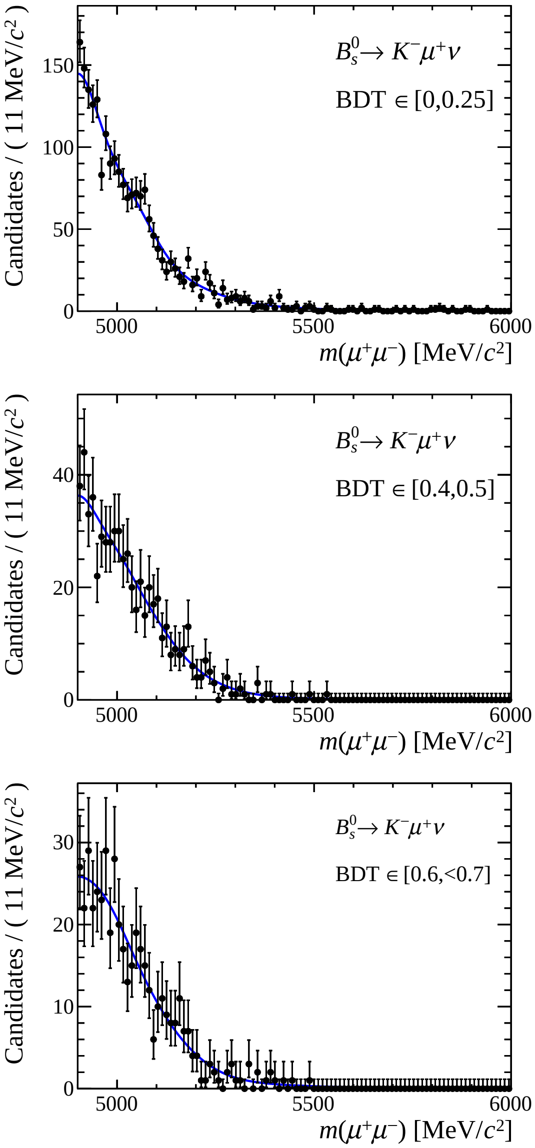


Figure 6.13: Invariant mass distribution of the $B_s^0 \to K^-\mu^+\nu_\mu$ channel and shape fitted with one-dimensional kernel density estimators using Gaussian kernels, from the sum of all simulation. The distributions are shown separately for BDT bins 1 to 6 (from left to right, from top to bottom).

From this an estimate for the branching fraction can be obtained:

$$\mathcal{B}(B^0 \to \pi^0\mu^+\mu^-) = (0.82 \pm 0.36) \cdot 10^{-8} \tag{6.10.7}$$

MC samples of Run 1 and Run 2 events, split equally between the polarities are used for the two decays with both muons required to be in the detector acceptance. The full breakdown of the numbers for the MC-driven estimates integrated over the entire BDT range, including the computed yields, are listed in Table 6.17 and Table 6.18.

Table 6.17: Estimated $B^+ \to \pi^+\mu^+\mu^-$ efficiencies and yields for the Run 1 and Run 2 samples.

	β_u [10^{11}]	ϵ_{gen} [10^{-1}]	ϵ_{sel} [10^{-3}]	ϵ_{pid} [10^{-1}]	ϵ_{trig} [10^{-1}]	N_{exp}
Run 1	7.25 ± 0.19	2.803 ± 0.010	5.07 ± 0.05	8.0247 ± 0.0030	9.566 ± 0.022	13.8 ± 1.8
Run 2	21.67 ± 0.46	2.8728 ± 0.0026	5.672 ± 0.029	8.3432 ± 0.0011	9.593 ± 0.011	49 ± 6
$\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-) = (1.75 \pm 0.22) \cdot 10^{-8}$						

Table 6.18: Estimated $B^0 \to \pi^0\mu^+\mu^-$ efficiencies and yields for the Run 1 and Run 2 samples.

	β_u [10^{11}]	ϵ_{gen} [10^{-1}]	ϵ_{sel} [10^{-3}]	ϵ_{pid} [10^{-1}]	ϵ_{trig} [10^{-1}]	N_{exp}
Run 1	7.25 ± 0.19	2.827 ± 0.011	4.72 ± 0.05	8.0153 ± 0.0030	9.519 ± 0.023	6.0 ± 2.7
Run 2	21.67 ± 0.46	2.8767 ± 0.0028	5.236 ± 0.028	8.3717 ± 0.0011	9.577 ± 0.011	21 ± 9
$\mathcal{B}(B^0 \to \pi^0\mu^+\mu^-) = (0.82 \pm 0.36) \cdot 10^{-8}$						

The expected numbers of events in bins of BDT from the same MC-driven estimates are listed in Table 6.19.

The invariant mass distributions in bins of BDT, fitted on MC are shown in Fig. 6.14. The same mass shape is used for both $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ decays, as the dimuon mass distribution is identical.

Table 6.19: Expected $B^+ \to \pi^+\mu^+\mu^-$ yields (left) and expected $B^0 \to \pi^0\mu^+\mu^-$ yields (right) per BDT bin for Run 1 and Run 2.

BDT range	Run 1	Run 2
All	13.8 ± 1.8	49 ± 6
0.0-0.25	6.1 ± 0.8	22.9 ± 2.9
0.25-0.4	2.7 ± 0.4	9.9 ± 1.3
0.4-0.5	1.45 ± 0.19	5.0 ± 0.6
0.5-0.6	1.17 ± 0.16	4.0 ± 0.5
0.6-0.7	0.95 ± 0.13	3.1 ± 0.4
0.7-1.0	1.36 ± 0.18	4.5 ± 0.6

BDT range	Run 1	Run 2
All	6.0 ± 2.7	21 ± 9
0.0-0.25	1.9 ± 0.8	6.7 ± 3.0
0.25-0.4	1.0 ± 0.4	3.6 ± 1.6
0.4-0.5	0.66 ± 0.29	2.3 ± 1.0
0.5-0.6	0.63 ± 0.28	2.2 ± 1.0
0.6-0.7	0.62 ± 0.27	2.2 ± 1.0
0.7-1.0	1.3 ± 0.6	4.4 ± 1.9

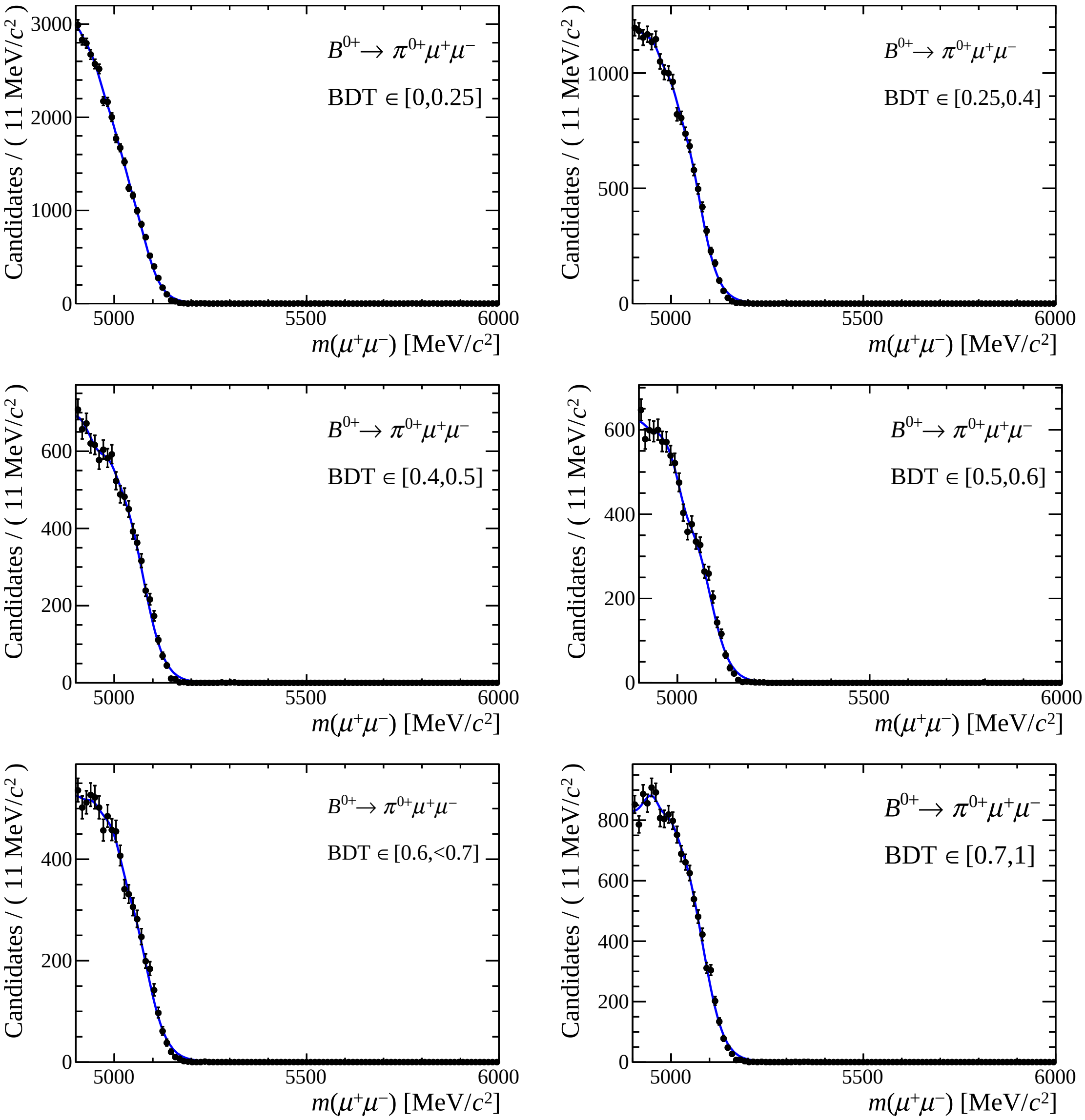


Figure 6.14: Invariant mass distribution of the combined $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$ channels and shape fitted with one-dimensional kernel density estimators using Gaussian kernels, from the sum of all simulation. The distributions are shown separately for BDT bins 1 to 6 (from left to right, from top to bottom).

$\boldsymbol{\Lambda_b^0 \to p\mu^-\nu_\mu}$

The decay $\Lambda_b^0 \to p\mu^-\nu_\mu$ can fake the $B^0_{(s)} \to \mu^+\mu^-$ decay when the proton is misidentified as a muon. This can happen due to noise or *punch-through* protons (described earlier in Sect. 4.3.1).

The branching fraction of the decay has been measured by the LHCb collaboration [115] to be

$$\mathcal{B}(\Lambda_b^0 \to p\mu^-\nu_\mu) = (4.1 \pm 1.0) \cdot 10^{-4}. \tag{6.10.8}$$

For the hadronisation fraction, the LHCb measurement [116] of

$$r_\Lambda(p_\mathrm{T}) \equiv \frac{f_{\Lambda_b^0}}{f_u + f_d}(p_\mathrm{T}) = A\left[p_1 + \exp\left(p_2 + p_3 \times p_\mathrm{T}\right)\right], \tag{6.10.9}$$

with

$$\begin{aligned} A &= 1.000 \pm 0.061 \\ p_1 &= (7.93 \pm 1.41) \cdot 10^{-2} \\ p_2 &= -1.022 \pm 0.047 \\ p_3 &= (-0.107 \pm 0.002)\,\mathrm{GeV}^{-1} \end{aligned}$$

is used.

By assuming $f_d = f_u$ the relation $\beta_\Lambda = \beta_u \cdot 2r_\Lambda(p_\mathrm{T})$ is obtained. Since the $2r_\Lambda(p_\mathrm{T})$ factor is p_T dependent, a per-event weight is assigned and included in the selection efficiency computation. That way, β_u can simply be used in the normalisation.

MC samples of Run 1 and Run 2 events, equally split in the two magnetic polarities are used to study the background. The events have been generated requiring $m(p\mu) > 4500\,\mathrm{MeV}/c^2$ at generator level.

The full breakdown of the numbers for the MC-driven estimate integrated over the entire BDT range is listed in Table 6.20.

Table 6.20: Numbers entering into the computation of the $\Lambda_b^0 \to p\mu^-\nu_\mu$ peaking background and estimated background yields for the Run 1 and Run 2 samples. Note, that the selection efficiency ϵ_sel also contains the effect of the reweighting due to the Λ_b^0 hadronization.

	β_u $[10^{11}]$	ϵ_gen $[10^{-2}]$	ϵ_sel $[10^{-2}]$	ϵ_pid $[10^{-5}]$	ϵ_trig $[10^{-1}]$	N_exp
Run 1	7.25 ± 0.19	1.322 ± 0.007	0.9 ± 1.0	5.5 ± 1.1	8.91 ± 0.17	1.8 ± 2.0
Run 2	21.67 ± 0.46	1.3225 ± 0.0024	1.0 ± 1.0	7.4 ± 1.0	8.91 ± 0.17	7 ± 8
$\mathcal{B}(\Lambda_b^0 \to p\mu^-\nu_\mu) = (4.1 \pm 1.0) \cdot 10^{-4}$						

The expected numbers of events in bins of BDT from the same MC-driven estimates are listed in Table 6.21. It has been checked that the very high yield uncertainties resulting from the form factor uncertainties have only a negligible impact on the signal branching fraction.

The invariant mass distributions in bins of BDT are fitted on MC, again with one-

Table 6.21: Expected $\Lambda^0_b \to p\mu^-\nu_\mu$ yields per BDT bin for Run 1 and Run 2.

