

Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Rowina Caspary

born in Freiberg (Germany)

2020

Search for the lepton flavour violating decay

$\tau^- \rightarrow \mu^- \mu^+ \mu^-$ with the LHCb experiment

This Master thesis has been carried out by Rowina Caspary

at the

Physikalisches Institut

under the supervision of

Prof. Dr. Stephanie Hansmann-Menzemer

Zusammenfassung:

In dieser Arbeit wird die erwartete Sensitivität für die Messung des Verzweigungsverhältnisses des Lepton-Flavour-verletzenden Zerfalls $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ berechnet. Der Zerfall kann im minimal erweiterten Standardmodell der Teilchenphysik nur durch Neutrinooszillationen mit einem Verzweigungsverhältnis $\mathcal{O}(10^{-55})$ auftreten. Theorien jenseits des Standardmodells sagen jedoch Werte bis zu 10^{-7} voraus. Die vom LHCb-Experiment gemessene Obergrenze liegt bei $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq 4.6 (5.6) \times 10^{-8}$. In dieser Arbeit wird die Messung auf Daten erweitert, die mithilfe des LHCb-Experiments bei Proton-Proton Kollisionen in den Jahren 2016 – 2018 bei einer Schwerpunktenenergie von $\sqrt{s} = 13$ TeV aufgenommen wurden und einer integrierten Luminosität von 5.5 fb^{-1} entsprechen. Eine multivariate Analyseverfahren wird genutzt, um Ereignisse als Signal oder Untergrund zu klassifizieren. Dabei werden neue Variablen, die die Isolation der Myonen quantifizieren, verwendet. Die Analyse wird getrennt für Signalereignisse mit zwei, beziehungsweise drei identifizierten Myonen und für die drei Datennahmejahre optimiert. Der Zerfall $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ wird als Referenz genutzt, um die erwartete Sensitivität für die sechs Unterklassen getrennt zu bestimmen. Mithilfe der CL_S -Methode werden die Sensitivitäten kombiniert. Es wird erwartet, dass eine Obergrenze von $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq 2.96 (3.58) \times 10^{-8}$ für ein Konfidenzintervall von 90% (95%) erreicht wird.

Abstract:

In this thesis, the expected sensitivity on the branching fraction of the lepton flavour violating decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ is estimated using 5.5 fb^{-1} of data collected at the LHCb detector in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV in the years 2016 – 2018. The minimal extended standard model of particle physics including neutrino oscillation predicts it at a branching fraction of $\mathcal{O}(10^{-55})$. However, some beyond standard model theories predict branching fractions up to 10^{-7} . While the previous measurement done by LHCb set an upper limit of $4.6 (5.6) \times 10^{-8}$, novel techniques are applied in this work in order to improve the sensitivity of the measurement. A multivariate classifier using new muon isolation variables is trained to distinguish signal and background to enhance the signal sensitivity. For each year, the data is separated into two sub-samples depending on the number of muon candidates triggered by the LHCb muon system. The $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ channel is used as reference channel to estimate an expected limit for all six sub-samples separately. The CL_S method is applied to combine them and results in $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq 2.96 (3.58) \times 10^{-8}$ at 90% (95%) confidence level.

Contents

1	Introduction	6
2	Theory	8
2.1	Standard Model	8
2.2	Lepton flavour violation	10
3	Data acquisition	12
3.1	Large Hadron Collider	12
3.2	LHCb detector	15
3.3	The LHCb trigger system	21
3.4	Monte Carlo simulation	22
3.5	Event reconstruction	23
4	Analysis outline	24
4.1	Analysis strategy	24
4.2	Signal channel	26
4.3	Reference and control channel	29
5	Analysis tools	30
5.1	ROOT analysis framework	30
5.2	<i>sPlot</i> technique	30
5.3	TMVA toolkit	31
5.4	CL_S technique	33
6	Data preselection	38
6.1	Selection variables	38
6.2	Stripping line selection	41
6.3	Production rates	42
6.4	Trigger requirements and event preselection	44
6.5	Monte Carlo correction	48
6.6	Signal channel preparation	51
7	Multivariate classifier training and application	54
7.1	Variable selection	54
7.2	Training optimisation	57
7.3	Application of the results	62
7.4	Correction in classifier output	64
7.5	Classifier performance validation	68

8	Expected limit	69
8.1	Punzi Figure of Merit	69
8.2	Limit extrapolation	75
8.3	Single event sensitivity	76
8.4	Results	81
9	Conclusion and Outlook	82
I	Appendix	85
A	Bibliography	86

1 Introduction

The ultimate goal of all particle physicists is the complete description of the underlying rules describing all of our universe and everything that happens within, from the big bang to the unknown future. While this goal itself might never be reached completely, huge steps towards it have been taken by the science community during the past centuries. These results converged ultimately in what is called the Standard Model of particle physics. This extensive model describes the particle components of the universe and the interactions between them. Ever since, the objective of particle physics has been to investigate the Standard Model for weaknesses and extend it for a better description of nature. So far almost every experiment confirmed the predictions made by the Standard Model. Nevertheless, one of the most recent additions is the fact that neutrinos possess mass and are able to oscillate between different states, resulting in the minimal extended Standard Model. This extension adds the possibility of *lepton flavour violating decays*, like the decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$, which is predicted with a very low branching fraction of $\mathcal{O}(10^{-55})$ [1]. At the same time, many theories beyond the Standard Model allow for this decay with a branching fraction at the level of $\mathcal{O}(10^{-10} - 10^{-7})$. Examples are:

- supersymmetric low-scale seesaw models [2]
- Little Higgs models [3]
- the Minimal Supersymmetric Standard Model [4]

The current best experimental limit on the branching fraction is $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) < 2.1 \times 10^{-8}$ at 90% confidence level, observed by the Belle Collaboration with data from the BELLE detector [5]. The best limit reached by LHCb (Large Hadron Collider beauty) at CERN is 4.6×10^{-8} at 90% confidence level [6]. With the new dataset taken by LHCb in the years 2016 – 2018 corresponding to a total integrated luminosity of 5.57 fb^{-1} [7], a higher sensitivity than that of BELLE is expected to be reached.

If the decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ can be observed in the LHCb data, this would be an unambiguous sign for new physics phenomena beyond the Standard Model and would broaden our understanding of physics considerably. On the other hand, if no evidence can be found, a new limit can be set, which would solidify the rules for the description of nature and put theories beyond the Standard Model to a more strict test.

The goal of this thesis is to use the power of multivariate analysis techniques in order to distinct $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ signal events from background events and compute signal selection efficiencies for the 2016 – 2018 dataset taken by LHCb to calculate

the sensitivity of the analysis.

In chapter 2 this thesis gives a short introduction into the Standard Model of physics and the physics background of the investigated processes, followed by an introduction to the general setup of the LHCb experiment at CERN in Chapter 3. Furthermore, an overview of the data acquisition process both for data measured by the LHCb experiment and simulated data is given. Chapter 4 introduces the signal and reference channel. Chapter 5 describes the analysis strategy and the relevant tools used in this thesis. In Chapter 6, the preparation of data is explained and subsequently in Chapter 7 the use of the multivariate analysis toolkit TMVA to separate signal and background is described. Chapter 8 presents the final results. Finally, in Chapter 9, they are summarized and discussed.

The analysis outlined in this thesis is performed within the context of the LHCb experiment, which is dedicated to the study of b and c hadron decays. The LHCb collaboration is an international collaboration composed of approximately 1400 scientists from 85 universities and laboratories and 19 countries.

The analysis team working on the $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ decay consists of F. Archilli¹, M. Calvi², S. Capelli², R. Caspary¹, M. Chrzaszc³, D. Fazzini², G. Frau¹ and M. Martinelli². This thesis relies on work done by other members of this group in a few instances, specifically in Section 6.3 and 8.3, which will be explicitly honoured at the beginning of the individual chapters.

¹University of Heidelberg, Physikalisches Institut Heidelberg, Germany

²The University of Milano Bicocca, Italy and INFN Milano

³Polish Academy of Sciences, Institute of Nuclear Physics (IFJ PAN) and University of Zurich

2 Theory

The aim of this thesis is the search for the lepton flavour violating decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$, which is highly suppressed in the minimal extended Standard Model of particle physics (SM) including neutrino oscillations. Various theories beyond the SM allow for the process with a much higher branching fraction, making it an interesting process to study in order to search for physics phenomena beyond the SM. In this section, the theoretical background for the $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ process is given together with an introduction to the Standard Model of particle physics. The process of lepton flavour conservation, which is violated in $\tau^- \rightarrow \mu^- \mu^+ \mu^-$, is explained.

2.1 Standard Model

The Standard Model (SM) of particle physics is a relativistic, renormalisable quantum field theory describing the fundamental constituents of the universe, the *elementary particles*, and the *forces*, which allow for interactions between them. It provides a full description of all interactions in nature except for gravity at the most basic level. It successfully describes the overwhelming majority of the current experimental data taken in laboratory experiments.

In the SM, all known matter is built up of twelve matter particles and twelve anti-matter particles. The latter mirror the matter particles by having the exactly opposite charges, but the same mass. All elementary matter and anti-matter particles are fermions of half-integer spin. Responsible for interactions between the particles, and between each other, are five *gauge bosons* of spin 1, as well as the Higgs boson with spin 0. All elementary matter particles can be defined as mass or flavour eigenstates. For quarks and charged leptons, the mass eigenstate corresponds mostly to one flavour eigenstate. The particle content of the SM is summarised in Tab. 2.1 for fermions and Tab. 2.2 for bosons with their most important properties. The convention $c = 1$ and $\hbar = 1$ is used throughout the thesis.

The twelve fundamental matter particles are divided into six quarks and six leptons. The leptons are the electron, e^- , the muon, μ^- and the tau, τ^- , as well as their respective neutrinos, the electron-neutrino, ν_e , the muon-neutrino, ν_μ , and the tau-neutrino, ν_τ . The six quarks are the up-quark, u , the down-quark, d , the strange-quark, s , the charm-quark, c , the bottom-quark or beauty-quark, b , and the top-quark or truth-quark, t . All twelve fermions, as well as their anti-matter counterparts, interact via the weak interaction. The weak interaction implies the exchange of a W^\pm boson in so-called charged-current (CC) interactions and of a Z boson in so-called neutral-current (NC) interactions. All electrically charged particles interact via the electromagnetic (EM) interaction, via an exchange of neutral

Table 2.1: The fermions of the Standard Model with their individual properties [8].

generation:	1st		2nd		3rd	
leptons	e^-	ν_e	μ^-	ν_μ	τ^-	ν_τ
mass ¹ [MeV]	0.511	$< 10^{-6}$	105.66	$< 10^{-6}$	1776.86	$< 10^{-6}$
electric charge	-1	0	-1	0	-1	0
colour charge	no	no	no	no	no	no
spin	1/2	1/2	1/2	1/2	1/2	1/2
lepton number	1	1	1	1	1	1
lepton flavour	$\mathcal{L}_e = 1$	$\mathcal{L}_e = 1$	$\mathcal{L}_\mu = 1$	$\mathcal{L}_\mu = 1$	$\mathcal{L}_\tau = 1$	$\mathcal{L}_\tau = 1$
quarks	u	d	s	c	b	t
mass [MeV]	2.2	4.7	95	1,275	4,180	173,000
electric charge	+2/3	-1/3	-1/3	+2/3	-1/3	+2/3
colour charge	yes	yes	yes	yes	yes	yes
spin	1/2	1/2	1/2	1/2	1/2	1/2
baryon number	1/3	1/3	1/3	1/3	1/3	1/3

¹ The neutrino flavour eigenstates are highly mixed mass eigenstates so that an exact mass can not be assigned to a flavour eigenstate. The upper limit of all three mass eigenstates is 1.1 eV.

photons γ . Quarks carry colour charge and can therefore interact via the strong interaction by exchanging gluons, g . Gluons also carry colour charge and are thus the only self-interacting gauge bosons. Quarks usually do not exist individually, but are bound by gluons in composite particles that are colour-neutral, have integer charge and are collectively called hadrons. Mesons, like pions or kaons, but also tetraquarks, consist of an even number of quarks and anti-quarks. Baryons, like the proton or the neutron, but also pentaquarks, consist of an odd number of quarks and anti-quarks. The only quark that never hadronises into a composite particle is the top-quark, as it decays too fast.

In the SM, particles acquire mass through their interaction with the Higgs field. All massive particles are therefore able to interact with the Higgs boson, H . This includes all fermions and gauge bosons except for gluons and photons. However, in the original theoretical representation of the SM, neutrinos are massless, although experimental evidence shows that they are not [9].

Table 2.2: The bosons of the Standard Model with their individual properties [8].

	W	Z	photon	gluon	Higgs-boson
symbol	W^\pm	Z	γ	g	H
rest mass [GeV]	80.379	91.188	0	0	125.10
electric charge	± 1	0	0	0	0
colour charge	no	no	no	yes	no
spin	1	1	1	1	0

2.2 Lepton flavour violation

Symmetries and conservation laws are central to the development of the Standard Model of particles physics. According to Noether's theorem, any symmetry of the *Lagrangian* is linked to a conserved current [10]. The Lagrangian is a mathematical function used to describe the dynamics of a quantum field theory. By implication, any conserved quantity demands the invariance of the Lagrangian under a transformation. For example, energy conservation is linked to time invariance of the Lagrangian, and momentum conservation to space invariance (translation symmetry) [11]. The conservation of electrical charge, colour charge and weak isospin is linked to gauge invariance under a certain symmetry group each. An overview of conserved quantities and their corresponding symmetries is given in Tab. 2.3.

Table 2.3: Conserved quantities and their corresponding symmetries of the Lagrangian in particle physics.

conserved quantity	symmetry
mass-energy	time invariance
linear momentum	translation symmetry
angular momentum	rotation symmetry
electrical charge	U(1) gauge invariance
colour charge	SU(3) gauge invariance
weak isospin	SU(2) _L gauge invariance

There are further discrete symmetries under which the Lagrangian is invariant, which do not lead to conserved quantities. An example is CPT-invariance, which means that the Lagrangian is invariant under the product of charge conjugation (C), parity transformation (P) and time reversal (T).

There are also apparently conserved quantities in the Standard Model that are not linked to symmetries of the Lagrangian, but purely experimentally established. This means that no violation has been observed yet. An example is the baryon number B , which is $1/3$ for every quark and $-1/3$ for every anti-quark. The equivalent number for leptons is called lepton number L . It is 1 for every lepton and -1 for every anti-lepton. Additionally, there are three lepton flavour numbers, L_e , L_μ , and L_τ . Table 2.4 shows an overview over the lepton flavour numbers of all leptons. The sum of the three lepton flavour numbers equals the lepton number: $L = L_e + L_\mu + L_\tau$. Violation of the lepton flavour number has only been observed in neutrino oscillations so far. This opens up the possibility of lepton flavour number violation in any leptonic decay involving neutrinos in higher order diagrams. The decay discussed in this thesis, $\tau^- \rightarrow \mu^- \mu^+ \mu^-$, violates both L_τ and L_μ conservation. The decay process is shown in Fig. 2.1.¹

¹For any decay mentioned in this thesis, the charge conjugated process is implied, too.

Table 2.4: Lepton flavour numbers of the SM leptons.

particle	L_e	L_μ	L_τ
e^-	1	0	0
e^+	-1	0	0
ν_e	1	0	0
$\bar{\nu}_e$	-1	0	0
μ^-	0	1	0
μ^+	0	-1	0
ν_μ	0	1	0
$\bar{\nu}_\mu$	0	-1	0
τ^-	0	0	1
τ^+	0	0	-1
ν_τ	0	0	1
$\bar{\nu}_\tau$	0	0	-1

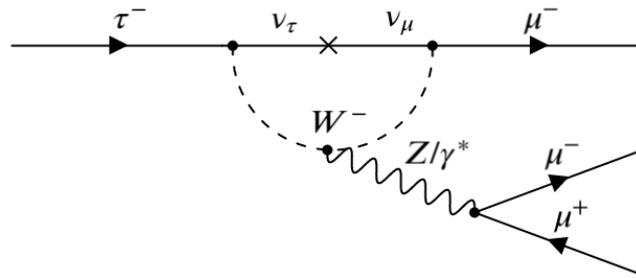


Figure 2.1: Violation of lepton flavour number via neutrino oscillation in the decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$.

3 Data acquisition

In this thesis, data taken with the LHCb experiment at the LHC accelerator in the years 2016 – 2018 is analysed and compared to simulated data. This chapter introduces the LHC accelerator complex and the LHCb experiment with the subdetectors relevant for analysing the decay. An overview of the LHCb trigger system and the simulation of data for LHCb is given. Furthermore, the event selection and reconstruction mechanisms that both measured and simulated data undergo before being used for this thesis are detailed.

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest and most powerful accelerator to date. Situated at the *Conseil européen pour la recherche nucléaire* (CERN) on the border of Switzerland and France, it consists of a 27 kilometre ring of superconducting magnets around the beam lines, where two high-energy particle beams are accelerated in opposite direction close to the speed of light. The LHC works in ultra-high vacuum and at ultra-low temperatures. It accelerates protons as well as heavy ions coming from pre-accelerators. They collide in the detectors which are situated at four collision points and at various other stations around the LHC ring. The whole accelerator complex can be seen in Fig. 3.1.

This thesis concentrates on proton-proton collisions. The protons are produced by stripping the electrons from hydrogen atoms. They are then injected into the proton synchrotron booster from the Linear Accelerator 2 (LINAC2) at an energy of 50 MeV. The booster accelerates them to 1.4 GeV and feeds them to the Proton Synchrotron (PS). At an energy of 25 GeV they are sent to the Super Proton Synchrotron (SPS) to be accelerated to 450 GeV. Finally they are transferred in two directions into the LHC which accelerates them up to 6.5 TeV in a time span of 20 minutes. Under normal conditions, the beams circulate in the LHC beam pipes for many hours before new beams need to be injected. An overview over the LHC accelerators and the energy they accelerate the protons to is shown in Table 3.1.

The CERN accelerator complex *Complexe des accélérateurs du CERN*

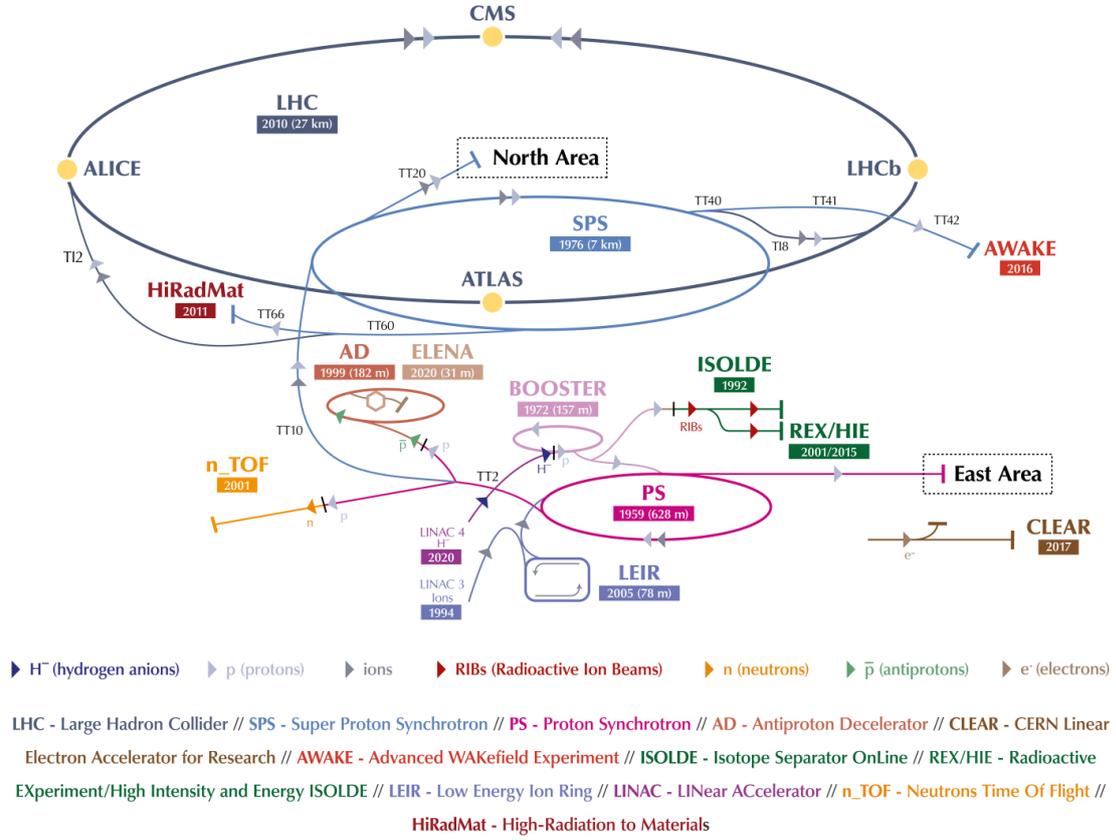


Figure 3.1: The CERN accelerator complex[12]

Table 3.1: The boost energies of the protons inside the different stages of the CERN accelerator complex [13].

accelerator	proton energy
LINAC2	50 MeV
booster	1.4 GeV
PS	25 GeV
SPS	450 GeV
LHC	6.5 TeV

The two most important features of an accelerator are its centre-of-mass energy, \sqrt{s} , and its instantaneous luminosity, \mathcal{L} . The luminosity determines the event rates for a given process. The number of interactions N is the product of the luminosity integrated over time t and the *cross section* σ of the collision:

$$N = \sigma \int \mathcal{L}(t) dt \quad (3.1)$$

The LHC was up to now active in two periods, which are called *Run 1* and *Run 2*. *Run 1* lasted from 2010 – 2012 with an integrated luminosity at LHCb of 3.23 fb^{-1} and a centre-of-mass energy of 7 and 8 TeV. *Run 2* lasted from 2015 – 2018 with an integrated luminosity at LHCb of 5.9 fb^{-1} and a centre-of-mass energy of 13 TeV. The data for this thesis was taken in the years 2016 – 2018 of *Run 2*. In these years the LHC reached a peak luminosity for proton-proton collisions of $\approx 1.2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [13], of which LHCb recorded an integrated luminosity of 5.57 fb^{-1} [7] at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. The integrated luminosity corresponding to the data sets recorded by LHCb in the different years can be found in Fig. 3.2 together with the corresponding beam energies.

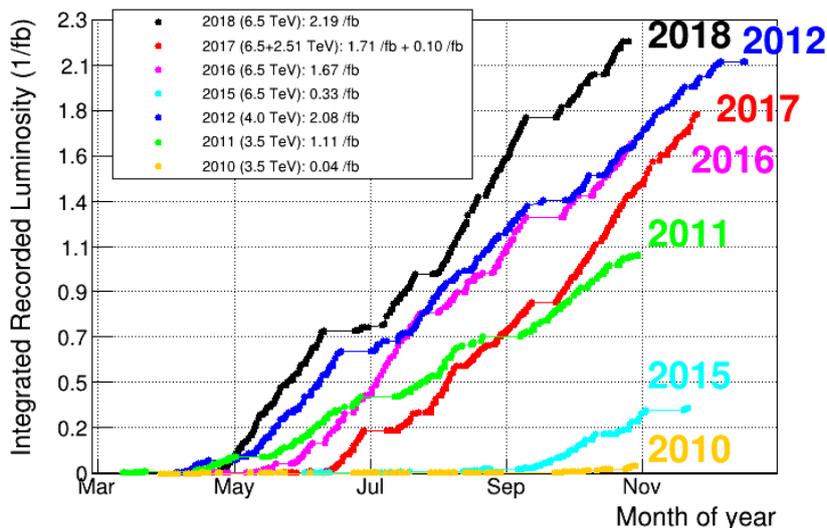


Figure 3.2: Integrated recorded luminosity of LHCb for the years 2010 – 2018 [7].

There are seven experiments at the LHC making use of particles produced in the proton-proton collisions. The four major experiments of the LHC, ATLAS (A Toroidal LHC ApparatuS), ALICE (A Large Ion Collider Experiment), CMS (Compact Muon Solenoid) and LHCb (LHC beauty) are stationed at the four interaction points. TOTEM (TOTal Elastic and diffractive cross section Measurement) is spread across almost half a kilometre around the CMS interaction point. LHCf (LHC forward) is made of two detectors sitting along the LHC beamline on both sides of the ATLAS collision point. MOEDAL (MONopole and Exotics Detector At the LHC) is deployed around the same intersection region as the LHCb detector.

3.2 LHCb detector

The LHCb experiment [14] is the smallest of the four large experiments at the LHC and is dedicated to precision measurements of particles containing b and c quarks, collectively known as B and D hadrons. A cross section of the LHCb detector can be found in Fig. 3.3. LHCb is a forward detector with the single subdetectors stacked behind each other, optimised for the detection of B and D hadrons.

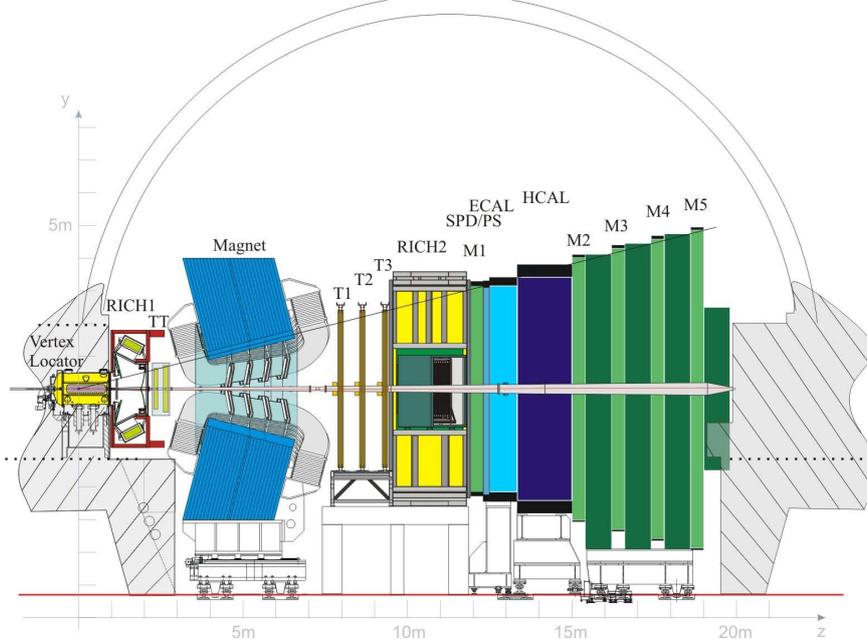


Figure 3.3: Cross section of the LHCb detector [15]

Figure 3.4 shows an overview over the coordinate system used in LHCb. The z axis is defined parallel to the beam axis. The direction along the z axis pointing from the interaction point towards the muon stations M1 – M5 is called *downstream*, while the opposite direction is called *upstream* [16]. The z direction labels the axis downstream and parallel to the beam pipe and the y axis points up towards the surface. The x axis, which lies parallel to the ground and points away from the centre of the LHC accelerator ring, points out of the figure.

It is sometimes useful to use momentum and energy information only from the two axes perpendicular to the beam axis, the x and y axis, as due to momentum conservation the total momentum in the collision in these directions amounts to zero. Momentum and energy reconstructed from the information in these directions only are called transverse momentum p_T

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (3.2)$$

and transverse energy E_T

$$E_T^2 = m^2 + p_T^2 \hat{=} E^2 - p_z^2. \quad (3.3)$$

The azimuthal angle in the x-y-plane is called ϕ . For the horizontal angle θ in the z-x- or z-y-plane, the pseudorapidity η is introduced:

$$\eta = -\ln[\tan(\theta/2)] \quad (3.4)$$

The distribution of particles is more equal in dependency of η than of θ . The pseudo-rapidity approximates the rapidity for high energy particles.

