Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Chishuai Wang

born in Zhejiang, P.R.China

2021

A feasibility study for the search of the lepton flavour violating decay $\tau^- \to \phi(K^+K^-)\mu^-$ at the LHCb experiment

This Master thesis has been carried out by Chishuai Wang

at the

Physikalisches Institut

under the supervision of

Prof. Dr. Ulrich Uwer

and

Dr. Flavio Archilli

Zusammenfassung:

In einem erweiterten Standardmodell, welches Neutrinomassen einbezieht, sind Lepton-Flavour-verletzende (LFV) Zerfälle möglich, aber stark unterdrückt, mit Zerfallsbreiten kleiner als $\mathcal{O}(10^{-40})$. Daher wäre jede Beobachtung von LFV ein deutlicher Hinweis auf neue Physik.
Die hier vorgestellte Machbarkeitsstudie sucht nach dem LFV-Zerfall $\tau^- \to \phi(K^+K^-)\mu^-$. Dazu werden Daten genutzt, die einer integrierten Luminosität von 5.6 fb^{-1} entsprechen, gemessen mit dem LHCb-Experiment in den Jahren 2016-2018 bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$ in pp-Kollisionen. Die τ-Leptonen werden bei dem Zerfall von $D_s^\pm$ und $D^\pm$-Mesonen, sowie b-Hadronen produziert. Diese Studie beinhaltet eine Ereignisselektion einschließlich einer multivariaten Klassifikation. Das Suchergebnis wird von dem Hintergrundzerfall $D_s^+ \to \phi(K^+K^-)\mu^+\nu_\mu$ dominiert. Für den gesuchten Zerfall wird kein Signal gemessen. Mithilfe des Referenzkanals $D_s^- \to \phi(\mu^+\mu^-)\pi^-$ wird ein einseitiges 90% Konfidenzintervall für die Zerfallsbreite ermittelt

$$\mathcal{B}(\tau \to \phi\mu) < 9.4 \times 10^{-6} \tag{0.1}$$

Um diesen Grenzwert zu optimieren, ist eine Methode zur effektiven Hintergrundreduzierung notwendig.

Abstract:

In the Standard Model with an extension to account for neutrino masses, Lepton Flavour Violating (LFV) decays are possible but highly suppressed with branching fractions smaller than $\mathcal{O}(10^{-40})$. Therefore, any observation of LFV decays would be a strong indication of new physics.
The feasibility study presented here performs a search for the LFV decay $\tau^- \to \phi(K^+K^-)\mu^-$ using data corresponding to an integrated luminosity of 5.6 fb^{-1} collected by the LHCb experiment during the years 2016-2018 at a center-of-mass pp collision energy of $\sqrt{s} = 13\,\text{TeV}$. The τ leptons are produced in the decays of $D_s^\pm, D^\pm$ mesons and b-hadrons. This study involves an event selection procedure including multivariate classification. The search result is dominated by the background decay $D_s^+ \to \phi(K^+K^-)\mu^+\nu_\mu$. No signal for the decay of interest is discovered. Using the reference channel $D_s^- \to \phi(\mu^+\mu^-)\pi^-$, an upper limit of the branching fraction is evaluated at 90% confidence level:

$$\mathcal{B}(\tau \to \phi\mu) < 9.4 \times 10^{-6} \tag{0.2}$$

Improving this limit will require a method to effectively reduce the backgrounds.

Contents

1 Introduction

Within the framework of quantum field theory, the Standard Model of particle physics summarizes our current understanding of elementary particles and fundamental interactions. With scientific facilities like the Large Hadron Collider (LHC), physicists are able to test the Standard Model and make attempts to search for potential new effects beyond it. One of such attempts is the search of Lepton Flavour Violating (LFV) decays. In the Standard Model, LFV decays are highly suppressed and only possible through neutrino oscillation. Its branching fraction is predicted to be so small that it is pratically unreachable with today's experiments. Such a process will be highly sensitive to contributions from possible new physics. Any sign of LFV would be a strong indication for new physics.

The main purpose of this work is to perform a feasibility study for the search of the LFV decay $\tau^- \to \phi(K^+K^-)\mu^-$ with the data collected by the Large Hadron Collider beauty (LHCb) experiment during the years from 2016 to 2018. The data sample is generated from proton-proton collisions with a center-of-mass energy of 13 TeV and its integrated luminosity reaches 5.6 fb^{-1}. The LHCb experiment is operated by the LHCb collaboration, which includes over 1400 members representing many different institutes from all over the world.

For a study of this kind namely a search for a very rare decay, no signal is expected and very often no signal is found. In the situation that no signal discovery is claimed, the analysis still provides important information using a statistical interpretation of the data sample. The result is usually presented as an upper limit on the signal branching fraction. In this analysis, given the well-known branching fraction $\mathcal{B}(\phi \to K^+K^-)$, the result is presented as a constraint on the branching fraction $\mathcal{B}(\tau \to \phi\mu)$.

Several upper limits on $\mathcal{B}(\tau \to \phi\mu)$ have been published by experiments like BaBar and Belle. In 2009, the BaBar experiment set an upper limit of 1.9×10^{-7} at 90% confidence level with a total data sample corresponding to an integrated luminosity of about 451 fb^{-1} [1]. And in 2011, the Belle experiment set an upper limit of 8.4×10^{-8} at 90% confidence level with 854 fb^{-1} of data [2].

In LHCb, copious τ leptons are produced through c-meason ($D_s^{\pm}, D^{\pm}$) and b-hadron decays. LFV processes in τ decays have been studied at LHCb before. A very competitive result has been reached for $\tau^- \to \mu^+\mu^-\mu^-$, the limit on branching fraction was set to $\mathcal{B}(\tau^- \to \mu^+\mu^-\mu^-) < 4.6 \times 10^{-8}$ with 90% confidence level [3]. In addition, decay modes $\tau^- \to \bar{p}\mu^+\mu^-$ and $\tau^- \to p\mu^-\mu^-$ have also been probed. The LHCb measurements represent the first direct experimental limits on these two channels [4]. With these results, LHCb has proved to be able to probe LFV in τ decays. For the decay $\tau^- \to \phi(K^+K^-)\mu^-$, some preliminary studies were done at LHCb, but so far no result has been published.