BDT range	Run 1	Run 2
All	1.8 ± 2.0	7 ± 8
0.0-0.25	0.6 ± 0.7	1.8 ± 1.9
0.25-0.4	0.29 ± 0.32	1.1 ± 1.2
0.4-0.5	0.18 ± 0.20	0.8 ± 0.8
0.5-0.6	0.20 ± 0.22	0.9 ± 0.9
0.6-0.7	0.19 ± 0.21	0.9 ± 1.0
0.7-1.0	0.32 ± 0.35	1.6 ± 1.8

dimensional kernel density estimators using Gaussian kernels. The fit results are shown in Fig. 6.15 for the sample.

$B^+_c \to J/\psi\mu^+\nu_\mu$

Partially reconstructed decays of the B^+_c meson could be a dangerous background to the $B^0_{(s)} \to \mu^+\mu^-$ signal, due to the larger mass. That said, the hadronisation fraction of a b quark to a B^+_c is about two orders of magnitude lower than to a B^+ meson (though with large uncertainties).

The decay $B^+_c \to J/\psi\mu^+\nu_\mu$, with $J/\psi \to \mu^+\mu^-$, could pass the signal selection in the case that a good vertex is reconstructed between the muon from the semileptonic decay and the oppositely charged muon from the J/ψ decay. Given the above topology, the muon isolation is expected to be rather effective in rejecting this background.

A simple J/ψ veto is also used to further reject these events. The cut consists of vetoing events in which one of the two candidate muons, coupled to any other oppositely-charged muon in the event (selected with $\text{ProbNN}_\mu > 0.3$), falls in a window of $(\text{m}(\mu\mu) - \text{m}(J/\psi)) < 30\,\text{MeV}/c^2$.

This veto is expected to reject 68.7% (65.7% in Run 2) of $B^+_c \to J/\psi\mu^+\nu_\mu$ events (in the whole BDT range), with a small signal loss of 0.2% in Run 1 and 0.3% in Run 2.

The veto is included in all of the estimates for the other exclusive backgrounds given in this section, although they are rather unaffected. The majority of the vetoed $B^+_c \to J/\psi\mu^+\nu_\mu$ events fall in the low BDT region because of the long track isolation.

The branching fraction of this decay including the hadronisation fraction ratio has recently been measured [117] to be:

$$\frac{f_c}{f_u + f_d} \cdot \mathcal{B}(B^+_c \to J/\psi\mu^+\nu_\mu) = (7.07 \pm 0.15 \pm 0.24) \cdot 10^{-5} \text{ for 7 TeV,} \tag{6.10.10}$$

and

$$\frac{f_c}{f_u + f_d} \cdot \mathcal{B}(B^+_c \to J/\psi\mu^+\nu_\mu) = (7.36 \pm 0.08 \pm 0.30) \cdot 10^{-5} \text{ for 13 TeV.} \tag{6.10.11}$$

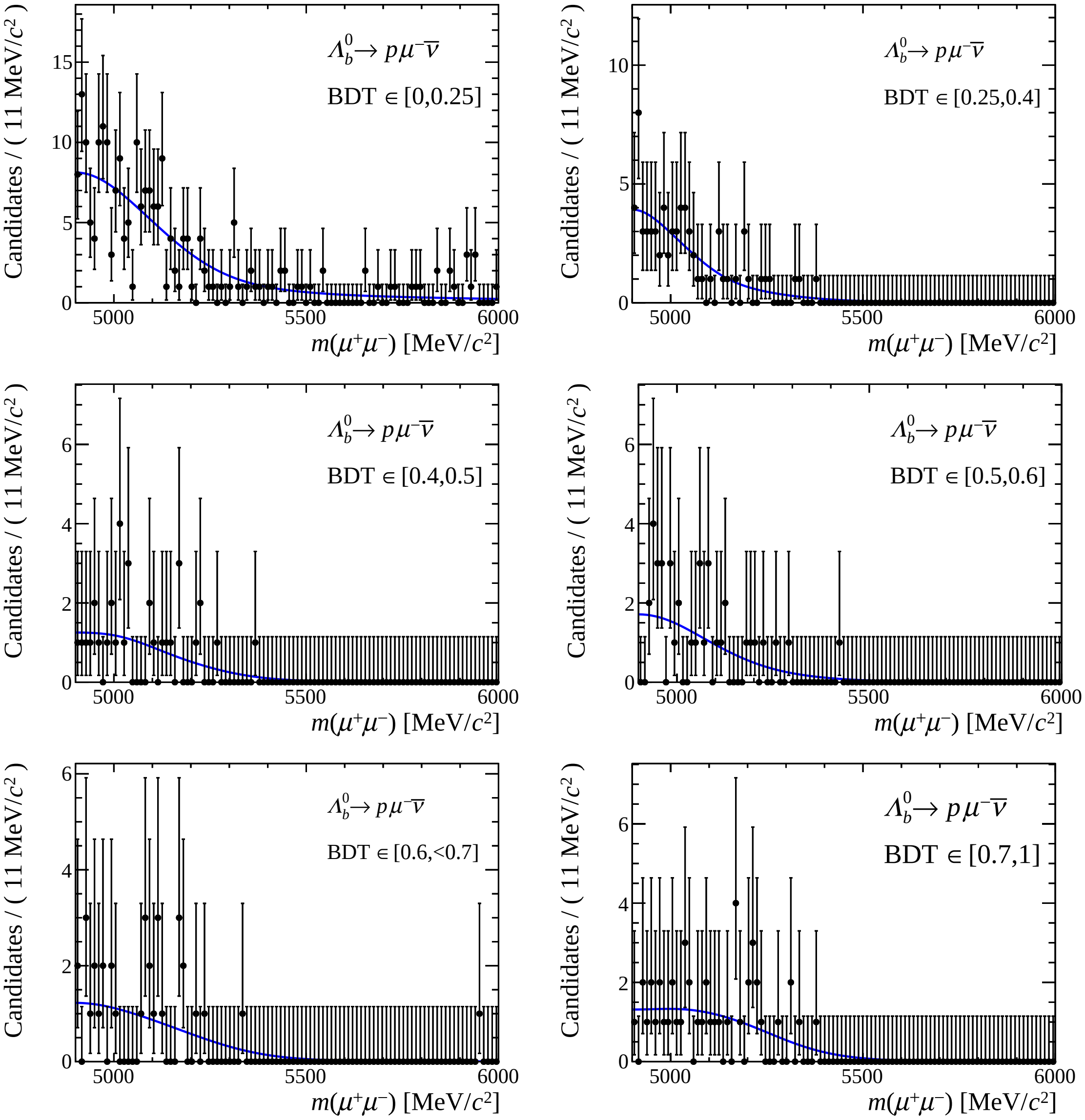


Figure 6.15: Invariant mass distribution of the $\Lambda_b^0 \to p\mu^-\nu_\mu$ channel and shape fitted from the sum of all simulation. The distributions are shown separately for BDT bins 1 to 6 (from left to right, from top to bottom).

in the kinematic region $4\,\mathrm{GeV}/c < p_T < 25\,\mathrm{GeV}/c$ and $2.5 < \eta < 4.5$ of the B_c^+ meson.

Assuming $f_u = f_d$, and f_c/f_u is similar for 7 TeV and 8 TeV, one can then estimate the effective branching fraction of the whole decay chain as:

$$\begin{aligned}
\mathcal{B}_{\text{eff}} &= 2\frac{f_c}{f_u + f_d} \cdot \mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu) \cdot \mathcal{B}(J/\psi \to \mu^+\mu^-) && (6.10.12)\\
&= (8.43 \pm 0.34) \cdot 10^{-6} \text{ for Run 1,} && (6.10.13)\\
&= (8.8 \pm 0.4) \cdot 10^{-6} \text{ for Run 2.} && (6.10.14)
\end{aligned}$$

A correction factor accounting for the ratio of acceptances between $B^+ \to J/\psi K^+$ and $B_c^+ \to J/\psi\mu^+\nu_\mu$ under the kinematic selection of Ref. [117] is calculated from MC generations without acceptance cuts. It is found to be consistent with one ($\alpha = 1.001 \pm 0.008$) and thus dropped.

To study this background, MC samples are again used, with an even split of magnet polarities. The full breakdown of the numbers for the MC-driven estimate integrated over the entire BDT range is listed in Table 6.22. The expected numbers of events in bins of BDT from the same MC-driven estimates are listed in Table 6.23, along with the computed yield.

Table 6.22: Numbers entering into the computation of the $B_c^+ \to J/\psi\mu^+\nu_\mu$ peaking background and estimated background yields for the Run 1 and Run 2 samples. Note, that the reconstruction and selection efficiency ϵ_{sel} also contains the PID efficiency measured on data.

	β_u $[10^{11}]$	ϵ_{gen} $[10^{-3}]$	ϵ_{sel} $[10^{-3}]$	ϵ_{pid} $[10^{-1}]$	ϵ_{trig} $[10^{-1}]$	N_{exp}
Run 1	7.25 ± 0.19	1.7018 ± 0.0023	3.95 ± 0.06	7.9880 ± 0.0031	9.50 ± 0.04	31.1 ± 1.6
Run 2	21.67 ± 0.40	1.9295 ± 0.0017	3.413 ± 0.023	8.3544 ± 0.0011	9.596 ± 0.014	101 ± 5
$\mathcal{B}_{\text{eff}}(B_c^+ \to J/\psi\mu^+\nu_\mu) = (8.43 \pm 0.34) \cdot 10^{-6}$ for Run 1						
$\mathcal{B}_{\text{eff}}(B_c^+ \to J/\psi\mu^+\nu_\mu) = (8.8 \pm 0.4) \cdot 10^{-6}$ for Run 2						

The invariant mass distributions in bins of BDT are fitted on MC with one-dimensional kernel density estimators using Gaussian kernels. The fit results are shown in Fig. 6.16 for the sample.

Table 6.23: Expected $B_c^+ \to J/\psi\mu^+\nu_\mu$ yields per BDT bin for Run 1 and Run 2.

BDT range	Run 1	Run 2
All	31.1 ± 1.6	101 ± 5
0.0-0.25	24.4 ± 1.3	79 ± 4
0.25-0.4	3.27 ± 0.23	11.7 ± 0.6
0.4-0.5	1.54 ± 0.14	4.59 ± 0.27
0.5-0.6	0.86 ± 0.10	2.85 ± 0.19
0.6-0.7	0.61 ± 0.08	1.56 ± 0.12
0.7-1.0	0.43 ± 0.07	0.85 ± 0.08

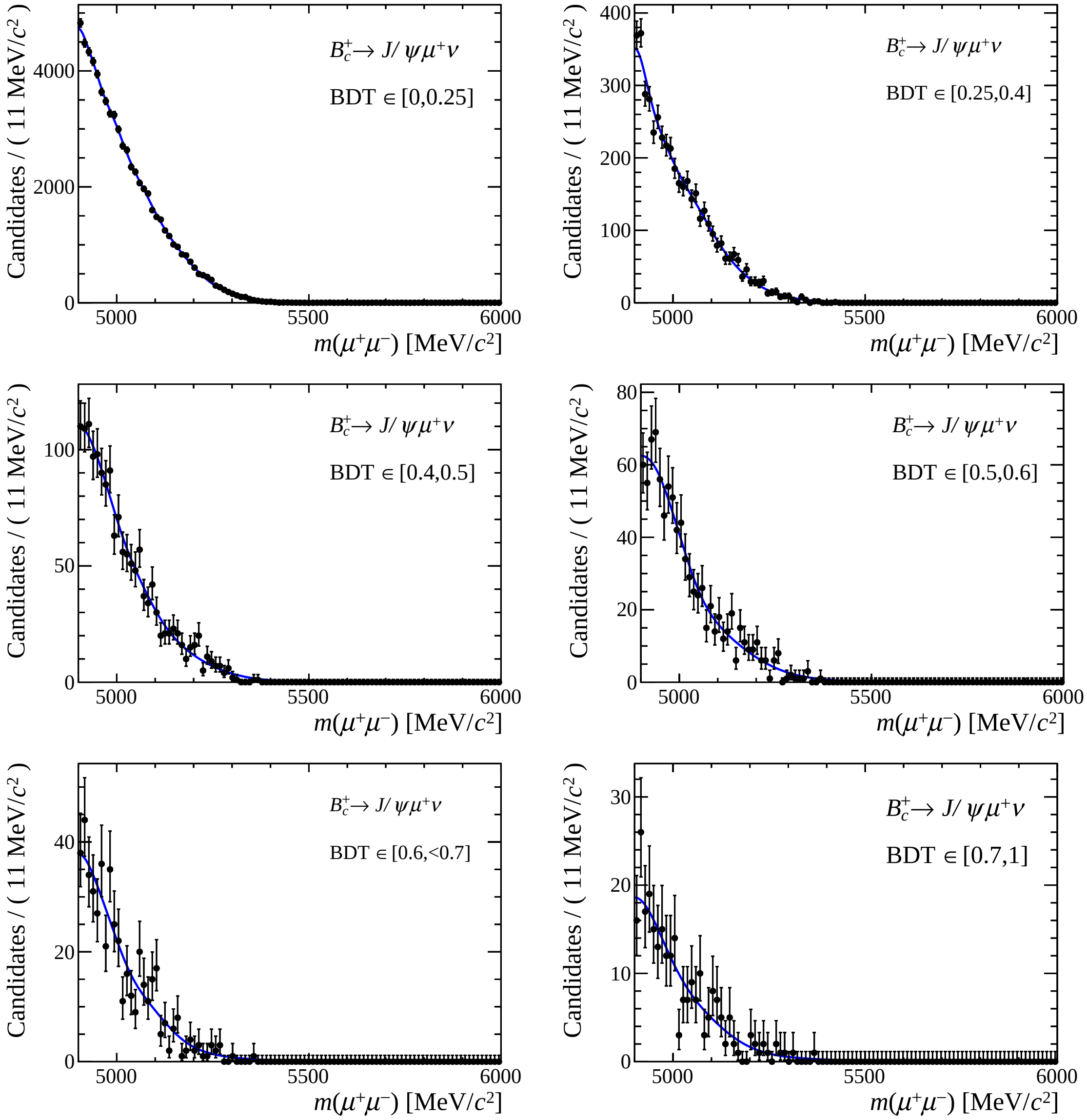


Figure 6.16: Invariant mass distribution of the $B_c^+ \to J/\psi\mu^+\nu_\mu$ channel and shape fitted from the sum of all simulation. The distributions are shown separately for BDT bins 1 to 6 (from left to right, from top to bottom).