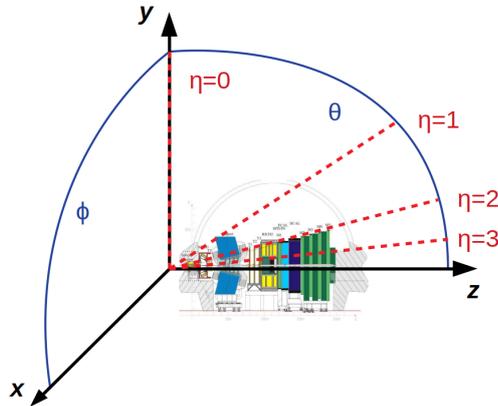


Figure 3.4: Overview of the coordinate system used at LHCb.

3.2.1 The HeRSChEL detector

Not labeled in Fig. 3.3 is the subdetector called *High Rapidity Shower Counters for LHCb* (HERSCHEL) [17]. It is not located in the LHCb cavern itself, but in the LHC tunnel on both sides of the interaction point. It consists of twenty square plastic scintillators just outside the vacuum pipe. Its purpose is to detect activities corresponding to particles produced with an angle too small to be detected by the rest of the LHCb detector. The layout is shown in Fig. 3.5.

HERSCHEL was installed for *Run 2* to enhance the capabilities of LHCb in diffractive physics and to increase the sensitivity in high rapidity regions (up to $\eta \approx 5$). This enables the detection of showers induced by very forward particles that interact in the beam pipe or in other machine parts.

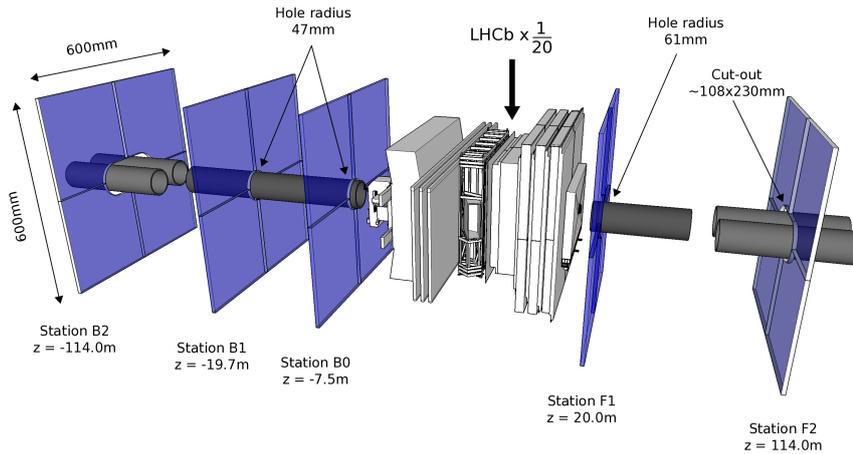


Figure 3.5: Layout of the active areas of the HERSCHEL stations around the LHCb detector taken from [17], with HERSCHEL stations magnified by a factor of 20 with respect to the rest of the LHCb detector, and the z axis not to scale.

3.2.2 The VELO detector

The next subdetector is the *Vertex Locator* (VELO) [18]. VELO is built to completely cover the angular acceptance of detectors downstream. Its main task is to identify the decay vertex of *B* and *D hadrons*. The sensitive detector is only moved mechanically towards the beam once the beam is stable to prevent it from fatal radiation damage. The modules of the lower half can be seen in Fig. 3.6.

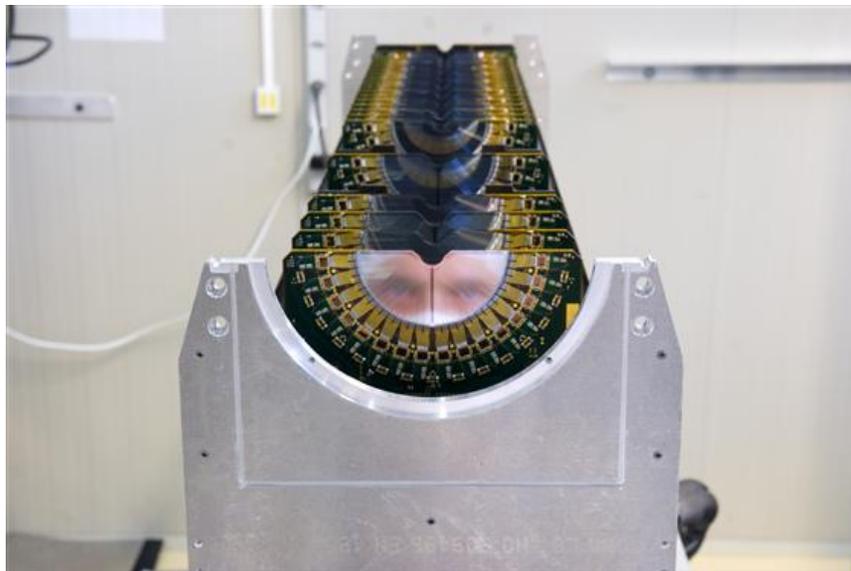


Figure 3.6: The modules of the lower half of the VELO detector [19]

By measuring the distance between the collision point of the protons (primary vertex) and the decay point of the B or D hadrons (secondary or decay vertex), the VELO detector is able to determine the decay length of B and D hadrons within $10\ \mu\text{m}$. This allows for measurements of particle decay times and impact parameters with high precision. The impact parameter (IP) of a track describes the minimal distance between the primary vertex and a reconstructed track. This is visualised in Fig. 3.7. Tracks originating from a secondary vertex have a non-zero impact parameter. Track segments composed solely of VELO hits are called VELO tracks.

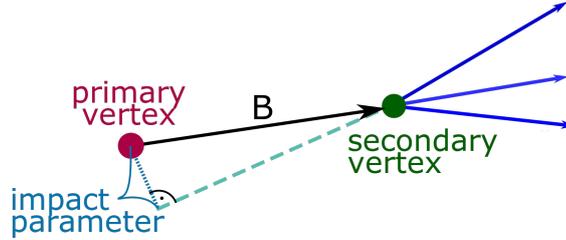


Figure 3.7: Visualisation of the Impact Parameter in the decay of a B meson.

3.2.3 The RICH detector

After the VELO comes the *Ring Imaging CHerenkov* detector (RICH) [20]. It consists of two detectors built for particle identification. The two detectors, lying on either side of the magnet, are responsible for identifying a range of different charged particles by measuring emissions from Cherenkov radiation. Cherenkov radiation occurs whenever a charged particle passes through a certain medium faster than light in the same medium. RICH1 is filled with a silica aerogel as well as the dense gaseous C_4F_{10} , while RICH2 is filled with the dense gaseous CF_4 . The schematic layout of the RICH1 detector is shown in Fig. 3.8.

Passing through the gas with the refractive index n with a velocity larger than the speed of light in the gas, c/n , the charged particle emits photons in a cone of aperture θ around the particles direction of propagation. The velocity of the particle, $v = \beta c$, can be calculated directly from the angle θ using the formula:

$$\cos(\theta) = \frac{1}{\beta n} \quad (3.5)$$

Together with the momentum information from the tracking system, this can be used to determine the particle mass and therefore identify the particle type.

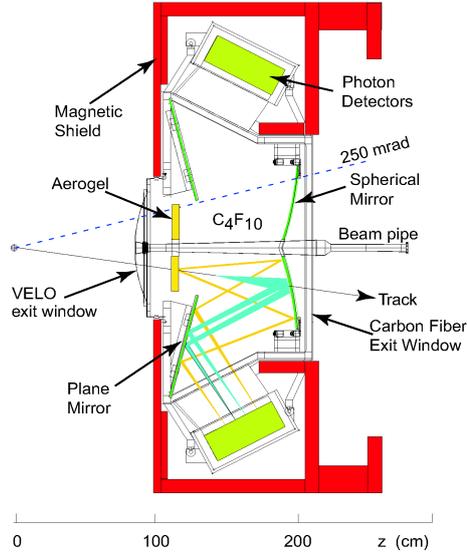


Figure 3.8: Schematic layout of the RICH1 detector in the side view taken from [21].

3.2.4 The LHCb magnets

The powerful LHCb magnets [22], responsible for bending the path of charged particles in order to measure mass and momentum, weigh both 27 tonnes and are mounted inside a 1,450 tonne steel frame. The magnetic field integrated over the track length of 10 meters equals $\int B dl = 4 \text{ Tm}$. The magnet is responsible for bending the path of the particles to reconstruct their momentum and mass. The LHCb detector has two configurations called *magnet up* (MagUp) and *magnet down* (MagDown) [23]. They correspond to the dipole magnetic field being along the positive or negative y axis. The detector configuration regularly swaps between the two polarities in order to minimise systematic effects on the tracking detector.

3.2.5 The tracking system

The tracking system is responsible for detecting the traces of charged particles in the detector. It assigns momentum as well as measures the impact parameter with high resolution. The tracking station upstream of the dipole magnet, called Trigger Tracker (TT) [24], as well as the central part of each of the three tracking stations T1 – T3, called Inner Tracker (IT) [25], consist of silicon microstrip detectors. The main task of the TT is to measure the trajectories of low-momentum particles before they are bent out of the detector acceptance by the magnets and therefore do not reach T1 – T3, and to assign transverse momentum information to tracks.

The outer part of the three tracking stations T1 – T3 is called Outer Tracker (OT) [27]. It is situated further from the beam pipe and made up of thousands of gas-filled straw tubes. The tracking stations are separated into inner and outer trackers because the particle flux is much higher closer to the beam pipe. The tracking stations measure momentum and impact parameter with different resolution depending on the distance to the beam pipe.

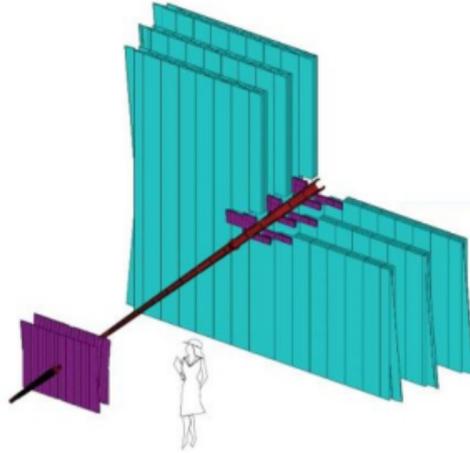


Figure 3.9: The LHCb tracking system with the silicon tracker in purple and the outer tracker in cyan [26].

As shown in Fig. 3.10, tracks reconstructed from hits in the entire tracking system are called LONG tracks, in contrast to VELO tracks reconstructed solely from VELO hits. An overview of the tracking system can be seen in Fig. 3.9.

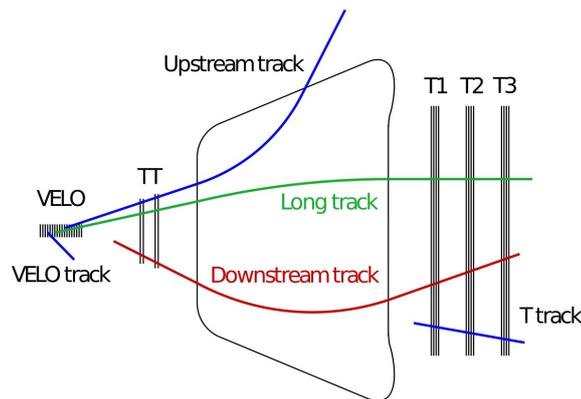


Figure 3.10: VELO and LONG tracks in LHCb, together with downstream, upstream and T tracks.

3.2.6 The calorimeters

The LHCb calorimeter system [28] consists of the Scintillating Pad Detector (SPD), the PreShower (PS), the Electromagnetic CALorimeter (ECAL) and the Hadronic CALorimeter (HCAL). Its main responsibility is the identification and measurement of the energy and momentum of hadrons, electrons and photons with high transverse energy, providing information for online selection. Additionally, it distinguishes prompt and secondary photons. Prompt means the particles are produced directly in the collision, while secondary photons are produced in the decay of other

particles, like the π^0 . The separation between photons and electrons at the first trigger level is made by the SPD and the PS, which are separated by a lead layer. The charged electrons deposit energy in the scintillator material, while the neutral photons generally do not interact. After the distinction between photons and electrons is made, the ECAL is employed to measure the energy of both. Neutral and charged hadrons deposit most of their energy in the HCAL and only a small part of it in the ECAL.

3.2.7 The muon system

Located at the far end of the detector, the muon system consists of five rectangular muon *stations*, $M1 - M5$ [29]. Each station is filled with a mixture of carbon dioxide, argon and tetrafluoromethane and connected to wire electrodes to read out the results of reactions between muons and the gas. The muon system is responsible for identifying and measuring the path of muons, which are present in the final states of many B and D *hadron* decays and therefore vitally important to the LHCb physics program. The muon detector uses the penetrative power of muons to provide a robust muon trigger for the online selection.

The muon detector has an angular acceptance of 20 - 306 mrad in the bending and 16 - 258 mrad in the non-bending plane. This provides a geometrical acceptance of about 20% for muons from B and D *hadron* decays relative to the full solid angle.

3.3 The LHCb trigger system

The LHCb detector is designed to work with a maximum luminosity of $2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. The expected collision rate at this luminosity is 40 MHz. The LHCb trigger system [30] consists of two levels. It exploits the fact that b-flavoured hadrons are relatively heavy and long lived. An overview is shown in Fig. 3.11.

The first level trigger L0 is implemented in hardware. It selects candidates with high transverse momentum p_T or high transverse energy E_T . The L0 muon and dimuon triggers search for muon candidates with a high transverse momentum and trigger on hits on a straight line through the five muon stations. They allow for up to eight muon candidates per event. The L0 calorimeter trigger searches for hadron, electron, photon and neutral pion candidates which deposit energy in the calorimeters. For each particle type it chooses only the candidate with the highest transverse energy measured in the calorimeters. The L0 pile-up system is a component of the VELO detector. It rejects events with too many proton-proton collisions (primary vertices). About 10% of events accepted by L0 are triggered by more than one L0 trigger type. The L0 trigger reduces the event rate from 40 MHz to around 1 MHz. The next trigger level, the High Level Trigger (HLT), is implemented in software and consists of a C++ application running on the events accepted by L0. The application responsible for running the HLT is called MOORE [31]. It is running on an Event Filter Farm. Its main task is to confirm the hardware trigger decision and

to fully reconstruct the events.

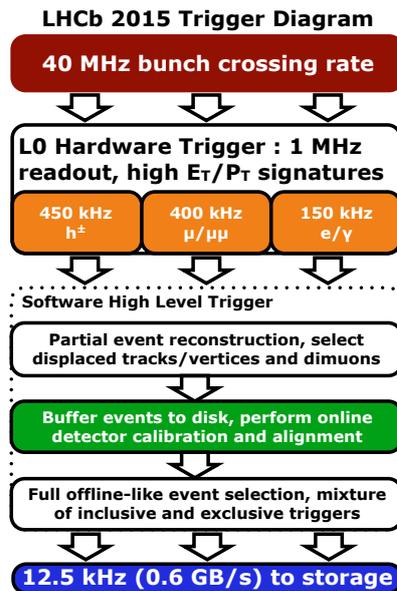


Figure 3.11: The LHCb trigger flow for *Run 2* [32].

HLT contains two layers. The first is called HLT1. It reduces the L0 data output rate of 1 MHz to a few tens of kHz. It subjects the events to different sequences of algorithms, called "trigger lines", dependent on the decision issued by L0. It confirms the L0 candidates by adding information from either the VELO or the tracking system and applies cuts on the transverse momentum and the impact parameter. Sometimes, the HLT1 also looks for extra particles, for example for two muons from a good vertex.

The next layer is called HLT2. It runs on the output of HLT1. First, it performs a full pattern recognition on the remaining events. Tracks associated to stable particles are used to form composite particles. HLT2 selects events which contain specific resonances likely to be produced in B decays and fully reconstructed B hadron decays. The HLT2 trigger reduces the data output rate of 1 MHz to a rate of about 12.5 kHz. It stores the information about the trigger decisions.

Once a signal candidate is reconstructed in the offline analysis, two different cases are distinguished. Trigger On Signal (TOS) means that the particle triggering the specific trigger decision belongs to the reconstructed signal. Trigger Independent of Signal (TIS) means that an additional particle not belonging to the reconstructed decay triggered the decision.

3.4 Monte Carlo simulation

There are multiple reasons why it is helpful to have simulated data. It can be used to generate an expectation from a model, to compare data with a theory and see

whether it deviates significantly or to determine strategies how to analyse the data. The aim of Monte Carlo simulation (MC) is to randomly produce simulated data which behaves the same way as data taken in the experiment. The simulation of LHCb MC events is complex and consists of multiple steps. Responsible for the generation of MC events is the LHCb software package GAUSS [33]. In the first step, proton-proton collisions are simulated using the event generator PYTHIA [34]. The decay of the resulting unstable particles is then simulated using the EVTGEN library [35]. Subsequently, the simulated events are going through a simulation of the detector, produced with GEANT4 [36]. The detector response to the simulated events, including the response of the hardware trigger L0, is simulated with BOOLE [37]. The output from BOOLE is then processed by MOORE and follows the same event flow as the detector output from the L0 trigger.

3.5 Event reconstruction

Triggered, raw data and simulated data coming from MOORE are reconstructed to transform the detector hits into objects such as tracks and clusters. This is done by the BRUNEL application. The reconstructed events coming from BRUNEL are saved into files through a filtering process called stripping. The stripping is done using the LHCb application DAVINCI. Each stripping line applies a series of cuts on the reconstructed data that removes a large number of candidates which are uninteresting for the specified process. The output from DAVINCI is stored in files accessible by ROOT which contain the information of each event. The complete LHCb data flow is shown in Fig. 3.12.

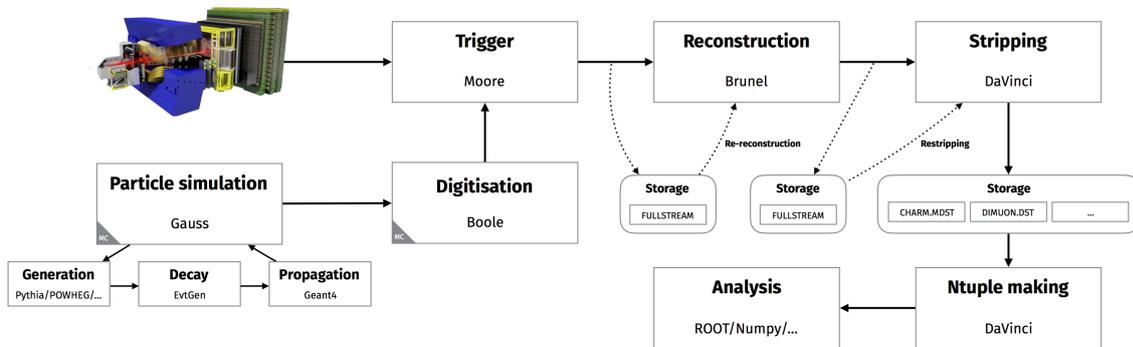


Figure 3.12: The LHCb data flow taken from [38].

4 Analysis outline

The aim of this study is the search for the lepton flavour violating decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$, which is highly suppressed in the minimal extended Standard Model of particle physics (SM) including neutrino oscillations. Monte Carlo samples are prepared and analysed in order to study the decay characteristics of $\tau^- \rightarrow \mu^- \mu^+ \mu^-$. A more abundant, but kinematically very similar decay, namely $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$, is studied in data and in simulation with the objective of checking for potential differences between the Monte Carlo samples and data taken by the experiment. This so-called control channel is as well used as reference to derive the limit on the branching fraction, as will be explained later. In this section, an overview over the analysis strategy is given and the characteristics of the signal and reference channel are introduced. The tools used for this analysis are briefly mentioned here and further explained in the next chapter.

4.1 Analysis strategy

The MC and data samples used in this analysis pass the reconstruction and stripping described in the last chapter. Stripping lines define a number of loose cuts referring to the wanted decay characteristics. Stripping lines specific for the signal and control channel are used to obtain data samples, separately for the three years 2016 – 2018. The samples obtained are called *signal data sample* and *control data sample*, and *MC signal sample* and *MC control sample*, respectively.

There are different production processes for both the signal and the control channel. As each of these processes is simulated separately, with the resulting number of candidates per process not necessarily representing the true ratio expected from data, it is necessary to correct the production rate of the MC samples.

Then, specific trigger lines corresponding to characteristics of the signal decays are chosen. Any event not triggered by these trigger lines is removed in every sample. These events are less likely to stem from the signal decay. In order to further concentrate on interesting events, preselection cuts are applied to all samples. These are chosen mostly equivalent to the previous analysis performed on *Run 1* data [6]. The control channel is used to compare Monte Carlo and data distributions in selected variables. As the $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ channel is kinematically very similar to the $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ channel, any difference between Monte Carlo and data found in this channel is expected to appear similarly in the signal channel. The $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ signal distributions in the control data sample are obtained using the *sWeighting* procedure, which removes the background on a statistical basis. A fit to the MC control sample is used to optimise the *sWeighting* procedure

on the control data sample. For more information on *sWeighting*, see Sec. 5.2. The *sWeighted* control data distributions are normalised and compared to the normalised MC control distributions. Weights are applied to the MC control events to match the control data distributions. These weights are stored in dependency of the selected variables and then applied to the MC signal samples. This is done separately for all three years, 2016 – 2018. These weights are necessary, because the MC needs to reproduce the data as good as possible so that the efficiencies of applied cuts can be calculated using the simulated samples.

The stripping line chosen for the signal channel requires at least two muons to be identified by the LHCb muon system as good muons. Therefore, the signal data and MC samples are separated into two sub-samples each, which are called the 2μ sub-sample and the 3μ sub-sample. They are mutually exclusive and hold all events where exactly two or three muons are accepted as good muons respectively.

The network is trained separately on the 2μ sub-sample and the 3μ sub-sample as well as on a merged sample of both. Figure 4.1 shows schematically the preparation of all samples until they are handed to the multivariate classifier.

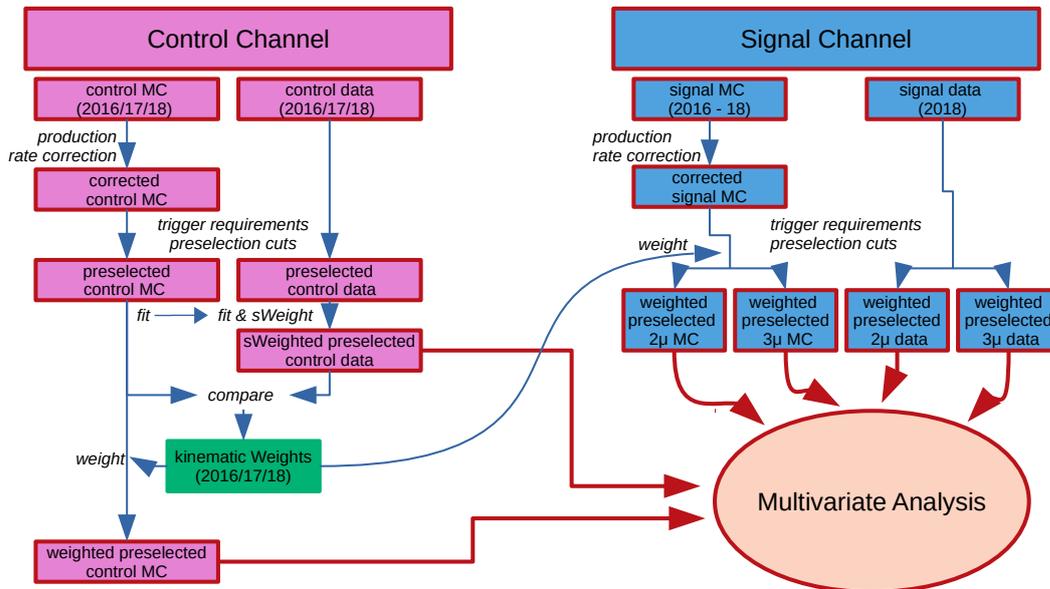


Figure 4.1: Schematic overview over the preparation of the samples used in the multivariate analysis.

The best training when applied to the different sub-samples is chosen in the end and applied to all samples.

The output distributions of this training classifier in the control channel are used to again compare the control data sample and the MC control sample for all three years separately. A second set of weights is produced to match the data distributions to the MC distributions. The weights correct mainly for residual discrepancies and differences in variable correlations between the MC and the data samples. Again, these weights are applied to the MC signal samples of all three years.

An additional particle identification cut is placed on the good muons of the signal samples in order to further reduce background. This particle identification cut is optimised together with the cut on the multivariate classifier output for the 2μ and the 3μ sub-samples separately. For the 2μ sub-sample an additional cut on the particle identification of the third particle, which is not accepted as a good muon by the muon chambers, is applied. As the classifier is composed out of kinematic variables, it is uncorrelated with the particle identification variables. The cuts are optimised separately for each year. The efficiencies of these cuts on the MC signal samples of all years are calculated separately for the 2μ and the 3μ sub-samples. A fit through the mass distribution of each of the signal data samples is used to estimate the expected background in the signal region. The efficiencies are used together with the expected background to calculate and combine expected upper limits on the branching fraction for all three years. For this, the $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ channel is used as reference channel.

4.2 Signal channel

The signal process $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ is the decay of a single tau into three muons. No other particles are expected in the final state. The process is highly suppressed in the Standard Model, but neutrino oscillations allow for it with a very tiny branching fraction of $\mathcal{O}(10^{-55})$ [1]. Theories describing phenomena beyond the SM allow for the process with a much higher branching fraction [2][3][4]. The branching fraction is mainly limited by the properties of the neutral particle responsible for the lepton flavour violation, as shown in Fig. 4.2.

In beyond SM theories the neutrino can be replaced by another neutral particle responsible for the lepton flavour violation, as shown for a yet unknown particle called X^0 in Fig. 4.3a. This could for example be a heavy neutrino. Another possible Feynman diagram for the signal process with an extended particle content is shown in Fig. 4.3b. This includes, additionally to the X^0 , a new charged particle called here X^- .

There are five B and D hadron decays involving τ^- in the final state considered as dominant τ^- production modes in this thesis. The τ^- can be produced in the decay of a D_s^\pm , which can be either (i) prompt, or (ii) coming from a B meson. It can also be produced in the decay of a D^\pm , which can also be either (iii) prompt or (iv) coming from a B meson. Finally, the τ^- can be produced (v) prompt in the decay

of a B meson. An overview of the branching fractions of these five sub-channels with their total branching fraction \mathcal{B}_{tot} can be found in Tab. 4.1. The ratio \mathcal{B}_{tot} is calculated by multiplying the individual branching fractions for the decay steps and adding the two branching fractions if there are separate values for both charges. In case of $\mathcal{B}_{\bar{b} \rightarrow D^+}$ no branching fraction has been measured so far, so the shown value is obtained by multiplying the branching fraction $\mathcal{B}_{\bar{b} \rightarrow D_s^+}$ with $(|V_{cd}|^2/|V_{cs}|^2)$, where V_{cd} and V_{cs} are the CKM matrix elements describing the transition $c \rightarrow d$ and $c \rightarrow s$ respectively.