The subsequent chapters of this thesis are organized as follows: Chapter 2 will provide a brief introduction of the theories related to the decay of interest. Chapter 3 presents an introduction of the LHCb experiment. In chapter 4, the decay channels involved in the analysis are discussed, including the signal channel, the reference channel and some backgrounds. Chapter 5 summarizes some preparation work done on the data and Monte Carlo samples before they are used in the analysis. These preparations include a weighting procedure for the different production channels in the simulation, the data blinding procedure and some preliminary selections. In chapter 6, the correction on the Monte Carlo kinematics is discussed. A multivariate analysis and particle identification selections are presented in chapter 7 and chapter 8. They are applied to further separate the background and the potential signal. In chapter 9, an estimation of signal and background shapes is described. In chapter 10, the evaluation of the normalisation factor is presented. With the signal/background shapes and the normalisation factor as input, an upper limit on the branching fraction $\mathcal{B}(\tau \to \phi\mu)$ is evaluated in chapter 11, which is the main result of this study. In the end, chapter 12 serves to conclude the analysis.

This analysis is conducted within the LHCb collaboration and includes the usage of data processing and simulation frameworks/modules provided by the LHCb collaboration. The data/simulation sample collection and processing rely on the collective work of the collaboration. In addition, the analysis is performed in parallel with the analysis of $\tau^- \to \mu^+\mu^-\mu^-$ of LHCb Run2, which is carried out by a group ($\tau^- \to \mu^+\mu^-\mu^-$ Run2 proponents) within the LHCb collaboration. The evaluation of $\tau/D_s/D$ productions (described in chapter 4), the production weighting and the selection of the reference channel (described in chapter 5) and the evaluation of the reference channel efficiency (described in chapter 10) are the collective work of the group and the author.

The figures and numbers presented in this thesis are not officially approved by LHCb, and should be considered as preliminary results.

2 Theories introduction

In this chapter, the basic content of Standard Model with its contribution to the LFV decay $\tau^- \to \phi(K^+K^-)\mu^-$ as well as possible contributions from physics beyond the Sandard Model are introduced.

2.1 Standard Model and massive neutrinos

In the past decades, the Standard Model has been very successful in explaining most of the particle physics experiment results. As shown in Fig. 2.1, the Standard Model consists of three generations of quarks and leptons, the gauge bosons which mediate electroweak/strong interactions and the Higgs boson. The quark sector of the Standard Model consists of three up-type quarks, up (u), charm (c), top (t) and three down-type quarks, down (d), strange (s), bottom (b). These quarks have their corresponding antiquarks. In the lepton sector of the Standard Model, there are charged (anti-)leptons of three flavours (e, μ, τ) and their corresponding (anti-)neutrinos.

The interactions of these particles are described within a gauge theory. The gauge symmetry of the Standard Model is described as $SU(3)_c \times SU(2)_L \times U(1)_Y$ in the language of group theory. The symmetry $SU(3)_c$ stands for the gauge field of the strong interaction, which is mediated through massless gluons. The symmetry $SU(2)_L \times U(1)_Y$ is a crucial part of the electroweak theory. The gauge group $SU(2)_L$ indicates that the corresponding interaction only involves left-handed fermions or right-handed antifermions.

Through spontaneous symmetry breaking of the Higgs mechanism, gauge symmetry $SU(2)_L \times U(1)_Y$ results in three massive gauge bosons i.e. $Z, W^{\pm}$ for the weak interaction and one massless photon for the electromagnetic interaction.

Originally in the Standard Model, neutrinos were supposed to be massless. However, the discovery of neutrino oscillation [6] provided a solid evidence that at least two types of neutrinos are actually massive. Nowadays, the Standard Model is usually extended to include massive neutrinos into consideration.

For massive neutrinos, the mass eigenstates and flavour eigenstates are different and they can transform into one another through a unitary mixing matrix i.e. the Pontecorvo–Maki–Nakagawa–Sakata matrix (PMNS matrix), as shown in Eq. 2.1. This allows a neutrino of certain lepton flavour to oscillate into another flavour.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{2.1}$$

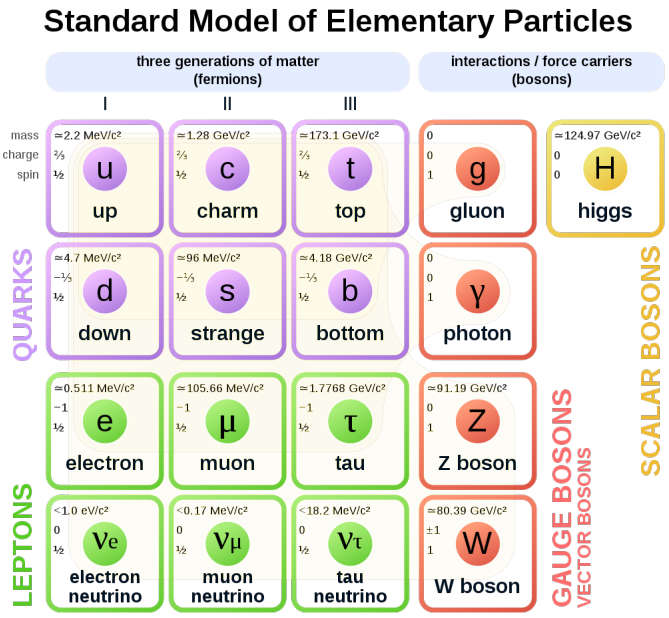


Figure 2.1: The Standard Model of elementary particles. The graph was taken from [5].

In the Standard Model, without neutrino oscillation, lepton flavour is known to be conserved. In the presence of neutrino oscillation, it serves as the only source of lepton flavour violation. LFV decays can happen through the so-called Penguin diagrams. Fig. 2.2 shows two possible Penguin diagrams for the $\tau^- \to \phi(K^+K^-)\mu^-$ decay. These processes are highly suppressed due to the smallness of the neutrino mass and the unitarity of the PMNS matrix. The neutrino masses come into the loop diagram as perturbation terms, which can be evaluated order by order. When considering all of the possible neutrinos, the interference of the diagrams comes into play. The sum of amplitudes of diagrams with different neutrinos takes the following form:

$$\mathcal{M} \sim \sum_i U^*_{\tau i} U_{\mu i} f(\frac{m_{\nu_i}}{m_W}) \tag{2.2}$$

with neutrino masses much less than the mass scale of weak interaction bosons $m_{\nu_i} \ll m_W$ [7]. The function f stands for the contribution of loop diagram with a certain neutrino mass. The zero order term vanishes with the unitarity condition:

$$\mathcal{M}^{(0)} \sim \sum_i U^*_{\tau i} U_{\mu i} = 0 \tag{2.3}$$

and the higher order terms are also very small due to the fact that the masses of neutrinos are extremely small. In general, the branching fractions for charged lepton flavour violation (cLFV) decays caused by massive neutrinos are around $\mathcal{O}(10^{-40})$