6.10.2 Non semileptonic backgrounds, $B^0_{(s)} \to h^+h^{(\prime)-}$

The double misidentification probability is estimated by convoluting the kaon and pion misidentification curves, explained in Sect. 6.9.1, with the simulated p and p_{T} spectra of the two hadrons from $B^0_{(s)} \to h^+h^{(\prime)-}$ decays, after the full selection is applied.

The double misidentification probabilities for Run 1 and Run 2 data are given in Table 6.24, separately for the four exclusive decay modes, being $B^0 \to K^+\pi^-$, $B^0 \to \pi^+\pi^-$, $B^0_s \to \pi^+K^-$ and $B^0_s \to K^+K^-$.

The average double misidentification for the inclusive $B^0_{(s)} \to h^+h^{(\prime)-}$ decays, $\epsilon_{hh\to\mu\mu}$, is also shown in the last column. It is obtained by weighting the values of the single modes according to their relative production rates[5].

The total $\epsilon_{hh\to\mu\mu}$ also includes a correction to account for the BDT-dependence of the $h\mu$ procedure.

Table 6.24: Double misidentification probability in units of 10^{-6} for Run 1 and Run 2 data; the PID$_{\mu,4}$ selection is used for Run 1 and 2015, and the PID$_{\mu,8}$ selection for Run 2.

	$B^0 \to \pi\pi$	$B^0 \to K\pi$	$B^0_s \to K\pi$	$B^0_s \to KK$	$\epsilon_{hh\to\mu\mu}$
Run 1	15.91 ± 0.58	7.31 ± 0.39	7.41 ± 0.40	3.94 ± 0.29	18.4 ± 2.1
Run 2	20.43 ± 0.28	7.26 ± 0.20	7.29 ± 0.20	3.11 ± 0.14	9.41 ± 0.72

Using the values of $\epsilon_{hh\to\mu\mu}$, together with the $B^0 \to K^+\pi^-$ TIS yields corrected by the expected fraction of each mode, the number of $B^0_{(s)} \to h^+h^{(\prime)-}$ double misidentified events is evaluated and summarised in Table 6.25.

Table 6.25: Number of expected $B^0_{(s)} \to h^+h^{(\prime)-}$ events per run.

	Run 1	Run 2
$N_{hh\to\mu\mu}$	20.1 ± 2.6	33.4 ± 2.9

In order to take into account the BDT dependence (due to correlation between PID and BDT via the momentum of the selected candidates) of the double misidentification, it is evaluated individually for each BDT bin. The number of expected events in each BDT bin is then computed. The estimates of $B^0_{(s)} \to h^+h^{(\prime)-}$ double misidentification background per BDT bin, are summarised in Table 6.26 for Run 1 and Run 2 data.

The MC mass distributions of the four $B^0_{(s)} \to h^+h^{(\prime)-}$ modes are summed up with a weight that accounts for the branching fractions, f_s/f_d, and selection efficiencies, including the full PID selection. The weights equalize the statistics of the four modes, prior to taking the sum. The resulting MC events are shown in Fig. 6.17 together with the `RooKeysPdf` fit curve.

---

[5] The following branching fractions are used [71]: $\mathcal{B}(B^0 \to K^+\pi^-) = (19.6 \pm 0.5) \cdot 10^{-6}$, $\mathcal{B}(B^0 \to \pi^+\pi^-) = (5.12 \pm 0.19) \cdot 10^{-6}$, $\mathcal{B}(B^0_s \to \pi^+K^-) = (5.6 \pm 0.6) \cdot 10^{-6}$, $\mathcal{B}(B^0_s \to K^+K^-) = (25.2 \pm 1.7) \cdot 10^{-6}$, and $f_s/f_d = 0.244 \pm 0.012$ (13 TeV value).

Table 6.26: Number of $B^0_{(s)} \to h^+h^{(\prime)-} \to \mu^+\mu^-$ events as a function of the BDT bin for Run 1 and Run 2 data.

BDT range	Run 1	Run 2
All	20.1 ± 2.6	33.4 ± 2.9
0-0.25	6.1 ± 1.4	10.7 ± 1.7
0.25-0.4	3.8 ± 0.9	6.2 ± 1.0
0.4-0.5	1.57 ± 0.20	2.90 ± 0.26
0.5-0.6	1.61 ± 0.21	2.82 ± 0.26
0.6-0.7	1.65 ± 0.21	2.68 ± 0.24
0.7-1.0	5.3 ± 0.7	8.1 ± 0.7

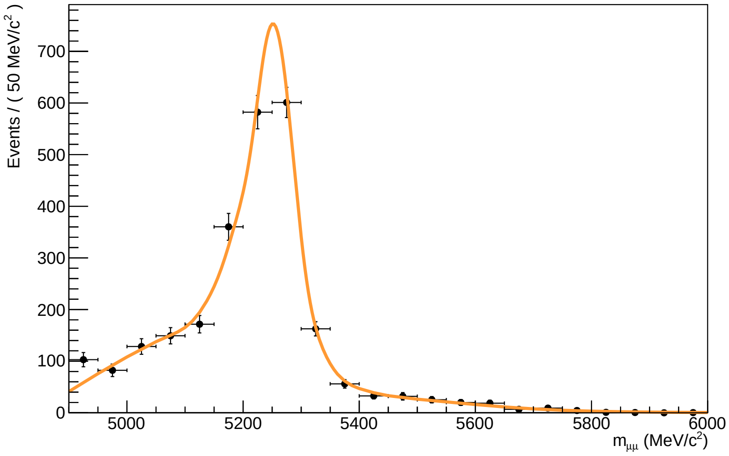


Figure 6.17: $B^0_{(s)} \to h^+h^{(\prime)-} \to \mu^+\mu^-$ invariant mass and fit curve.

6.10.3 Background summary tables

The background estimates in bins of BDT and in the mass range of 4900 to 6000 MeV/c^2 for $B^0_{(s)} \to h^+h^-$, $B^0 \to \pi^-\mu^+\nu_\mu$, $B^0_s \to K^-\mu^+\nu_\mu$, $B^+ \to \pi^+\mu^+\mu^-$ and $B^0 \to \pi^0\mu^+\mu^-$, $\Lambda^0_b \to p\mu^-\nu_\mu$, and $B^+_c \to J/\psi\mu^+\nu_\mu$, are summarized in the following Tables 6.27 and 6.28 for Run 1 and Run 2 data, respectively.

Table 6.27: Exclusive background yields per BDT bin with their total estimated uncertainties for Run 1 data.

BDT range	$B^0_{(s)} \to h^+h^{(\prime)-}$	$B^0 \to \pi^-\mu^+\nu_\mu$	$B^0_s \to K^-\mu^+\nu_\mu$	$B^+ \to \pi^+\mu^+\mu^-$	$B^0 \to \pi^0\mu^+\mu^-$	$\Lambda^0_b \to p\mu^-\nu_\mu$	$B^+_c \to J/\psi\mu^+\nu_\mu$
All	20.1 ± 2.6	46 ± 4	11.0 ± 2.4	13.8 ± 1.8	6.0 ± 2.7	1.8 ± 2.0	31.1 ± 1.6
0.0-0.25	6.1 ± 1.4	14.7 ± 1.4	2.0 ± 0.5	6.1 ± 0.8	1.9 ± 0.8	0.6 ± 0.7	24.4 ± 1.3
0.25-0.4	3.8 ± 0.9	7.4 ± 0.9	1.30 ± 0.33	2.7 ± 0.4	1.0 ± 0.4	0.29 ± 0.32	3.27 ± 0.23
0.4-0.5	1.57 ± 0.20	5.6 ± 0.7	0.73 ± 0.23	1.45 ± 0.19	0.66 ± 0.29	0.18 ± 0.20	1.54 ± 0.14
0.5-0.6	1.61 ± 0.21	3.9 ± 0.6	1.6 ± 0.4	1.17 ± 0.16	0.63 ± 0.28	0.20 ± 0.22	0.86 ± 0.10
0.6-0.7	1.65 ± 0.21	4.8 ± 0.7	1.40 ± 0.35	0.95 ± 0.13	0.62 ± 0.27	0.19 ± 0.21	0.61 ± 0.08
0.7-1.0	5.3 ± 0.7	9.1 ± 1.1	4.0 ± 0.9	1.36 ± 0.18	1.3 ± 0.6	0.32 ± 0.35	0.43 ± 0.07

Table 6.28: Exclusive background yields per BDT bin with their total estimated uncertainties for Run 2 data.

BDT range	$B^0_{(s)} \to h^+h^{(\prime)-}$	$B^0 \to \pi^-\mu^+\nu_\mu$	$B^0_s \to K^-\mu^+\nu_\mu$	$B^+ \to \pi^+\mu^+\mu^-$	$B^0 \to \pi^0\mu^+\mu^-$	$\Lambda^0_b \to p\mu^-\nu_\mu$	$B^+_c \to J/\psi\mu^+\nu_\mu$
All	33.4 ± 2.9	193 ± 13	25 ± 5	49 ± 6	21 ± 9	7 ± 8	101 ± 5
0.0-0.25	10.7 ± 1.7	63 ± 5	3.7 ± 1.0	22.9 ± 2.9	6.7 ± 3.0	1.8 ± 1.9	79 ± 4
0.25-0.4	6.2 ± 1.0	34.4 ± 2.8	3.1 ± 0.8	9.9 ± 1.3	3.6 ± 1.6	1.1 ± 1.2	11.7 ± 0.6
0.4-0.5	2.90 ± 0.26	22.3 ± 2.0	2.5 ± 0.7	5.0 ± 0.6	2.3 ± 1.0	0.8 ± 0.8	4.59 ± 0.27
0.5-0.6	2.82 ± 0.26	18.9 ± 1.8	3.6 ± 0.9	4.0 ± 0.5	2.2 ± 1.0	0.9 ± 0.9	2.85 ± 0.19
0.6-0.7	2.68 ± 0.24	18.1 ± 1.8	3.7 ± 0.9	3.1 ± 0.4	2.2 ± 1.0	0.9 ± 1.0	1.56 ± 0.12
0.7-1.0	8.1 ± 0.7	36.0 ± 3.1	8.6 ± 1.9	4.5 ± 0.6	4.4 ± 1.9	1.6 ± 1.8	0.85 ± 0.08

6.11 Normalisation and Branching Fraction Results

The signal branching fractions are normalised to the two decay channels $B^0 \to K^+\pi^-$ and $B^+ \to J/\psi K^+$ (with $J/\psi \to \mu^+\mu^-$). Their branching fractions are taken from the PDG [71].

Therefore, the signal branching fractions are expressed as

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = \overbrace{\underbrace{\frac{\mathcal{B}_{\text{norm}}}{N_{\text{norm}}} \times \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}}}_{\alpha_d} \times \frac{f_d}{f_s}}^{\alpha_s} \times N_{B^0_s \to \mu^+\mu^-} \tag{6.11.1}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = \alpha_d \times N_{B^0 \to \mu^+\mu^-} \tag{6.11.2}$$

where α_s and α_d are the normalisation factors for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, respectively, N and ϵ are the yields and efficiencies of the normalisation channels, and f_s/f_d is the ratio of hadronisation fractions. Here and in the following we assume that $f_d = f_u$, as this is known to hold to good accuracy.

The detection efficiencies are split according to the detection stages: detector acceptance (i.e. generation level), reconstruction and selection, and trigger, such that

$$\frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} = \frac{\epsilon^{\text{Acc}}_{\text{norm}}}{\epsilon^{\text{Acc}}_{\text{sig}}} \times \frac{\epsilon^{\text{RecSel}|\text{Acc}}_{\text{norm}}}{\epsilon^{\text{RecSel}|\text{Acc}}_{\text{sig}}} \times \frac{\epsilon^{\text{Trig}|\text{RecSel}}_{\text{norm}}}{\epsilon^{\text{Trig}|\text{RecSel}}_{\text{sig}}}, \tag{6.11.3}$$

where the efficiency for each subsequent stage is estimated for candidates that pass the previous stages.

Geometrical detector acceptance

The detector acceptance is defined as the fraction of the candidates that have all their decay products within the geometrical detector acceptance of LHCb.

The total detector acceptance, however, will also be affected by the magnetic field and by the interactions with the detector material; these effects will be evaluated as part of the reconstruction efficiency discussed in the next section.

The geometrical detector acceptances for the signal and normalisation channels, estimated with MC, are listed in Table 6.29.

The decay products are required to fly into the LHCb detector acceptance, defined by the polar angle in the range of $[10, 400]$ mrad, which is chosen to be larger than the physical LHCb detector acceptance in order to allow for the recovery of particles by the magnetic field.

The geometrical acceptance efficiencies in Table 6.29 are similar for the signal and the $B^0 \to K^+\pi^-$ decay, but lower for $B^+ \to J/\psi K^+$ decays. This is expected due to the kinematic distribution of the final decay products. The additional particle in the final state of the $B^+ \to J/\psi K^+$ decay makes it less likely that all of the decay products will stay within the acceptance range.

In the production of the $B^0_s \to \mu^+\mu^-\gamma$ MC, an acceptance cut applied to the other signal channels was missed. For this reason, the $B^0_s \to \mu^+\mu^-\gamma$ efficiency at generation stage is much higher than the other channels. This cut is instead taken into account in the next selection stage, meaning that the RecSel efficiency (see Table 6.30) is significantly lower than the other signal channels. Overall, the product of the $\epsilon^{\mathrm{Acc}} \times \epsilon^{\mathrm{Rec|Sel}}$ is similar for each of the signal channels, as expected.