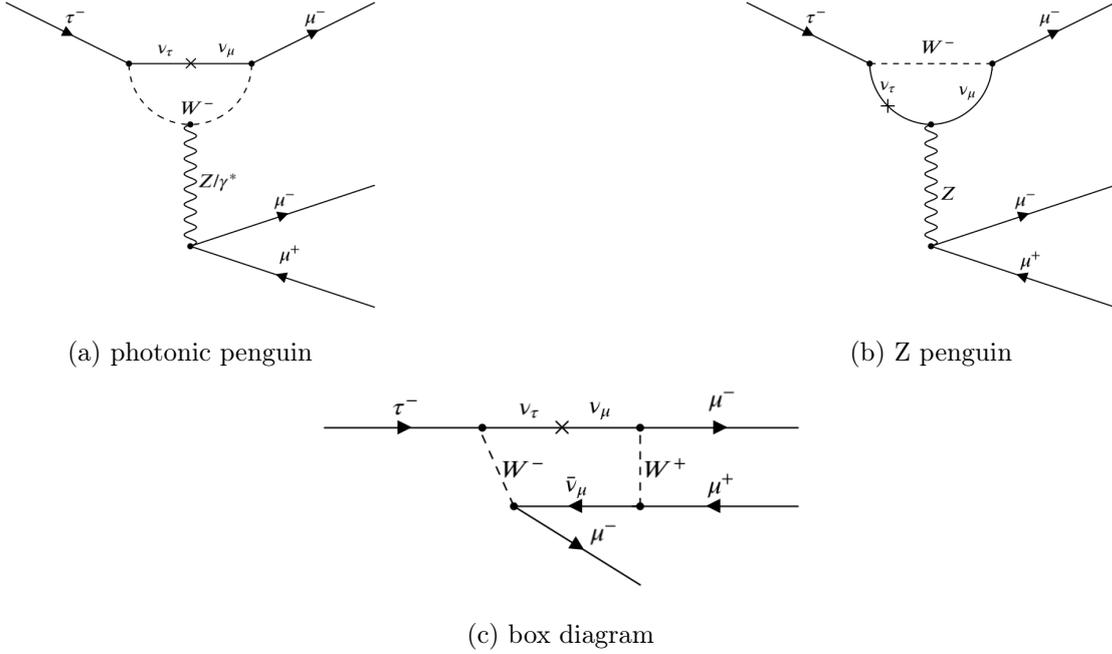


Figure 4.2: Three possible Feynman diagrams for the signal process involving neutrino oscillation.

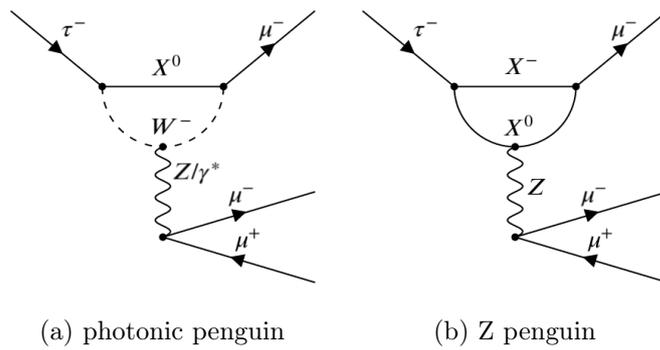


Figure 4.3: Two possible Feynman diagrams for the signal process in beyond Standard Model theories involving unknown particles $X^{0/-}$.

Table 4.1: Branching fractions for the τ production channels [39].

	channel	$\mathcal{B}_{\bar{b} \rightarrow D/D_s}$ [%]	$\mathcal{B}_{\bar{b}/D/D_s \rightarrow \tau}$ [%]	\mathcal{B}_{tot} [%]
(i)	$D_s^- \rightarrow \tau^-$	-	5.48 ± 0.23	5.48 ± 0.23
(ii)	$\bar{b} \rightarrow D_s^- \rightarrow \tau^-$	14.7 ± 2.1	5.48 ± 0.23	1.36 ± 0.21
	$\bar{b} \rightarrow D_s^+ \rightarrow \tau^+$	10.1 ± 3.1		
(iii)	$D^- \rightarrow \tau^-$	-	0.12 ± 0.03	0.12 ± 0.03
(iv)	$\bar{b} \rightarrow D^- \rightarrow \tau^-$	22.7 ± 1.6	0.12 ± 0.03	0.028 ± 0.006
	$\bar{b} \rightarrow D^+ \rightarrow \tau^+$	0.5 ± 0.5^1		
(v)	$\bar{b} \rightarrow \tau^\pm$	-	2.41 ± 0.23	2.41 ± 0.23

¹ obtained by multiplying $\mathcal{B}_{\bar{b} \rightarrow D_s^+}$ with $(|V_{cd}|^2/|V_{cs}|^2)$

4.3 Reference and control channel

The process $D_s^- \rightarrow \phi(\mu^+\mu^-)\pi^-$ is used both as reference and as control channel. As in the signal channel process, three particles are expected in the final state, but only two of them are muons, while the third one is a charged pion. This process has an easily identifiable final state with a high branching fraction [8] of

$$\begin{aligned} \mathcal{B}(D_s^- \rightarrow \phi(\mu^+\mu^-)\pi^-) &= \mathcal{B}(D_s^- \rightarrow \phi(K^+K^-)\pi^-) \times \frac{\mathcal{B}(\phi \rightarrow \mu^+\mu^-)}{\mathcal{B}(\phi \rightarrow K^+K^-)} \\ &= (2.24 \pm 0.08) \times 10^{-2} \times \frac{(2.86 \pm 0.19) \times 10^{-4}}{(49.2 \pm 0.5) \times 10^{-2}} \\ &= (1.30 \pm 0.10) \times 10^{-5}. \end{aligned}$$

Additionally, the final state is kinematically similar to that of the signal process. The Feynman diagram of this process can be found in Fig. 4.4.

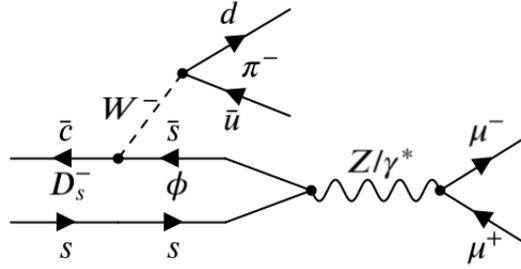


Figure 4.4: Feynman diagram of the reference and control channel.

There are two possible production channels for the reference process. The D_s^\pm can either be produced promptly in the detector (vi), or it can result from a B meson (vii). The branching fractions for the D_s^\pm production can be found in Tab. 4.2.

Table 4.2: Branching fractions for the D_s^\pm production [39].

	channel	$\mathcal{B}_{\bar{b} \rightarrow D_s}$ [%]	\mathcal{B}_{tot} [%]
(vi)	D_s^\pm	-	1
(vii)	$\bar{b} \rightarrow D_s^-$	14.7 ± 2.1	24.8 ± 3.7
	$\bar{b} \rightarrow D_s^+$	10.1 ± 3.1	

5 Analysis tools

The aim of this thesis is to estimate the sensitivity and give an expected upper limit on the branching fraction of the decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ using the decay $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ as reference channel. The measured and simulated data samples for both decay channels are preselected by applying a number of trigger conditions and selection cuts. The simulated sample is weighted to correct for discrepancies between Monte Carlo and data using the $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ channel as control channel. For this, the $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ signal events must be selected. A multivariate analysis tool is used to train binary decision trees to distinguish between signal events and background events. Statistical tools are used to determine the optimal cut on the output of the multivariate classifier and calculate the expected upper limit from the efficiency of the resulting cut. The tools used for this process are introduced in this chapter.

5.1 Root analysis framework

The ROOT analysis framework [40] is an object oriented framework written specifically for data analysis in high energy particle physics. It is based on the C++ programming language. The first version, version 0.5, was released in 1995, with version 1.0 being released in 1997. ROOT can be accessed via a user interface or the command line. It is especially built to deal with high data quantities and focuses on saving data in the form of so-called *trees* or *ntuples* with the substructures *branches* and *leaves*. Data can then be presented in the form of histograms in one, two or three dimensions, fitted, evaluated, minimised or manipulated in many different ways, allowing an easy form of visualisation for the properties of huge data sets.

5.2 *sPlot* technique

The *sPlot* technique [41] is a statistical method used to unfold various distributions from different sources. In this thesis, these sources are the signal decay and the combinatorial background. A fit to a so-called *discriminating variable* is used to statistically separate signal and background. In this case, this variable is the invariant mass distribution of the mother particle. The signal can be described by a Gaussian distribution, while the combinatorial background is described by a declining exponential. The knowledge about the shapes of the distributions in the discriminating variable is used in the *sPlot* technique to reconstruct distributions of

other variables, for which the distributions for the different sources are not known. These variables are called *control variables*. In this thesis these are e.g. the number of tracks, the transverse momentum of the mother particle and muon isolation variables.

A fit to the mass distribution determines the contribution of signal and background as function of the mass. Based on this fit, the *sPlot* technique assigns a weight, the so-called *sWeight*, to each event. By weighting the events with the *sWeights*, the true distributions of the control variables are reproduced on average.

5.3 TMVA toolkit

The most dominant source of background to the signal channel comes from partially reconstructed B meson decays. Selection cuts are applied to the signal data sample in order to remove as much as possible of this background. In order to find the optimal separation between signal and background events, the correlation between observables has to be taken into account. Multivariate analysis (MVA) is a technique performing studies across multiple dimensions and especially taking into account the relationships between input variables. An MVA is used in this work as a classification technique to classify events as either signal or background events. The output classifier function is trained to optimise simultaneously both signal significance and background rejection while avoiding overtraining. Overtraining occurs when due to lack of statistics or an overly complex classifier function the classifier does not only find a general pattern to identify the signal but trains on fluctuations of a specific training sample. The classifier is therefore not able to identify signal in a different data set with the same performance.

The toolkit for multivariate analysis (TMVA) [42] is integrated in the ROOT framework. It uses a *factory* to find the optimal classification function. The factory object is created at the beginning of the program and provides member functions to specify the training and test data sets, to register the discriminating input and to book the multivariate methods. It calls for training, testing and the evaluation of the method and creates output weight files after the training phase for each method. Background and signal samples are handed to the factory in the form of two ROOT trees. The factory then assesses the preliminary properties of the trees by calculating the correlation coefficients and by ranking the input variables according to their separation potential. After normalising the input variables, the factory copies both trees and splits the copies randomly into a training and a test tree. The fraction of events used for training can be chosen by the user. Then the factory trains on the training samples. It applies the output weights to the test samples. The performances on the training and test samples are monitored to avoid overtraining.

The chosen method for this thesis is the *boosted decision tree* (BDT). The BDT is a binary and tree structured classifier for signal and background. For all training variables the optimal separation cut is determined using the GiniIndex $p \cdot (1 - p)$, where the purity p gives the ratio of signal events over all events in the node. The

events are split in two samples using the cut with the maximal increase from the GiniIndex of the parent node to the sum of the GiniIndices of each of the daughter nodes weighted by the relative fractions of events. The higher the value of the option *nSteps* is set, the smaller steps are taken between different tested cuts, and the more cuts are therefore tested. A higher value of *nSteps* can therefore lead to higher precision, but is also more prone to fluctuations.

The two samples are separated in two so-called nodes, where a new separation criterion is determined for each node and the samples are split again, until the number of events in a node falls under a certain number that is defined as input option *MinNodeSize* in percentage of the total number of events or until the maximum depth of the tree, *MaxDepth*, is reached.

The resulting leaf nodes are then classified as either signal or background with respect to the majority of events in the leaf node. An example for a simple tree is shown in Fig. 5.1. An event in a signal leaf node is assigned the value +1, while an event in a background leaf node is assigned the value -1 or 0, depending on the method.

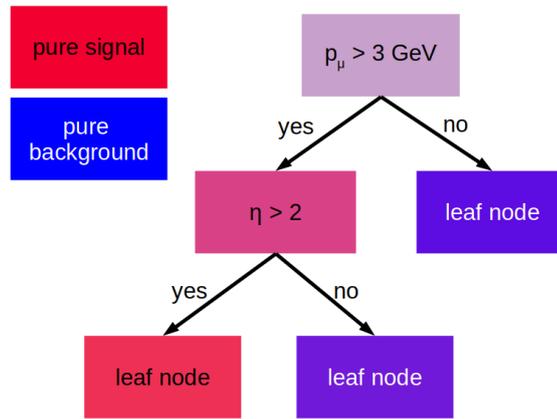


Figure 5.1: A schematic overview over a simple BDT.

Single trees can easily be visualised, but are unstable with respect to statistical fluctuations. Therefore a whole forest of slightly different trees is built. Two different BDT algorithms are tested for this thesis. These two proved generally successful in previous analyses and are compared to find the better performing one for this specific analysis.

When the *adaptive boost algorithm* (AdaBoost) is used, events that were misclassified during the training of a decision tree are given a higher event weight in the training of the following tree. Each tree is therefore trained on a modified event sample, where previously misclassified events are boosted by a common weight α , which is derived from the misclassification rate r_{mis} of the previous tree:

$$\alpha = \left(\frac{1 - r_{mis}}{r_{mis}} \right)^{\beta_{Ada}}, \quad (5.1)$$

The learning rate β_{Ada} is a parameter chosen by the user to influence the strength of

the response to a misclassification. The output value for each event x is calculated from all output values $h_i(x)$ using equation 5.2 and lies between 0 and +1.

$$y_{AdaBoost}(x) = \frac{1}{N_{Trees}} \sum_{i=1}^{N_{Trees}} \ln(\alpha_i) \cdot h_i(x) \quad (5.2)$$

With the *gradient boost algorithm* (GradBoost), the model response, $y_{GradBoost}(x)$, for each event x is calculated as a weighted sum from the single decision tree response functions, $f(x; \vec{\alpha}_j)$:

$$y_{GradBoost}(x; P) = \sum_{j=0}^M \beta_j f(x; \vec{\alpha}_j); \quad P \in \{\beta_j; \vec{\alpha}_j\}_0^M, \quad (5.3)$$

where α_j are the parameters determining each tree, e.g. the splitting variables and splitting values, while β_j are the weights applied to each tree. Each new tree is meant to correct for the shortcomings of the previous tree by adjusting the parameters P such, that the deviation between the model response $y_{GradBoost}(x)$ and the true value $y_{true}(x)$, which is -1 or $+1$, is minimised. The robustness of GradBoost can be enhanced by reducing the learning rate β_{Grad} of the algorithm.

The possibility of overtraining can be reduced by limiting the number of trees, $nTrees$, the minimum percentage of events in the leaf nodes, $MinNodeSize$, the number of steps to test a variable range for the optimal cut, $nSteps$, the maximum depth of the trees, $MaxDepth$, or by varying the number of events chosen for training, $nTraining$. Overtraining is checked via a Kolmogorov-Smirnov test (KS test) on both background and signal. The KS test is a test of the agreement of two continuous, one-dimensional probability distributions. In this case, the output distributions of the test and the training sample are compared. The KS test quantifies the distance between the distributions and gives back a statistical probability, p_{KS} , that the samples are drawn from the same underlying distribution function. The TMVA repeats the KS test on a random sub-sample of events 1000 times and reports the number of times the individual tests are > 0.5 . The optimal response of the repeated KS tests, r_{KS} , is 0.5. However, as this is very hard to reach, it is often enough to rely on a comparison of the output distributions by eye. The samples should have a similar output distribution if no overtraining occurs.

The BDT also gives a variable ranking at the end of each training. The importance is derived by counting how often the variable has been used to split, weighted by the square of the separation gain the split had achieved and the numbers of events in the node. The separation gain is the amount by which the GiniIndex increased from the parent node to the sum of the GiniIndices of each of the daughter nodes weighted by the relative fractions of events.

5.4 CL_S technique

The aim of this thesis is to calculate the expected limit on the branching fraction of the decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$. The CL_S method is used to calculate this limit. This

section introduces the procedure used to calculate and combine the limits. It follows the outline of [43].

The expected number of $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ decays, $N_{exp}(\tau^- \rightarrow \mu^- \mu^+ \mu^-)$, can be computed via

$$N_{exp}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) = \mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \times N_{tot}(\tau^-), \quad (5.4)$$

where $N_{tot}(\tau^-)$ is the total number of τ^- decays and $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-)$ the branching fraction of the decay.

Not all of these events are observed, as each experiment is subjected to an efficiency, ϵ :

$$N_{exp,obs}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) = \epsilon \times N_{exp}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \quad (5.5)$$

The observed number of events can therefore be used to estimate the branching fraction, if the efficiency and the total number $N_{tot}(\tau^-)$ are known:

$$\mathcal{B}_{est}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) = \frac{N_{obs}(\tau^- \rightarrow \mu^- \mu^+ \mu^-)}{\epsilon \times N_{tot}(\tau^-)} \quad (5.6)$$

If no signal events are observed, an upper limit (UL) can be set on the branching fraction:

$$\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) < \frac{UL(\tau^- \rightarrow \mu^- \mu^+ \mu^-)}{\epsilon \times N_{tot}(\tau^-)} = \alpha \times UL(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \quad (5.7)$$

The proportionality factor α is called normalisation factor of single event sensitivity. However, even if no signal events are expected to be measured, the expectation of observed events in the signal region is not zero. The expected number of background events, B , has to be taken into account.

The task of the observer is now to distinguish whether the measured data is compatible with the null hypothesis H_0 , that there is no signal and only background is measured ($\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) = 0$), or an alternative hypothesis H_n , that signal appears additionally to the background with a branching fraction depending on the parameter n ($\mathcal{B}_n(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \geq 0$). The best way to distinct between an alternative hypothesis H_n and the null hypothesis H_0 is the computation of the likelihood ratio:

$$\lambda(N_{obs}) = \frac{L_{H_n}(N_{obs})}{L_{H_0}(N_{obs})} \quad (5.8)$$

The likelihood $L_{H_n}(N_{obs})$ gives the probability that N_{obs} events are observed if the alternative hypothesis H_n is true. Equivalently, the likelihood $L_{H_0}(N_{obs})$ gives the probability that N_{obs} events are observed if the null hypothesis H_0 is true. In the case of counting experiments, the Poisson probability is used:

$$L_{H_n}(N_{obs}) = p(N_{obs}|H_n) = \frac{(B + S_n)^{N_{obs}}}{N_{obs}!} \times e^{-(B+S_n)}$$

$$L_{H_0}(N_{obs}) = p(N_{obs}|H_0) = \frac{B^{N_{obs}}}{N_{obs}!} \times e^{-B}$$

where B is the expected number of background events and $S_n = \mathcal{B}_n(\tau^- \rightarrow \mu^- \mu^+ \mu^-)/\alpha$ the number of signal events expected to be observed given the hypothesis. As the likelihood ratio is used in order to distinguish between the signal and the null hypothesis after the measurement, it is called the *test statistics* Q .

In case the observation is done in more than one sample, for example separately for each year, the combined test statistics is the product of the likelihood ratio in each sample i :

$$Q(\vec{N}_{obs}) = \prod_{i=1}^{n_{samples}} Q_i = \prod_{i=1}^{n_{samples}} \frac{L_{H_n}(N_{obs,i})}{L_{H_0}(N_{obs,i})} \quad (5.9)$$

In case of the Poissonian likelihood, this can be simplified to:

$$Q(\vec{N}_{obs}) = \prod_{i=1}^{n_{samples}} e^{-S_{n,i}} \left(1 + \frac{S_{n,i}}{B_i}\right)^{N_{obs,i}} \quad (5.10)$$

If the observed test statistics Q_{obs} is bigger than a critical value Q_{crit} , the alternative hypothesis H_n is rejected. If it is smaller, the null hypothesis H_0 is rejected. As shown in Fig. 5.2, the critical value Q_{crit} is chosen such, that the probability to reject H_n when it is indeed true, is $1 - c$, where c is the confidence level. It is usually set to 90% or 95%.

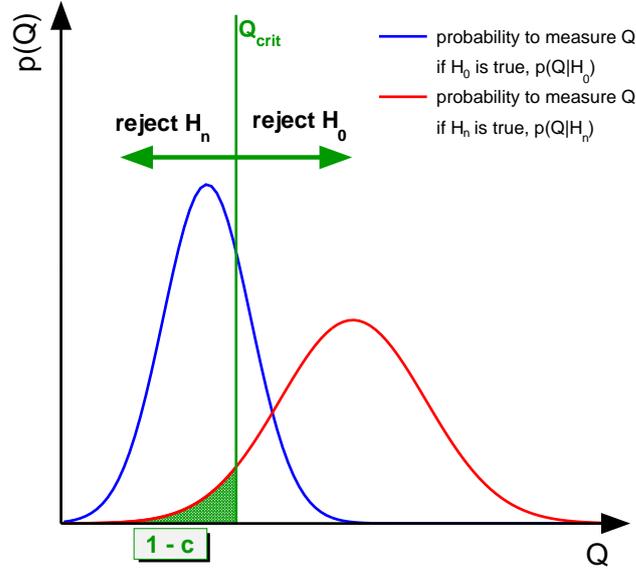


Figure 5.2: Determination of the significance and confidence level depending on the critical value Q_{crit} .

Alternatively to the likelihood ratio, the CL_S value can be used as critical value. This is calculated using the confidence levels CL_{S+B} and CL_B . The probability to observe Q_{obs} or an even larger value for Q if the alternative hypothesis H_n is true is

called CL_{S+B} :

$$CL_{S+B} = P(Q \leq Q_{obs}|H_n) \quad (5.11)$$

If not an observed limit is determined, but an expected limit, the expected test statistics, Q_{exp} , will be used instead of the observed one, Q_{obs} . This is defined as the median of the distribution of the test statistics given the null hypothesis, $P(Q|H_0)$:

$$0.5 = P(Q \leq Q_{exp}|H_0) \quad (5.12)$$

The confidence level CL_{S+B} equals the probability that H_n is incorrectly rejected, even though it is true:

$$CL_{S+B} = P(Q \leq Q_{exp}|H_n) \quad (5.13)$$

It can be determined by producing the expected distribution of the test statistics assuming H_n is true, as shown in Fig. 5.3.

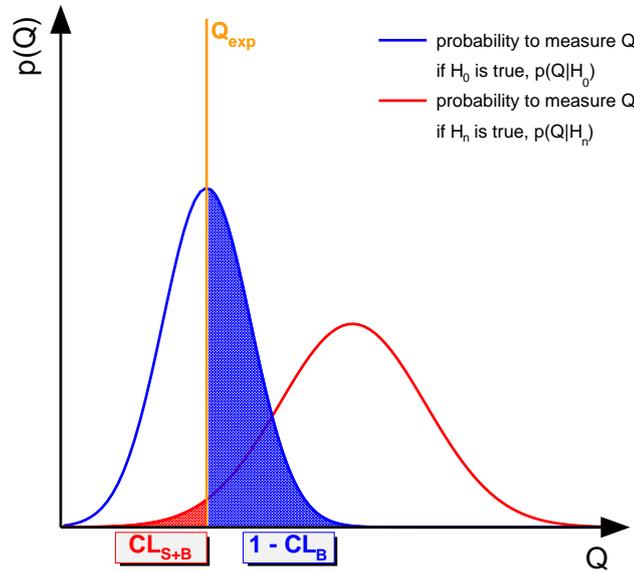


Figure 5.3: Determination of the confidence levels depending on the expected test statistics Q_{exp} .

At the same time, the probability to observe Q_{exp} or an even larger value for Q if the null hypothesis H_0 is true is called $1 - CL_B$:

$$1 - CL_B = P(Q \geq Q_{exp}|H_0) \quad (5.14)$$

In case of an expected limit, this value is equal to 0.5, as Q_{exp} is determined as the median of the test statistics distribution assuming H_0 is true.

If the experiment is not sensitive enough to distinguish between H_0 and H_n , this procedure leads to a high rejection rate for H_n . Instead, CL_S is defined as:

$$CL_S = \frac{CL_{S+B}}{CL_B} \quad (5.15)$$

The alternative hypothesis H_n is now rejected, if $CL_S \leq 1 - c$.
In the case of this thesis, instead of testing a single alternative hypothesis, a limit on the branching fraction is supposed to be determined. The limit is set equal to the smallest branching fraction that can be rejected with the desired confidence level using the CL_S method.

6 Data preselection

For the measurement of the lepton flavour violating decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$, the kinematically similar channel $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ is used as control channel to correct discrepancies between the data sample and the Monte Carlo sample. This chapter gives an overview of the variables relevant for this analysis. It describes the production rate correction for the Monte Carlo samples, which relies on work done by other members of this group. Furthermore, this chapter gives an overview of the trigger requirements and preselection cuts applied to both the control and the signal channel. The signal data and Monte Carlo samples are each separated into two sub-samples according to their number of muons identified by the muon system.

6.1 Selection variables

The trigger and reconstruction algorithms provide a number of variables which can be used to select and categorise events. The relevant variables are shortly described here.

The **invariant mass** of the three daughter tracks of the signal and control candidates is computed as follows:

$$m_{reco}^2 = \left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2 = \left(\sum_i \left(\sqrt{m_i^2 + \vec{p}_i^2}\right)\right)^2 - \left(\sum_i \vec{p}_i\right)^2, \quad (6.1)$$

where \vec{p}_i is the reconstructed momentum of each track and m_i the mass of the associated particle, in this case a muon or a pion.

If the assumptions on the identity of the daughter particles are correct and the momentum is fully reconstructed, the resulting reconstructed mass of the mother particle, m_{reco} , will be close to the mass of the τ^- or D_s^- , depending on the channel. However, if the assumption is incorrect, the reconstructed mass will be shifted in first approximation by the mass difference between the assumed particle and the actual daughter particle. The reconstructed invariant mass of the mother particle is called m_τ for the signal channel and m_{D_s} for the control and reference channel. In a similar way, the invariant mass of each pair of particles can be reconstructed. The invariant mass of two tracks matched to muons is called $m_{\mu^+\mu^-}$ if the muons are of opposite charge or $m_{\mu^+\mu^+}$ if they are of the same charge, regardless of whether this charge is positive or negative.

For the **particle identification (PID)**, three variables are used. These are *ProbNN*, *RichDLL* and *isMuon*. *ProbNN* and *RichDLL* make the assumption that a track corresponds to a certain particle, for example a muon or a pion

An artificial neural network is used to determine whether a track corresponds to a certain particle with a probability between 0 and 1. This variable is called *ProbNN* [44]. As null hypothesis for it the track is identified with a pion, so that $ProbNN_{particle}$ gives the probability that the track corresponds to a certain particle with respect to it corresponding to a pion.

The data from the RICH subdetector is processed using the global likelihood approach [45], by comparing various particle type hypotheses for each of the tracks. By calculating the differences in the log-likelihood values for a given particle type hypothesis and a pion hypothesis, the *RichDLL* variable for the given particle type or ghost is computed.

$$RichDLL_{particle} = \Delta(\ln(L)) = \ln(L_{particle}) - \ln(L_{\pi}) = \ln(L_{particle}/L_{\pi}) \quad (6.2)$$

To identify a track to be a muon requires signals in a subset of muon stations. The higher the momentum is, the more muon stations are required by the algorithm. The requirements are listed in Tab. 6.1. Any track fulfilling these requirements is labeled as $isMuon = 1$. For more information on the *isMuon* variable, see [46].

Table 6.1: Required stations with hits for *isMuon* depending on the track momentum [46].

p [GeV]	required stations
$p < 3$	always false
$3 < p < 6$	M2 & M3
$6 < p < 10$	M2 & M3 & (M4 or M5)
$p > 10$	M2 & M3 & M4 & M5

Additionally to identifying the type of particle corresponding to a track, the *ProbNN* and *RichDLL* variables can also be used to identify so-called *ghost tracks*. This means that random hits are matched by the algorithm as a track although they are not caused by a particle.

The **kinematic variables** of a track are the track momentum p and transverse momentum p_T , the pseudo-rapidity η , the energy E and the decay length λ_{decay} . The decay length is the distance between the primary vertex and the decay vertex or end vertex. It can also be used to estimate the lifetime $\tau = \lambda_{decay}/c$, where c corresponds to the velocity of light in vacuum. Additional useful variables are the Distance Of Closest Approach (DOCA) and $\cos(\alpha)$. The DOCA measures the minimal distance of two reconstructed tracks. This can be very useful to identify whether they stem from the same vertex. The angle α is the angle between the momentum \vec{p} of a particle and the vector connecting the primary vertex and the decay vertex $\vec{x}_{decay} - \vec{x}_{primary}$. For well reconstructed candidates, the angle α should be small, thus $\cos \alpha$ should be close to one.