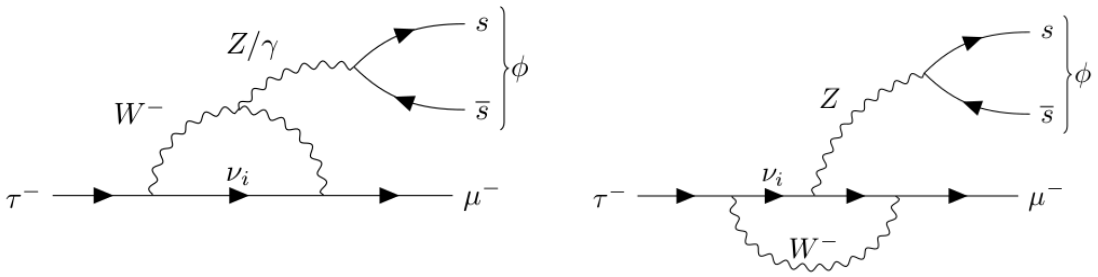


Figure 2.2: Two possible feynman diagrams for $\tau^- \to \phi(K^+K^-)\mu^-$ through neutrino oscillation.

or less, which is far below the sensitivity of nowaday's experiments [8]. Thus, any sign of cLFV decays will be a strong indication for new physics beyond the Standard Model.

2.2 Physics beyond Standard Model

While the branching fractions of LFV decays are tiny within the framework of neutrino oscillation, some possible new physics theories predict much larger branching fractions up to $\mathcal{O}(10^{-8})$, which is accessible for today's experiments [2]. Thus the search of LFV decays is an effective way to search for possible new physics effects. Some of these potential new physics models are briefly introduced in the following paragraphs.

One of the possible new physics models contributing to LFV is the leptoquark model. Leptoquarks are hypothetical particles. They are assumed to be able to interact with both quarks and leptons. They carry both the baryon number and the lepton number [9]. The existence of leptoquarks are predicted by many beyond the Sandard Model theories e.g. the Pati-Salam model [10] and some grand unification theories [11, 12]. Fig. 2.3 shows the Feynmann diagram of the decay $\tau^- \to \phi(K^+K^-)\mu^-$ mediated by a hypothetical leptoquark. Some recent results of lepton univerisality tests at LHCb have shown some tensions with the Standard Model prediction. These studies include the measurement of $R(K^{(*)})$ (the ratio of branching fractions between $B^+ \to K^{(*)}\mu^+\mu^-$ and $B^+ \to K^{(*)}e^+e^-$) [13, 14], $R(J/\psi)$ (ratio of branching fractions between $B_c^+ \to J/\psi\tau^+\nu_\tau$ and $B_c^+ \to J/\psi\mu^+\nu_\mu$) [15], and $R(D^*)$ (ratio of branching fractions between $\bar{B}^0 \to D^{*-}\tau^+\nu_\tau$ and $\bar{B}^0 \to D^{*-}\mu^+\nu_\mu$) [16, 17]. Similar studies were also performed at Belle and Babar [18, 19]. In addition, deviations from the Standard Model were also found in the angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ by LHCb [20, 21]. Although it is too early to consider these results as evidences for physics beyond the Standard Model, they provide hints for the possible existence of new physics models including leptoquark.

In addition, a model with a possible additional neutral gauge boson (Z') is dis-

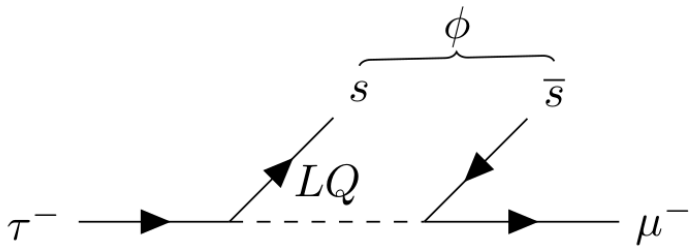


Figure 2.3: Feynmann diagram of $\tau^- \to \phi(K^+K^-)\mu^-$ mediated by a possible leptoquark(LQ).

cussed in the attempts to explain the aforementioned tensions [22, 23]. This additional Z' boson is assumed to have a mass much larger than the eletroweak symmetry breaking scale and has generic couplings with leptons and quarks. In the Sandard Model, flavour changing neutral current (FCNC) transition is forbidden at tree level, while this hypothetical heavy neutral boson may couple directly with leptons of different flavours. Hence it may contribute to LFV decays at tree level [23].

Supersymmetry (SUSY) framework also provides opportunities for LFV processes. SUSY is a well-known beyond the Sandard Model theory: it associates fermions with their superpartner bosons and bosons with their superpartner fermions. It helps to solve several important puzzles of particle physics including hierachy problem and the possible unification of particle physics with gravity [9]. LFV decays are discussed in different versions of SUSY e.g. Minimal Supersymmetric extension of the Sandard Model (MSSM) and MSSM enlarged with right handed neutrinos (the neutrion masses are generated with the Seesaw mechanism) [24, 25].

3 LHCb experiment

This chapter is dedicated to a brief introduction of the LHCb experiment, including its detector, its data processing procedures and the corresponding software frameworks.

3.1 LHCb project

The LHC near Geneva is currently the largest collider ever built in the world. It has a circular tunnel with a circumference of 27km. It provides proton-proton collisions with a center-of-mass energy $\sqrt{s}$ up to 14 TeV [26].

The LHCb experiment is one of the large experiments operated at LHC by the European Organization for Nuclear Research (CERN). Exploiting the large amount of $b\bar{b}$ hadrons produced in the high energy pp collisions at LHC, it was designed to study heavy flavour physics. Its research topics include CP asymmetries and rare decays in beauty and charm hadrons [27, 28]. The LHCb experiment has started operation in 2010. Fig. 3.1 shows the recorded luminosity of pp collisions by the LHCb experiment in the past several years. From 2010 to 2012, LHCb performed its Run1 operation at $\sqrt{s} = 7$ and 8 TeV, which reached a total integrated luminosity around 3 fb^{-1}. In Run2 which started from 2015 and ended in 2018, the LHC operated at higher collision energy $\sqrt{s} = 13$ TeV, which resulted in a higher $b\bar{b}$ production cross section. The integrated luminosity of data collected in Run2 reached around 6 fb^{-1}. Since the end of Run2, the LHCb experiment has been going through an upgrade. Several subdetectors will be replaced and some will be removed before the start of future Run3 and Run4. In the future, the LHCb experiment will operate at an even higher luminosity, which will produce higher signal yields. For the analysis discussed in this thesis, the data collected in LHCb Run2 will be used.