Reconstruction and selection efficiencies

The reconstruction efficiency is the number of candidates that are successfully reconstructed, divided by the number of candidates in the detector acceptance. The selection efficiency is then the percentage of reconstructed decay candidates that then pass the signal selection (see Sect. 6.5).

Both efficiencies depend on the characteristics of the decay channel, including the number of particles in the final state, their kinematic distributions, the track reconstruction efficiency, and the particle identification efficiency.

The reconstruction and selection efficiencies are combined into $\epsilon^{\mathrm{Rec|Sel}}$. This efficiency is evaluated on simulated samples, considering only candidates that have already passed the detector acceptance cuts.

The efficiencies for each channel are in good agreement between the years 2015, 2016, 2017, and 2018 (Run 2). In general, the Run 1 efficiencies do not agree as well, but since only the ratio of signal and normalisation efficiencies matters in the final calculation, variations in the efficiencies cancel out. The efficiency for $B^0_s \to \mu^+\mu^-\gamma$ is much lower than that of the other signal channels, as the acceptance cut which was omitted in generation is included here, as mentioned earlier.

The efficiency to reconstruct a particle track is measured on $J/\psi \to \mu^+\mu^-$ data using the tag-and-probe method [118]. The efficiency with which tracks are reconstructed is determined in bins of momentum and pseudo-rapidity, and applied as a correction factor to the reconstruction and selection efficiencies. The RecSel efficiencies are listed in Table 6.30.

PID efficiency correction factors

The acceptance of the muon detector, and the efficiency of the ISMUON algorithm in identifying true muon tracks, are included in the reconstruction and selection efficiencies measured on simulation.

These efficiencies have also been determined on data and simulation using the tag-and-probe technique on the $B^+ \to J/\psi K^+$ sample. From these independent measurements, correction factors are derived and applied to the simulated reconstruction and selection efficiencies.

Table 6.29: Geometrical detector acceptance, estimated as the fraction of decays contained in the polar angle region of $[10, 400]$ mrad. MCXX represents MC created with the data taking conditions of a given year. The numbers in the last column are weighted and averaged according to the fraction of MagUp and MagDown data in each year (∼50% of each).

Channel	$\epsilon^{\text{Acc}}_{\text{MagUp}}$	$\epsilon^{\text{Acc}}_{\text{MagDown}}$	ϵ^{Acc}
$B^0_s \to \mu^+\mu^-$			
MC11	$(18.413 \pm 0.047)\%$	$(18.340 \pm 0.063)\%$	$(18.377 \pm 0.039)\%$
MC12	$(18.726 \pm 0.033)\%$	$(18.609 \pm 0.032)\%$	$(18.668 \pm 0.023)\%$
MC15	$(19.318 \pm 0.054)\%$	$(19.273 \pm 0.052)\%$	$(19.296 \pm 0.037)\%$
MC16	$(19.439 \pm 0.068)\%$	$(19.401 \pm 0.066)\%$	$(19.420 \pm 0.047)\%$
MC17	$(19.274 \pm 0.063)\%$	$(19.271 \pm 0.058)\%$	$(19.273 \pm 0.043)\%$
MC18	$(19.375 \pm 0.067)\%$	$(19.371 \pm 0.068)\%$	$(19.373 \pm 0.048)\%$
$B^0 \to \mu^+\mu^-$			
MC11	$(18.329 \pm 0.053)\%$	$(18.329 \pm 0.053)\%$	$(18.329 \pm 0.038)\%$
MC12	$(18.646 \pm 0.045)\%$	$(18.735 \pm 0.045)\%$	$(18.691 \pm 0.032)\%$
MC15	$(19.288 \pm 0.057)\%$	$(19.276 \pm 0.059)\%$	$(19.282 \pm 0.041)\%$
MC16	$(19.394 \pm 0.071)\%$	$(19.369 \pm 0.070)\%$	$(19.382 \pm 0.050)\%$
MC17	$(19.370 \pm 0.070)\%$	$(19.357 \pm 0.069)\%$	$(19.364 \pm 0.049)\%$
MC18	$(19.406 \pm 0.067)\%$	$(19.440 \pm 0.068)\%$	$(19.423 \pm 0.048)\%$
$B^0_s \to \mu^+\mu^-\gamma$			
MC11	$(47.300 \pm 0.120)\%$	$(47.380 \pm 0.130)\%$	$(47.340 \pm 0.088)\%$
MC12	$(47.300 \pm 0.120)\%$	$(47.380 \pm 0.130)\%$	$(47.340 \pm 0.088)\%$
MC15	$(47.580 \pm 0.120)\%$	$(47.240 \pm 0.120)\%$	$(47.410 \pm 0.085)\%$
MC16	$(47.340 \pm 0.130)\%$	$(47.310 \pm 0.130)\%$	$(47.325 \pm 0.092)\%$
MC17	$(47.615 \pm 0.120)\%$	$(47.505 \pm 0.120)\%$	$(47.560 \pm 0.085)\%$
MC18	$(47.250 \pm 0.130)\%$	$(47.150 \pm 0.125)\%$	$(47.200 \pm 0.090)\%$
$B^+ \to J/\psi K^+$			
MC11	$(16.379 \pm 0.030)\%$	$(16.440 \pm 0.058)\%$	$(16.410 \pm 0.033)\%$
MC12	$(16.712 \pm 0.049)\%$	$(16.691 \pm 0.068)\%$	$(16.702 \pm 0.042)\%$
MC15	$(17.368 \pm 0.046)\%$	$(17.377 \pm 0.047)\%$	$(17.373 \pm 0.033)\%$
MC16	$(17.302 \pm 0.041)\%$	$(17.382 \pm 0.042)\%$	$(17.342 \pm 0.029)\%$
MC17	$(17.398 \pm 0.057)\%$	$(17.389 \pm 0.058)\%$	$(17.394 \pm 0.040)\%$
MC18	$(17.398 \pm 0.059)\%$	$(17.309 \pm 0.059)\%$	$(17.354 \pm 0.042)\%$
$B^0 \to K^+\pi^-$			
MC11	$(17.730 \pm 0.036)\%$	$(17.740 \pm 0.031)\%$	$(17.735 \pm 0.024)\%$
MC12	$(19.040 \pm 0.075)\%$	$(18.916 \pm 0.069)\%$	$(18.978 \pm 0.051)\%$
MC15	$(19.579 \pm 0.053)\%$	$(19.599 \pm 0.052)\%$	$(19.589 \pm 0.037)\%$
MC16	$(19.679 \pm 0.070)\%$	$(19.641 \pm 0.069)\%$	$(19.660 \pm 0.049)\%$
MC17	$(19.748 \pm 0.065)\%$	$(19.612 \pm 0.065)\%$	$(19.680 \pm 0.046)\%$
MC18	$(19.734 \pm 0.068)\%$	$(19.575 \pm 0.069)\%$	$(19.655 \pm 0.048)\%$

Table 6.30: Reconstruction and selection efficiencies for $B^0_{(s)} \to \mu^+\mu^-$ (γ) and the normalisation channels, evaluated on MC simulated samples after full reconstruction and selection. The J/ψ veto has been applied on $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$ and $B^0 \to K^+\pi^-$. The PID$_\mu$ cut has not been applied to the signal channels. For $B^0_s \to \mu^+\mu^-\gamma$, only events above the 4.9 GeV threshold are considered.

Channel	N^{Acc}	$N^{\text{AccRecSel}}$	$\epsilon^{\text{RecSel\|Acc}}$
$B^0_s \to \mu^+\mu^-$			
MC11	$534,499$	$180,540$	$(33.78 \pm 0.06)\%$
MC12	$2,080,642$	$655,129$	$(31.487 \pm 0.032)\%$
MC15	$1,011,344$	$333,341$	$(32.96 \pm 0.05)\%$
MC16	$2,011,123$	$667,217$	$(33.176 \pm 0.033)\%$
MC17	$2,012,366$	$668,064$	$(33.198 \pm 0.033)\%$
MC18	$1,008,582$	$335,158$	$(33.23 \pm 0.05)\%$
$B^0 \to \mu^+\mu^-$			
MC11	$508,999$	$171,101$	$(33.62 \pm 0.07)\%$
MC12	$498,027$	$154,132$	$(30.95 \pm 0.07)\%$
MC15	$1,072,675$	$347,595$	$(32.40 \pm 0.05)\%$
MC16	$2,006,533$	$654,455$	$(32.616 \pm 0.033)\%$
MC17	$2,071,197$	$675,209$	$(32.600 \pm 0.033)\%$
MC18	$1,047,670$	$349,967$	$(33.40 \pm 0.05)\%$
$B^0_s \to \mu^+\mu^-\gamma$			
MC11	157,582	17,113	$(10.86 \pm 0.080)\%$
MC12	328,682	34,676	$(10.55 \pm 0.050)\%$
MC15	311,838	35,612	$(11.42 \pm 0.060)\%$
MC16	620,694	71,690	$(11.55 \pm 0.040)\%$
MC17	618,046	71,199	$(11.52 \pm 0.040)\%$
MC18	1,255,800	146,049	$(11.63 \pm 0.029)\%$
$B^+ \to J/\psi K^+$			
MC11	$762,312$	$140,048$	$(18.37 \pm 0.040)\%$
MC12	$5,047,318$	$840,275$	$(16.648 \pm 0.017)\%$
MC15	$2,072,461$	$359,980$	$(17.370 \pm 0.026)\%$
MC16	$4,182,716$	$742,742$	$(17.757 \pm 0.019)\%$
MC17	$10,006,407$	$1,777,045$	$(17.759 \pm 0.012)\%$
MC18	$9,912,257$	$1,758,921$	$(17.745 \pm 0.012)\%$
$B^0 \to K^+\pi^-$			
MC11	$775,505$	$201,069$	$(25.93 \pm 0.050)\%$
MC12	$8,581,113$	$2,068,318$	$(24.103 \pm 0.015)\%$
MC15	$2,009,237$	$513,096$	$(25.537 \pm 0.031)\%$
MC16	$4,006,284$	$1,023,331$	$(25.543 \pm 0.022)\%$
MC17	$4,008,116$	$1,024,970$	$(25.572 \pm 0.022)\%$
MC18	$4,109,859$	$1,048,447$	$(25.511 \pm 0.022)\%$

The muon identification efficiencies for the signals modes $B^0_{(s)} \to \mu^+\mu^-$ also include the ProbNN particle identification requirements. These are not included in the simulated reconstruction and selection efficiencies listed in Table 6.30, but are included in the signal selection through the correction factors:

$$C^{\text{ch}} = \frac{\epsilon_{\text{Data}}^{\text{isMuon(+ProbNN)|RecSel}}}{\epsilon_{\text{MC}}^{\text{isMuon|RecSel}}}. \tag{6.11.4}$$

These factors are computed for each channel as ratios between the efficiencies on simulation and data, and are listed in Table 6.31. The MC results are computed by simply finding the fraction of selected events that pass the isMuon requirement cut, while the data results are computed using maps from PIDCalib.

Table 6.31: Muon identification efficiencies after full selection for signal and control channel decays estimated on both simulation and data using PIDCalib. In addition to the isMuon efficiency, the signal channel efficiencies on data also include the ProbNN cut efficiency. The correction factors for the reconstruction and selection efficiencies are calculated as the ratio of the data and MC numbers in this table.

	$B_s^0 \to \mu^+\mu^-$		$B^0 \to \mu^+\mu^-$	
Year	$\epsilon_{\text{MC}}^{\text{isMuon\|RecSel}}$	$\epsilon_{\text{Data}}^{\text{isMuon(+ProbNN)\|RecSel}}$	$\epsilon_{\text{MC}}^{\text{isMuon\|RecSel}}$	$\epsilon_{\text{Data}}^{\text{isMuon(+ProbNN)\|RecSel}}$
2011	$(96.01 \pm 0.05)\%$	$(81.1 \pm 0.5)\%$	$(96.13 \pm 0.05)\%$	$(80.7 \pm 0.5)\%$
2012	$(96.28 \pm 0.02)\%$	$(81.0 \pm 0.4)\%$	$(96.27 \pm 0.05)\%$	$(80.7 \pm 0.4)\%$
2015	$(96.03 \pm 0.03)\%$	$(79.7 \pm 0.6)\%$	$(96.01 \pm 0.03)\%$	$(79.6 \pm 0.6)\%$
2016	$(95.96 \pm 0.02)\%$	$(83.8 \pm 0.2)\%$	$(95.95 \pm 0.02)\%$	$(83.5 \pm 0.2)\%$
2017	$(95.99 \pm 0.02)\%$	$(84.6 \pm 0.2)\%$	$(95.97 \pm 0.02)\%$	$(84.3 \pm 0.2)\%$
2018	$(96.02 \pm 0.03)\%$	$(83.9 \pm 0.2)\%$	$(96.01 \pm 0.03)\%$	$(83.6 \pm 0.2)\%$
	$B^+ \to J/\psi K^+$		$B_s^0 \to \mu^+\mu^-\gamma$	
Year	$\epsilon_{\text{MC}}^{\text{isMuon\|RecSel}}$	$\epsilon_{\text{Data}}^{\text{isMuon\|RecSel}}$	$\epsilon_{\text{MC}}^{\text{isMuon\|RecSel}}$	$\epsilon_{\text{Data}}^{\text{isMuon(+ProbNN)\|RecSel}}$
2011	$(94.62 \pm 0.06)\%$	$(95.92 \pm 0.07)\%$	$(96.01 \pm 0.05)\%$	$(81.5 \pm 0.5)\%$
2012	$(94.90 \pm 0.02)\%$	$(95.65 \pm 0.05)\%$	$(96.28 \pm 0.02)\%$	$(81.7 \pm 0.4)\%$
2015	$(94.68 \pm 0.04)\%$	$(91.93 \pm 0.06)\%$	$(96.03 \pm 0.03)\%$	$(79.6 \pm 0.6)\%$
2016	$(94.61 \pm 0.03)\%$	$(94.90 \pm 0.02)\%$	$(95.96 \pm 0.02)\%$	$(83.0 \pm 0.2)\%$
2017	$(94.61 \pm 0.02)\%$	$(94.64 \pm 0.02)\%$	$(95.99 \pm 0.02)\%$	$(85.2 \pm 0.2)\%$
2018	$(94.60 \pm 0.02)\%$	$(95.38 \pm 0.02)\%$	$(96.02 \pm 0.03)\%$	$(85.3 \pm 0.3)\%$

Trigger efficiencies

For the selection of the signal and $B^+ \to J/\psi K^+$ samples, no specific trigger line is required. This means that the samples contain both TIS and TOS events (see Sect. 3.2.9 for definitions). The events are predominantly triggered by muon and dimuon lines at every trigger level, so the trigger efficiency for both modes can be calibrated with $B^+ \to$

$J/\psi K^+$ data. These trigger efficiencies are calculated for events that have already been accepted, reconstructed and selected. The additional hadron in the $B^+ \to J/\psi K^+$ decay can introduce a bias when determining the signal efficiencies, which is avoided by using only the TOS efficiencies for muon and dimuon triggers.