The total number of all tracks in an event is called *nTracks*. The tracks belonging to the selected decay, which are therefore associated with the daughter particles, are called *selected tracks*. Any other tracks coming from the secondary vertex but not involved in the reconstruction of the mother particle are called *non-isolating*

tracks. Further tracks in the event that come from different secondary vertices are called *isolating tracks*. For the correct reconstruction of the event it is necessary to distinguish isolating tracks from non-isolating tracks. There are three different **isolation variables** used in this thesis. These are the *track isolation*, the *cone isolation* and the *muon isolation*.

Each track has an assigned *track isolation* value. It is defined as the number of LONG tracks close to the muon, where close is specified by a set of cuts on five track-specific variables. The sum of the track isolation variables of the three daughter tracks of a candidate is the *track isolation* of the candidate [47].

The *cone isolation* of a track is defined as:

$$\frac{p_T(\text{track}) - p_T(\text{cone})}{p_T(\text{track}) + p_T(\text{cone})} \quad (6.3)$$

where $p_T(\text{track})$ is the sum of the transverse momenta of all daughter tracks of the chosen track or the transverse momentum of the track itself, and $p_T(\text{cone})$ the sum of the transverse momenta of all tracks within a cone around the reconstructed track.

Inspired by the track isolation variable, *muon isolation* variables as described in [47] are used, which are either based on LONG tracks or VELO tracks. The variables are the output of a multivariate classifier using a BDT to train $b\bar{b} \rightarrow \mu^+\mu^- X$ background against $B_s \rightarrow \mu^+\mu^-$ signal. For each reconstructed muon, the isolation with regard to the closest track that is not a daughter of the signal candidate is calculated. Several criteria for isolation are used, resulting in three different variables for LONG and VELO each. The muon isolation variables based on LONG tracks are called LONGMAX1, LONGMAX2 and LONGMAX3 while the muon isolation variables based on VELO tracks are called VELOMAX1, VELOMAX2 and VELOMAX3. Each of these six muon isolation variables exists for any of the three reconstructed muons in the event. It is possible to combine the muon isolation variables for all three muons to get an isolation variable for a signal candidate. Alternatively, one can use the isolation variable of one of the three muons, as the isolations of the three tracks are strongly correlated. This can be seen in Fig. 6.1.

There are three variables corresponding to the quality of the **track reconstruction** and particle matching. These are called *track quality*, $vertex\chi^2$ and $IP\chi^2$. The *track quality* is determined from the track fit. The *track quality* of a signal candidate equals the maximum track χ^2 of the daughter tracks. The quality of the vertex fit is called $vertex\chi^2$ and gives the statistical significance of the vertex fit. Similarly, the statistical significance of the impact parameter IP is given by the variable $IP\chi^2$. Sometimes, these variables are divided by the number of degrees of freedom ndf of the fit.

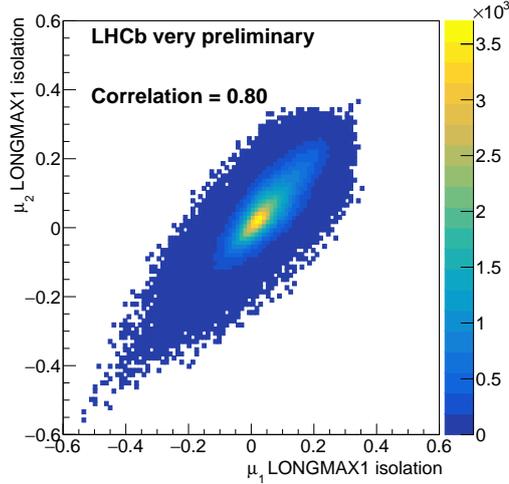


Figure 6.1: Correlation of the muon isolation of two muons of the MC signal candidate.

6.2 Stripping line selection

The reconstructed events coming from the BRUNEL application are saved into files through a filtering process called stripping. The stripping is done using the LHCb application DAVINCI. Each stripping line applies a series of cuts on the reconstructed candidates that removes a large number of events that are uninteresting for the specified process.

Table 6.2: Cuts applied by the stripping lines [48], [49].

	control channel	signal channel
μ^\pm		
$\sum \text{isMuon}$	2	≥ 2
μ^\pm, π^\pm		
p_T	> 300 MeV	> 300 MeV
track ghost probability	< 0.45	< 0.45
track χ^2/ndf	< 3	< 3
IP χ^2/ndf	> 9	> 9
mother (D_s^\pm or τ^\pm)		
m_{mother}	1968.34 ± 250 MeV	1776.86 ± 200 MeV
Vertex χ^2	< 15	< 15
IP χ^2	< 225	< 225
λ_{decay}	> 100 μm	> 100 μm

The stripping lines place a cut on the number of identified muons for both samples. For the control channel, exactly two muons are required, while for the signal channel

two or three muons are accepted. This looser requirement allows also for a muon candidate which is not detected as such, as it might be out of the acceptance of the muon system.

A wide signal window around the expected mother particle mass is defined. The window has a width of 500 MeV for the control channel and a width of 400 MeV for the signal channel. It cuts away events that are widely out of range.

A minimal requirement is placed on the transverse momentum of the daughter particles. The quality of the track is ensured by requiring a small track ghost probability. Cuts are placed on the χ^2 of both the track reconstruction and the impact parameter, as well as on the χ^2 of the mother vertex reconstruction and the impact parameter of the reconstructed mother. The decay time is also restricted. These cuts ensure a good reconstruction of the event. An overview over the set of cuts applied by the stripping lines can be found in Tab. 6.2.

6.3 Production rates

As was explained in Section 4.2, both the signal and the control channel have different production processes. Each of these processes is simulated separately, with the resulting number of candidates per process not necessarily representing the true ratio expected for data. Therefore, it is necessary to weight the candidates in a way that the ratio of candidates from each of these sub-channels corresponds to the production ratio in the experiment.

To calculate the fraction of events which come from each sub-channel, f_i , it is necessary to know the total production cross section of each sub-channel. The D and D_s samples used for this analysis contain only D mesons in the detector acceptance, thus the measured event rates need to be corrected for this. The final values for σ^{tot} are given in Tab. 6.3.

Table 6.3: Prompt charm cross section within the range $1 < p \text{ [GeV/c]} < 8$ and $2 < y < 4.5$ for D_s and D in acceptance and corrected [50] and total \bar{b} cross section [51].

channel	$\sigma^{acceptance}$ [μb]	σ^{tot} [μb]
D_s	353 ± 76	1732 ± 373
D	834 ± 78	4054 ± 379
\bar{b}	-	495 ± 52

Multiplying σ^{tot} with the branching fraction $\mathcal{B}_{tot,i}$ from Tab. 4.1 and Tab. 4.2, the cross section of each production channel can be calculated. The relative production fraction $f_{prod,i}$ is calculated as follows:

$$f_{prod,i} = \frac{\sigma_i^{tot} \times \mathcal{B}_{tot,i}}{\sum_i (\sigma_i^{tot} \times \mathcal{B}_{tot,i})} \quad (6.4)$$

Table 6.4: Branching fraction, cross sections and production fractions in 4π for the signal and control channel.

	channel	$\sigma^{tot} [\mu\text{b}]$	$\mathcal{B}_{tot} [\%]$	$\sigma^{tot} \times \mathcal{B}_{tot} [\mu\text{b}]$	f_{prod}
	signal				
(i)	$D_s^- \rightarrow \tau^-$	1732 ± 373	5.48 ± 0.23	94.9 ± 20.8	69.09 ± 3.74
(ii)	$\bar{b} \rightarrow D_s^- \rightarrow \tau^-$ $\bar{b} \rightarrow D_s^+ \rightarrow \tau^+$	990 ± 104	1.36 ± 0.21	13.5 ± 2.5	9.80 ± 0.53
(iii)	$D^- \rightarrow \tau^-$	4054 ± 379	0.12 ± 0.03	4.86 ± 1.2	3.54 ± 0.19
(iv)	$\bar{b} \rightarrow D^- \rightarrow \tau^-$ $\bar{b} \rightarrow D^+ \rightarrow \tau^+$	990 ± 104	0.028 ± 0.007	0.6 ± 0.1	0.20 ± 0.01
(v)	$\bar{b} \rightarrow \tau^-$ $\bar{b} \rightarrow \tau^+$	990 ± 104	2.41 ± 0.23	23.9 ± 3.4	17.37 ± 0.94
	control				
(vi)	D_s^-	1732 ± 373	1	1732 ± 373	87.8 ± 2.70
(vii)	$\bar{b} \rightarrow D_s^-$ $\bar{b} \rightarrow D_s^+$	990 ± 104	24.8 ± 3.74	245.52 ± 45.16	12.42 ± 0.38

An overview over the values can be found in Tab. 6.4.

At the simulation level, a number of acceptance cuts is applied to the events. Their efficiency, ϵ_{CUT} , is obtained by comparing the number of all generated events with those which pass the cuts. The relative event fraction f_i can be then computed in the following way:

$$f_i = \frac{\epsilon_{CUT,i} \times f_{prod,i}}{\sum_i (\epsilon_{CUT,i} \times f_{prod,i})} \quad (6.5)$$

The resulting values are shown in Tab. 6.5.

Table 6.5: Production fractions, cut efficiencies and relative fractions for all production modes.

	channel	$f_{prod} [\%]$	$\epsilon_{CUT} [\%]$	$f_i [\%]$
	signal			
(i)	$D_s^- \rightarrow \tau^-$	69.09 ± 3.74	11.1 ± 0.3	71.85 ± 1.16
(ii)	$\bar{b} \rightarrow D_s^- \rightarrow \tau^-$	9.80 ± 0.532	9.5 ± 0.4	8.72 ± 0.14
(iii)	$D^- \rightarrow \tau^-$	3.54 ± 0.19	11.2 ± 0.3	3.80 ± 0.06
(iv)	$\bar{b} \rightarrow D^- \rightarrow \tau^-$	0.20 ± 0.01	9.8 ± 0.4	0.18 ± 0.003
(v)	$\bar{b} \rightarrow \tau^\pm$	17.37 ± 0.94	9.6 ± 0.4	15.55 ± 0.25
	control			
(vi)	D_s^-	87.8 ± 2.70	12.2 ± 0.4	88.51 ± 0.60
(vii)	$\bar{b} \rightarrow D_s^\pm$	12.42 ± 0.38	11.2 ± 0.5	11.49 ± 0.08

In order to make sure that each sub-channel is represented in the correct fraction, the candidates are weighted with *fractions weights*. These are calculated for each

magnet polarity and year. The ratio N_i/f_i , where N_i is the number of candidates in a specific sub-channel and f_i its expected fraction, is normalised to the channel with the smallest ratio N_{min}/f_{min} . Thus the weight is obtained:

$$w_{fraction,i} = \frac{f_i}{f_{min}} \times \frac{N_{min}}{N_i} \quad (6.6)$$

6.4 Trigger requirements and event preselection

In order to cancel the systematic uncertainties coming from the trigger selection, candidates from both the signal and the control channel have to satisfy similar trigger requirements and preselection cuts. Selection requirements are applied at all three stages of the trigger system. The trigger is expected to be fired on the signal. This means that the signal candidate or its daughters must have triggered them. The relevant hardware trigger lines are called *LOMuon* and *LODiMuon*. The L0 muon trigger searches for straight-line tracks in the five muon stations. The track direction is used to estimate the transverse momentum of a muon candidate. For the *LOMuon* trigger the muon candidate with the largest p_T must be above the *LOMuon* threshold, while for the *LODiMuon* trigger the product of the largest and second largest p_T values of two muon candidates must be above the *LODiMuon* threshold. These thresholds differ for the different years, and can also slightly vary over the year. An overview over the thresholds that are used for MC events is shown in Tab. 6.6. For data events, the thresholds are similar, but can vary over time. Additionally, a maximum number of SPD hits is required in the L0 trigger lines.

Table 6.6: L0 trigger thresholds. [30]

L0 trigger	p_T threshold			SPD threshold
	2016	2017	2018	
Muon	>1.8 GeV	>1.35 GeV	>1.75 GeV	< 450
DiMuon	> 2.25 GeV ²	>1.69 GeV ²	>3.24 GeV ²	< 900

The HLT1 trigger lines are called *DiMuonLowMass* and *TrackMuon*. Both require the event to be triggered by either *LOMuon* or *LODiMuon*. Other requirements of the two trigger lines are shown in Tab. 6.7.

Table 6.7: HLT1 trigger requirements.

	DiMuonLowMass	TrackMuon
$p(\mu)$	$> 3 \text{ GeV}$	$> 6 \text{ GeV}$
$p_T(\mu)$	$> 0 \text{ GeV}$	$> 1.1 \text{ GeV}$
χ^2/ndf	< 4	< 3
<i>ProbNNghost</i>	< 0.2	< 0.2
$\text{IP}\chi^2$	> 4	> 35
isMuon	true	true
other	opposite charge or $m_{\mu^+\mu^+} > 220 \text{ MeV}$	VELO track compatible with hits in muon stations

While the trigger requirements for L0 and HLT1 are the same for both the signal and the control channel, the HLT2 trigger requirements depend on the number of muons that are accepted as good muons in the event. As can be seen in Tab. 6.2, exactly two daughters need to be identified as muons by the muon system for the control channel, while the signal channel demands two or three daughters to be identified as muons. In events with exactly three muons identified by the muon system, the only HLT2 trigger requirement is the *TriMuonTau23Mu* line. In events with exactly two muons identified by the muon system, there are two possible trigger lines. They are called *DiMuonDetached* and *TopoMuMu2Body*. They are not mutually exclusive. The *TopoMuMu2Body* trigger line relies on the output of a binary decision tree. The cuts of the HLT2 trigger lines can be found in Tab. 6.8.

Table 6.8: HLT2 trigger requirements.

DiMuonDetached	$p_T > 600 \text{ MeV}$ vertex $\chi^2 < 9$ $p_T(\text{at least one } \mu) > 300 \text{ MeV}$ decay length $\chi^2 > 7$ $\text{IP}\chi^2 > 9$
TopoMuMu2Body	BDT classifier > 0.99
TriMuonTau23Mu	vertex $\chi^2 < 25$ $\lambda_{\text{decay}} > 45 \mu\text{m}$ $ m_\tau - m_{\tau,\text{pdg}} \leq 225 \text{ MeV}$ $m_{\mu\mu} > 2 \times m_{\mu,\text{pdg}} + 14 \text{ MeV}$

The trigger requirements for both channels are shown in Tab. 6.9. They are applied to all samples. As the *HLT2DiMuonDetached* trigger line does not exist for the year 2016, events with exactly two good muons from these years always have to pass the *TopoMuMu2Body* trigger requirement. This concerns the signal channel events with exactly two muons identified by the muon system, as well as all control channel events for 2016. Table 6.10 shows the efficiencies of the trigger cuts for each year and both signal and control channel.

Table 6.9: Trigger requirements for all samples.

L0	Muon (TOS) or DiMuon (TOS)
HLT1	DiMuonLowMass (TOS) or TrackMuon (TOS)
$\sum \text{isMuon} = 2$: HLT2	DiMuonDetached (TOS) or TopoMuMu2Body (TOS)
$\sum \text{isMuon} = 3$: HLT2	TriMuonTau23Mu (TOS)

Table 6.10: Trigger cut efficiencies for the signal channel dependent on the number of good muons, and for the control channel.

number of good muons	2016	2017	2018
		signal channel	
2	$0.07 \pm (1.02 \times 10^{-3})$	$0.22 \pm (1.70 \times 10^{-3})$	$0.19 \pm (1.58 \times 10^{-3})$
3	$0.59 \pm (1.29 \times 10^{-3})$	$0.67 \pm (1.24 \times 10^{-3})$	$0.58 \pm (1.28 \times 10^{-3})$
		control channel	
2	$0.15 \pm (8.93 \times 10^{-3})$	$0.32 \pm (1.14 \times 10^{-3})$	$0.28 \pm (1.17 \times 10^{-3})$

In order to further remove background, a set of preselection cuts is applied to both samples. These cuts are chosen in accordance with the analysis of *Run 1* [6]. They are rather loose cuts to remove as much background as possible without removing possible signal candidates. For better comparability, similar cuts are applied on the signal and on the control channel. They are shown in Tab. 6.11.

Table 6.11: Preselection cuts applied to the signal channel.

variable	cut	reason
$m_{\mu+\mu-}$	> 450 MeV	remove $D_s^- \rightarrow \eta(\mu\mu\gamma)\mu\nu_\mu$
$m_{\mu+\mu+}$	> 250 MeV	remove clone tracks
$\ m_{\mu+\mu-} - 1020 \text{ MeV}\ $	> 20 MeV	remove $\phi(1020)$
$\cos(\alpha)$	> 0.99	avoid bad reconstruction
decay time	> -0.01 && < 0.025 ns	avoid bad reconstruction
m_τ (data only)	$\neq 1776.86 \pm 20$ MeV	blind signal region

Cuts are applied on the invariant mass of each pair of oppositely charged muons in order to remove background from the $\phi(1020)$ resonance and from the decay $D_s \rightarrow \eta(\mu\mu\gamma)\mu\nu_\mu$. The effect can be seen for all three years in Fig. 6.2. A cut on the invariant mass of the pair of same charged muons is meant to remove clone tracks that are caused by one particle, but interpreted as two tracks. Cuts on $\cos(\alpha)$ and the decay time are meant to remove badly reconstructed events. The efficiencies of the preselection cuts on the signal and on the control channel are shown in Tab. 6.12.

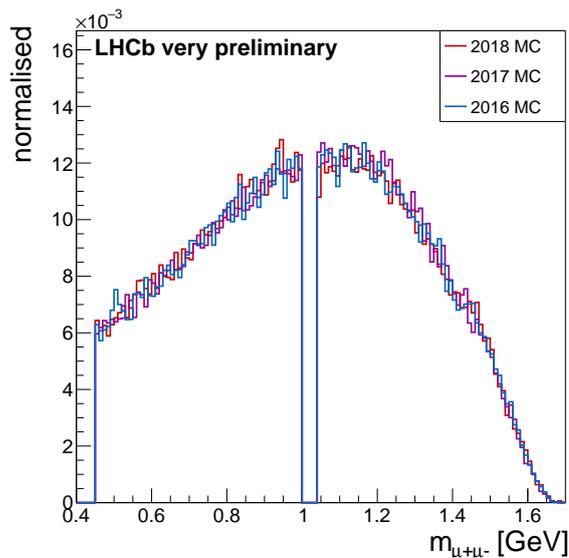


Figure 6.2: Distribution of the invariant mass of the first two muons of opposite charge for the simulated signal candidates of all three years.

Table 6.12: Preselection cut efficiencies for the signal channel dependent on the number of good muons.

number of good muons	2016	2017	2018
2	$0.22 \pm (5.56 \times 10^{-4})$	$0.22 \pm (5.23 \times 10^{-4})$	$0.22 \pm (5.66 \times 10^{-4})$
3	$0.54 \pm (9.53 \times 10^{-4})$	$0.54 \pm (9.61 \times 10^{-4})$	$0.53 \pm (9.44 \times 10^{-4})$

Additionally, a cut is applied on the signal data sample around the τ^\pm mass in order to *blind* the region during the analysis process. This is done in order to avoid any bias stemming from premature knowledge. Overall, the data is separated into three regions: the blinded signal region, the inner sidebands, which are used for training the boosted decision tree, and the outer sidebands, which are used for determining the efficiency of the cuts. The three regions are shown in Tab. 6.13.

Table 6.13: Regions of the signal channel data.

region	range [MeV]
left outer sidebands	1600 - 1747.18
left inner sideband	1747.18 - 1756.86
signal region	1756.86 - 1796.86
right inner sideband	1796.86 - 1807.18
right outer sideband	1807.18 - 1950

For the control channel, some cuts on $\cos(\alpha)$ and on the decay time are applied. The invariant mass of the two oppositely charged muons is requested to be consistent with the ϕ resonance and the invariant mass of all three tracks to be consistent with the D_s . Particle identification cuts using the $ProbNN$ variable are applied on all three daughter particles in order to reduce background. The preselection cuts of the control channel are shown in Tab. 6.14. Their cut efficiencies are shown in Tab. 6.15.

Table 6.14: Preselection cuts applied to the control channel.

variable	cut
m_{D_s}	1968.34 ± 50 MeV
$ProbNNmu(\mu)$	> 0.2
$ProbNNpi(\pi)$	> 0.2
$m_{\mu+\mu-}$	1020 ± 20 MeV
$\cos(\alpha)$	> 0.99
decay time	> -0.01 && < 0.025 ns

Table 6.15: Preselection efficiencies for the control channel sample, with the mass cut and the particle identification cut calculated separately from the rest of the preselection cuts.

	2016	2017	2018
ϵ_{presel}	$0.92 \pm (6.46 \times 10^{-4})$	$0.92 \pm (6.27 \times 10^{-4})$	$0.92 \pm (6.71 \times 10^{-4})$
ϵ_{mass}	$0.99 \pm (7.31 \times 10^{-4})$	$0.99 \pm (4.92 \times 10^{-4})$	$0.99 \pm (5.58 \times 10^{-4})$
ϵ_{ID}	$0.97 \pm (1.13 \times 10^{-3})$	$0.96 \pm (8.51 \times 10^{-4})$	$0.96 \pm (9.58 \times 10^{-4})$

Furthermore, some cuts are applied on the MC samples only. Badly reconstructed candidates are rejected exploiting matching of the reconstructed MC tracks to generated MC tracks. Events where the daughter tracks or the reconstructed mother are matched to the wrong particle are excluded. Also excluded are events where the mother particle decay has not been fully reconstructed and thus the reconstructed mass of the mother particle is shifted by more than 100 MeV from the true mass.

6.5 Monte Carlo correction

The simulation is often not ideally reproducing certain variables. An example is the number of tracks per event, $nTracks$, which is often underestimated. This has multiple causes, the most relevant of which is that diffraction is not well described in the simulation [34]. The comparison of data and simulation of the control channel is used to correct the simulation of the signal channel. It is necessary for the Monte Carlo simulation to give the best possible representation of the data, in order to be used to correctly evaluate efficiencies. The control channel has a very similar kinematic distribution to the signal channel and is therefore expected to have the same

systematic discrepancies. Any correction depending on kinematic variables should therefore be very similar for signal and control channel.

Since the control data sample also contains background events, the statistical procedure of *sWeighting* [41] is used to compare the relevant variable distributions of the $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ candidates in simulated samples and data samples for each of the three years. The control data is fitted in the full mass range as defined in Tab. 6.14 with an exponential background distribution and a Double Crystal Ball (DCB) signal distribution [52]. As starting values for the signal distribution, a fit to the Monte Carlo sample is used. The results of the fits for all three years can be found in Fig. 6.3.

The fits are used to assign an *sWeight* to each event, which can be used to obtain the signal distribution of any variable by weighting the events with it. For more information on the *sWeighting* procedure, see Sec. 5.2.

The data distributions of relevant variables weighted with the *sWeights* produced from the fit are compared with the simulated distributions weighted with the *fractions weights* of the same variables for each year.

Especially relevant here are the muon isolation variables, which are meant to heavily feature in the separation of background and signal. As these are defined for three muons in the signal channel, but only for two muons in the control channel, it is attempted to achieve correction in the muon isolation variables by correcting in other variables.

The distributions are compared bin-by-bin. This is first done in bins of the track multiplicity, $nTracks$, and then in three-dimensional bins of the mother particle momentum, p , transverse momentum, p_T , and the natural logarithm of the χ^2 of its impact parameter, $\ln(IP\chi^2)$. Weights are calculated for each candidate depending on these variables such that the weighted MC sample distributions and the *sWeighted* data sample distributions of the control channel agree in these variables.

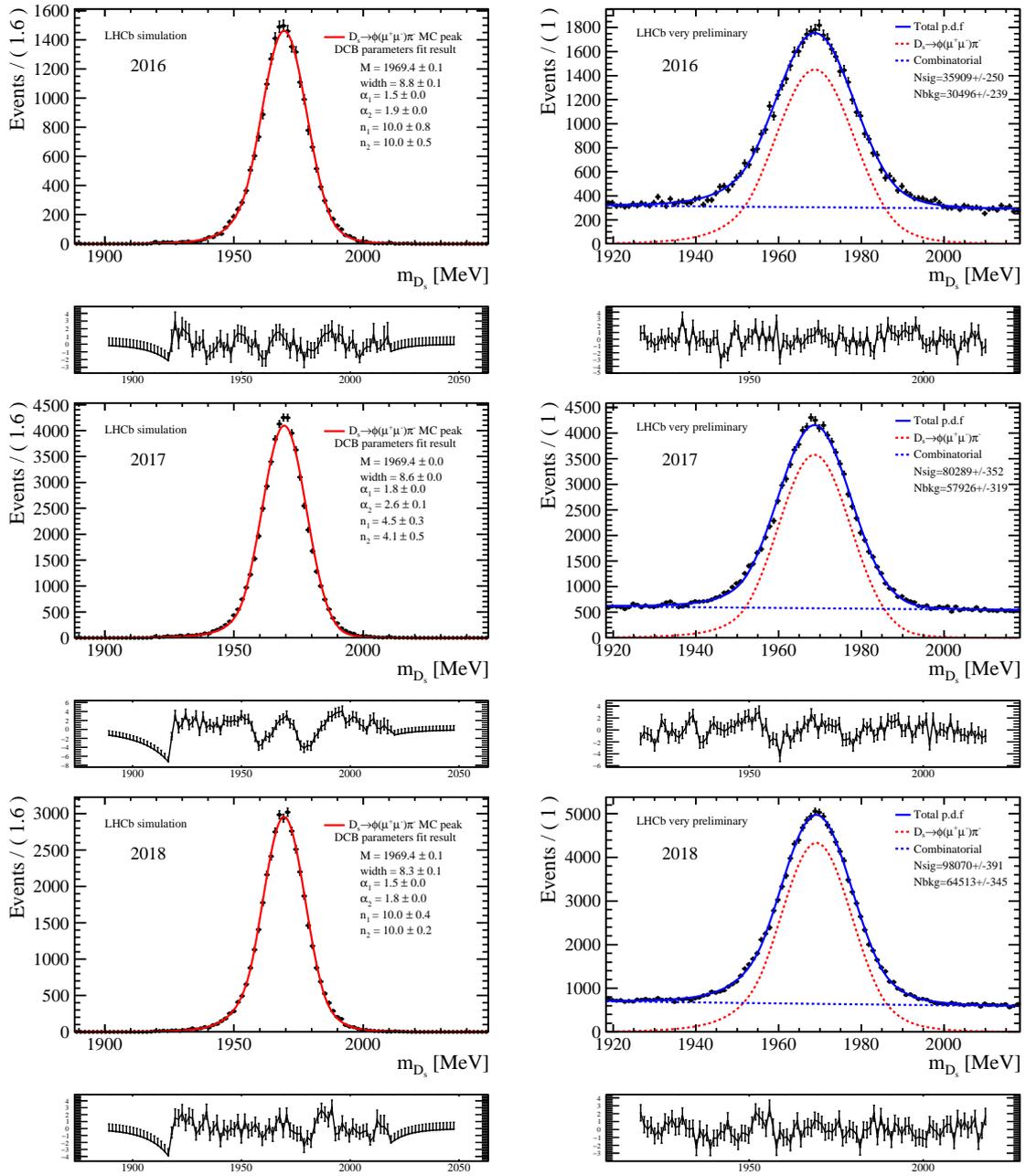


Figure 6.3: Invariant mass of simulated candidates with the signal model (red) and the background model (dotted blue line) overlaid for simulation (left) and data (right) for the years 2016 (top), 2017 (middle) and 2018 (bottom).

As is shown exemplary for 2018 in Fig. 6.4, good agreement is achieved in all presented variables.

The weights calculated for the MC control sample can now be stored depending on the values of $nTracks$, p , p_T and $\ln(IP\chi^2)$ for each year and applied to the MC signal samples. They are called *correction weights*. Any residual discrepancy is later reduced with a second MC correction depending on the classifier output value of the boosted decision tree training.