3.2 LHCb detector

In the high energy pp collisions of LHC, the $b\bar{b}$ hadrons are mainly produced within the same cone forward (or backward), and thus the subdetectors of LHCb are arranged as a forward single-arm spectrometer, which covers the angular range of $\pm$250mrad vertically and of $\pm$300mrad horizontally. A right-handed coordinate is usually assigned to identify the directions i.e. the z axis for the direction of the beam, the y axis for the vertical direction and the x axis for the horizontal direction [28].

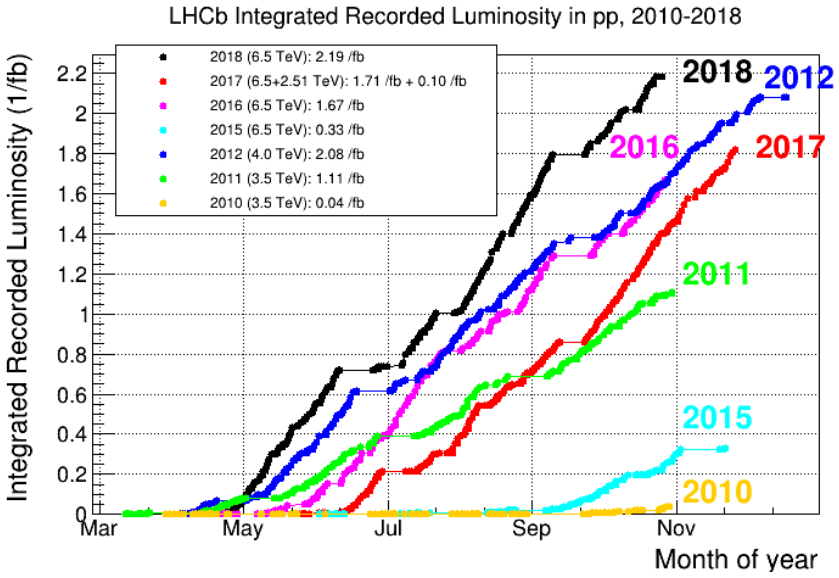


Figure 3.1: LHCb integrated recorded luminosity.The graph was taken from [29]

Fig. 3.2 shows the layout of the LHCb detector in Run1 and Run2, which consists of a Vertex Locator(Velo), an upstream Trigger Tracker(TT), a spectrometer magnet, three downstream Tracking stations (T1-T3), two Ring Imaging Cherenkov counters (RICH1,RICH2), a calorimeter system and a Muon detection system (M1-M5). The downstream tracking stations contain an inner part (Inner Tracker, IT) and an outer part (Outer Tracker, OT). The calorimeter system contains a Scintillator Pad Detector and a Preshower (SPD/PS), a shashlik Electromagnetic Calorimeter(ECAL) and a Hadronic Calorimeter (HCAL). These subdetectors are discussed further in the following part.

Significant changes are being made to the detector during the upgrade (Upgrade I). The VELO will be replaced by a new silicon pixel detector. The TT station will be replaced by a silicon micro-strip detector, the Upstream Tracker (UT). The main tracking stations (T1-T3) will be replaced by the Scintillating Fibre (SciFi) Tracker. Some changes will be made to improve the RICH. The SPD/PS and M1 will be removed. In addition, the readout electronics will be replaced to allow the readout at 40 MHz corresponding to the crossing rate of LHC beams.

3.2.1 Vertex locator

The vertex locator is designed to measure the track coordinates near the pp interaction vertex. It is important to detect the displacement of the secondary vertices as a distinctive feature of $b\bar{b}$ or $c\bar{c}$ hadron decays [28]. The VELO is composed by silicon modules. It provides an excellent measurement of the vertices and impact parameters (IP). The flight distance and the decay time derived from it are also measured by the VELO. The performance of the software trigger of LHCb depends on the re-

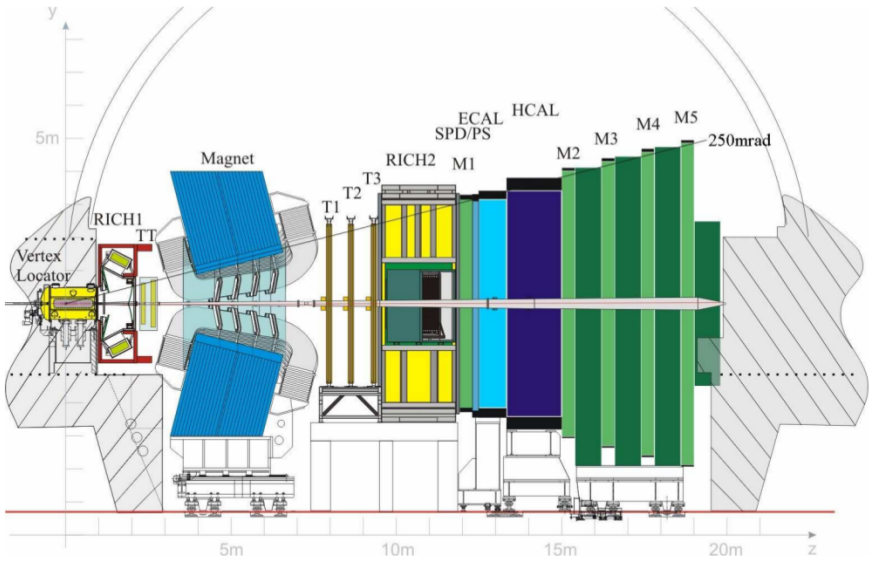


Figure 3.2: LHCb detector layout during Run1 and Run2. The graph was taken from [28].

construction of vertices in the VELO. To achieve high resolution, the VELO sensors are installed very close to the LHC beams (around 7 mm). However, during the LHC beam injection, a larger aperture is required. Thus the detector is constructed as two retractable halves. The two halves of VELO are retracted during the beam injection and closed during physics running. For the geometrical coverage, the two halves are designed to partially overlap when closed [30].

Fig. 3.3 shows the cross sections of the VELO detector. The detector is composed of 25 VELO stations and each station has a right and a left module. For a track in the LHCb acceptance, it is expected to cross at least three VELO stations. Except for 2 stations shown in the figure, each VELO module has one R-measuring sensor and one ϕ-measuring sensor. The sensors are constructed in the circular shape. The strips of the R-sensors are arranged azimuthally while the strips of the ϕ-sensors are arranged radially [31].

3.2.2 Magnet

The dipole magnet is installed to measure the charged particle momenta. The magnet contains two trapezoidal coils placed at the top and the bottom. Each coil is composed of 15 monolayer pancakes. These two coils generate a mostly vertically aligned magnetic field [32]. The orientation of the magnectic field is changed regularly. Thus, data samples and Monte Carlo samples are prepared for both magnetic orientations: MagDown and MagUp.