The trigger efficiencies are determined from data with the `TISTOS` method [83], which works by exploiting the overlap between the TIS and TOS categories, in order to determine the various efficiencies as:

$$\varepsilon_{\mathrm{TIS}} = \frac{N_{\mathrm{TISTOS}}}{N_{\mathrm{TOS}}} \tag{6.11.5}$$

$$\varepsilon_{\mathrm{TOS}} = \frac{N_{\mathrm{TISTOS}}}{N_{\mathrm{TIS}}} \tag{6.11.6}$$

$$\varepsilon_{\mathrm{trig}} = \frac{N_{\mathrm{trig}}}{N_{\mathrm{TIS}}}\frac{N_{\mathrm{TISTOS}}}{N_{\mathrm{TOS}}} \tag{6.11.7}$$

where N_X is the number of (background subtracted) candidates triggered within the category X. The category "Trig" refers to all triggered events by a given trigger strategy regardless of the category.

Efficiencies are determined in bins of kinematic variables in order to mitigate the correlation effects. Maps are created from $B^+ \to J/\psi K^+$ data as a function of

- the maximum p_{T} of the two muons, with bins $[0, 1.5, 4, 5, 30]$ GeV
- the product of the p_{T} of the two muons, with bins $[0, 5, 7.5, 12, 150]$ GeV^2.

These maps are then convoluted with the MC samples for signal channels, while the $B^+ \to J/\psi K^+$ efficiency is determined directly from data. The trigger efficiencies for the muonic channels are shown in Table 6.32.

The $B^+ \to J/\psi K^+$ efficiencies can also be used to determine the L0 and HLT1 trigger efficiency of $B^0 \to K^+\pi^-$. The $B^0 \to K^+\pi^-$ candidates are selected requiring the L0 and HLT1 trigger to be TIS as they are triggered mostly due to particles from the other b hadron. For this reason the L0 and HLT1 TIS can be calibrated with any b hadron decay. The $B^0 \to K^+\pi^-$ efficiency depends on the exact configuration of the L0 and HLT1 trigger, meaning that is varies over the years. At the HLT2 level, the candidates are required to be TOS. The total $B^0 \to K^+\pi^-$ efficiencies per year are summarised in Table 6.33.

6.12 Branching Fraction Results

The data are split by data taking period into Run 1 and Run 2 and also split into bins of the BDT. The resulting data categories are fitted simultaneously with an unbinned extended maximum likelihood fit. The lowest BDT bin $[0.0, 0.25]$ is excluded from the fit since it's background-dominated.

Table 6.32: Trigger efficiencies for the muonic channels in data as obtained through `TISTOS` (first uncertainty is statistical and the second systematic).

Year	$B^0_s \to \mu^+\mu^-$	$B^0 \to \mu^+\mu^-$
2011	$(96.00 \pm 0.60 \pm 2.50)\%$	$(95.90 \pm 0.60 \pm 2.70)\%$
2012	$(95.70 \pm 0.40 \pm 2.10)\%$	$(95.60 \pm 0.40 \pm 2.20)\%$
2015	$(95.70 \pm 0.70 \pm 1.40)\%$	$(95.70 \pm 0.70 \pm 1.40)\%$
2016	$(96.99 \pm 0.26 \pm 1.68)\%$	$(97.00 \pm 0.25 \pm 1.72)\%$
2017	$(98.13 \pm 0.24 \pm 1.56)\%$	$(98.09 \pm 0.24 \pm 1.60)\%$
2018	$(96.62 \pm 0.23 \pm 1.75)\%$	$(96.54 \pm 0.23 \pm 1.84)\%$
Year	$B^+ \to J/\psi K^+$	$B^0_s \to \mu^+\mu^-\gamma$
2011	$(88.50 \pm 0.70 \pm 1.50)\%$	$(95.51 \pm 0.60 \pm 2.96)\%$
2012	$(87.90 \pm 0.40 \pm 1.00)\%$	$(95.31 \pm 0.39 \pm 2.47)\%$
2015	$(88.20 \pm 0.80 \pm 1.60)\%$	$(95.53 \pm 0.70 \pm 1.85)\%$
2016	$(92.19 \pm 0.27 \pm 1.53)\%$	$(96.94 \pm 0.25 \pm 1.88)\%$
2017	$(90.83 \pm 0.29 \pm 1.34)\%$	$(97.94 \pm 0.24 \pm 1.82)\%$
2018	$(88.78 \pm 0.27 \pm 1.79)\%$	$(96.35 \pm 0.22 \pm 2.17)\%$

Table 6.33: TIS trigger efficiencies for $B^0 \to K^+\pi^-$, with the first uncertainty being statistical and the second systematic.

Year	$B^0 \to K^+\pi^-$
2011	$(3.84 \pm 0.04 \pm 0.30)\%$
2012	$(4.53 \pm 0.02 \pm 0.19)\%$
2015	$(8.15 \pm 0.08 \pm 0.27)\%$
2016	$(7.43 \pm 0.03 \pm 0.43)\%$
2017	$(7.46 \pm 0.03 \pm 0.29)\%$
2018	$(6.89 \pm 0.02 \pm 0.34)\%$

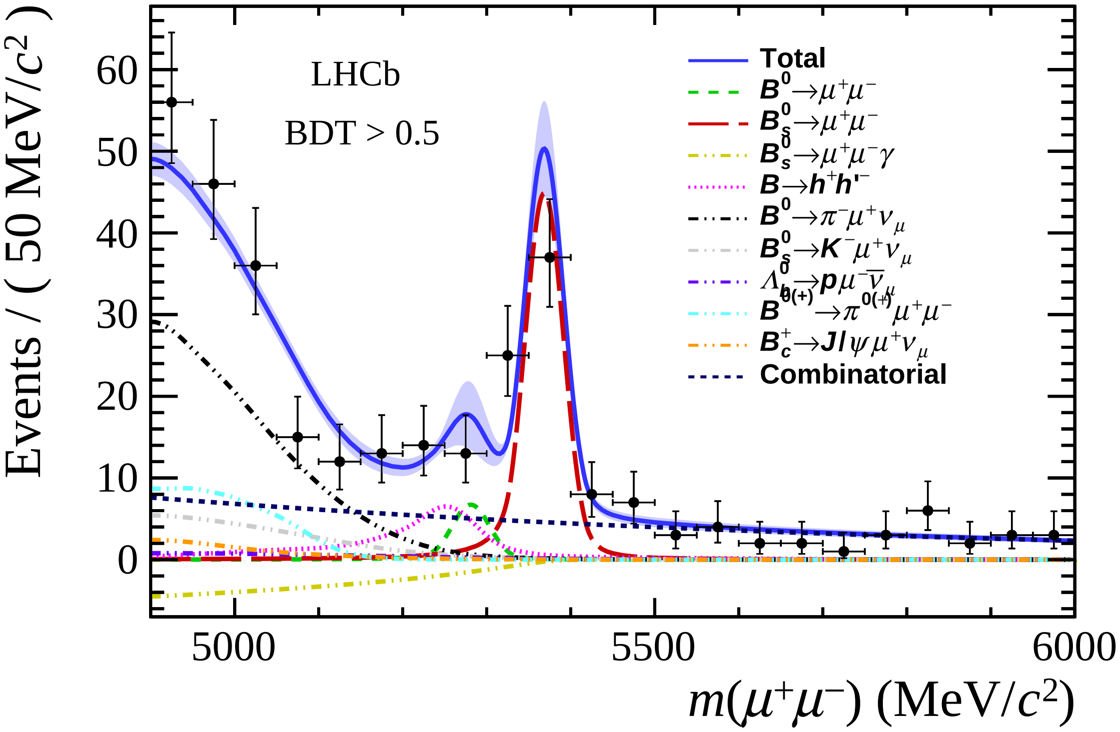


Figure 6.18: Mass distribution of signal candidates (black dots) for BDT > 0.5. The result of the fit is overlaid (blue line) and the different components detailed in the legend. The $B_s^0 \to \mu^+\mu^-\gamma$ component is left free to float in the fit.

The distributions of the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ components, described by individual Double Crystal Ball functions, are characterized by the certain parameters which are constrained in the fit, including the mean and mass resolution.

Shapes from physical backgrounds are determined on simulation using Gaussian Kernel Functions (RooKeys) within each category. The combinatorial background is free to float and modeled by a single exponential whose slope is common between the BDT categories, but independent for Run 1 and Run 2.

The fit result for the most sensitive region BDT > 0.5 is shown in Fig. 6.18. From this fit, the following branching fractions are obtained:

$$\begin{aligned}
\mathcal{B}(B_s^0 \to \mu^+\mu^-) &= (3.09^{+0.48}_{-0.45}) \times 10^{-9} \\
\mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.20^{+0.85}_{-0.75}) \times 10^{-10} \\
\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) &= (-2.53^{+1.59}_{-1.56}) \times 10^{-9} \\
\mathcal{R} = \frac{\mathcal{B}(B^0 \to \mu^+\mu^-)}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)} &= 0.039^{+0.030}_{-0.025}
\end{aligned}$$

where $\mathcal{R}$ is obtained by repeating the fit using $\mathcal{R} \times \mathcal{B}(B_s^0 \to \mu^+\mu^-)$ as a fit parameter in place of $\mathcal{B}(B^0 \to \mu^+\mu^-)$.

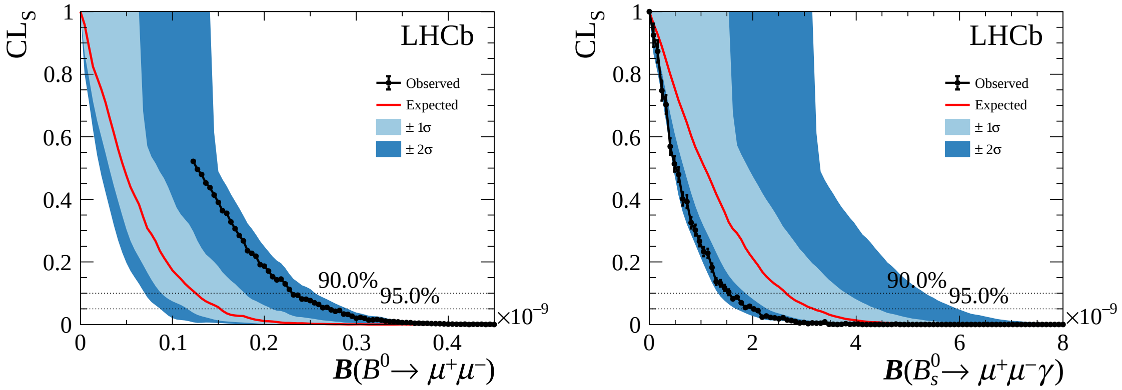


Figure 6.19: CL$_\mathrm{s}$-value distributions for the signal parameters $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$. The observed CL$_\mathrm{s}$-value distributions are compared to the expected values from background-only pseudo experiments ($\mathcal{B} = 0$). The red line corresponds to the expected upper limit curve.

The significance of the $B_s^0 \to \mu^+\mu^-$ signal is 10.3σ, while the significance of the $B^0 \to \mu^+\mu^-$ signal is 1.7σ. The $B^0 \to \mu^+\mu^-$ signal is therefore consistent with the background-only hypothesis, meaning that the signal yields are consistent with 0 within the given uncertainties. This means that this branching fraction result is instead converted into a limit with a given confidence level (CL). This gives the top limit on the branching fraction, *i.e.*, what the maximum branching fraction could be given the results of the fit. As the $B^0 \to \mu^+\mu^-$ significance is only 1.7σ, the CL$_\mathrm{s}$ method is used to set a limit on the branching fraction, which is found to be $\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.3(2.6) \times 10^{-10}$ at 90 %(95 %) CL.

The $B_s^0 \to \mu^+\mu^-\gamma$ signal is also consistent with the background-only hypothesis, so the same method is applied. An upper limit on the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction is found at $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) < 1.5(2.0) \times 10^{-9}$ at 90 % (95 %) CL. The measured upper limits are visualised in Fig. 6.19.