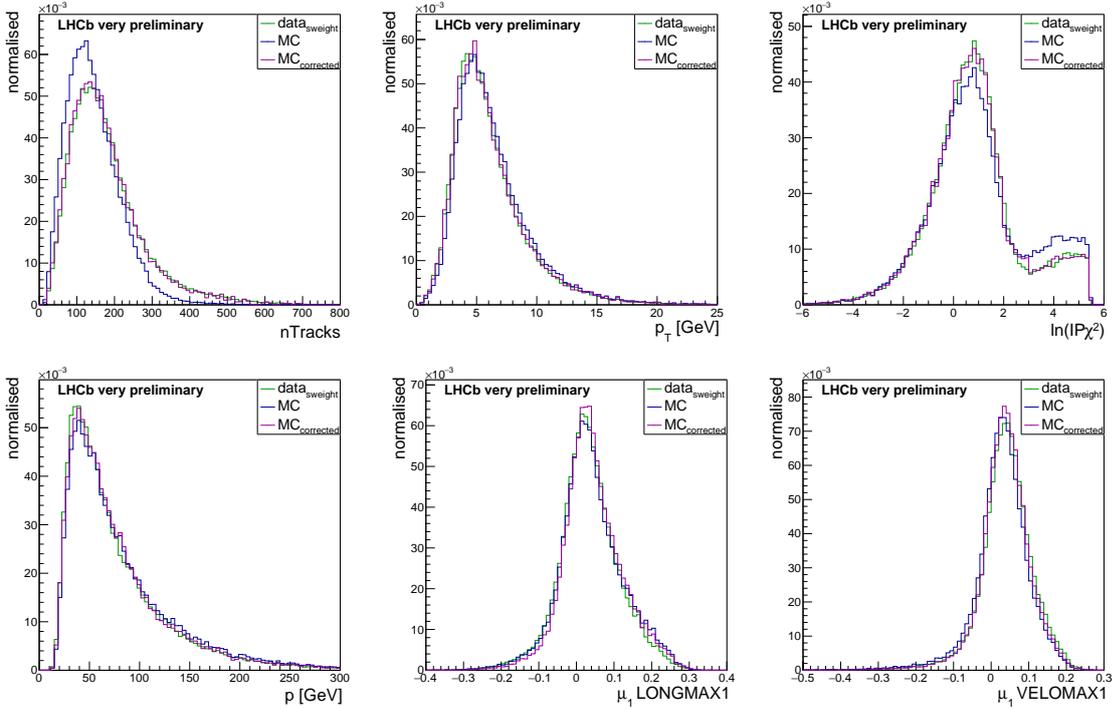


Figure 6.4: Corrected and uncorrected 2018 MC control sample distribution together with the *sWeighted* 2018 control data sample distribution for different variables of the mother particle associated with D_s , the track multiplicity of the event and the muon isolation variables LONGMAX1 and VELOMAX1 for one muon.

6.6 Signal channel preparation

As can be seen in Tab. 6.2, the stripping line for the signal channel requires only two muons to be *isMuon*, which are called *good muons* from now on. Any muon not fulfilling the $isMuon = 1$ requirement in the signal channel is called *non-good muon*. The signal channel sample can now be split into two mutually exclusive samples. The 2μ sub-sample contains all candidates which have exactly two good muons and one non-good muon, while the 3μ sub-sample contains all candidates which have exactly three good muons. The additional sample adds statistics to the analysis and can therefore improve the significance of the measurement. In order to have similar

treatment for both samples, only information of two good muons can be used for both samples. This leads to better comparability with the control and reference channel, which requires exactly two good muons.

Figure 6.5 shows the momentum distribution of muons and non-good muons of the MC signal 2μ sub-sample in comparison. As can be seen, the momentum of non-good muons is distributed with a much lower peak than that of the good muons. About 13% of non-good muons do not pass the requirement of a minimum momentum of 3 GeV in the muon detector. As can be seen in Tab. 6.1, the remaining muons must not have left hits in enough muon chambers to be identified as good muons. Also shown in Fig. 6.5 is the acceptance of good muons and non-good muons in the muon system. As can be seen, about 70% of non-good muons are outside of the geometrical acceptance region of the muon system and could therefore not be identified as good muons. The 2μ sub-sample is taken into account as an independent sub-sample from the 3μ sub-sample. Efficiencies are calculated for both samples separately.

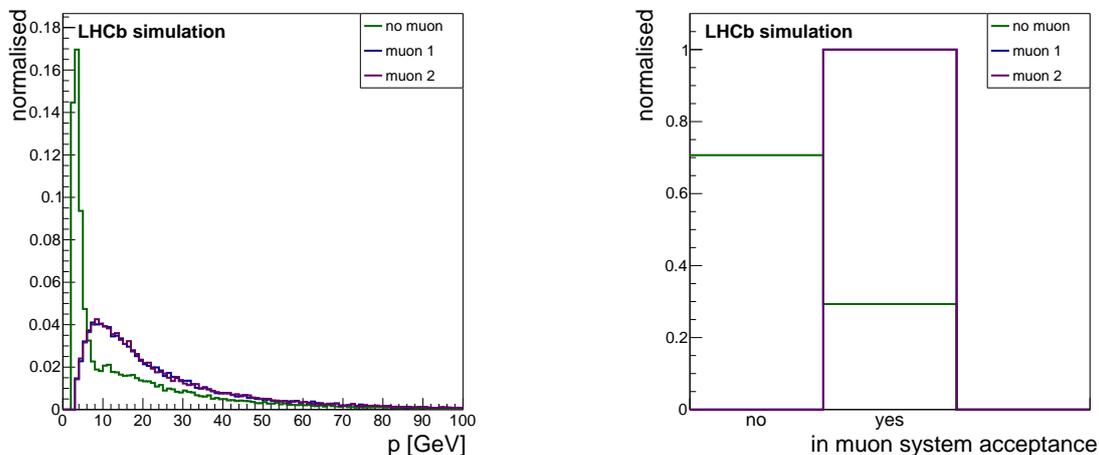


Figure 6.5: Distribution of momentum (left) and detector acceptance (right) for good muons and non-good muons of the signal channel MC 2μ sub-sample

After all preselection and trigger cuts are applied to the signal channel, the *correction weights* that are calculated using the MC control sample are stored for each MC signal sample candidate. It is used to weight the events in all future dealings together with the *fraction weights*.

The multivariate classifier trains to distinguish signal events from background events. For the training, the background is taken from the inner sidebands of the 2018 signal data sample, while the signal is taken from the MC signal sample from the years 2016 – 2018 in order to have more statistics. This is not optimal for training with the 2μ sub-sample, as there is some difference in the distributions of the Monte Carlo samples for the year 2016 and the years 2017/18, as can be seen in the top row of Fig. 6.6. This is caused by the missing trigger line for the year 2016, leading to stricter trigger requirements. However, a classifier trained with these samples still provides significant separation power for signal and background. Additionally, it has

a similar performance when applied to all three years, leading to less dependency on the specific sample and more robustness. This is verified later.

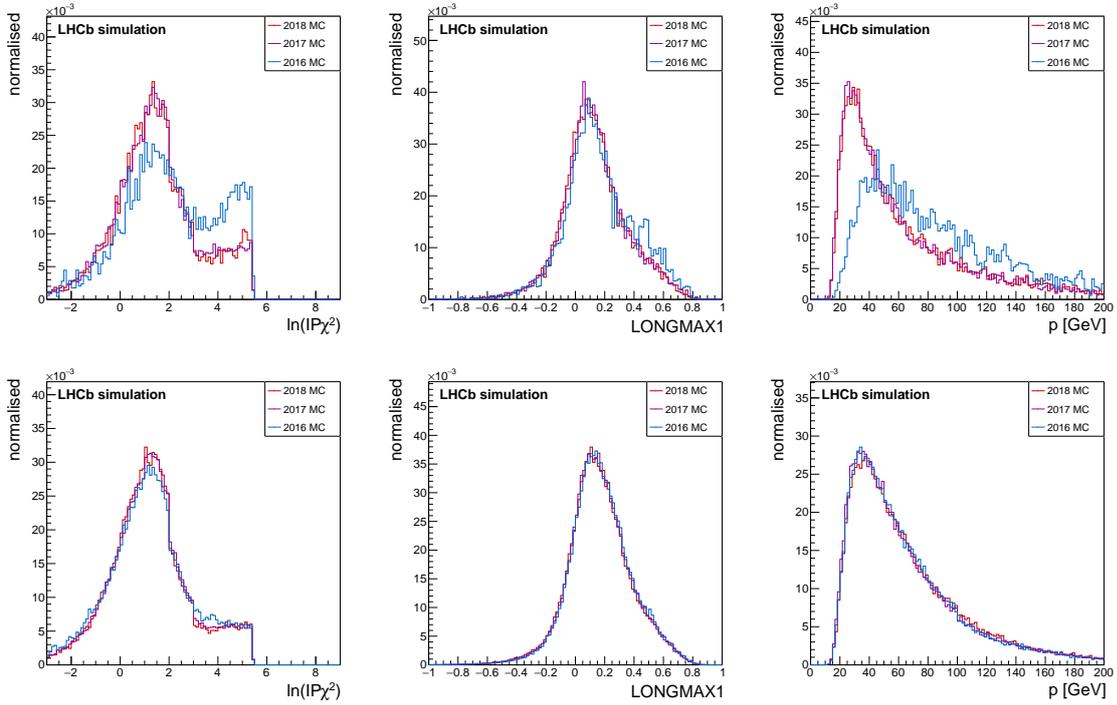


Figure 6.6: $\ln(IP\chi^2)$, the LONGMAX1 muon isolation variable and the momentum, p , in comparison for 2μ (top) and 3μ (bottom) MC sub-sample for the different years.

The same problem does not persist with the 3μ sub-sample, as can be seen in the bottom row of Fig. 6.6. Here, the distributions for all three years prove to be sufficiently similar, even though there are some minimal differences.

The classifier trained on 2018 data against 2016 – 2018 MC events is applied to all samples. The calculation of efficiencies and estimation of expected background for the final limit is done for all three years and both sub-sample separately, so that expected limits are calculated in total in six bins. These limits are then combined into one expected limit.

7 Multivariate classifier training and application

For the optimal separation of the signal decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ from combinatorial background, a multivariate analysis classifier is trained on a chosen set of variables. The preselected signal and control samples are weighted to achieve the correct fractions from the different production channels and to correct for discrepancies between data and simulation. The signal sample is separated into two sub-samples. The so-called 2μ sub-sample contains all events with exactly two muons identified by the muon system, while the so-called 3μ sub-sample contains all events with exactly three muons identified by the muon system. The multivariate classifier is trained on distinguishing signal and background in both samples separately, as well as on a merged sample. The best training result is applied to all samples, as well as to the preselected and weighted control samples from the decay $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$. The output distributions of the control data and MC samples are compared and used to determine a third set of weights, correcting for residual discrepancies between data and MC distributions. This chapter describes the training of the multivariate classifier, as well as the application of the results and the calculation of the third set of weights for the MC samples.

7.1 Variable selection

The main source of background to the $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ decay is combinatorial background. That means, that tracks in an event coming from different decays are mismatched as belonging to the signal decay, as they coincidentally share many characteristics with the signal decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$. For the optimal separation of the signal decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ from combinatorial background, a boosted decision tree (BDT) algorithm is trained. It produces a variable classifying events as either signal-like or background-like. This *classifier output* can then be cut upon in order to reduce the combinatorial background. The TMVA toolkit is used for the training. A selection of variables is given to the training algorithm.

For better comparability with the control channel $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$, which only contains two muons, two muons are chosen in each event of the signal channel $\tau^- \rightarrow \mu^- \mu^+ \mu^-$. They are referred to as *chosen muons*. Variables referring to these two muons are compared to the same variables of the control channel muons, while the variables referring to the third muon are compared to the pion variables from the control channel.

In events with exactly two *good muons*, which are identified as muons by the muon

system, the two good muons are the chosen muons. They do not necessarily have to be of opposite charge. It is shown later that candidates with two good muons of the same charge in the 2μ sub-sample do not lower the efficiency of the training mechanism. In events with exactly three good muons, a pair of oppositely charged muons is chosen. This leads to better comparability with the control channel again, where the two muons are always of opposite charge. Of the two muons with the same charge in the event, the muon with higher transverse momentum, p_T , is chosen, as muons with a higher transverse momentum are expected to be more isolated. This way, the isolation variables can more easily be used as distinguishing variables. The training is done separately in the 2μ sub-sample and the 3μ sub-sample. Additionally, training is done on a merged sample of both. For comparison with the *Run 1* analysis, the 3μ sub-sample is trained additionally using the full information from all three muons. Therefore, four different training setups are used:

- for comparison with the *Run 1* analysis, the 3μ sub-sample is trained using the full information from all three muons. The muon isolation information from all three muons is summed to obtain a final muon isolation variable. This setup is called the $3\mu_{3\mu}$ setup.
- the 3μ sub-sample is trained using two oppositely charged muons as chosen muons. Of the two muons with the same charge, the one with higher transverse momentum is chosen. The muon isolation information from one of these muons is used as isolation variable, as it is highly correlated with the isolation information from the other muon, as seen in Fig. 6.1. This setup is called the $3\mu_{2\mu}$ setup.
- the 2μ sub-sample is trained using the two good muons as chosen muons. The muon isolation information from one of these muons is used as muon isolation variable. This setup is called the $2\mu_{2\mu}$ setup.
- a merged sample with all signal channel candidates is trained. In events with two good muons, those are used as chosen muons, while in events with three good muons, a pair of opposite charge is chosen, with the one with higher transverse momentum of the two same-charged muons included. The muon isolation information from one of the chosen muons is used as muon isolation variable. This setup is called the $(2 + 3)\mu_{2\mu}$ setup.

The best training setup is chosen and applied to all samples.

Table 7.1 provides an overview over the number of candidates that passed the trigger requirements and preselection cuts in the 2μ and 3μ sub-samples and are used in training. In order to distinguish signal and background, the BDT is given the simulated MC events from each sub-sample as signal and the data events of the inner sidebands from each sub-sample as background. The *BDT signal sample* and the *BDT background sample* for each training are both separated randomly in a *BDT signal/background training sample* and a *BDT signal/background testing sample*.

Table 7.1: Number of candidates in all data and Monte Carlo sub-samples used in training.

	data (2018)	MC (2016 – 18)
2μ	1,154,162	39,670
3μ	164,073	361,190

The training is done on the BDT signal and background training samples. The BDT signal and background testing samples are used to test the performance of the training algorithm. As can be seen, the 3μ sub-sample has almost ten times the amount of MC candidates as the 2μ sample, while it has only around a seventh the amount of data candidates. This is not surprising, as events where two muons are identified as good muons are more likely to happen in the detector, while the generated MC sample contains only events with three muons.

After a series of tests, where training parameters like the maximum depth of the tree and the number of training events are varied, 15 variables are chosen that are most often identified as variables with the best separation power by the TMVA algorithm. These 15 variables are used for the training. Table 7.2 gives an overview over the chosen variables. These include two muon isolation variables, called LONGMAX1 and VELOMAX1. The chosen muons are called μ_1 and μ_2 , while the third muon is called μ_{not} . In case of the $3\mu_{3\mu}$ setup, no choice was made and the three muons are called μ_1 , μ_2 and μ_3 .

Table 7.2: Variables given to the TMVA algorithm for training.

$IP\chi^2$
LONGMAX1
μ_1 track quality
μ_2 track quality
μ_3/μ_{not} track quality
endvertex χ^2
μ_3/μ_{not} cone isolation
μ_1 cone isolation
μ_2 cone isolation
$\mu_1 p_T$
$\mu_2 p_T$
$\mu_3/\mu_{not} p_T$
τp_T
VELOMAX1
λ_{decay}

To adjust for the correct relative production rates, the MC events are weighted by the *fraction weights*. Additionally, they are weighted by the *correction weights* to compensate for discrepancies between data and simulation.

7.2 Training optimisation

In order to achieve the best separation between signal and background, two kinds of boosted decision trees are trained. These are the gradient boosted decision tree (GradBoost) and the adaptive boosted decision tree (AdaBoost). More information on them can be found in Chapter 5. The training settings for both are varied in order to achieve the best results. At the same time, the number of candidates used for the training, $n_{Training}$, is varied, while the rest of the candidates, the test sample, is used to test the training and evaluate the performance. The number of candidates used for the training is the same for background and signal sample. The performance is measured on the test sample by integrating over the Receiver Operating Characteristic (ROC) curve, which plots the signal efficiency ϵ_S over the background rejection $1 - \epsilon_B$ for different values of the cut on the classifier output. The goal is to achieve the highest possible value for the integral over the ROC curve. At the same time, the Kolmogorov-Smirnov test is used on both signal and background to check that no overtraining occurs.

Table 7.3: Optimised training settings for all setups with the best performance for each setup marked in light grey.

setting	$3\mu_{3\mu}$	$3\mu_{2\mu}$	$2\mu_{2\mu}$	$(2+3)\mu_{2\mu}$
AdaBoost				
$n_{Training}$	30,700	30,700	25,230	149,430
$MaxDepth$	5	4	5	4
$MinNodeSize$ [%]	7	10	7	6
$nSteps$	15	9	38	26
$\int ROC$	0.962	0.949	0.937	0.941
r_{KS} (signal)	0.152	0.690	0.017	0.226
r_{KS} (background)	0.048	0.914	0.576	0.453
GradientBoost				
$n_{Training}$	30,700	30,700	25,230	149,430
$MaxDepth$	5	5	7	5
$MinNodeSize$ [%]	9	9	5	6
$nSteps$	16	10	20	27
$\int ROC$	0.959	0.951	0.938	0.940
r_{KS} (signal)	0.210	0.142	0.037	0.227
r_{KS} (background)	0.621	0.546	0.799	0.661

The parameters optimised to achieve the best training setting are the maximum depth allowed for the tree, $MaxDepth$, the minimum size allowed for the leaf nodes in percentage of the total amount of candidates, $MinNodeSize$, and the number of steps to test a variables range for the optimal cut, $nSteps$. An overview over the optimised values together with the achieved ROC integral values, $\int ROC$, and Kolmogorov-Smirnov values, r_{KS} , can be found in Tab. 7.3. Other settings that

are the same for all training samples can be found in Tab. 7.4. The variables are normalised before training as this leads to slightly better results and guarantees that the algorithm can optimise over the whole variable range. For more information on training settings, see [42].

Table 7.4: Other training settings that were the same for all training setups. [42]

setting	value	explanation
nTrees	300	number of trees trained
β_{Ada} (AdaBoost only)	0.3	learning rate for the AdaBoost algorithm
β_{Grad} (GradBoost only)	0.06	learning rate for the GradBoost algorithm

The output distributions of the TMVA training algorithm for all training setups can be seen in Fig. 7.2 , where the distributions for training and testing are compared and the Kolmogorov-Smirnov values as calculated by the algorithm are shown.

The ROC curves for the better performing of the two boosted decision trees for each optimised training setup in comparison can be seen in Fig. 7.1.

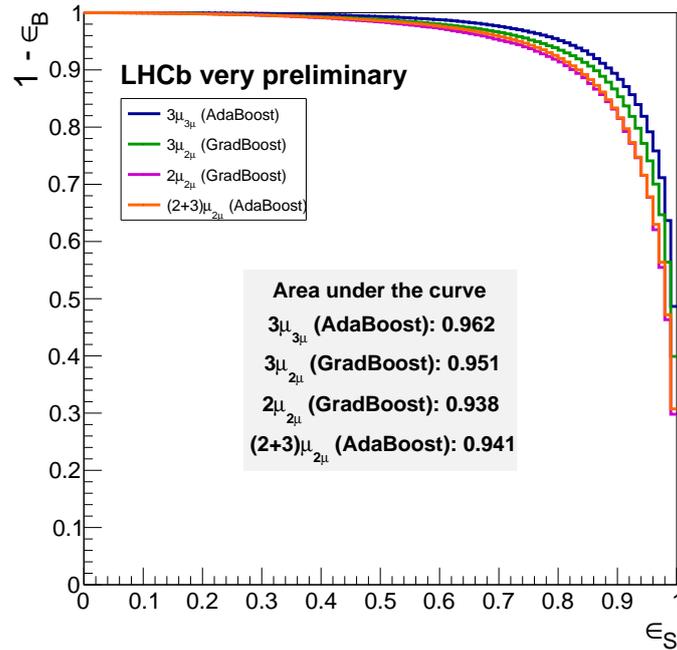


Figure 7.1: Signal efficiency against background rejection (ROC curve) in comparison for all four training setups.

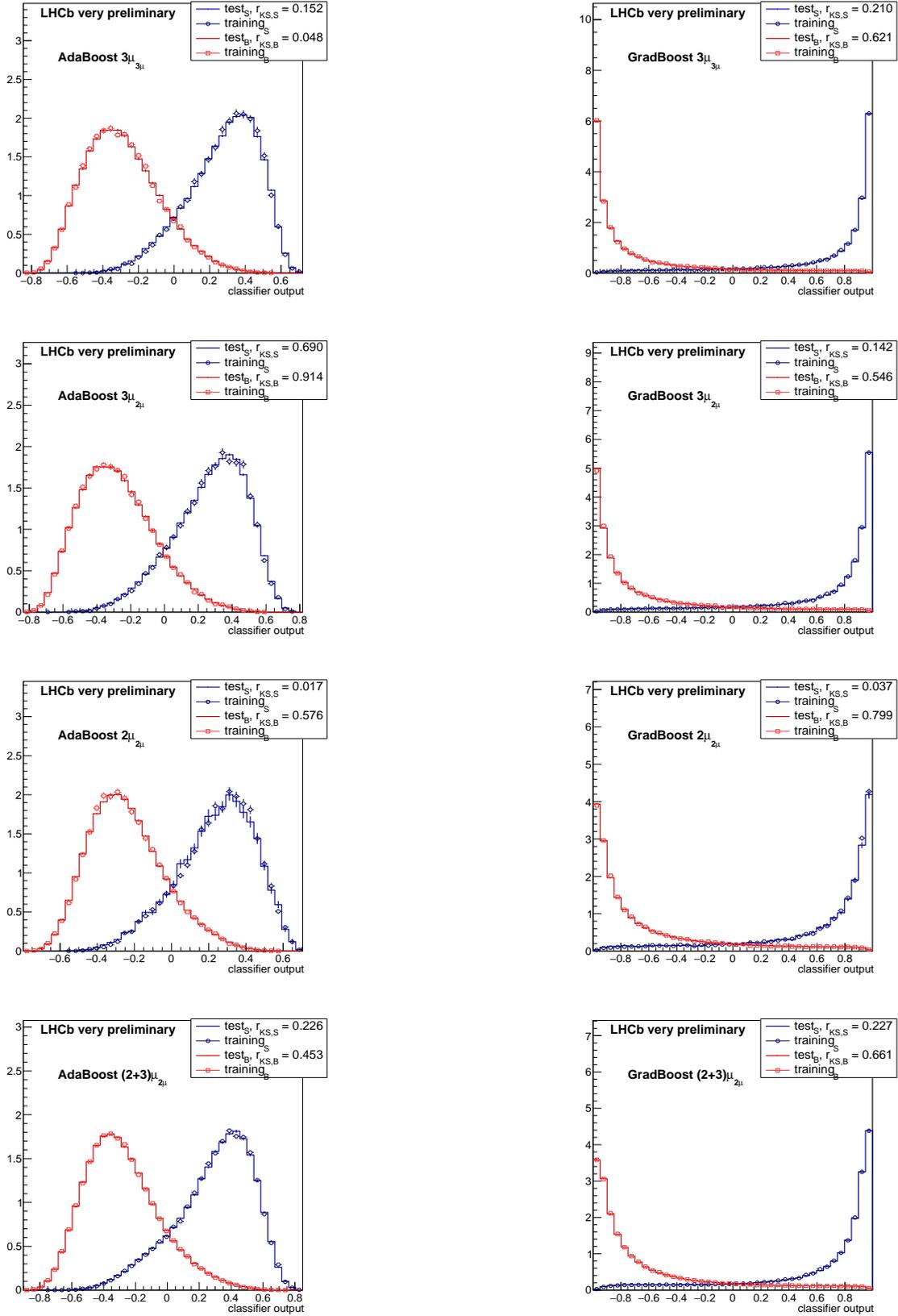


Figure 7.2: Training and test distributions for signal (blue) and background (red) for each training setup for the training with AdaBoost (left) and GradBoost (right).

As can be seen, the best result is achieved with the $3\mu_{3\mu}$ training setup. This is achieved using the AdaBoost algorithm. The best result using only the BDT isolation information of two muons is given by the GradBoost algorithm with the $3\mu_{2\mu}$ training setup. A comparison between the ROC curves achieved by this analysis and the ROC curves achieved by the different training methods of the same analysis with *Run 1* data can be found in Fig. 7.3. Various training algorithms were used and compared in *Run 1*. More information on them can be found in [6]. Figure 7.3 shows that there is already an improvement of about 10% with respect to the best algorithm of *Run 1*, even though the comparison is done at the same level of selection. The higher performance of the *Run 2* classifier is mainly caused by the influence of the muon isolation variables, which strongly shape the optimised classifier.

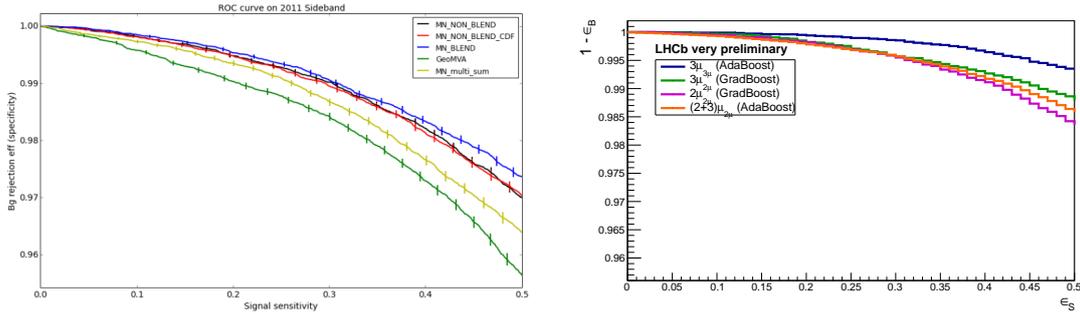


Figure 7.3: Signal efficiency against background rejection (ROC curve) for the published *Run 1* dataset [6] and this in comparison. Various training algorithms that were used in the *Run 1* analysis are shown in comparison.

In order to determine whether it is a problem to use candidates where both good muons have the same charge for the 2μ sub-sample, a training is performed with the $2\mu_{2\mu}$ setup using only candidates with oppositely charged muons. These are 83.60% of the 2μ MC sub-sample candidates, and 95.32% of the 2μ sub-sample data candidates. The performance of this training when applied to the 3μ sub-sample is compared to the training with the $2\mu_{2\mu}$ setup using all candidates, including those where both chosen muons have the same charge. Except for the number of candidates used for training, which is reduced to 13,260 candidates given the reduced number of total candidates, the same training settings are used as in the $2\mu_{2\mu}$ setup. Figure 7.4 shows the results of applying the $2\mu_{2\mu}$ training setup to the 3μ sample. The training performance using only oppositely charged muon candidates is shown, as well as the performance using all candidates. As can be seen, the addition of candidates where both good muons are of the same charge does not lower the performance and no relevant difference can be seen in the output distributions. Therefore, all 2μ sub-sample candidates can be used in training.

In order to find out which training setup should be used to determine the classifier output which is cut upon for removing background, the classifiers are applied to both the 2μ and the 3μ sub-sample.