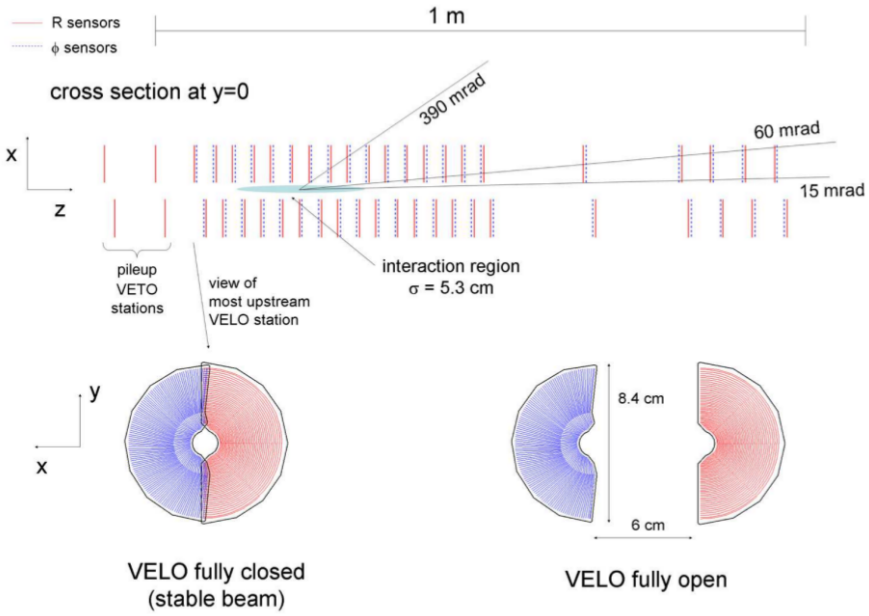


Figure 3.3: VELO schematic graph. The red stands for R-sensors and the blue stands for ϕ-sensors. The graph was taken from [28]

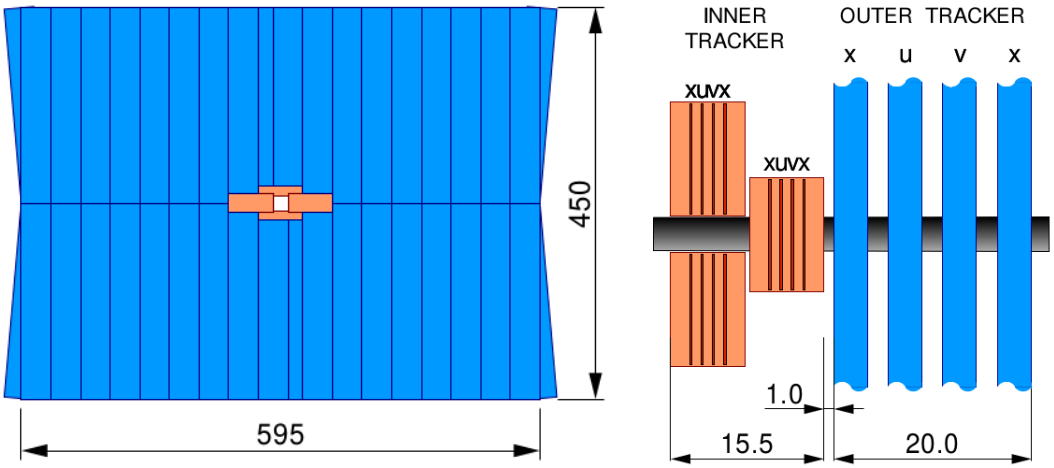
3.2.3 Trigger Tracker

The TT detector is constructed as a Silicon Tracker (ST) detector. It consists of two stations (four layers). With the beam pipe as z axis, the first two layers (labelled as TTaX, TTaU) are placed around $z = 232$cm and the second two (labelled as TTbV, TTbX) are placed around $z = 262$cm ($z = 0$cm corresponds to the collision point within the VELO). The strips in TTaX and TTbx are arranged along the vertical axis. In TTaU and TTbV, the strips are tilted by ± 5 degree [33]. With the vertically arranged strips, TT provides a measurement of the horizontal transverse coordinate of the charged particles. Along with the vertical magnetic field, this will contribute to the momentum measurement.

3.2.4 Downstream tracking stations

The tracking stations T1-T3 consist of two parts, the outer tracker (OT) and the inner tracker (IT). A schematic view of T1-T3 tracking stations is shown as Fig. 3.4.

Same as TT, IT is constructed from silicon microstrip detectors. In each of the 3 stations T1-T3, IT covers the part closest to the beam pipe which has a shape of a cross. As shown in Fig. 3.4, the IT at each station has four independent detector boxes i.e. above, below and on two sides of the beam pipe. Each box is composed of four layers (labelled as X,U,V,X). Similar to TT, the two x layers have vertically arranged cells while the cells in u and v are tilted for ± 5 degrees [34].



(a) view from front (b) view from top

Figure 3.4: One of the downstream tracking station T1-T3 layout (graph from [34]). The red part stands for IT while the blue part stands for OT. The numbers shown in the graph are in cm.

The OT detector is a drift-time detector constructed from gas-tight straw-tubes. The momentum resolution with a single hit in OT can reach $\delta p/p \approx 0.4\%$. Each straw tube (diameter 4.9 mm) measures the drift time of the electrons generated when a charged particle hits the detector. The electrons are collected in the anode wires in each tube.

As shown in Fig. 3.4(b), each tracking station of OT has four layers. Same as IT, they are labelled as X,U,V,X. The two X layers are arranged vertically and U, V are arranged tiltedly by ± 5 degrees with respect to the vertical axis. Each layer consists of 14 long modules (labelled as F type) and 8 short modules (labelled as S type) with half the length of the long module. As shown in Fig. 3.4(a), the short modules are placed above and below the beam pipe, and the long modules are installed on the side. The short modules are read out from the outer end while the long modules are splitted in the middle to be read out from both outer end [35].