As shown in the results, the fit prefers a branching fraction of $B_s^0 \to \mu^+\mu^-$ consistent with the SM and a negative value for the branching fraction of $B_s^0 \to \mu^+\mu^-\gamma$, the latter suggesting an under-fluctuation of the background in the left sideband. To investigate the impact of the $B_s^0 \to \mu^+\mu^-\gamma$ under-fluctuation on $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ signals, the fit is also performed by fixing the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction to 0. From this fit, the following branching fractions are obtained:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.15^{+0.49}_{-0.45}) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (0.87^{+0.81}_{-0.72}) \times 10^{-10}$$

which are compatible with the values obtained in the nominal fit. In addition, no significant deviations are observed in the pulls of the left sideband ([4900,5200] MeV/c^2) that could justify a significant mismodelling of the semileptonic backgrounds. With this configuration, the upper limit on $\mathcal{B}(B^0 \to \mu^+\mu^-)$ is also evaluated using $\mathrm{CL_s}$ which is found to be $\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.0(2.4) \times 10^{-10}$ at 90 %(95 %) CL.

6.12.1 Statistical and systematic uncertainties

The statistical uncertainty is derived by repeating the fit after fixing all the fit parameters, except for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions, the background yields and the slope of the combinatorial background, to their expected values. The systematic uncertainties of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$ are dominated by the uncertainty on f_s/f_d and the knowledge of the exclusive backgrounds, respectively. The branching fractions can be written as

$$\begin{aligned}
\mathcal{B}(B^0_s \to \mu^+\mu^-) &= (3.09^{+0.46}_{-0.43}(\text{stat})^{+0.15}_{-0.11}(\text{syst})) \times 10^{-9} \\
\mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.20^{+0.83}_{-0.74}(\text{stat}) \pm 0.14(\text{syst})) \times 10^{-10} \\
\mathcal{B}(B^0_s \to \mu^+\mu^-\gamma) &= (-2.5 \pm 1.4(\text{stat}) \pm 0.8(\text{syst})) \times 10^{-9} \\
\mathcal{R} &= 0.039^{+0.030}_{-0.024}{}^{+0.006}_{-0.004}
\end{aligned}$$

Chapter 7

Summary and Future outlook

This final chapter will discuss the results presented in this thesis, as well as their interpretation and implications with respect to previous measurements. The future prospects for these measurements will also be discussed.

7.1 Summary

In this thesis, the full Run 1 and Run 2 data sample from the LHCb experiment was analysed to measure the $B_s^0 \to \mu^+\mu^-$ branching fraction, and to search for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays. Prior to the analysis, feasibility studies were conducted to ensure that the $B_s^0 \to \mu^+\mu^-\gamma$ decay *could* be searched for using the partial reconstruction method within the $B_{(s)}^0 \to \mu^+\mu^-$ analysis framework.

The branching fractions of the $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ (with $m_{\mu\mu} > 4.9\,\text{GeV}/c^2$) decays, as well as the ratio $\mathcal{R}$ of the $B^0 \to \mu^+\mu^-/B_s^0 \to \mu^+\mu^-$ rates, have been measured to be;

$$\begin{aligned}
\mathcal{B}(B_s^0 \to \mu^+\mu^-) &= (3.09\,^{+\,0.46\,+\,0.15}_{-\,0.43\,-\,0.11}) \times 10^{-9} \\
\mathcal{B}(B^0 \to \mu^+\mu^-) &= (1.20^{+0.83}_{-0.74} \pm 0.14) \times 10^{-10} \\
\mathcal{R} &= 0.039\,^{+\,0.030\,+\,0.006}_{-\,0.024\,-\,0.004} \\
\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) &= (-2.5 \pm 1.4 \pm 0.8) \times 10^{-9}
\end{aligned}$$

where the first uncertainties are statistical and the second are systematic.

For the $B_s^0 \to \mu^+\mu^-$ branching fraction, this result represents a significant improvement in precision compared to the previous LHCb measurement, due to the reduced systematic uncertainty. This $B_s^0 \to \mu^+\mu^-$ branching fraction is in fact the most precise single-experiment measurement to date. The improvements are primarily due to a new precise value of the hadronisation fraction ratio, as well as a more precise calibration of the particle misidentification rates.

The $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ signals are consistent with the background-only hypothesis, meaning that the signal yields are consistent with 0 within the given uncertainties. Therefore, upper limits on the branching fractions are set to

$$\begin{aligned}\mathcal{B}(B^0 \to \mu^+\mu^-) &< 2.6 \times 10^{-10}\\ \mathcal{B}(B^0_s \to \mu^+\mu^-\gamma) &< 2.0 \times 10^{-9}\end{aligned}$$

at 95% CL, the latter with $m_{\mu\mu} > 4.9\,\mathrm{GeV}/c^2$. The limit on the $B^0 \to \mu^+\mu^-$ decay is the most stringent to date from a single experiment.

The upper limit on the $B^0_s \to \mu^+\mu^-\gamma$ branching fraction represents the first limit set on this decay. It is important to emphasise that the limit is only set for the high invariant mass region, being $m_{\mu\mu} > 4.9\,\mathrm{GeV}/c^2$, and there has been no attempt in this work to extrapolate the result to the full branching fraction (over the full $m_{\mu\mu}$ range).

All of the results presented in this thesis are compatible with the SM, and the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ results are consistent with previous branching fraction measurements and limits respectively. As such, the results place further constraints on the possible new physics contributions to these decays. The results have recently been published by the LHCb collaboration and can be found in Refs. [5] and [6].

7.1.1 Future Outlook

With the imminent start of Run 3 for the LHC comes several opportunities for improvements on the measurements presented in this thesis. The results are statistically limited, meaning that an increased LHCb dataset will allow us to make more precise measurements for all three of the signals discussed here.

The current dataset (Run 1 and Run 2 collectively) totals 6 fb^{-1} at a centre-of-mass energy of 13 TeV. With the upgrades that are currently being implemented in preparation for Runs 3 and 4, it should be possible to collect a further 41 fb^{-1} in total over these runs. This will be made possible by the implementation of a totally software based trigger for taking data with an instantaneous luminosity around five times larger than it is currently. Since the $B^0_s \to \mu^+\mu^-$ decay has already been observed, the aim of future measurements will be to achieve a precision measurement to probe New Physics. With the increased dataset comes the opportunity to increase this precision, as well as hopefully achieving observation of the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-\gamma$ channels. A significant $B^0 \to \mu^+\mu^-$ signal, and the following calculation of the ratio between the B^0_s and B^0 decay modes, will also provide a test of minimal flavour violation.

The systematic uncertainty on the $B^0_s \to \mu^+\mu^-$ measurement will have to be decreased to take full advantage of the increased data sets, which relies on more precise measurement of the relative production fraction of B^0_s and B^0 mesons, fs/fd.

Following the feasibility studies conducted in Chapter 5, it would also be interesting to conduct a search for $B^0_s \to \mu^+\mu^-\gamma$ using the same technique used in this thesis, but with the invariant mass region extended down to $4.5\,\mathrm{GeV}/c^2$. This would require careful investigation of any additional backgrounds that could appear in this region.

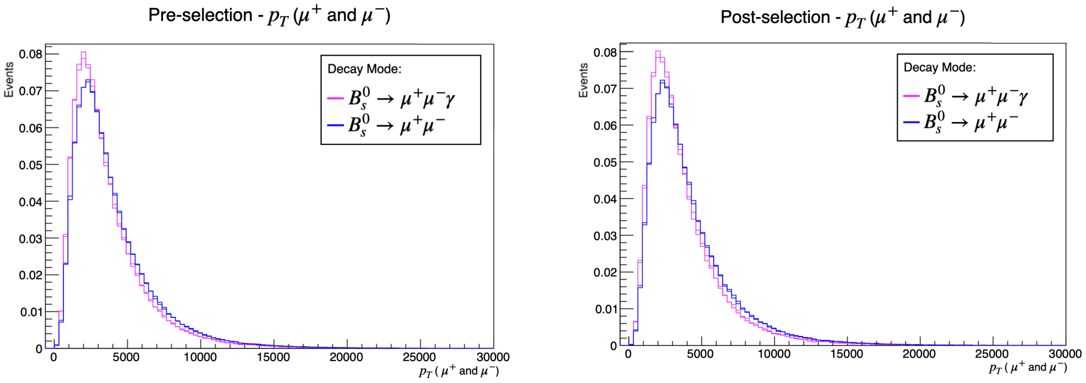
Following Run 4, a second upgrade of the LHCb detector is also being planned, which will increase the instantaneous luminosity even more.

With all these planned upgrades we should finally be able to confirm whether there is new physics at play in these B meson decays, or not. Either way, we take a step towards understanding the fundamental nature of the universe.

Appendix A

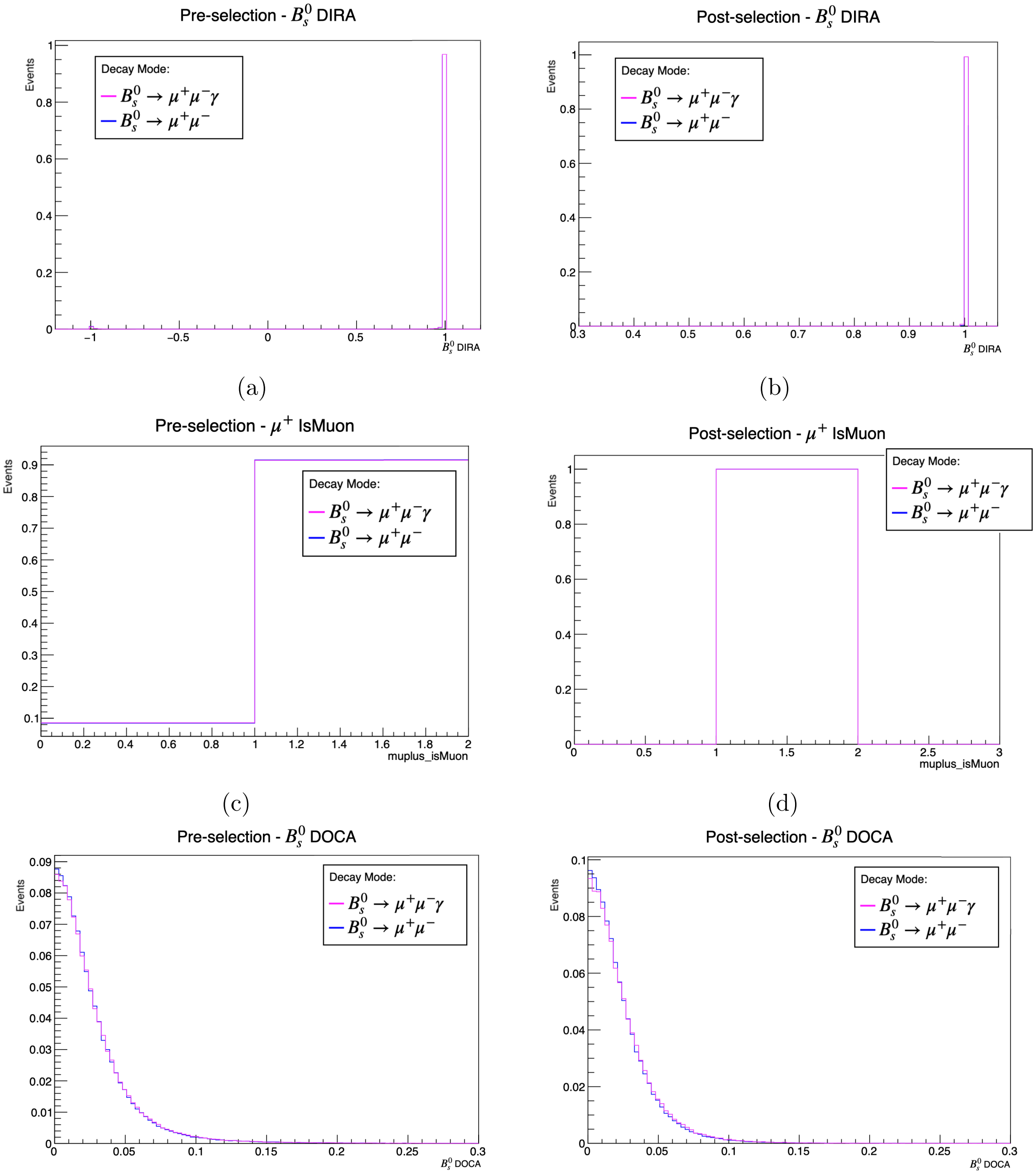
Selection variable shape comparisons for $B_s^0 \to \mu^+\mu^-\gamma$ feasibility studies

This section contains the distribution shape comparisons of for $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays, for each of the variables used in the selection process. Variables showing significant differences between the two signals are included in the main text (Chapter 5). For variable definitions see Sections 5.3.3 and 5.3.2.



(a) (b)

Figure A.1: Signal distributions of $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ for p_{T} of the two muons prior to selection (a) and after selection (b). Plots normalised to unit area.