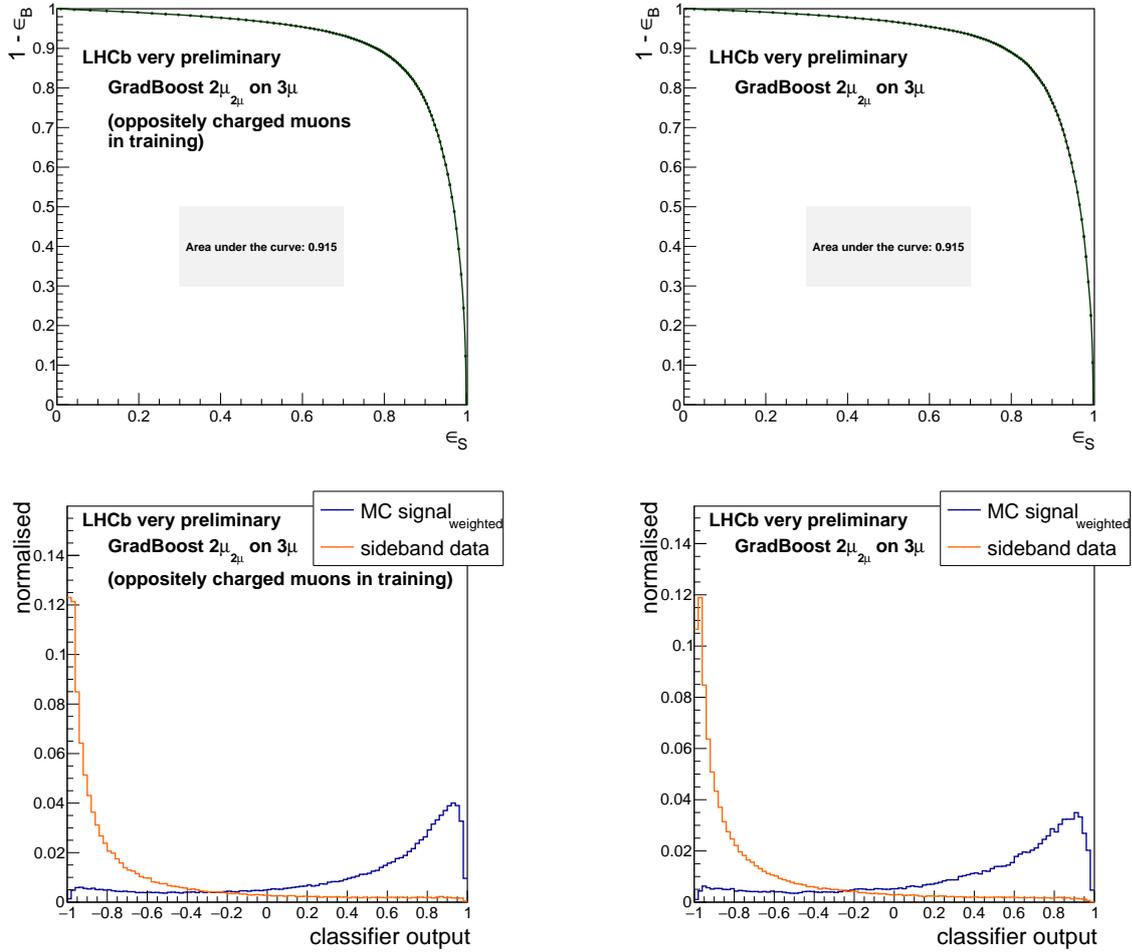


Figure 7.4: ROC curves and classifier output distributions for applying the GradBoost $2\mu_{2\mu}$ training done with all candidates (right) or only with events with oppositely charged muons (left) to the 3μ sub-sample.

7.3 Application of the results

The classifiers of the three training setups with two chosen muons are applied to the 2μ and the 3μ sub-sample. The performance is again checked by integrating over the ROC curve. The results can be found in Tab. 7.5. It has to be taken into account that a slight bias might appear as the candidates used for the training itself are now also taken into account when applying the training results.

Table 7.5: Performances (integral over ROC curve) when applying the classifiers of different training setups to the 2μ and 3μ sub-samples.

	2μ	3μ
$2\mu_{2\mu}$ <i>AdaBoost</i>	0.918	0.913
$2\mu_{2\mu}$ <i>GradBoost</i>	0.919	0.915
$3\mu_{2\mu}$ <i>AdaBoost</i>	0.878	0.923
$3\mu_{2\mu}$ <i>GradBoost</i>	0.881	0.927
$(2+3)\mu_{2\mu}$ <i>AdaBoost</i>	0.892	0.919
$(2+3)\mu_{2\mu}$ <i>GradBoost</i>	0.892	0.917

As can be seen, the $2\mu_{2\mu}$ setup with GradBoost shows better results when applied to the 2μ sample than both the $3\mu_{2\mu}$ and the $(2+3)\mu_{2\mu}$ setup. The improvement that can be gained with the other training setups in the application to the 3μ sample is negligible in comparison. The $2\mu_{2\mu}$ setup with the GradBoost algorithm is therefore chosen to be used to produce the *classifier output*, as it shows the best performance. This classifier output is used to correct the MC a second time. A cut on the classifier output is later determined in order to optimally separate signal and background. The variables for this setup can be found in Tab. 7.6, ordered by their importance as determined by the TMVA algorithm. The output distributions when applying the $2\mu_{2\mu}$ setup to the 2μ sub-sample, to the 3μ sub-sample and to the control channel $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ can be found in Fig. 7.5. The control data sample events are weighted with *sWeights* again, in order to extract the signal distribution statistically.

Table 7.6: Variable ordering for the chosen $2\mu_{2\mu}$ setup with the GradBoost algorithm.

rank	variable	rank	variable	rank	variable
1	$IP\chi^2$	6	μ_2 track quality	11	$\mu_1 p_T$
2	LONGMAX1	7	$\mu_{not} p_T$	12	λ_{decay}
3	μ_1 cone isolation	8	μ_{not} track quality	13	$\mu_2 p_T$
4	μ_1 track quality	9	μ_2 cone isolation	14	τp_T
5	endvertex χ^2	10	μ_{not} cone isolation	15	VELOMAX1

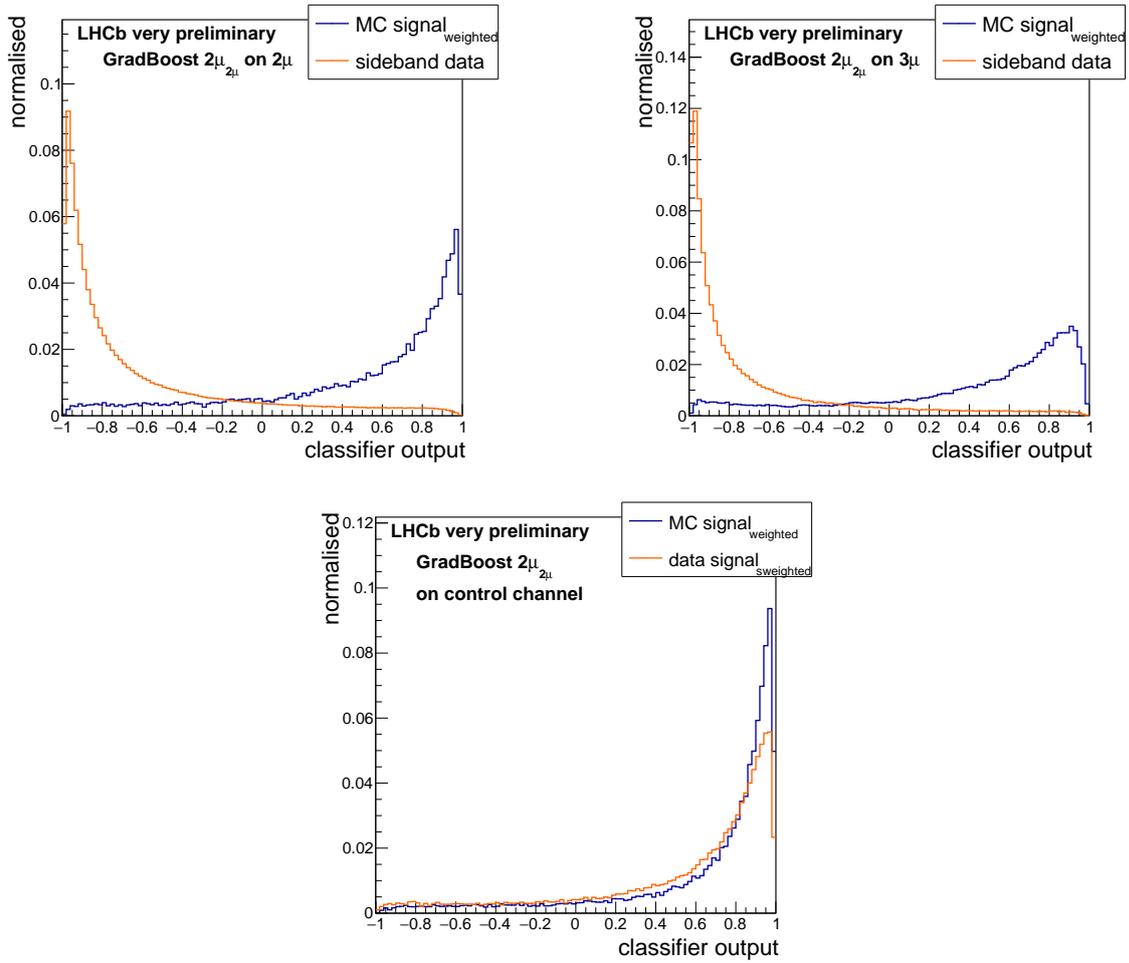


Figure 7.5: Output distributions of applying the $2\mu_{2\mu}$ GradBoost training to the 2μ and the 3μ sub-samples of the signal channel (above) and the 2018 control sample (below).

7.4 Correction in classifier output

After weighting the Monte Carlo events with *correction weights*, as described in Sec. 6.5, the distributions of Monte Carlo variables are adapted better to the data distributions. This is necessary, as the Monte Carlo events are used to calculate the necessary efficiencies for the final limit on the branching fraction of the $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ decay.

However, while the distributions are adapted in some of the most important variables, minor residual discrepancies remain. As can be seen in Fig. 7.5, the classifier output distribution is not yet the same for Monte Carlo and data. As a cut is applied on the classifier output later and the efficiency of this cut needs to be correctly predicted by the MC, it is necessary to remove this discrepancy. This is done by calculating another set of weights, called *BDT correction weights*, using the differences between the Monte Carlo distribution and the data distribution of the classifier output.

The strategy to correct for this discrepancy is the same as used in Sec. 6.5. The control channel $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ is used to compare the distributions. The data events are weighted with the *sWeights* obtained in the fit as seen in Fig. 6.3 in order to obtain the signal distribution. This signal distribution is then compared bin-by-bin to the distribution of the Monte Carlo, weighted with the *fraction weights* and *correction weights*. In each bin, *BDT correction weights* are calculated such, that the Monte Carlo distribution weighted with all three sets of weights approaches the data distribution. This can be seen in Fig. 7.6.

The weights are calculated and stored for all three years separately and applied to the MC signal sample.

To make sure other variable distributions are not influenced negatively by these weights, the five most important training variables as well as the VELOMAX1 muon isolation variable are plotted before and after the weighting with the *BDT correction weights* exemplary for the 2018 control samples. It is shown in Fig. 7.7 that the *BDT correction weights* only minimally influence the most important variables. While the distributions of the muon cone isolation and track quality are respectively under- and overestimated by the MC, the *BDT correction weights* do not noticeably change that. The only variable where a noticeable change can be seen when weighting with the *BDT correction weights* is the variable which is most important for the classifier, $\ln(IP\chi^2)$. The weighting boosts events with a higher $IP\chi^2$. As events with a low classifier output are boosted, that means that MC events with a high $IP\chi^2$ are sorted more often as background than data events. This is, however, not a big deviation and therefore not problematic.

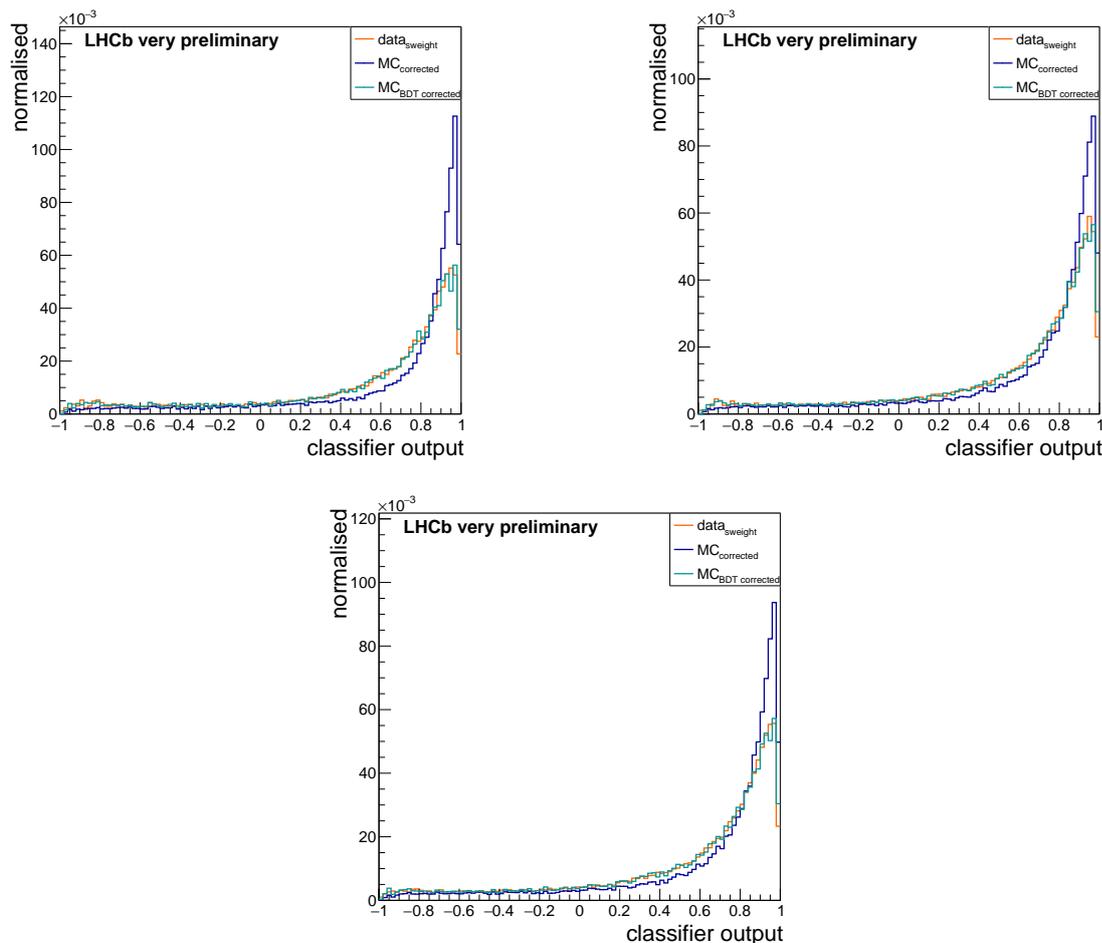


Figure 7.6: Classifier output distributions of control MC with and without *BDT correction weights* in comparison with *sWeighted* control data for 2016 (upper left), 2017 (upper right) and 2018 (bottom)

The *BDT correction weights* affect not only the variables themselves, but also slightly the correlation between them. This is visible in Fig. 7.8, which shows the correlation of 2018 MC control sample events between $\ln(IP\chi^2)$ and the LONGMAX1 isolation variable, as well as the correlation between the LONGMAX1 isolation variable and the μ_1 track quality.

By comparing the correlation before and after applying the *BDT correction weights*, as shown in the top (before) and the middle (after), it can be seen, that the correlation between the variables decreases slightly. This is further visualised in the bottom plots, which show the difference in correlation before and after applying the *BDT correction weights*.

With the *fraction weights*, the *correction weights* and the *BDT correction weights* stored for the signal channel and the *classifier output* from the $2\mu_{2\mu}$ setup with GradBoost training stored for each signal candidate, it is now possible to search for the optimal cut on the *classifier output* to separate signal and background.

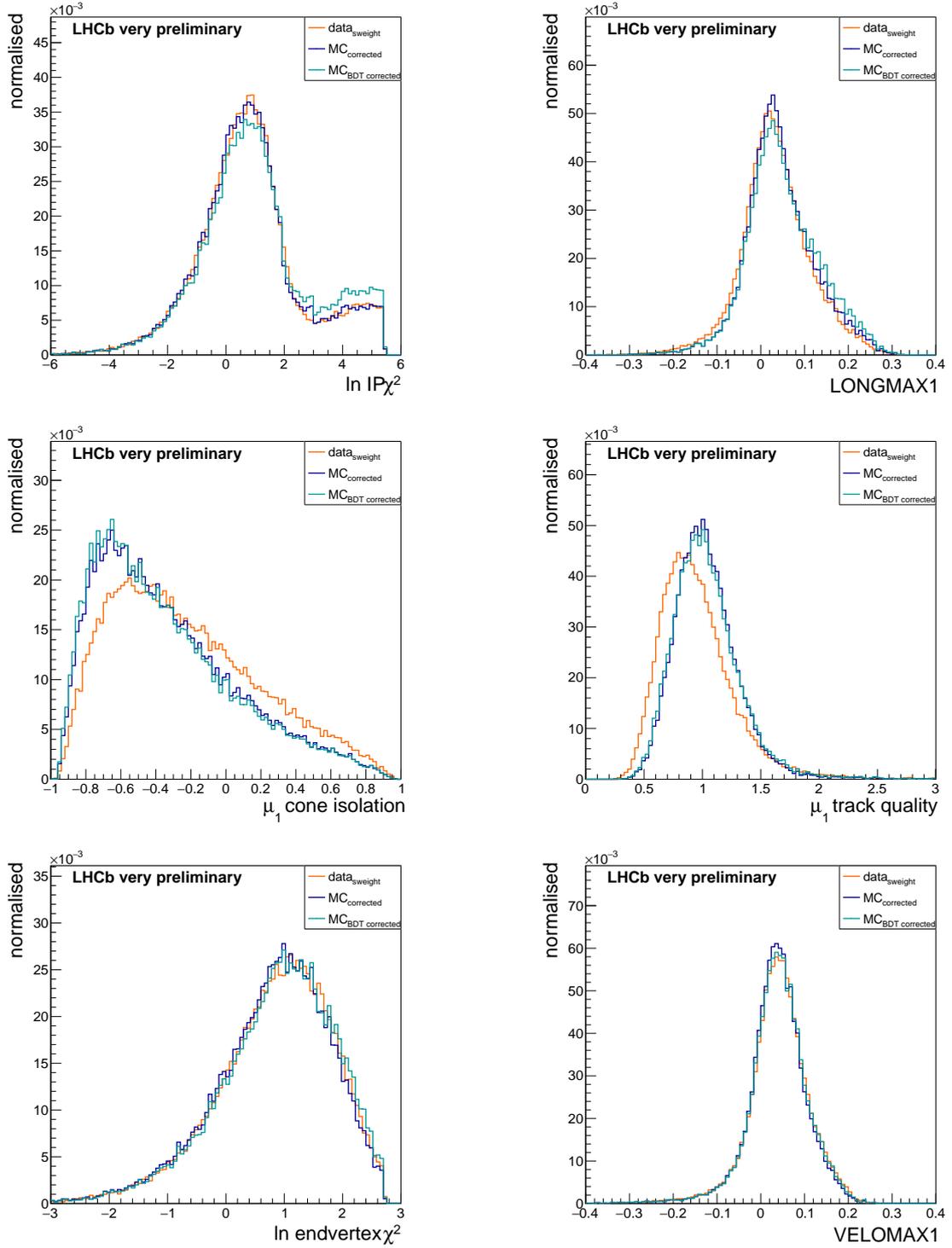


Figure 7.7: The distributions of the five most important training variables and the VELO-MAX1 muon isolation variable for the 2018 MC control samples before (purple) and after (light blue) weighting with the *BDT correction weights* in comparison to the same distributions of the *sWeighted* 2018 control data samples (orange).

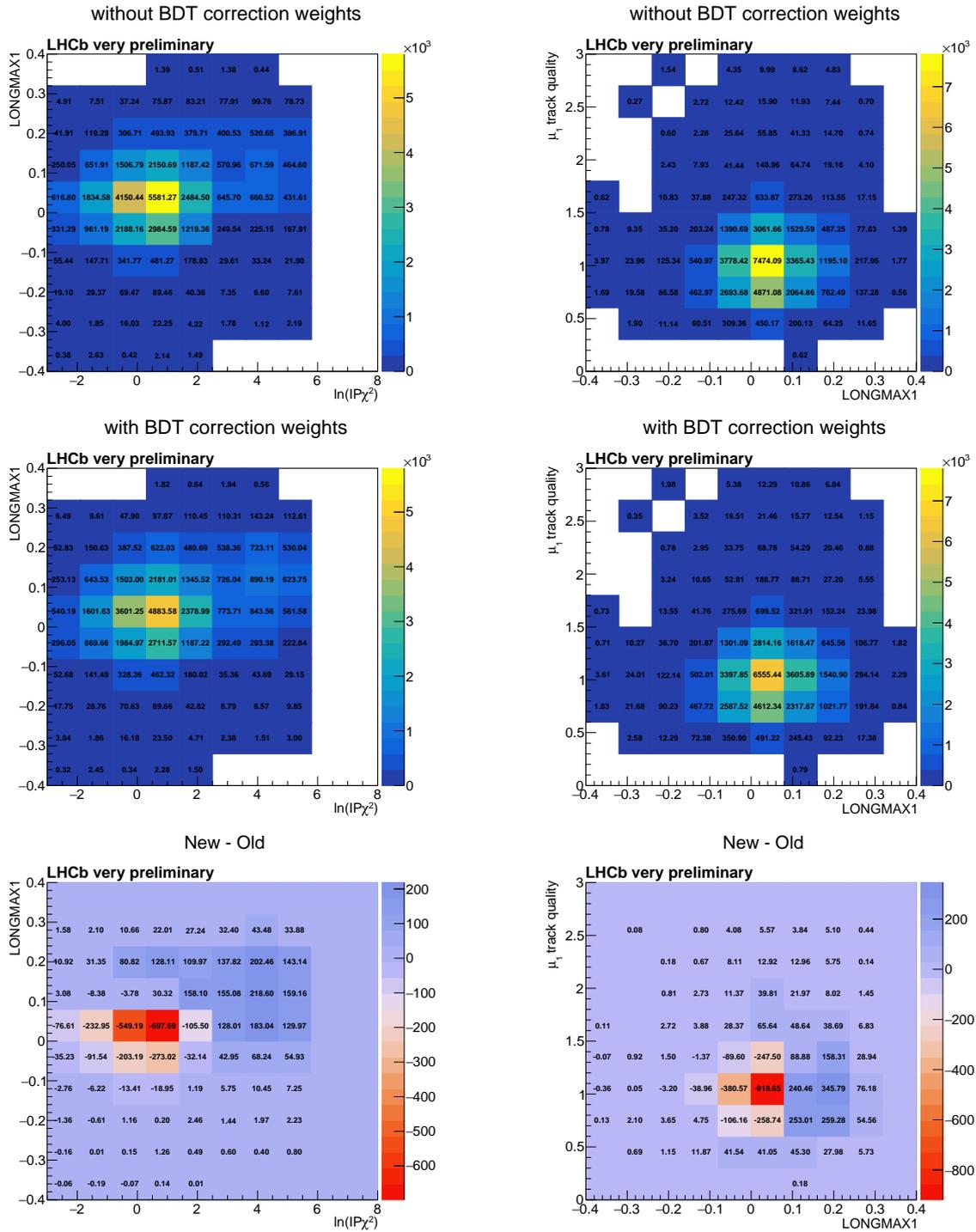


Figure 7.8: Correlation between $\ln(IP\chi^2)$ and LONGMAX1 (left) as well as between LONGMAX1 and μ_1 track quality (right) before (top) and after (middle) weighting with the *BDT correction weights*, as well as the difference between the histograms before and after (bottom).

7.5 Classifier performance validation

The final expected limit is calculated in six bins. These are the 2μ and 3μ sub-samples for each of the three years, separately. The efficiencies for the calculation of the limit are estimated on the weighted Monte Carlo samples for each bin separately, while the expected background is estimated by fitting the outer sidebands of the signal channel data in each of the six bins.

The classifier is applied to the final samples and the weights used on the Monte Carlo events. The performance, measured by the integral over the ROC curve, is used to confirm that the classifier is not overly biased towards a certain year or sub-sample, despite being trained on 2016 – 2018 signal against 2018 background.

The low number of Monte Carlo events for the 2μ sub-sample leads to fluctuations in the performance of this sub-sample. Nevertheless, Tab. 7.7 shows the robustness of the classifier for the different samples, years and magnet polarities.

Table 7.7: Performance (integral over the ROC curve) of the final classifier on each year, sample and magnet polarity.

year	magnet polarity	2μ	3μ
2018	down	0.913	0.916
2018	up	0.912	0.915
2017	down	0.908	0.916
2017	up	0.911	0.915
2016	down	0.901	0.912
2016	up	0.915	0.914

8 Expected limit

The aim of this thesis is to estimate an expected limit on the branching fraction of the decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$. The kinematically similar channel $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ is used as control channel to correct discrepancies in the Monte Carlo. Both channels have to pass a set of trigger requirements and are pre-selected to remove background. A multivariate analysis classifier is trained to obtain the optimal separation variable between the $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ signal and combinatorial background. This is done separately on two sub-samples, called 2μ and 3μ sub-sample. They are defined by the number of daughter candidates that are identified as muons by the muon system in the event and are mutually exclusive. The final optimised classifier is applied to all signal samples. This chapter describes the determination of the optimal separation cut on the classifier output and the calculation of its efficiency. The efficiency is used as input to calculate the single event sensitivity, α , which gives the proportionality factor between the measured number of events and the branching fraction. The single event sensitivity is used to determine an expected limit under the assumption that no signal event is measured. The acceptance, retention rate and reconstruction efficiency needed to compute α are calculated by other members of the group. The previous measurement limit, based only on *Run 1* data, is extrapolated to the current statistics and compared to the estimation.

8.1 Punzi Figure of Merit

The multivariate analysis training on the 2μ sub-sample of the signal channel produces a classifier that is applied to all signal and control samples. The aim is to find the optimal cut on the *classifier output*, which leads to the highest sensitivity of the analysis. The *Punzi Figure of Merit* is used to determine the optimal cut in six bins, namely two sub-samples and three years each.

The Punzi Figure of Merit is calculated using the following equation:

$$P = \frac{\epsilon_S}{a/2 + \sqrt{B}}, \quad (8.1)$$

where $\epsilon_S = N_{cut}/N_{tot}$ is the signal efficiency. Of N_{tot} signal candidates before the cut, N_{cut} pass the cut. B is the number of background candidates left after applying the cut. All these yields are determined in the signal region of the signal channel. The variable a , corresponding to the aimed significance of the observation, is set to three.

While the signal efficiency can be determined from Monte Carlo, it is more difficult to obtain the number of expected background events in the signal region for

a certain cut, as the signal region of the signal data sample is still blinded. It is obtained by fitting the outer sidebands of the signal data samples and extrapolating the number of expected background candidates from the fit. In order to have a smooth distribution to fit, it is necessary to apply particle identification conditions on the daughter particles. For all good muons, a minimum value for the particle identification variable $ProbNNmu$ is required. As this is not defined for non-good muons, a minimum requirement for the $RichDLLmu$ variable of the non-good muons is set. This only concerns the 2μ sub-sample, as all muons in the 3μ sub-sample are good muons. The distribution of the $ProbNNmu$ and $RichDLLmu$ variable for good and non-good muons are shown in Fig. 8.1.