3.2.5 RICH

There are two RICH detectors installed in LHCb. RICH1 is installed upstream of the magnet and RICH2 is installed downstream. The two RICH detectors play important roles in the charged particle identification (PID). Based on the fact that the particles with higher momentum are mostly generated in lower angular range, the two RICH detectors are arranged as such: RICH1 is used to cover the low momentum range (around 1$\sim$60 GeV/c) and the full angular range (around 25-300 mrad) while RICH2 is used for the high momentum range (around 15$\sim$100 GeV/c) with a smaller angular coverage (around 15-120 mrad). The particle identification of RICH uses the Cherenkov photon emission. Fig. 3.5 shows the correlation between the emission

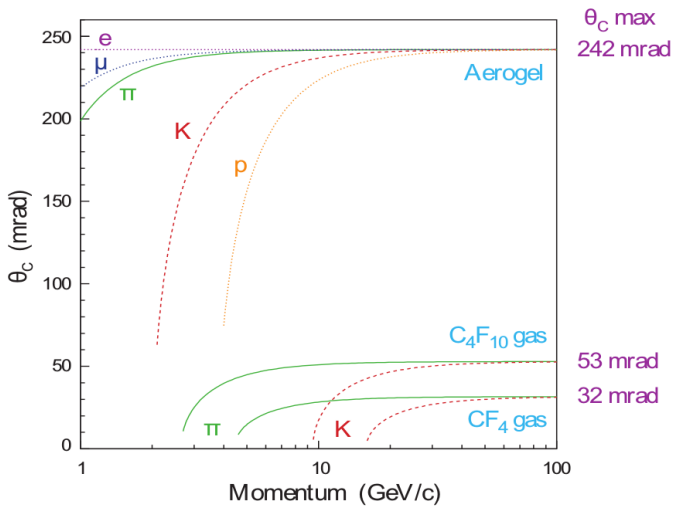


Figure 3.5: Cherenkov angle-momentum relationship in several materials used by RICH. The graph was taken from [28]. Aerogel was used in RICH1 during Run1, but not used anymore during Run2.

angle of the Cherenkov photon (θ_c) and momentum of charged particles (π, K etc.) in the different materials used by RICH. C_4F_{10} which has good identification ability for low momentum particles is used for RICH1, while CF_4 is used for RICH2 [36].

Combining with the track information provided by the tracking system, RICH detectors are able to use the Cherenkov angle information for PID. The RICH detectors usually work in high occupancy condition. An overall RICH event likelihood is calculated for a given set of hypotheses for assuming each track to be one of μ, π, K etc. The likelihood is optimized to find the best hypothesis for each track. As the final result for a single track, the difference of the overall event likelihood is calculated for changing that track's hypothesis [36].

3.2.6 Calorimeter system

The calorimeter system is used to measure the energy deposition and the positions of electrons, photons and hadrons. It also provides particle identification information for these particles. As mentioned before, the calorimeter system consists of SPD, PS, ECAL and HCAL. The PS detector can contribute to the differentiation of eletrons against charged pions and SPD contributes to the identification between charged particles and neutral particles e.g. π^0. Due to the fact that the hits density varies largely across the surface of the calorimeter, it is divided into different regions for different cell sizes. From the inner side to the outer side, SPD/PS and ECAL are divided into three regions and HCAL is divided into two regions. The inner region has finer cells, while the outer region has larger cells [28].

The calorimeter system plays an important role in the hardware trigger. Based on the energy deposited and the position of the energy cluster, a variable called

transverse energy E_T can be calculated. The transverse energy is evaluated in clusters of 2×2 cells. It is calculated as follows:

$$E_T = \sum_{i=1}^{4} E_i \sin\theta_i \tag{3.1}$$

where E_i stands for the energy measured in cell i and θ_i stands for the angle between the z axis and the line connecting the center of cell i and the proton-proton interaction point [37]. The transverse energy (E_T) is a crucial variable used in the hardware trigger system. Further discussion on the hardware trigger can be found in section 3.4.

3.2.7 Muon detection system

The muon system is used for detecting and selecting muons with high transverse momentum p_T in the trigger and the muon identification. It consists of 5 stations (M1-M5). M1 is placed before the calorimeters and used to improve the p_T measurement for the hardware trigger. M2-M5 are placed downstream of the calorimeter. Except for the inner part of M1 which is constructed as triple-GEM detectors, the other parts of the five muon stations are constructed as multi-wire proportional chambers. In between M2-M5, there are iron absorber layers used for selecting penetrating muons [38].

Some PID information can be derived from the muon system: A binary variable IsMuon is define by finding hits in the field of interest (FOI) derived from the track extrapolation. Based on the momentum of a track, it is expected to go through a certain number of muon stations if it is assumed to be a muon. IsMuon is triggered if the expected hits are found. Given the extrapolation of the track, the hit distribution in the muon stations will be different for the muon and the no-muon hypothesis. The distribution of a distance variable is calculated and a difference of the log-likelihood for the muon and the no-muon hypothesis is derived as muDLL. This muDLL variable can also be combined with the PID information from RICH and the calorimeters [38].

3.3 LHCb data flow and its softwares

Given the 40MHz bunch crossing rate provided by LHC, the LHCb experiment collects a large amount of data during running. A procedure of data flow and a set of software packages are designed and built to process this huge amount of data. Fig. 3.6 shows the data flow of LHCb during Run2 and the software packages used at each data processing step.

The data from the LHCb detector are firstly filtered with triggers before they are stored. The trigger system of LHCb consists of the hardware trigger and the software trigger. A more detailed description of the trigger system will be found

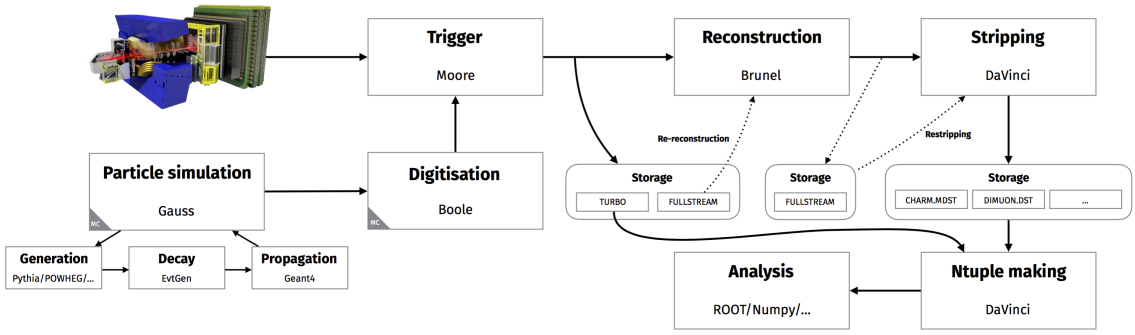


Figure 3.6: LHCb data flow in Run2. The graph was taken from [39].

in section 3.4. The trigger system has been redesigned during the long shutdown between Run1 and Run2. The trigger process is handled by the MOORE software package [40]. At the software trigger level, some reconstructions have already been conducted online. In Run1, online reconstruction performed at software trigger level was poor in quality and thus an offline reconstruction was necessary. The offline reconstruction is performed by the BRUNEL software package [41]. This procedure is called fullstream. However, in Run2, the online system was able to perform a real-time alignment and calibration, and the quality of online reconstruction has been improved to the same level as offline reconstruction. These online reconstructed data are selected online and are stored [42]. This procedure is called the Turbo stream [43]. In Run2, the Turbo stream runs in parallel with the fullstream data flow. In this analysis, the data processed through the fullstream are used.