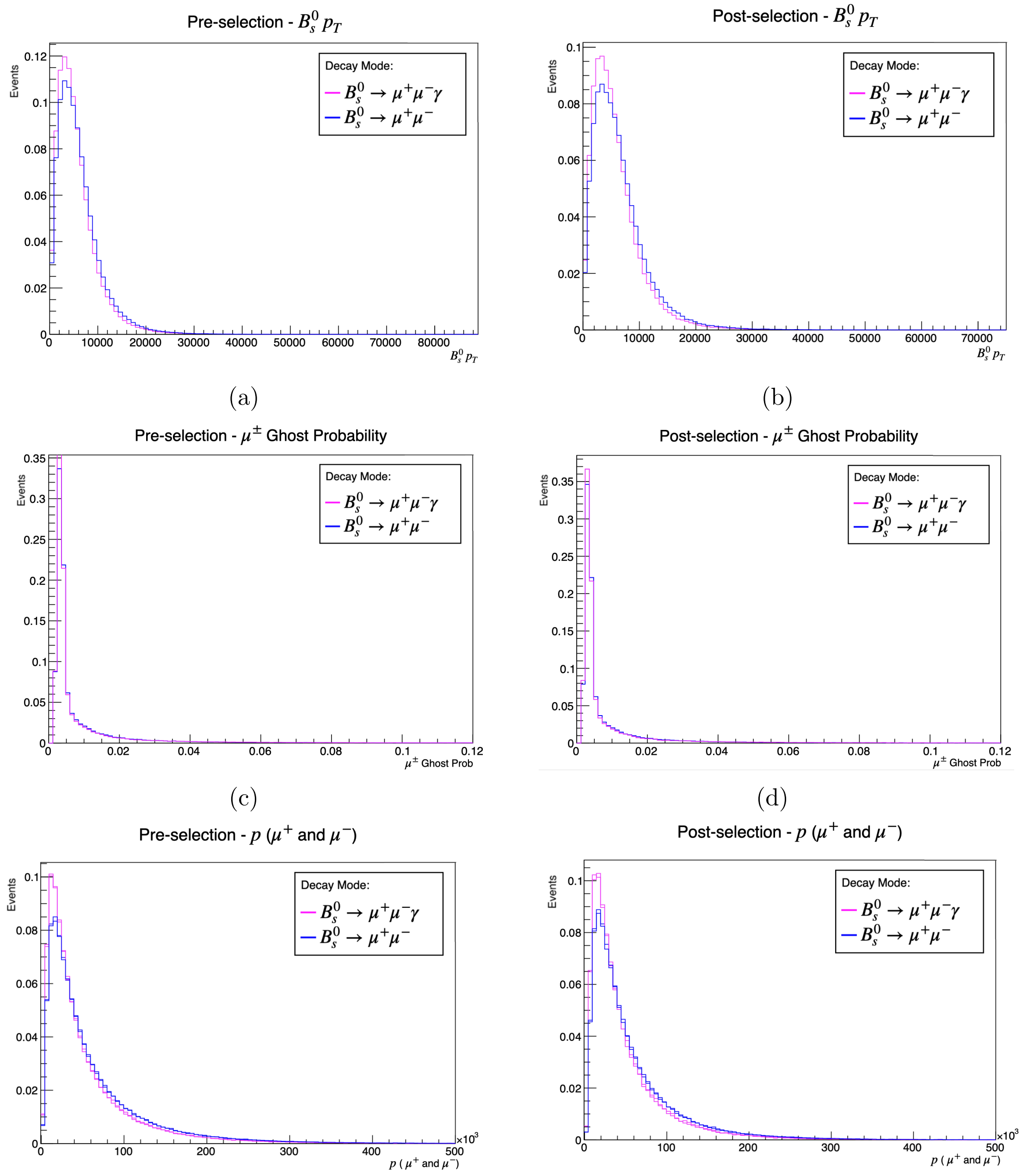


(a) (b)

(c) (d)

(e) (f)

Figure A.2: Signal distributions of $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ for DIRA of the B candidate prior to selection (a) and after selection (b); isMuon output of signal muons prior to selection (c) and after selection (d); and lastly, DOCA of the B candidate prior to selection (e) and after selection (f). Plots normalised to unit area.



(a) (b)

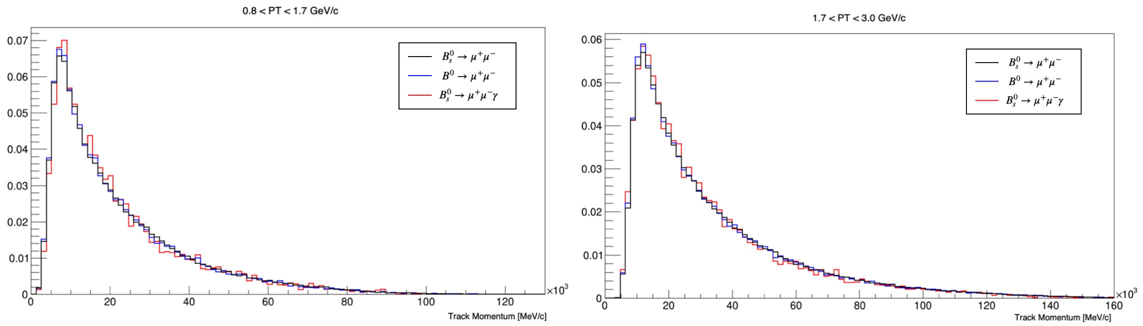
(c) (d)

(e) (f)

Figure A.3: Signal distributions of $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ for p_{T} of the B candidate prior to selection (a) and after selection (b); Ghost Probability output of signal muons prior to selection (c) and after selection (d); and lastly, p of the two signal muons prior to selection (e) and after selection (f). Plots normalised to unit area.

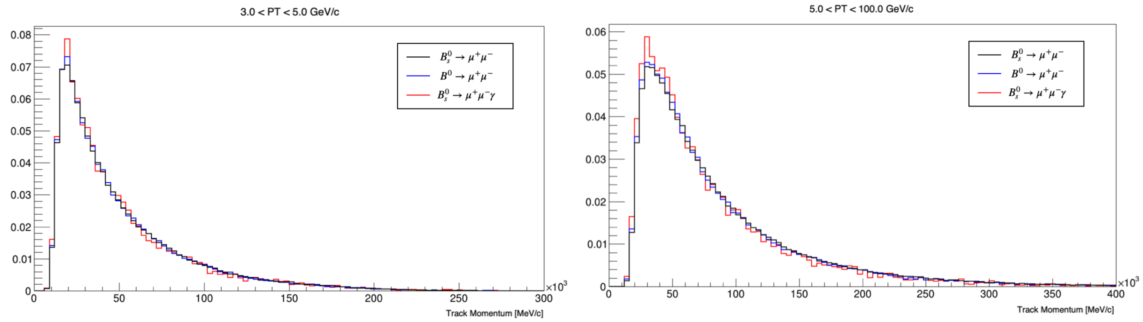
Appendix B

MC momentum distributions for the three signal channels



(a) p_T bin 1

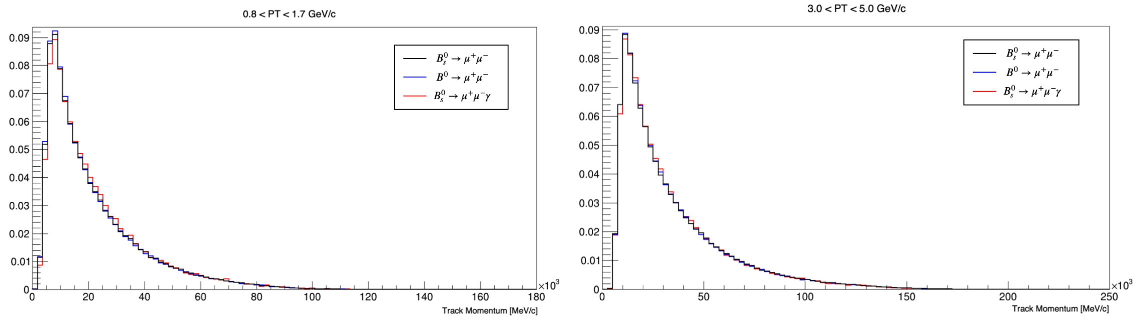
(b) p_T bin 2



(c) p_T bin 3

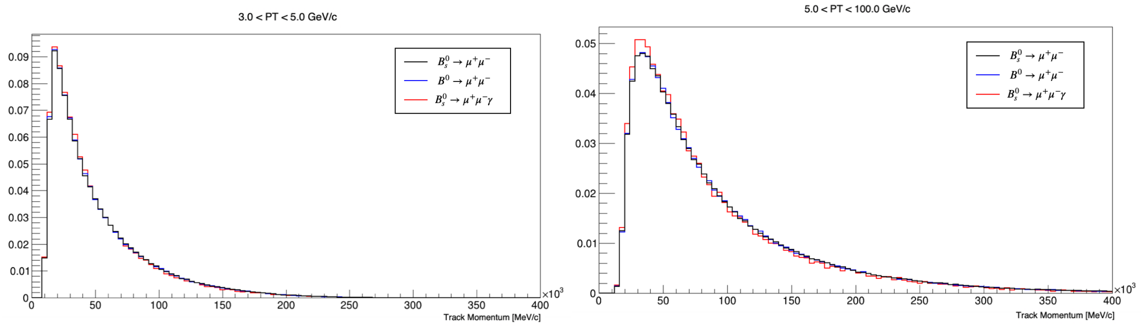
(d) p_T bin 4

Figure B.1: Normalised Run 1 MC momentum distributions for $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$, in four p_T bins.



(a) p_T bin 1

(b) p_T bin 2



(c) p_T bin 3

(d) p_T bin 4

Figure B.2: Normalised Run 2 MC momentum distributions for $B_s^0 \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$, in four p_T bins.

Bibliography

[1] The LHCb Collaboration and CMS collaboration, *Observation of the rare* $B_s^0 \to \mu^+\mu^-$ *decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68–72.

[2] F. Archilli, *Observation of* $B_s^0 \to \mu^+\mu^-$ *and first measurement of its effective lifetime with LHCb Run-2 data*, `https://cds.cern.ch/record/2253469`, (2017).

[3] The 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) 191801, `arXiv:1703.05747`.

[4] The LHCb collaboration, *Combination of the ATLAS, CMS and LHCb results on the* $B_{(s)}^0 \to \mu^+\mu^-$ *decays*, LHCb-CONF-2020-002. CERN-LHCb-CONF-2020-002, CERN, Geneva, 2020.

[5] LHCb Collaboration, R. Aaij, C. Abellán Beteta, T. Ackernley *et al.*, *Measurement of the* $B_s^0 \to \mu^+\mu^-$ *decay properties and search for the* $B^0 \to \mu^+\mu^-$ *and* $B_s^0 \to \mu^+\mu^-\gamma$ *decays* , 2021. All figures and tables, along with machine-readable versions and any supplementary material and additional information, are available at https://cern.ch/lhcbproject/Publications/p/LHCb-PAPER-2021-008.html. V2 adds the arXiv identifier for the simultaneously-submitted LHCb-PAPER-2021-007, arxiv:2108.09284.

[6] LHCb Collaboration, R. Aaij, C. Abellán Beteta, T. Ackernley *et al.*, *Analysis of neutral B-meson decays into two muons* , CERN, Geneva, 2021. All figures and tables, along with machine-readable versions and any supplementary material and additional information, are available at https://cern.ch/lhcbproject/Publications/p/LHCb-PAPER-2021-007.html (LHCb public pages).

[7] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013.

[8] C. S. Unnikrishnan, *Bosons, fermions, spin, gravity, and the spin-statistics connection*, 2019.

[9] The LHCb Collaboration, R. Aaij *et al.*, *Observation of the Resonant Character of the $Z(4430)^-$ State*, Phys. Rev. Lett. **112** (2014) 222002.

[10] The LHCb Collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115** (2015) 072001.

[11] J. Iliopoulos, *Introduction to the standard model of the electro-weak interactions*, 2013.

[12] Wikimedia Commons, *Standard Model of Elementary Particles - Wikimedia Commons, the free media repository*, 2020. [Online; accessed 9-March-2021].

[13] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cim. **19** (1961) 154.

[14] J. Ellis, *Higgs Physics*, `https://cds.cern.ch/record/1638469`, 2013. doi: 10.5170/CERN-2015-004.117.

[15] M. Bargiotti *et al.*, *Present knowledge of the Cabibbo-Kobayashi-Maskawa matrix*, Riv. Nuovo Cim. **23N3** (2000) 1, `arXiv:hep-ph/0001293`.

[16] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics (RPP)*, Chin. Phys. **C40** (2017) 010001.

[17] T. Inami and C. S. Lim, *Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes $K_L \to \mu\bar{\mu}$, $K+ \to \pi^+\nu\bar{\nu}$ and $K^0 \leftrightarrow \bar{K^0}$*, Prog. Theor. Phys. **65** (1981) 297, [Erratum: Prog.Theor.Phys. 65, 1772 (1981)].

[18] A. Arbey, M. Battaglia, F. Mahmoudi, and D. Martínez Santos, *Supersymmetry confronts $B_s \to \mu^+\mu^-$: Present and future status*, Phys. Rev. **D87** (2013) 035026, `arXiv:1212.4887`.

[19] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, Reviews of Modern Physics **68** (1996) 1125–1244.

[20] A. J. Buras, R. Fleischer, J. Girrbach, and R. Knegjens, *Probing New Physics with the $B^0 \to \mu^+\mu^-$ Time-Dependent Rate*, JHEP **07** (2013) `arXiv:1303.3820`.

[21] X.-Q. Li, J. Lu, and A. Pich, *$B^0_{s,d} \to \ell^+\ell^-$ decays in the aligned two-Higgs-doublet model*, Journal of High Energy Physics (2014) .

[22] HPQCD, B. Colquhoun *et al.*, *B-meson decay constants: a more complete picture from full lattice QCD*, Phys. Rev. **D91** (2015) 114509, `arXiv:1503.05762`.

[23] C. Bobeth *et al.*, *$B_{s,d} \to \ell^+\ell^-$ in the Standard Model with Reduced Theoretical Uncertainty*, Phys. Rev. Lett. **112** (2014) 101801, `arXiv:1311.0903`.

[24] D. Melikhov and N. Nikitin, *Rare radiative leptonic decays* $B^0_{(d,s)} \to \ell^+\ell^-\gamma$, Phys. Rev. **D70** (2004) 114028, `arXiv:hep-ph/0410146`.

[25] F. Dettori, D. Guadagnoli, and M. Reboud, $B^0_s \to \mu^+\mu^-\gamma$ *from* $B^0_s \to \mu^+\mu^-$, Phys. Lett. **B768** (2017) 163, `arXiv:1610.00629`.