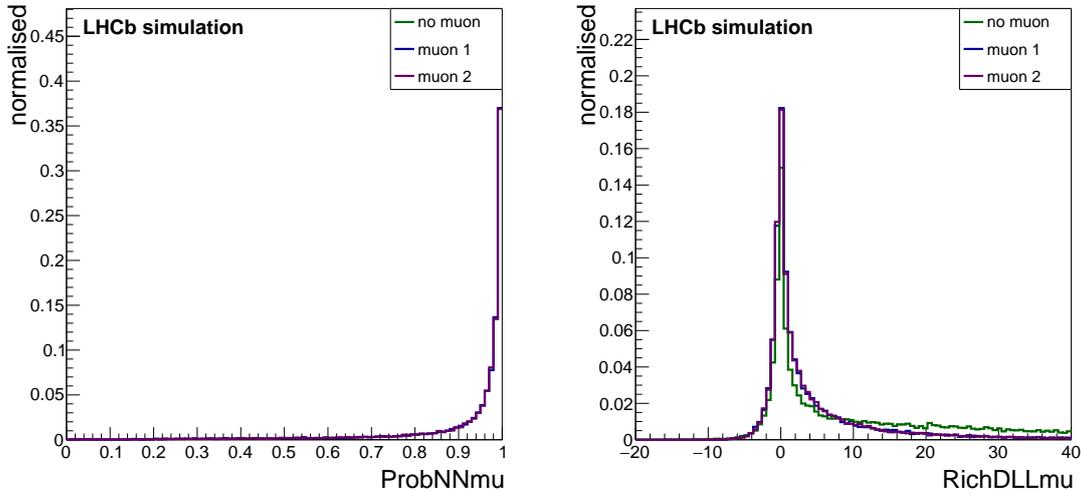


Figure 8.1: Monte Carlo distribution of the $ProbNNmu$ and $RichDLLmu$ variables for the 2μ sample muons. The $ProbNNmu$ distribution for the non-good muon is not shown, as it is not defined.

Figure 8.2 shows the effect of the $ProbNNmu$ cut exemplary on the 2018 3μ sub-sample. As can be seen, the cut effectively reduces background and removes a bump at the right end of the left outer sideband. This bump is most likely a peak produced by $D^\pm \rightarrow K^\pm \mu^+ \mu^-$ decays where one K is misidentified as a μ , moving the peak by the mass difference of ≈ 390 MeV. Additionally, the background, which before has an upwards slope, is much flatter after the cut. The raising background is caused by misidentified candidates from semileptonic $D_s^\pm \rightarrow \mu^\pm X$ decays, where X can be any possible combination of particles.

The effect of the particle identification cut on the 2018 2μ sub-sample can be seen exemplary in Fig. 8.3. It has a similar effect as the particle identification cut on the 3μ sub-sample. The main difference is the overwhelming surplus of data background before the cut, prompting a much stricter cut on the $RichDLLmu$ variable of the non-good muon.

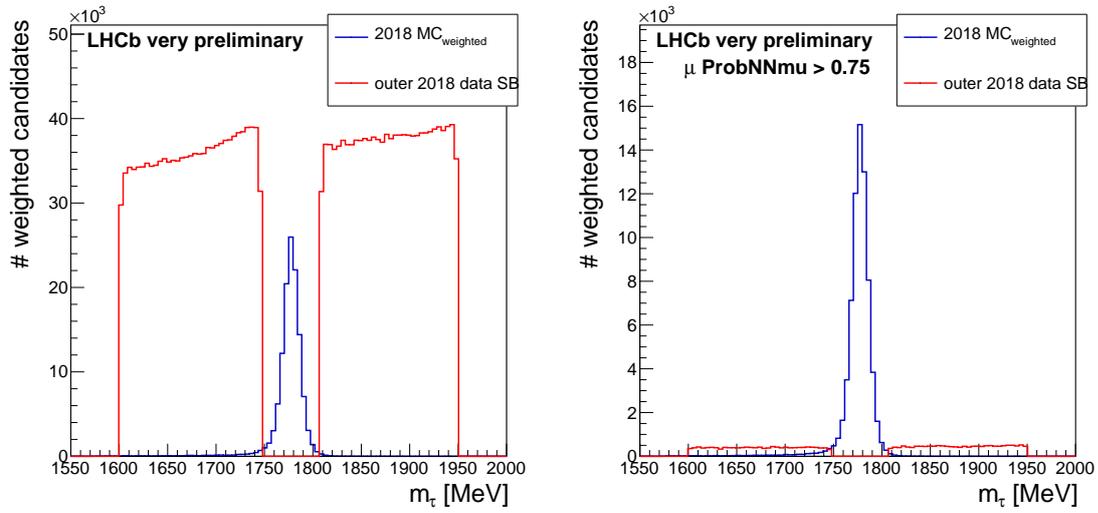


Figure 8.2: 2018 Monte Carlo and outer sidebands data distribution before (left) and after (right) the $ProbNN\mu$ cut is applied to the 2018 3μ sub-sample.

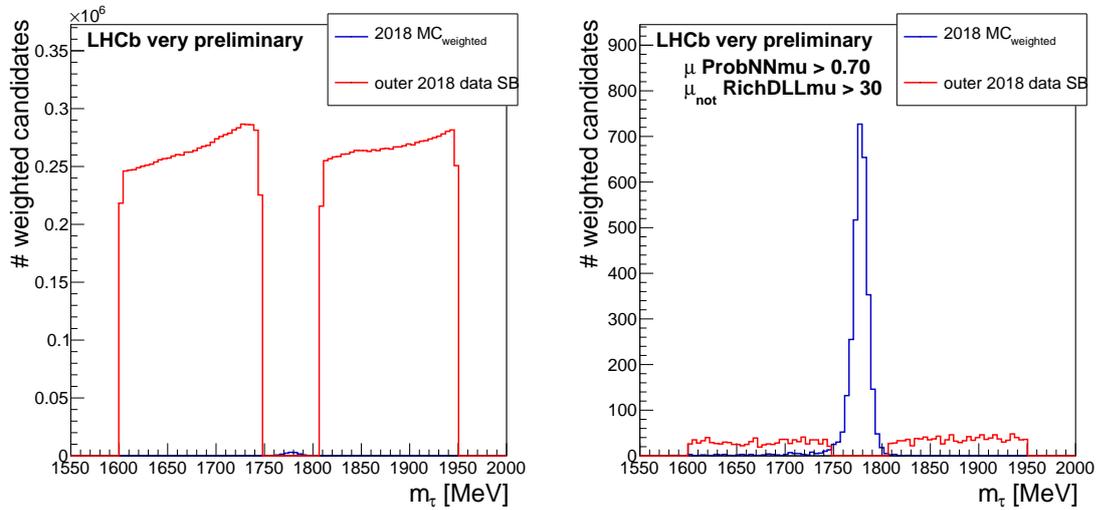


Figure 8.3: 2018 Monte Carlo and outer sidebands data distribution before (left) and after (after) the $ProbNN\mu$ and $RichDLL\mu$ cut is applied to the 2018 2μ sub-sample.

Together with the particle identification cut, a cut is also applied on the classifier output of the GradBoost $2\mu_{2\mu}$ training setting. For the 3μ sub-sample, this cut is optimised for each year two-dimensionally together with the *ProbNNmu* cut. As the result proves a very low correlation between the *ProbNNmu* variable and the classifier output, for the 2μ sub-sample the *ProbNNmu* cut on the two good muons is optimised first. Subsequently, the *RichDLLmu* cut on the non-good muons is optimised two-dimensionally together with the cut on the classifier output. All cuts are optimised by maximising the Punzi Figure of Merit. The distribution of the Punzi Figure of Merit for the two-dimensional cut optimisations can be found in Fig. 8.4. Table 8.1 shows the optimised cuts for all six bins.

Table 8.1: Optimised particle identification and classifier output cuts for all six bins.

	2μ			3μ		
	2016	2017	2018	2016	2017	2018
μ <i>ProbNNmu</i> >	0.6	0.7	0.7	0.75	0.8	0.75
μ_{not} <i>RichDLLmu</i> >	26	32	30	–	–	–
classifier output >	0.65	0.4	0.75	0.5	0.5	0.5

A sum of two exponentials is fitted to the outer sidebands of the signal data in order to determine the expected background in the signal region. The effect of the cut on the outer sidebands and the resulting fits can be found in Fig. 8.5. These exponentials are meant to model the combinatorial background, resulting in a falling exponential curve, and the tail of a Gaussian distribution resulting from leftover misidentified semileptonic $D_s^\pm \rightarrow \mu^\pm X$ decays. The expected background is determined in the blinded signal window of (1756.86 – 1796.86) MeV. The estimated expected background yields, B_{exp} , in each bin together with the particle identification and classifier output cut efficiency, $\epsilon_{ID, BDT}$, as well as the efficiency of the mass cut to the signal window, ϵ_{mass} are given in Tab. 8.2.

Table 8.2: Particle identification cut efficiencies and classifier output cut efficiency, $\epsilon_{ID, BDT}$, mass cut efficiency, ϵ_{mass} , and expected background, B_{exp} , for each of the six bins.

	2μ		
	2016	2017	2018
ϵ_{mass}	$0.93 \pm (3.63 \times 10^{-3})$	$0.93 \pm (2.22 \times 10^{-3})$	$0.93 \pm (2.37 \times 10^{-3})$
$\epsilon_{ID, BDT}$	$0.03 \pm (2.56 \times 10^{-3})$	$0.12 \pm (2.60 \times 10^{-3})$	$0.09 \pm (2.45 \times 10^{-3})$
B_{exp}	2.02 ± 1.42	14.34 ± 3.79	6.21 ± 2.49
	3μ		
	2016	2017	2018
ϵ_{mass}	$0.92 \pm (6.99 \times 10^{-4})$	$0.93 \pm (7.01 \times 10^{-4})$	$0.92 \pm (7.62 \times 10^{-4})$
$\epsilon_{ID, BDT}$	$0.33 \pm (1.26 \times 10^{-3})$	$0.31 \pm (1.29 \times 10^{-3})$	$0.35 \pm (1.41 \times 10^{-3})$
B_{exp}	181.02 ± 13.45	154.73 ± 12.44	201.95 ± 14.21

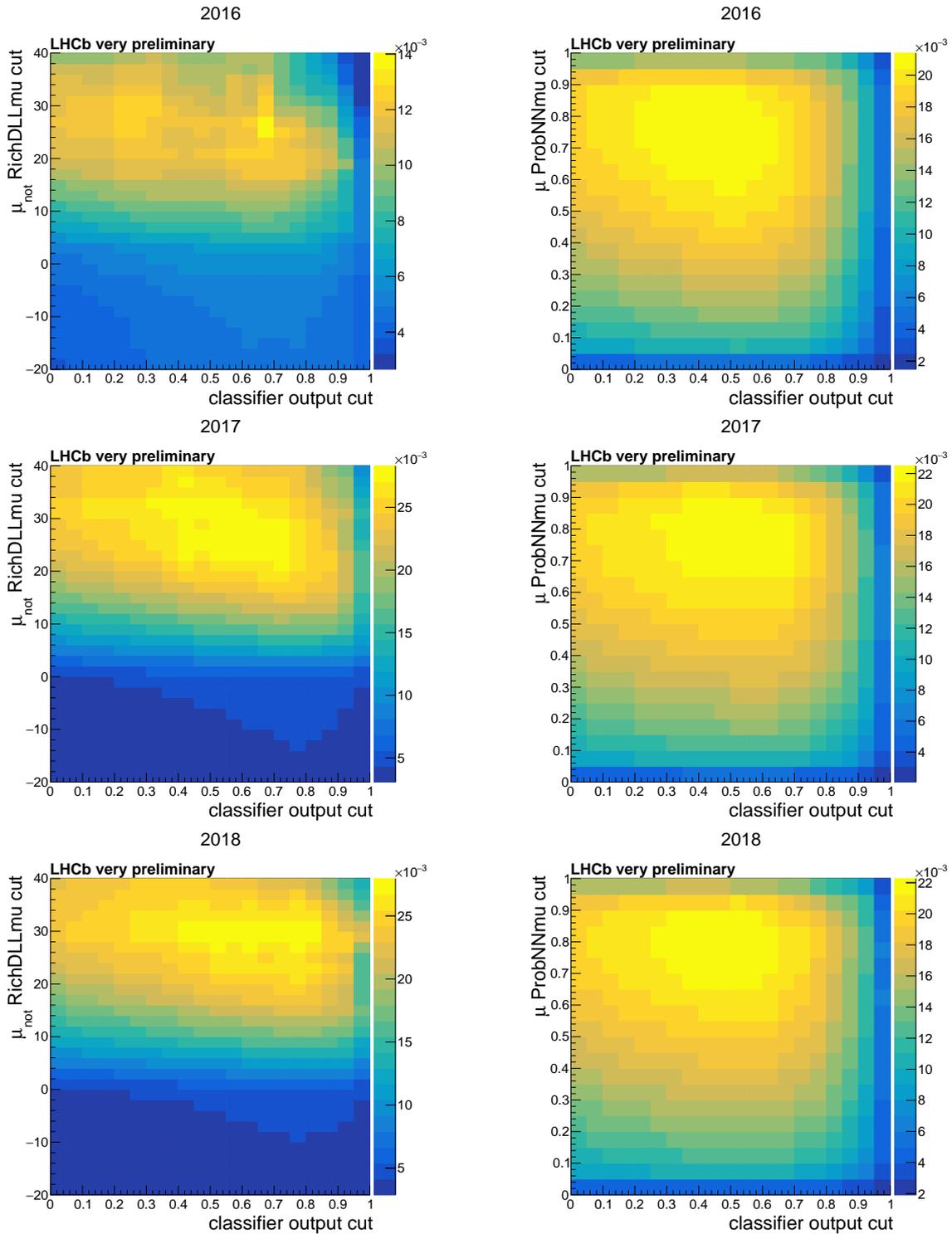


Figure 8.4: Distribution of the Punzi Figure of Merit for optimising the *RichDLLmu* cut together with the classifier output cut for the 2μ sub-sample (left) and for optimising the *ProbNNmu* cut together with the classifier output cut for the 3μ sub-sample (right) for 2016 (top), 2017 (middle) and 2018 (bottom).

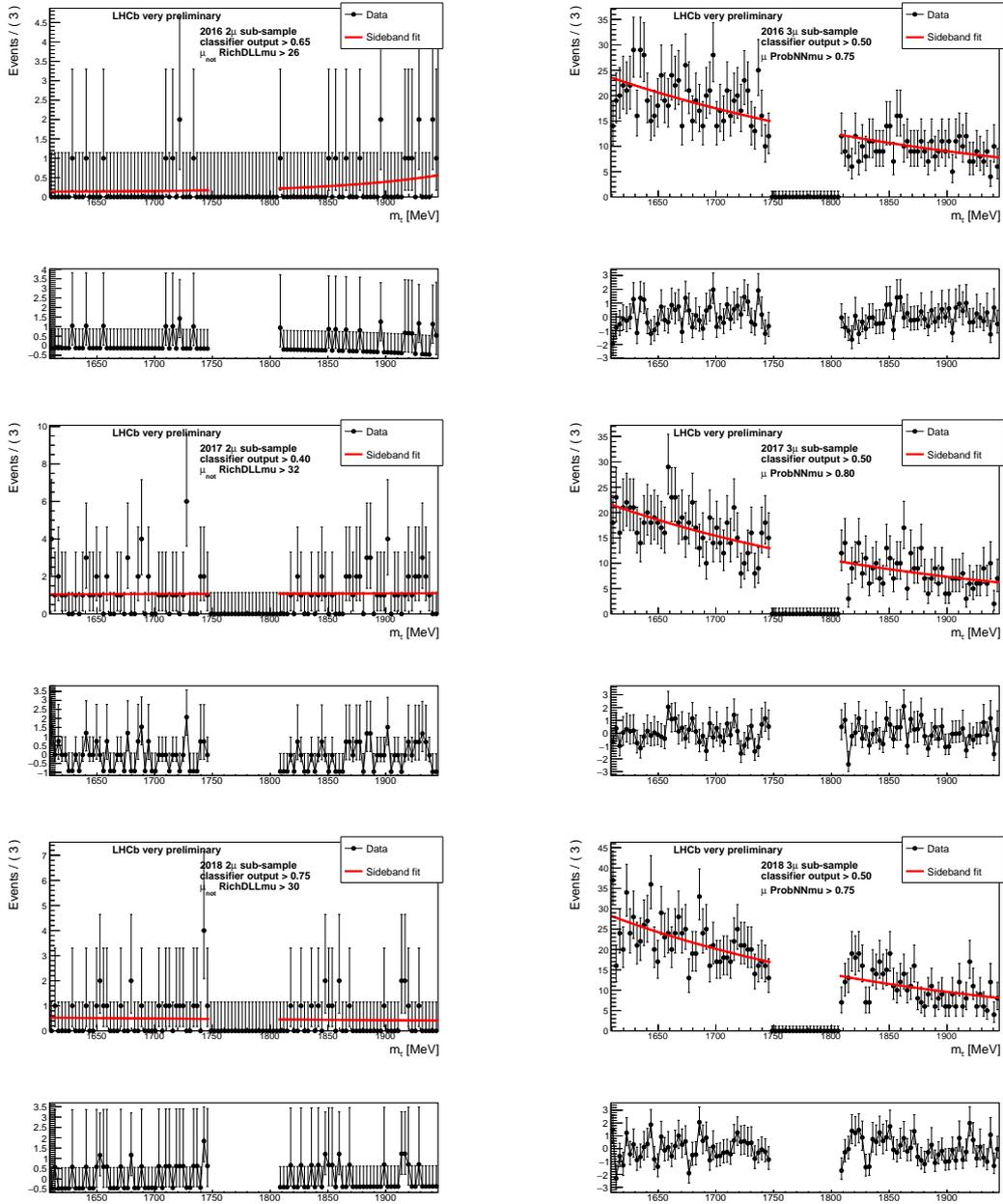


Figure 8.5: Fit to the outer sidebands of the signal data samples to estimate the expected background in the signal range for all six bins: 2 μ sub-sample (left) and 3 μ sub-sample (right) in the years 2016 (top), 2017 (middle) and 2018 (bottom).

8.2 Limit extrapolation

Using the signal efficiencies of the optimised cuts, the single event sensitivity, α , can be determined. It can be used to estimate an expected limit for the case of no events observed in the signal region. The complete formula for the calculation of α can be found in eq. 8.6. The upper limit (UL) that was expected for the *Run 1* dataset is at

$$UL_{Run1} = 5.0 (6.1) \times 10^{-8} \text{ 90\% (95\%)} [6] \quad (8.2)$$

The final limit was slightly better than that, as less background than expected was measured.

This expected limit of *Run 1* is extrapolated to the current statistics, to get an estimation of what value can be expected from the current statistics. This value is then compared to the expected limit calculated from the single event sensitivity.

The branching fraction \mathcal{B} is proportional to $\frac{N_{sig}}{N_{norm}}$, where N_{sig} is the number of observed signal events in the signal region of the signal channel, while N_{norm} is the number of observed signal events in the signal region of the reference channel. As reference channel, the $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ decay channel is used. The MC control and data samples used for the correction of the Monte Carlo are used with the same trigger requirements, weights and preselection cuts as reference MC and data samples.

The 90% or 95% confidence level defines the upper limit (UL) to the interval around the assumed branching fraction of the null hypothesis, $\mathcal{B} = 0$, such that the integrated probability for the true value to lie within this interval is 90% or 95%. This means that the probability for the true branching fraction to be higher than the limit is only 10% or 5%. The limit on the branching fraction is proportional to $\frac{N_{sig}^{UL}}{N_{norm}}$, where N_{sig}^{UL} is the upper limit on the number of observed signal events. It is proportional to the uncertainty on the number of expected background events, B_{exp} , as a higher uncertainty on the number of expected background events means the limit on the observed events that could be signal must be set higher. As a Poissonian distributed variable, the uncertainty on the number of expected background events is its square root, $\sqrt{B_{exp}}$. Therefore, the upper limit UL is proportional to:

$$UL \propto \frac{\sqrt{B_{exp}}}{N_{norm}} \quad (8.3)$$

The number of both the expected number of background events in the signal channel and the expected number of signal events in the reference channel are directly proportional to the integrated luminosity and the production cross section of the relevant mother particles. It is assumed that the cross sections for the combinatorial background processes, $\sigma_{comb\ bkg}$, scale with the same factor as the cross section for the reference channel, σ_{D_s} , as the main combinatorial background comes from

semileptonic D_s decays, so that σ_{D_s} can be used:

$$UL \propto \frac{\sqrt{\int_{Run} \mathcal{L} dt \cdot \sigma_{comb bkg}}}{\int_{Run} \mathcal{L} dt \cdot \sigma_{D_s}} \approx \frac{\sqrt{\int_{Run} \mathcal{L} dt \cdot \sigma_{D_s}}}{\int_{Run} \mathcal{L} dt \cdot \sigma_{D_s}} = \frac{1}{\sqrt{\int_{Run} \mathcal{L} dt \cdot \sigma_{D_s}}} \quad (8.4)$$

As all other values for the calculation of the limit are independent from luminosity and energy and therefore the same for both runs, it is possible to extrapolate the obtained limit for *Run 1* to the luminosity and energy of *Run 2*:

$$\begin{aligned} UL_{Run2,extrapolated} &= UL_{Run1} \times \frac{\sqrt{N_{norm,Run1}}}{\sqrt{N_{norm,Run2}}} = UL_{Run1} \times \frac{\sqrt{\int_{Run1} \mathcal{L} dt \cdot \sigma_{Run1}}}{\sqrt{\int_{Run2} \mathcal{L} dt \cdot \sigma_{Run2}}} \\ &= UL_{Run1} \times \frac{\sqrt{\int_{2011} \mathcal{L} dt \cdot \sigma_{2011} + \int_{2012} \mathcal{L} dt \cdot \sigma_{2012}}}{\sqrt{\int_{Run2} \mathcal{L} dt \cdot \sigma_{Run2}}} \end{aligned}$$

Table 8.3: D_s production cross sections and luminosities for *Run 1* and *Run 2*. The D_s cross section at 8 TeV is scaled from the cross section at 7 TeV using a factor of 8/7, equivalent to [6].

	2011	2012	2016 - 2018
$\int \mathcal{L} dt$	1.11 fb ⁻¹	2.08 fb ⁻¹	5.57 fb ⁻¹
σ_{D_s}	197 ± 31 μb [53]	225 ± 46 μb	353 ± 76 μb [54]

Using the values in Tab. 8.3, an expected limit for *Run 2* given the circumstances of *Run 1* can be extrapolated to:

$$UL_{Run2,extrapolated} = 2.95 (3.60) \times 10^{-8} \text{ 90\% (95\%)}$$

8.3 Single event sensitivity

To estimate the branching fraction \mathcal{B} for a decay for which N_{sig} signal events are observed, the *single event sensitivity*, α , is needed. It represents the proportionality factor between the number of *observed* signal events, N_{sig} , and the branching fraction, \mathcal{B} .

$$\mathcal{B}_{sig} = \alpha \times N_{sig} \quad (8.5)$$

The single event sensitivity can be determined using equation 8.6, where:

$$\begin{aligned} \mathcal{B}(\tau^- \rightarrow \mu^+ \mu^- \mu^+) &= \mathcal{B}(D_s^- \rightarrow \phi(\mu^+ \mu^-) \pi^-) \times \frac{f_{D_s}^\tau}{\mathcal{B}(D_s^- \rightarrow \tau^- \bar{\nu}_\tau)} \times \frac{\epsilon_{D_s}}{\epsilon_\tau} \times \frac{N_{sig}}{N_{norm}} \\ &= \alpha \times N_{sig} \end{aligned} \quad (8.6)$$

- $\mathcal{B}(D_s^- \rightarrow \phi(\mu^+\mu^-)\pi^-) = (1.30 \pm 0.10) \times 10^{-5}$ (see Sec. 4.3) is the branching fraction for the reference channel
- $f_{D_s}^\tau = (78.88 \pm 3.77)\%$ is the fraction of all produced τ^\pm that originated in the decay of a D_s . It equals the sum of f_{prod} from channel (i) and (ii) in Tab. 6.4.
- $\mathcal{B}(D_s^- \rightarrow \tau^-\bar{\nu}_\tau) = (5.48 \pm 0.23)\%$ [8] is the branching fraction of D_s into τ with the production of a neutrino
- ϵ_{D_s} is the overall efficiency for reference channel signal events
- ϵ_τ is the overall efficiency for the signal channel signal events, including the efficiencies ϵ_{mass} and $\epsilon_{ID,BDT}$ given in Tab. 8.2
- N_{norm} is the number of observed reference channel events in the signal region of $m_{D_s} \pm 50$ MeV around the D_s mass, $m_{D_s} = 1776.86$ MeV, from the reference data sample, determined from the fits in Fig. 6.3

The overall efficiencies of both signal channel sub-samples and of the control channel are an important input parameter for the calculation of α . They determine the fraction of events that are detected within the LHCb acceptance and pass all requirements for the channel of interest. A higher efficiency means more sensitivity of the experiment. As efficiencies are calculated on MC events, it is important that the MC is correctly simulating the data. Different efficiencies are included in the calculation of the overall efficiencies. An overview over them can be found in Tab. 8.4.

Table 8.4: Efficiencies for the 2μ and 3μ sub-sample of the signal channel and for the reference channel. The acceptance, ϵ_{acc} , the retention rate, ϵ_{ret} , and the reconstruction efficiency, ϵ_{reco} , are calculated by other members of the group. The trigger efficiency, ϵ_{trig} , is taken from Tab. 6.10. The preselection efficiency, ϵ_{presel} , for the signal channel is taken from Tab. 6.12. The preselection efficiency, ϵ_{presel} , the particle identification efficiency, ϵ_{ID} , and the mass cut efficiency, ϵ_{mass} , for the reference channel are taken from Tab. 6.15. The particle identification and classifier output cut efficiencies $\epsilon_{ID, BDT}$, as well as the mass cut efficiency, ϵ_{mass} , for the signal channel are taken from Tab. 8.2.