In the fullstream, the reconstructed data from BRUNEL are stored in DST or microDST format files. These data are further processed through a stripping selection, which is handled by the DAVINCI software package [44]. The stripping procedure contains a set of lines and each line corresponds to a set of selections. The data passing each stripping line are stored into the corresponding locations.

The data through Turbo stream and fullstream both need to be processed by DAVINCI to generate files which are easier to access for analysts. Different software packages are available at the analysis level e.g. ROOT [45] and Numpy.

The simulation data flow in LHCb shares most parts of the real data flow. As shown in Fig. 3.6, particle simulation is controlled by the GAUSS software package [46]. The simulation begins with the generator level simulation, which contains the generation of the particles and their decays. The generation is processed by PYTHIA [47, 48] etc. and the decays are controlled by EVTGEN [49]. The generated particles are then simulated to propagate through the detectors. The propagation is delegated to GEANT4 [50, 51]. The software package BOOLE [52] is in charged of the simulation of the subdetectors' response and their digitisation. The output of BOOLE is then processed similarly as real data, and goes through the trigger system etc.

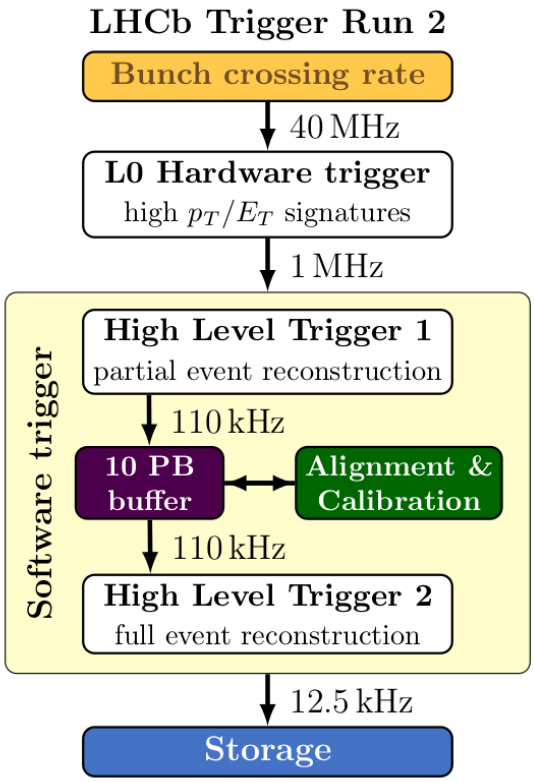


Figure 3.7: An overview of data flow for the LHCb trigger system during Run2. The graph was taken from [42].

3.4 LHCb trigger system

With the large amount of data generated during the proton-proton collisions of LHC, only part of the events which are potentially of interest are recorded. The trigger system of LHCb is dedicated to select such events and reduce the event rate during the process. As shown in Fig. 3.7, the LHCb trigger system during Run2 consists of three levels i.e. the hardware trigger (L0 trigger) and two levels of software High Level Triggers (HLT1 and HLT2). Each level consists of a group of so-called trigger lines, which are algorithms for trigger decisions [42].

Starting with the 40 MHz bunch crossing, the events at LHCb are firstly filtered with the L0 hardware trigger, which searches for high transverse momentum, energy signatures and reduces the event rate to around 1 MHz. As addressed in section 3.2.6, the calorimeter system plays an important part in this process. For the clusters with high E_T measured in HCAL, the E_T in the corresponding cluster of ECAL is added to it, and it is assumed to be a hadron candidate corresponding to the L0Hadron trigger line. Photons and electrons are usually stopped within the ECAL before they can reach the HCAL. The high E_T cluster measured in the ECAL with corresponding PS cells hits are assumed to be photon or electron candidates,

corresponding to the L0Photon and the L0Electron trigger line respectively. The difference between photons and electrons is whether or not a corresponding SPD cell hit can be found. The E_T of the candidates are compared to certain threshold to determine if the corresponding trigger line is triggered [37]. In the Muon station, muon candidates with high transverse momentum are searched. The L0Muon trigger line sets a threshold on the largest p_T while the L0DiMuon trigger line sets a threshold on the product of the largest and the second largest p_T. In addition, requirements on the maximum number of SPD hits are applied for most L0 trigger lines in order to reduce the event complexity [42].

After the hardware trigger, two levels of HLT trigger (HLT1 and HLT2) are applied. The HLT triggers reduce the event rate to around 12.5 kHz. It is the rate at which the data is stored. In HLT1, the events are partially reconstructed, including some tracks and primary vertices (PV, i.e. the proton-proton collision vertices). Due to the clear signature of muons, some muon identification can also be done at HLT1 level. In addition to the event filtering, HLT1 also provides input data for the alignment and calibrations. At HLT2, the constraints on the computing time are not so strict as in HLT1 and the full event reconstruction becomes possible. In addtion to the muon identification which is already available in HLT1, HLT2 is able to fully use the PID information from the detector.

4 Contributing channels

This chapter provides an overview of the decay channels which we are interested at, including the signal channel, the reference channel and some main background channels. The different sources of τ production at LHCb and the calculation of their contribution fractions are discussed in detail. The correct fractions will be used in the processing of Monte Carlo samples.

4.1 Signal τ production

The LFV signal channel $\tau^- \to \phi(K^+K^-)\mu^-$ can contain τ leptons from different sources. It is necessary to evaluate the contribution of these sources.

As described in section 3.2, the LHCb detectors are arranged in the forward region. Only particles entering the acceptance region of LHCb can be detected. The contributions of different τ production channels will first be analyzed in full space (4π solid angle) and later modified by an acceptance factor. The acceptance will be discussed in section 5.2.

The production of τ leptons at LHCb comes mainly from prompt $D_s^\pm, D^\pm$ decays and b hadron decays. The prompt production refers to the direct generation from the primary proton-proton collision vertex. For convenience, in the following part of this article, charge conjugation will be included implicitly and the $D_s^\pm$ and $D^\pm$ will be written simply as D_s and D unless stated otherwise.

The D_s, D mesons decay into τ leptons through the leptonic decay $D_{(s)}^+ \to \tau^+\nu_\tau$ as shown in Fig. 4.1. Branching fractions $\mathcal{B}(D_s \to \tau\nu_\tau) = (5.48 \pm 0.23)\%$ and $\mathcal{B}(D \to \tau\nu_\tau) = (1.2 \pm 0.27) \times 10^{-3}$ are taken from PDG [53].