[26] The LHCb Collaboration, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay* $B^0_s \to \phi\mu^+\mu^-$, JHEP **09** (2015) 179, `arXiv:1506.08777`.

[27] The LHCb Collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, `arXiv:2103.11769`.

[28] A. Kozachuk, D. Melikhov, and N. Nikitin, *Annihilation-type rare radiative* $B_{(s)} \to V\gamma$ *decays*, Physical Review D **93** (2016) .

[29] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, *On the Standard Model prediction for* $\mathcal{B}(B^0_{s,d} \to \mu^+\mu^-)$, Eur. Phys. J. **C72** (2012) 2172, `arXiv:1208.0934`.

[30] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, The European Physical Journal C **45** (2006) 97–107.

[31] R. J. Knegjens, *Strategies to Hunt for New Physics with Strange Beauty Mesons*, PhD thesis, Vrije U., Amsterdam, 2014.

[32] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, *Minimal flavour violation: an effective field theory approach*, Nuclear Physics B **645** (2002) 155–187.

[33] L.-S. Geng *et al.*, *Towards the discovery of new physics with lepton-universality ratios of* $b \to s\ell\ell$ *decays*, Physical Review D **96** (2017) .

[34] A. J. Buras *et al.*, *Patterns of flavour violation in the presence of a fourth generation of quarks and leptons*, Journal of High Energy Physics **2010** (2010) .

[35] M. Blanke *et al.*, *Rare K and B Decays in a warped extra dimension with custodial protection*, Journal of High Energy Physics **2009** (2009) 108–108.

[36] L. J. Hall and H. Murayama, *Geometric model of the generations*, Physical Review Letters **75** (1995) 3985–3988.

[37] G. G. Ross, L. Velasco-Sevilla, and O. Vives, *Spontaneous cp violation and non-abelian family symmetry in susy*, Nuclear Physics B **692** (2004) 50–82.

[38] S. Antusch, S. F. King, and M. Malinský, *Solving the susy flavour and cp problems with su(3) family symmetry*, Journal of High Energy Physics **2008** (2008) 068–068.

[39] K. Agashe and C. D. Carone, *Supersymmetric flavor models and the* $b \to \phi k_s$ *anomaly*, Physical Review D **68** (2003) .

[40] D. M. Straub, *Overview of Constraints on New Physics in Rare B Decays*, in *47th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, 237–242, 2012, `arXiv:1205.6094`.

[41] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[42] The ALICE Collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[43] The ATLAS Collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[44] The CMS Collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[45] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, 2016. General Photo.

[46] The LHCb collaboration, C. Elsässer, *$\bar{b}b$ production angle plots*, `https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb\_ProductionAngles.html`, 2020.

[47] The LHCb Collaboration, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30** (2015) 1530022, `arXiv:1412.6352`.

[48] The LHCb Collaboration, J. Alves, A.Ãugusto *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[49] The LHCb Collaboration, *LHCb magnet: Technical design report*, 2000.

[50] The LHCb Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb outer tracker: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2001.

[51] N. Brook *et al.*, *LHCb RICH1 Engineering Design Review Report Members of the LHCb RICH group*, 2020.

[52] LHCb Collaboration, S. Amato *et al.*, *LHCb RICH: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2000.

[53] The LHCb Collaboration, E. P. Olloqui *et al.*, *LHCb preshower (PS) and scintillating pad detector (SPD): Commissioning, calibration, and monitoring*, in *Journal of Physics: Conference Series*, **160**, 012046, IOP Publishing, 2009.

[54] LHCb Collaboration, S. Amato, D. Carvalho *et al.*, *LHCb calorimeters: Technical Design Report*, Technical design report. LHCb, CERN, Geneva, 2000.

[55] A. A. Alves *et al.*, *Performance of the LHCb muon system*, Journal of Instrumentation **8** (2013) P02022–P02022.

[56] M. Santimaria, *Search for the* $B_d^0 \to \mu^+\mu^-$ *decay and measurement of the* $B_s^0 \to \mu^+\mu^-$ *branching fraction and effective lifetime*, , PhD Thesis. Presented 16 Feb 2018.

[57] The LHCb collaboration, *Trigger Schemes*, `http://lhcb. web. cern. ch/lhcb/speakersbureau/html/TriggerScheme. html` (2020).

[58] J. P. Dufey *et al.*, *The LHCb trigger and data acquisition system*, IEEE Trans. Nucl. Sci. **47** (2000) 86.

[59] M. Adinolfi *et al.*, *LHCb data quality monitoring*, Journal of Physics: Conference Series **898** (2017) 092027.

[60] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP '96 Workshop*, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81.

[61] The LHCb collaboration, *BRUNEL, The LHCb Event Reconstruction program*, `http://lhcb-release-area. web. cern. ch/LHCb-release-area/DOC/brunel/` (2021).

[62] P. Mato, *GAUDI-Architecture design document*, LHCb note 98-064 (2020).

[63] The LHCb collaboration, *DAVINCI, The Physics Analysis Software for the LHCb experiment*, `http://lhcb-release-area. web. cern. ch/LHCb-release-area/DOC/davinci/` (2020).

[64] R. Hierk, M. Merk, M. Needham, and R. Van der Eijk, *Performance of the LHCb 00 track fitting software* , CERN, Geneva, 2000.

[65] E. Rodrigues, *Tracking definitions*, LHCb-2007-006, CERN-LHCb-2007-006, CERN, Geneva, 2007.

[66] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb*, LHCb-PUB-2014-044. CERN-LHCb-PUB-2014-044, CERN, Geneva, 2014.

[67] The LHCb Collaboration, $B_s^0 \to \mu^+\mu^-$ *event display*, `http://lhcb-public.web.cern.ch/en/Images2017/BsMuMuVertex.png`, 2016. [Online; accessed 06-May-2021].

[68] S. Tolk, *Discovery of Rare B Decays: First Observation of the* $B_s^0 \to \mu^+\mu^-$ *Decay, First Evidence of the* $B_0^s \to \mu^+\mu^-$ *Decay*, PhD thesis, Groningen U., 2016.

[69] LHCb MuonID group, F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020. 20 p, Comments: 17 pages, 10 figures.

[70] D. Derkach, M. Hushchyn, and N. Kazeev, *Machine learning based global particle identification algorithms at the lhcb experiment*, EPJ Web of Conferences **214** (2019) 06011.

[71] Particle Data Group, P. A. Zyla *et al.*, *Review of Particle Physics*, PTEP **2020** (2020) 083C01.

[72] R. W. Forty and O. Schneider, *RICH pattern recognition* , CERN, Geneva, 1998.

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

[74] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`.

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

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

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

[78] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide*, physics/0703039, CERN, Geneva, 2007.

[79] F. Archilli *et al.*, *Background studies for $B^0 \to \mu^+\mu^-$ analysis optimization*, LHCb-INT-2014-047. CERN-LHCb-INT-2014-047, CERN, Geneva, 2014.

[80] C. Jones, *ANN PID Retuning for Reco14 Data*, `https://indico.cern.ch/event/226062/contributions/475644/attachments/371741/517276/ANNPIDRetuning-Reco14-06052013.pdf`, 2013.

[81] R. Aaij *et al.*, *Strong constraints on the rare decays $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$*, Physical Review Letters **108** (2012) .

[82] C. Adrover *et al.*, *Search for the rare decays $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ with 1.02 fb $^{-1}$*, .

[83] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039, 2014.

[84] R. Aaij *et al.*, *Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B^0_s decay branching fractions*, Physical Review D **104** (2021) .

[85] The LHCb Collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP **04** (2013) 001, `arXiv:1301.5286`.

[86] C. Bobeth, M. Gorbahn, and E. Stamou, *Electroweak Corrections to $B_{s,d} \to \ell^+\ell^-$*, Phys. Rev. D **89** (2014) 034023, `arXiv:1311.1348`.

[87] T. Hermann, M. Misiak, and M. Steinhauser, *Three-loop QCD corrections to $B_s \to \mu^+\mu^-$*, JHEP **12** (2013) 097, `arXiv:1311.1347`.

[88] M. Beneke, C. Bobeth, and R. Szafron, *Enhanced electromagnetic correction to the rare B-meson decay $B_{s,d} \to \mu^+\mu^-$*, Phys. Rev. Lett. **120** (2018) 011801, `arXiv:1708.09152`.

[89] J. Aebischer *et al.*, *B-decay discrepancies after Moriond 2019*, Eur. Phys. J. C **80** (2020) 252, `arXiv:1903.10434`.

[90] K. Petridis and M. Santimaria, *New results on theoretically clean observables in rare B-meson decays from LHCb*, .

[91] F. Kruger and D. Melikhov, *Gauge invariance and form-factors for the decay $B \to \gamma l^+l^-$*, Phys. Rev. D **67** (2003) 034002, `arXiv:hep-ph/0208256`.

[92] C. Q. Geng, C. C. Lih, and W.-M. Zhang, *Study of $B_{(s,d)} \to l^+l^-\gamma$ decays*, Phys. Rev. D **62** (2000) 074017, `arXiv:hep-ph/0007252`.

[93] Y. Dincer and L. M. Sehgal, *Charge asymmetry and photon energy spectrum in the decay $B_{(s)} \to l^+l^-\gamma$*, Phys. Lett. B **521** (2001) 7, `arXiv:hep-ph/0108144`.

[94] S. Descotes-Genon and C. T. Sachrajda, *Universality of nonperturbative QCD effects in radiative B decays*, Phys. Lett. B **557** (2003) 213, `arXiv:hep-ph/0212162`.

[95] T. M. Aliev, A. Ozpineci, and M. Savci, *B(q) → lepton+ lepton- gamma decays in light cone QCD*, Phys. Rev. **D55** (1997) 7059, `arXiv:hep-ph/9611393`.

[96] BaBar, B. Aubert *et al.*, *Search for the decays $B^0 \to e^+e^-\gamma$ and $B^0 \to \mu^+\mu^-\gamma$*, Phys. Rev. D **77** (2008) 011104, `arXiv:0706.2870`.

[97] C. Adrover et al. , *Search for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with 3 fb^{-1} at LHCb*, LHCb-ANA-2013-032 (2013).

[98] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, 2016.

[99] W. Huang and G. Wormser, *Limitations regarding the present usage of PIDCalib to compute efficiencies and misIDs*, `https://indico.cern.ch/event/869049/`

`contributions/3663861/attachments/1960109/3257320/PIDCalib_run1run2.pdf`, 2019.

[100] M. Palutan, M. Rama, and M. Santimaria, *Ongoing studies of* $\pi \to \mu$ *and* $K \to \mu$ *misIDs for the* $B^0 \to \mu^+\mu^-$ *analysis*, `https://indico.cern.ch/event/913160/contributions/3843422/attachments/2028355/3393962/misids_28apr2020.pdf`, 2020.

[101] M. Palutan, M. Rama, and M. Santimaria, *Pion and kaon misid in muon selection. ongoing checks of pidcalib output*, `https://indico.cern.ch/event/887097/contributions/3747736/attachments/1987152/3311567/misidKpi_2dfit_vs_pidcalib.pdf`, 2020.

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

[103] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, CERN-LHCb-PUB-2014-039, CERN, Geneva, 2014.

[104] F. Archilli *et al.*, *Search for the* $B^0 \to \mu^+\mu^-$ *decay and measurement of the* $B^0_s \to \mu^+\mu^-$ *branching fraction and effective lifetime*, LHCb-ANA-2016-038 (2016).

[105] G. Mancinelli and J. Serrano, *Study of Muon Isolation in the* $B^0_s \to \mu^+\mu^-$ *Channel*, CERN-LHCb-INT-2010-011, CERN, Geneva, 2010.

[106] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, Journal of Instrumentation **8** (2013) P04022–P04022.

[107] The LHCb Collaboration, *PIDCalibPackage*, `https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage`, 2021. [Online; accessed 21-May-2021].

[108] M. Pivk and F. R. Le Diberder, *splot: A statistical tool to unfold data distributions*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **555** (2005) 356–369.

[109] S. Malde, *PIDCalib*, `https://indico.cern.ch/event/462919/contributions/1137595/attachments/1194293/1734684/PIDGBU.pdf`, 2015. [Online; accessed 21-May-2021].

[110] D. Scora and N. Isgur, *Semileptonic meson decays in the quark model: An update*, Phys. Rev. D **52** (1995) 2783, `arXiv:hep-ph/9503486`.

[111] N. Isgur, D. Scora, B. Grinstein, and M. B. Wise, *Semileptonic B and D Decays in the Quark Model*, Phys. Rev. D **39** (1989) 799.

[112] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, `arXiv:1909.12524`.

[113] LHCb Collaboration, R. Aaij *et al.*, *First observation of the decay* $B_s^0 \to K^- \mu^+ \nu_\mu$ *and measurement of* $|V_{ub}|/|V_{cb}|$, Phys. Rev. Lett. **126** (2020) 081804, `arXiv:2012.05143`.

[114] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays* $B/B_s \to (\pi, K)(\ell^+\ell^-, \ell\nu, \nu\bar{\nu})$ *in the perturbative QCD approach beyond the leading-order*, Phys. Rev. **D86** (2012) 114025, `arXiv:1207.0265`.

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

[116] The LHCb collaboration, R. Aaij *et al.*, *Measurement of b-hadron fractions in 13 TeV p p collisions*, Phys. Rev. **D100** (2019) 031102(R), `arXiv:1902.06794`.

[117] The LHCb collaboration, R. Aaij *et al.*, *Measurement of the* B_c^- *production fraction and asymmetry in 7 and 13 TeV pp collisions*, Phys. Rev. **D100** (2019) 112006, `arXiv:1910.13404`.

[118] The LHCb collaboration, *Measurement of the track reconstruction efficiency at LHCb*, Journal of Instrumentation **10** (2015) P02007.