	2016	2μ 2017	2018
ϵ_{acc}	$0.11 \pm (4.30 \times 10^{-3})$	$0.11 \pm (4.30 \times 10^{-3})$	$0.11 \pm (4.30 \times 10^{-3})$
ϵ_{ret}	$0.57 \pm (3.64 \times 10^{-4})$	$0.57 \pm (3.64 \times 10^{-4})$	$0.57 \pm (3.66 \times 10^{-4})$
ϵ_{reco}	$0.40 \pm (4.76 \times 10^{-4})$	$0.40 \pm (4.76 \times 10^{-4})$	$0.40 \pm (4.79 \times 10^{-4})$
ϵ_{presel}	$0.22 \pm (5.56 \times 10^{-4})$	$0.22 \pm (5.23 \times 10^{-4})$	$0.22 \pm (5.66 \times 10^{-4})$
ϵ_{trig}	$0.07 \pm (1.02 \times 10^{-3})$	$0.22 \pm (1.70 \times 10^{-3})$	$0.19 \pm (1.58 \times 10^{-3})$
ϵ_{mass}	$0.93 \pm (3.63 \times 10^{-3})$	$0.93 \pm (2.22 \times 10^{-3})$	$0.93 \pm (2.37 \times 10^{-3})$
$\epsilon_{ID, BDT}$	$0.03 \pm (2.56 \times 10^{-3})$	$0.12 \pm (2.60 \times 10^{-3})$	$0.09 \pm (2.45 \times 10^{-3})$
ϵ_{τ}	$(1.17 \pm 0.25) \times 10^{-5}$	$(1.31 \pm 0.27) \times 10^{-4}$	$(8.40 \pm 1.71) \times 10^{-5}$
	2016	3μ 2017	2018
ϵ_{acc}	$0.11 \pm (4.30 \times 10^{-3})$	$0.11 \pm (4.30 \times 10^{-3})$	$0.11 \pm (4.30 \times 10^{-3})$
ϵ_{ret}	$0.57 \pm (3.64 \times 10^{-4})$	$0.57 \pm (3.64 \times 10^{-4})$	$0.57 \pm (3.66 \times 10^{-4})$
ϵ_{reco}	$0.40 \pm (4.76 \times 10^{-4})$	$0.40 \pm (4.76 \times 10^{-4})$	$0.40 \pm (4.79 \times 10^{-4})$
ϵ_{presel}	$0.54 \pm (9.53 \times 10^{-4})$	$0.54 \pm (9.61 \times 10^{-4})$	$0.53 \pm (9.44 \times 10^{-4})$
ϵ_{trig}	$0.59 \pm (1.29 \times 10^{-3})$	$0.67 \pm (1.24 \times 10^{-3})$	$0.58 \pm (1.28 \times 10^{-3})$
ϵ_{mass}	$0.92 \pm (6.99 \times 10^{-4})$	$0.93 \pm (7.01 \times 10^{-4})$	$0.92 \pm (7.62 \times 10^{-4})$
$\epsilon_{ID, BDT}$	$0.33 \pm (1.26 \times 10^{-3})$	$0.31 \pm (1.29 \times 10^{-3})$	$0.35 \pm (1.41 \times 10^{-3})$
ϵ_{τ}	$(2.29 \pm 0.46) \times 10^{-3}$	$(2.52 \pm 0.51) \times 10^{-3}$	$(2.39 \pm 0.48) \times 10^{-3}$
	2016	reference 2017	2018
ϵ_{acc}	$0.12 \pm (3.31 \times 10^{-3})$	$0.12 \pm (3.31 \times 10^{-3})$	$0.12 \pm (3.31 \times 10^{-3})$
ϵ_{ret}	$0.48 \pm (3.25 \times 10^{-4})$	$0.48 \pm (3.18 \times 10^{-4})$	$0.47 \pm (3.38 \times 10^{-4})$
ϵ_{reco}	$0.17 \pm (3.54 \times 10^{-4})$	$0.17 \pm (3.47 \times 10^{-4})$	$0.17 \pm (3.70 \times 10^{-4})$
ϵ_{presel}	$0.92 \pm (6.46 \times 10^{-4})$	$0.92 \pm (6.27 \times 10^{-4})$	$0.92 \pm (6.71 \times 10^{-4})$
ϵ_{trig}	$0.15 \pm (8.93 \times 10^{-3})$	$0.32 \pm (1.14 \times 10^{-3})$	$0.28 \pm (1.17 \times 10^{-3})$
ϵ_{mass}	$0.99 \pm (7.31 \times 10^{-4})$	$0.99 \pm (4.92 \times 10^{-4})$	$0.99 \pm (5.58 \times 10^{-4})$
ϵ_{ID}	$0.97 \pm (1.13 \times 10^{-3})$	$0.96 \pm (8.51 \times 10^{-4})$	$0.96 \pm (9.58 \times 10^{-4})$
ϵ_{D_s}	$(1.27 \pm 0.33) \times 10^{-3}$	$(2.72 \pm 0.70) \times 10^{-3}$	$(2.36 \pm 0.60) \times 10^{-3}$

The combined efficiencies are shown in Tab. 8.5 for each signal channel bin and in Tab. 8.6 for each year of the reference channel.

Table 8.5: The final efficiencies ϵ_τ for each of the six bins.

	2016	2017	2018
2μ	$(1.17 \pm 0.25) \times 10^{-5}$	$(1.31 \pm 0.27) \times 10^{-4}$	$(8.40 \pm 1.71) \times 10^{-5}$
3μ	$(2.29 \pm 0.46) \times 10^{-3}$	$(2.52 \pm 0.51) \times 10^{-3}$	$(2.39 \pm 0.48) \times 10^{-3}$

Table 8.6: The final efficiencies ϵ_{D_s} for the reference channel.

	2016	2017	2018
	$(1.27 \pm 0.33) \times 10^{-3}$	$(2.72 \pm 0.70) \times 10^{-3}$	$(2.36 \pm 0.60) \times 10^{-3}$

The ratio $\frac{\epsilon_{D_s}}{\epsilon_\tau}$ needs to be corrected by a correction factor stemming from the track reconstruction efficiency, which is also calculated by other members of the group. The corrected efficiency ratios are given in Tab. 8.7.

Table 8.7: The final corrected efficiency ratios $\epsilon_{D_s}/\epsilon_\tau$ for each of the six bins.

	2016	2017	2018
2μ	109.52 ± 36.69	(20.73 ± 6.78)	28.08 ± 9.19
3μ	0.56 ± 0.18	1.08 ± 0.35	0.99 ± 0.32

Table 8.8: Number of observed reference channel events for each year using the fit in Fig. 6.3.

	2016	2017	2018
	35909 ± 250	80289 ± 352	98070 ± 391

Table 8.8 shows the number of observed reference channel events for each year. Together with the constants given below Eq. 8.6 and the efficiency ratios from Tab. 8.7, the single event sensitivity α can now be determined for each bin. The final single event sensitivities are given in Tab. 8.9.

Table 8.9: The final single event sensitivities α for each of the six bins.

	2016	2017	2018
2μ	$(5.72 \pm 0.78) \times 10^{-7}$	$(4.84 \pm 0.55) \times 10^{-8}$	$(5.37 \pm 0.62) \times 10^{-8}$
3μ	$(2.91 \pm 0.33) \times 10^{-9}$	$(2.52 \pm 0.28) \times 10^{-9}$	$(1.88 \pm 0.21) \times 10^{-9}$

Assuming the measured number of events in the signal region equals exactly the expected background given in Tab. 8.2, the CL_S method can be used to set an upper limit on the branching fraction for each bin. These limits are shown in Tab. 8.10.

Table 8.10: Upper limit on the number of signal events and the limit on the branching fraction at 90% and 95% confidence level assuming N_{exp} events are observed in the signal region.

	2μ		
	2016	2017	2018
N_{exp}	3	14	6
90%: $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq$	2.19×10^{-6}	3.81×10^{-7}	3.24×10^{-7}
95%: $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq$	3.00×10^{-6}	4.75×10^{-7}	4.02×10^{-7}
	3μ		
	2016	2017	2018
N_{exp}	181	155	202
90%: $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq$	6.65×10^{-8}	5.29×10^{-8}	4.55×10^{-8}
95%: $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq$	8.09×10^{-8}	6.51×10^{-8}	5.49×10^{-8}

The CL_S method can also be used to combine the information from all six bins in one combined expected limit for *Run 2*:

$$\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq 2.96 (3.58) \times 10^{-8} \text{ 90\% (95\%)}$$

The likelihood distributions used for the calculation of the confidence levels for these branching fractions are shown for the background-only and for the background+signal hypothesis in Fig. 8.6.

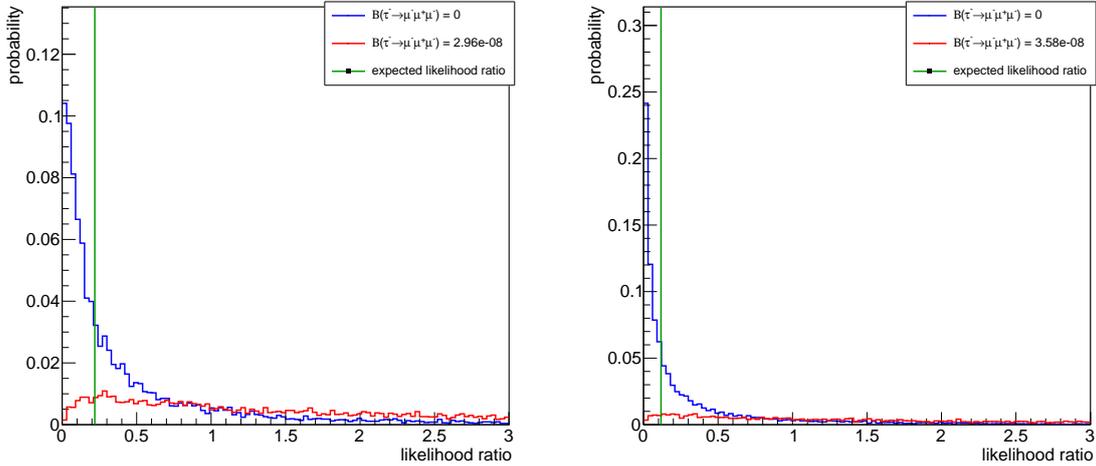


Figure 8.6: Likelihood distributions used for the calculation of the 90% (left) and 95% (right) confidence level.

8.4 Results

The upper limit set on the $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ decay in *Run 1*, was at

$$UL_{Run\ 1, observed} = 4.6 (5.6) \times 10^{-8} \text{ 90\% (95\%)}.$$

Due to fluctuations, this already presented a slight improvement to the expected limit calculated for *Run 1*, which was at:

$$UL_{Run\ 1, expected} = 5.0 (6.1) \times 10^{-8} \text{ 90\% (95\%)}.$$

Using only the naive extrapolation of the increased luminosity and cross section, the expected limit for *Run 2* should be at

$$UL_{Run\ 2, extrapolated} = 2.95 (3.60) \times 10^{-8} \text{ 90\% (95\%)}.$$

The expected limit calculated in this thesis is already very close to this limit, even exceeding it at 95% confidence level.

$$UL_{Run\ 2, expected} = 2.96 (3.58) \times 10^{-8} \text{ 90\% (95\%)}.$$

This happens despite the fact that during *Run 1*, the analysis was done in more than 30 bins of the particle identification variable *ProbNNmu* and the classifier output, leading to an improvement of the combined limit of almost 20% [43]. Despite not using this binning, the expected limit could be reached with this analysis thanks to the addition of the 2μ sub-sample and a noticeable improvement of the classifier of about 10%, in part through the new muon isolation variables used in this thesis. With a much higher luminosity of 33.2 fb^{-1} , the CMS detector observed a significantly less tight limit on the 2016 dataset produced by the LHC of:

$$UL_{CMS, 2016} = 8.0 \times 10^{-8} \text{ 90\% [55]}$$

With the binning method applied to this analysis together with further corrections to the single event sensitivity and some possible improvements on the output classifier for the 3μ sub-sample, as well as the combination of the *Run 1* limit with the *Run 2* limit, an improvement of at least 30% upon the expected limit calculated in this thesis can be expected to be observed in this analysis. This is expected to lead to an observed limit tighter than the current worldwide best limit calculated by the BELLE experiment:

$$UL_{best\ limit\ (BELLE)} = 2.1 \times 10^{-8} \text{ 90\%}$$

9 Conclusion and Outlook

This thesis estimates an expected limit on the branching fraction of the decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ using data collected by the LHCb experiment in proton-proton collisions at the LHC accelerator at CERN from the years 2016 – 2018 at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ and a total integrated luminosity of 5.57 fb^{-1} . The decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ violates lepton flavour conservation and is only predicted in the extended Standard Model of particle physics including neutrino oscillations with a very low branching fraction of $\mathcal{O}(10^{-55})$. However, many theories beyond the Standard Model allow for this decay with a branching fraction at the level of $\mathcal{O}(10^{-10} - 10^{-7})$, making it an interesting decay to probe for new physics.

In this thesis, it is outlined how data and simulated Monte Carlo samples of both the $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ signal channel and the $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ control channel are prepared and pre-selected. The signal channel samples are split into two samples called the 2μ and the 3μ sub-sample depending on the number of muon candidates triggered by the LHCb muon system. The control channel is used to correct the Monte Carlo sample in four kinematic variables. The Monte Carlo events are weighted such, that their variable distributions approximate the data sample distributions in these variables. The 2μ and 3μ sub-samples are then used to train binary decision trees to determine a separation variable. This variable classifies events in a range between -1 and 1 , where -1 means background-like and 1 means signal-like. The control channel is again used to correct the Monte Carlo sample. Weights are added to the Monte Carlo events such that the classifier output distributions of the Monte Carlo sample approach those of the data sample. A figure of merit is used to determine the most sensitive cut on particle identification for the muon candidates and the classifier output of the event for each year and sub-sample separately.

The $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ channel is used as reference channel for calculating a limit on the branching fraction for the signal channel decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$. Using the observed number of $D_s^- \rightarrow \phi(\mu^- \mu^+) \pi^-$ signal events and the estimated efficiencies for both the reference and the signal sub-samples for each year, the single event sensitivity α for all three years and both sub-samples can be calculated. They are given in Tab.9.1.

Table 9.1: The final single event sensitivities α for each of the six samples.

	2016	2017	2018
2μ	$(5.72 \pm 2.40) \times 10^{-7}$	$(4.84 \pm 2.00) \times 10^{-8}$	$(5.37 \pm 2.22) \times 10^{-8}$
3μ	$(2.91 \pm 1.20) \times 10^{-9}$	$(2.52 \pm 1.04) \times 10^{-9}$	$(1.88 \pm 0.78) \times 10^{-9}$

The single event sensitivity gives the proportionality factor between the branching

fraction and the number of observed signal events:

$$\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) = \alpha \times N_{sig}, \quad (9.1)$$

where an upper limit on the number of observed signal events can be determined from the number of observed events in the signal region, N_{obs} , given the number of expected background events, B_{exp} . The number of expected background events is determined by fitting the distribution of the signal data sample outside of the signal range. Assuming the branching fraction to be zero, the number of observed events is expected to be equal to the expected background. For this expectation, a limit on the branching fraction can be calculated for all three years and both sub-samples. These values are given in Tab. 9.2.

Table 9.2: Upper limit on the number of signal events and the limit on the branching fraction at 90% and 95% confidence level assuming N_{exp} events are observed in the signal region.

	2016	2017	2018
N_{exp}	3	14	6
90%: $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq$	2.19×10^{-6}	3.81×10^{-7}	3.24×10^{-7}
95%: $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq$	3.00×10^{-6}	4.75×10^{-7}	4.02×10^{-7}
	2016	2017	2018
N_{exp}	181	155	202
90%: $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq$	6.65×10^{-8}	5.29×10^{-8}	4.55×10^{-8}
95%: $\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq$	8.09×10^{-8}	6.51×10^{-8}	5.49×10^{-8}

A combined expected upper limit can be set to:

$$\mathcal{B}(\tau^- \rightarrow \mu^- \mu^+ \mu^-) \leq 2.96 (3.58) \times 10^{-8} \text{ 90\% (95\%)}$$

This is very close to the expected limit extrapolated from the expected limit calculated during *Run 1* [6] taking the increased luminosity and centre-of-mass energy in account:

$$UL_{Run 2, extrapolated} = 2.95 (3.60) \times 10^{-8} \text{ 90\% (95\%)}$$

The expected limit is significantly better than the limit reached during *Run 1*:

$$UL_{Run 1, observed} = 4.6 (5.6) \times 10^{-8} \text{ 90\% (95\%)}$$

As during *Run 1* an additional binning in particle identification and classifier output variables improved the limit by about 20%, a similar improvement is expected for this dataset with the introduction of binning. Additionally, the final analysis will provide further corrections to the single event sensitivity and improve upon the output

classifier. So far, the presented measurement represents a limit on the branching fraction tighter than from the previous measurement, but still above the current worldwide best limit, reached by the BELLE experiment:

$$UL_{best\ limit\ (BELLE)} = 2.1 \times 10^{-8} \text{ 90\%}$$

With further improvements on the classifier and combination with the limit reached in *Run 1*, it will be possible to improve upon the limit set before by the BELLE experiment. This result will further challenge theories beyond the Standard Model and provide knowledge about the elementary rules of the universe.

Part I

Appendix

A Bibliography

- [1] P. Blackstone, M. Fael, and E. Passemar. $\tau \rightarrow \mu\mu\mu$ at a rate of one out of 10^{14} tau decays? *Eur. Phys. J. C*, 80, Jun 2020. doi: 10.1140/epjc/s10052-020-8059-7. URL <https://link.springer.com/article/10.1140/2Fepjc%2Fs10052-020-8059-7>. arXiv:1912.09862 [hep-ph].
- [2] Amon Ilakovac, Apostolos Pilaftsis, and Luka Popov. Charged lepton flavor violation in supersymmetric low-scale seesaw models. *Physical Review D*, 87 (5), Mar 2013. ISSN 1550-2368. doi: 10.1103/physrevd.87.053014. URL <http://dx.doi.org/10.1103/PhysRevD.87.053014>. arXiv:1212.5939[hep-ph].
- [3] M. Raidal, A. van der Schaaf, and I. Bigi et al. Flavor physics of leptons and dipole moments. *Eur.Phys.J. C*, 57, Nov 2008. doi: 10.1140/epjc/s10052-008-0715-2. URL <https://link.springer.com/article/10.1140/2Fepjc%2Fs10052-008-0715-2>.
- [4] Ernesto Arganda and María J. Herrero. Testing supersymmetry with lepton flavor violating τ and μ decays. *Phys. Rev. D*, 73:055003, Mar 2006. doi: 10.1103/PhysRevD.73.055003. URL <https://link.aps.org/doi/10.1103/PhysRevD.73.055003>.
- [5] K. Hayasaka et al. Search for lepton-flavor-violating τ decays into three leptons with 719 million produced $\tau^+\tau^-$ pairs. *Physics Letters B*, 687(2-3):139–143, Apr 2010. ISSN 0370-2693. doi: 10.1016/j.physletb.2010.03.037. URL <http://dx.doi.org/10.1016/j.physletb.2010.03.037>.
- [6] J. Albrecht, M. Calvi, M. Chrzaszcz, L. Gavardi, J. Harrison, B. Khanji, G. Lafferty, E. Rodrigues, N. Serra, P. Seyfert, and T. Likhomanenko. Search for the lepton flavour violating decay $\tau \rightarrow \mu^+\mu^-\mu^-$, Jan 2014. URL <https://cds.cern.ch/record/1645011>.
- [7] The LHCb Collaboration. Lhcb operations plots webpage, n.d.. URL <https://lbgroups.cern.ch/online/OperationsPlots/index.htm>. Accessed: 2020-07-23.
- [8] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018. doi: 10.1103/PhysRevD.98.030001. URL <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [9] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998. doi: 10.1103/PhysRevLett.81.1562. URL <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.

- [10] M. Thomson. Modern particle physics, 2013.
- [11] F. Halzen and A. Martin. Quarks and leptons: an introductory course in modern particle physics, 1984.
- [12] E. Mobs. The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019, Jul 2019. URL <https://cds.cern.ch/record/2684277>. General Photo.
- [13] Communications CERN Education and Outreach Group. LHC Guide. CERN brochure, Mar 2017. URL <https://cds.cern.ch/record/2255762>.
- [14] The LHCb Collaboration. Lhcb detector performance. *International Journal of Modern Physics A*, 30(07):1530022, Mar 2015. ISSN 1793-656X. doi: 10.1142/S0217751x15300227. URL <http://dx.doi.org/10.1142/S0217751X15300227>.
- [15] Rolf Lindner. LHCb layout_2. LHCb schema_2. LHCb Collection., Feb 2008. URL <https://cds.cern.ch/record/1087860>.
- [16] Ch. Ilgner, M. Domke M. Lieng, M. Nedos, J. Sauerbrey, S. Schleich, B. Spaan, K. Warda, and J. Wishahi. The beam conditions monitor of the lhcb experiment, 2010.
- [17] K. Akiba et al. The herschel detector: high-rapidity shower counters for lhcb. *Journal of Instrumentation*, 13(04):P04017–P04017, Apr 2018. ISSN 1748-0221. doi: 10.1088/1748-0221/13/04/p04017. URL <http://dx.doi.org/10.1088/1748-0221/13/04/P04017>.
- [18] P.R. Barbosa-Marinho et al. *LHCb VELO (VERtex LOcator): Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2001. URL <https://cds.cern.ch/record/504321>.
- [19] Maximilien Brice. Assembling the last module of the vertex locator for LHCb. Assemblage en salle propre de l'expérience LHCb du 42e et dernier module du VELO (Localisateur du vertex). Mar 2007. URL <http://cds.cern.ch/record/1024838>.
- [20] S. Amato et al. *LHCb RICH: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2000. URL <https://cds.cern.ch/record/494263>.
- [21] The LHCb Collaboration. The LHCb detector at the LHC. *Journal of Instrumentation*, 3(08):S08005–S08005, aug 2008. doi: 10.1088/1748-0221/3/08/s08005. URL <https://doi.org/10.1088/1748-0221/3/08/s08005>.
- [22] S. Amato et al. *LHCb magnet: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2000. URL <https://cds.cern.ch/record/424338>.

- [23] M Vesterinen. Considerations on the LHCb dipole magnet polarity reversal. Technical Report LHCb-PUB-2014-006. CERN-LHCb-PUB-2014-006, CERN, Geneva, Apr 2014. URL <https://cds.cern.ch/record/1642153>. On behalf of the LHCb collaboration.
- [24] C. Abellan Beteta et al. Monitoring radiation damage in the LHCb Tracker Turicensis. Technical Report arXiv:1809.05063, Sep 2018. URL <https://cds.cern.ch/record/2638494>. 17 pages, 11 figures, 3 tables.
- [25] P.R. Barbosa-Marinho et al. *LHCb inner tracker: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2002. URL <https://cds.cern.ch/record/582793>. revised version number 1 submitted on 2002-11-13 14:14:34.
- [26] A. Scarabott. Test of lepton flavour universality using b s semileptonic decays, Jul 2020. URL <https://cds.cern.ch/record/2724839>. Presented 16 Jul 2020.
- [27] P.R. Barbosa-Marinho et al. *LHCb outer tracker: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2001. URL <https://cds.cern.ch/record/519146>.
- [28] S. Amato et al. *LHCb calorimeters: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2000. URL <https://cds.cern.ch/record/494264>.
- [29] P.R. Barbosa-Marinho et al. *LHCb muon system: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2001. URL <https://cds.cern.ch/record/504326>.
- [30] R. Aaij et al. Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. *JINST*, 14(arXiv:1812.10790. 04):P04013. 43 p, Dec 2018. doi: 10.1088/1748-0221/14/04/P04013. URL <https://cds.cern.ch/record/2652801>. 46 pages, 35 figures, 1 table. All figures and tables are available at <https://cern.ch/lhcbproject/Publications/LHCbProjectPublic/LHCb-DP-2019-001.html>.
- [31] CERN (Meyrin) LHCb Collaboration. Computing Model of the Upgrade LHCb experiment. Technical Report CERN-LHCC-2018-014. LHCb-TDR-018, CERN, Geneva, May 2018. URL <https://cds.cern.ch/record/2319756>.
- [32] The LHCb Collaboration. Lhcb trigger conference diagrams and plots webpage, n.d.. URL <https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbTriggerConferenceDiagramsPlots>. Accessed: 2020-10-20.

- [33] S. Miglioranzi, M. Clemencic, C. Corti, S. Easo, C.R. Jones, M. Pappagallo, and P. Robbe. The LHCb Simulation Application, Gauss: Design, Evolution and Experience. Technical Report LHCb-PROC-2011-006. CERN-LHCb-PROC-2011-006, CERN, Geneva, Jan 2011. URL <https://cds.cern.ch/record/1322402>.
- [34] T. Sjöstrand, S.n Mrenna, and P. Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852–867, Jun 2008. ISSN 0010-4655. doi: 10.1016/j.cpc.2008.01.036. URL <http://dx.doi.org/10.1016/j.cpc.2008.01.036>.
- [35] D.J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152–155, 2001. doi: 10.1016/S0168-9002(01)00089-4.
- [36] J. Allison. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006. doi: 10.1109/TNS.2006.869826. URL <https://cds.cern.ch/record/1035669>.
- [37] R. Antunes-Nobrega et al. *LHCb computing: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2005. URL <https://cds.cern.ch/record/835156>. Submitted on 11 May 2005.
- [38] LHCb Starterkit. The lhcb data flow, 2015-2019. URL <https://lhcb.github.io/starterkit-lessons/first-analysis-steps/dataflow.html>. Accessed: 2020-09-09.
- [39] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. ISSN 2050-3911. doi: 10.1093/ptep/ptaa104. URL <https://doi.org/10.1093/ptep/ptaa104>. 083C01.
- [40] R. Brun, F. Rademakers, and S. Panacek. ROOT, an object oriented data analysis framework. 2000. URL <https://cds.cern.ch/record/491486>.
- [41] M. Pivk and F.R. Le Diberder. : A statistical tool to unfold data distributions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555(1-2): 356–369, Dec 2005. ISSN 0168-9002. doi: 10.1016/j.nima.2005.08.106. URL <http://dx.doi.org/10.1016/j.nima.2005.08.106>.
- [42] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. Tmva 4. toolkit for multivariate data analysis with root. users guide, Mar 2007. URL <https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf>. arXiv:physics/0703039[Data Analysis, Statistics and Probability].
- [43] Paul Seyfert. The search for $\tau \rightarrow \mu\mu\mu$ at LHCb, Jan 2015. URL <https://cds.cern.ch/record/2002363>. Presented 14 Jan 2015.

- [44] L. Anderlini, S. Benson, O. Lupton, B. Sciascia, and V. Gligorov. Computing strategy for PID calibration samples for LHCb Run 2. Technical Report LHCb-PUB-2016-020. CERN-LHCb-PUB-2016-020, CERN, Geneva, Jul 2016. URL <https://cds.cern.ch/record/2199780>.
- [45] L. Lista. *Statistical Methods for Data Analysis in Particle Physics*. Springer, Cham, 2017. URL <https://link.springer.com/book/10.1007/978-3-319-62840-0>. doi: <https://doi.org/10.1007/978-3-319-62840-0>.
- [46] R. Aaij et al. Optimization of the muon reconstruction algorithms for LHCb Run 2. Technical Report LHCb-PUB-2017-007. CERN-LHCb-PUB-2017-007, CERN, Geneva, Feb 2017. URL <https://cds.cern.ch/record/2253050>.
- [47] F. Archilli et al. Search for the $B^0 \rightarrow \mu^+\mu^-$ decay and measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction and effective lifetime. May 2016. URL <https://cds.cern.ch/record/2156720>.
- [48] J Harrison, P. Seyfert, and M. Chruszcz. Webpage: Tau23muds2phipiline, n.d.. URL <http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/config/stripping34/leptonic/strippingtau23muds2phipiline.html>. Accessed: 2020-07-20.
- [49] J Harrison, P. Seyfert, and M. Chruszcz. Webpage: Tau23mutau23muline, n.d.. URL <http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/config/stripping34/leptonic/strippingtau23mutau23muline.html>. Accessed: 2020-07-20.
- [50] R. Aaij et al. Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 03(LHCB-PAPER-2015-041):159. 46 p, Oct 2015. doi: 10.1007/JHEP03(2016)159. URL <https://cds.cern.ch/record/2057627>. Includes comments of erratum *JHEP* 09 (2016) 013 and an additional erratum *JHEP* 05 (2017) 074. All figures and tables, along with any supplementary material and additional information, are available at <https://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCBProjectPublic/LHCB-PAPER-2015-041.html>.
- [51] Roel et al. Aaij. Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV. Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 10 (LHCB-PAPER-2015-037. LHCB-PAPER-2015-037. CERN-PH-EP-2015-222):172. 33 p, Sep 2015. doi: 10.1007/JHEP10(2015)172. URL <https://cds.cern.ch/record/2048812>. All figures and tables, along with any supplementary material and additional information, are available at <https://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCBProjectPublic/LHCB-PAPER-2015-037.html>. An erratum has been submitted, and the contents of the errata are reflected in this manuscript.

- [52] M.J. Oreglia. A study of the reactions $\psi' \rightarrow \gamma \gamma \psi$, ph.d. thesis, appendix d, Dec 1980. URL <https://www.slac.stanford.edu/cgi-bin/getdoc/slac-r-236.pdf>.
- [53] R. Aaij et al. Prompt charm production in pp collisions at $\sqrt{s} = 7$ TeV. *Nuclear Physics B*, 871(1):1–20, Jun 2013. ISSN 0550-3213. doi: 10.1016/j.nuclphysb.2013.02.010. URL <http://dx.doi.org/10.1016/j.nuclphysb.2013.02.010>.
- [54] LHCb collaboration. Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV, 2015.
- [55] CMS Collaboration. Search for the lepton flavor violating decay $\tau \rightarrow 3\mu$ in proton-proton collisions at $\sqrt{s} = 13$ tev, 2020.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Dresden
Heidelberg, den (Datum) 24.12.2020 Dorena Caspary