In the LHCb experiment, D_s, D mesons can come from the prompt generation or from b hadron decays. The prompt charm production cross sections at LHC with a pp collision energy of $\sqrt{s} = 13\,\text{TeV}$ were meausured in [54]. The measured cross

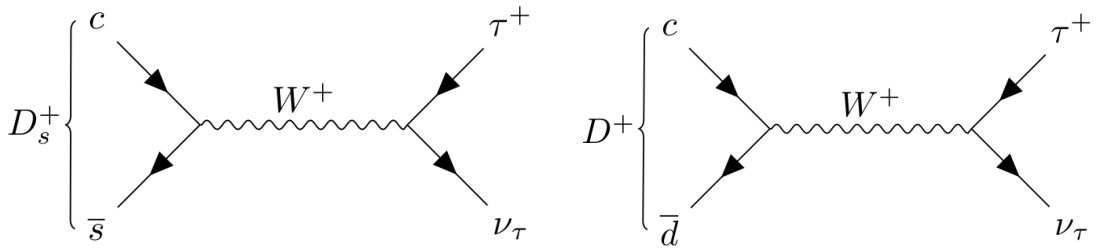


Figure 4.1: D_s, D mesons decay into τ leptons.

sections for p_T value within $1 < p_T < 8\,\text{GeV}/c$ and within the acceptance of the LHCb detector are:

$$\begin{aligned}\sigma^{\text{measured}}(pp\to D_sX) &= 353\pm 76\,\mu\text{b}\\ \sigma^{\text{measured}}(pp\to DX) &= 834\pm 78\,\mu\text{b}\end{aligned} \tag{4.1}$$

Since the LHCb detector is a forward detector, factors of acceptance and momentum range need to be considered, which lead to 4π cross sections:

$$\begin{aligned}\sigma^{4\pi}(pp\to D_sX) &= (1732\pm 373)\,\mu\text{b}\\ \sigma^{4\pi}(pp\to DX) &= (4054\pm 379)\,\mu\text{b}\end{aligned} \tag{4.2}$$

The D, D_s mesons can also come from b hadron decays. To evaluate these two contributions, the $b\bar{b}$ production cross section is taken from [55] as $\sigma^{4\pi}(pp\to b\bar{b}) = 495\pm 52\,\mu\text{b}$. Branching fractions for inclusive decays $b,\bar{b}\to D^{\pm}_{(s)}$ are taken as:

$$\begin{aligned}&\mathcal{B}(b\to D_s^+) = \mathcal{B}(\bar{b}\to D_s^-) = (14.7\pm 2.1)\%\\ &\mathcal{B}(b\to D_s^-) = \mathcal{B}(\bar{b}\to D_s^+) = (10.1\pm 3.1)\%\\ &\mathcal{B}(b\to D^+) = \mathcal{B}(\bar{b}\to D^-) = (23.7\pm 1.8)\%\\ &\mathcal{B}(b\to D^-),\quad \mathcal{B}(\bar{b}\to D^+)\qquad \text{unobserved}\end{aligned} \tag{4.3}$$

from PDG [53]. It is possible to roughly estimate the branching fraction of $b\to D^-(\bar{b}\to D^+)$ from $\mathcal{B}(b\to D_s^-)$:

$$\mathcal{B}(b\to D^-) = \mathcal{B}(\bar{b}\to D^+) \sim \frac{|V_{cd}|^2}{|V_{cs}|^2}\mathcal{B}(b\to D_s^-) \sim 0.5\% \tag{4.4}$$

from here below, we will assume $\mathcal{B}(b\to D^-) = \mathcal{B}(\bar{b}\to D^+) = (0.5\pm 0.5)\%$.

Considering the leptonic decay branching fractions of D, D_s, we derive:

$$\begin{aligned}\mathcal{B}(b\bar{b}\to D_s\to\tau) &= (1.36\pm 0.21)\%\\ \mathcal{B}(b\bar{b}\to D\to\tau) &= (0.029\pm 0.007)\%\end{aligned} \tag{4.5}$$

The last significant source of τ production is from prompt b hadron decays. The inclusive branching fraction for b hadrons decay into τ can be taken as $\mathcal{B}(b/\bar{b}\to\tau\nu_\tau X) = (2.41\pm 0.23)\%$ from PDG [53]. The b hadron production cross section $\sigma(pp\to b\bar{b})$ has been listed above.

To summarize the above derivation, Eq. 4.6 lists the cross sections of interest and their relations with the input variables. The variables on the right-hand side are summarized in Tab. 4.1.

$$\begin{cases}\sigma^{4\pi}(b\bar{b}\to\tau) = 2\sigma^{4\pi}(pp\to b\bar{b})\mathcal{B}(b/\bar{b}\to\tau)\\ \sigma^{4\pi}(b\bar{b}\to D_s\to\tau) = 2\sigma^{4\pi}(pp\to b\bar{b})(\mathcal{B}(b\to D_s^+)+\mathcal{B}(b\to D_s^-))\mathcal{B}(D_s\to\tau)\\ \sigma^{4\pi}(D_s\to\tau) = \sigma^{4\pi}(pp\to D_s)\mathcal{B}(D_s\to\tau)\\ \sigma^{4\pi}(b\bar{b}\to D\to\tau) = 2\sigma^{4\pi}(pp\to b\bar{b})(\mathcal{B}(b\to D^+)+\mathcal{B}(b\to D^-))\mathcal{B}(D\to\tau)\\ \sigma^{4\pi}(D\to\tau) = \sigma^{4\pi}(pp\to D)\mathcal{B}(D\to\tau)\\ \sigma^{4\pi}(\text{prompt } D_s) = \sigma^{4\pi}(pp\to D_s)\\ \sigma^{4\pi}(b\bar{b}\to D_s) = 2\sigma^{4\pi}(pp\to b\bar{b})(\mathcal{B}(b\to D_s^+)+\mathcal{B}(b\to D_s^-))\end{cases} \tag{4.6}$$

Table 4.1: The variables used to calculate the cross sections. To simplify the uncertainty estimation, these variables are considered independent in this thesis.

Variables	Value
$\sigma^{4\pi}(pp \to b\bar{b})$	(495 ± 52) µb
$\sigma^{4\pi}(pp \to D_s)$	(1732 ± 373) µb
$\sigma^{4\pi}(pp \to D)$	(4054 ± 379) µb
$\mathcal{B}(b/\bar{b} \to \tau)$	$(2.41 \pm 0.23)\%$
$\mathcal{B}(b \to D_s^+)$	$(14.7 \pm 2.1)\%$
$\mathcal{B}(b \to D_s^-)$	$(10.1 \pm 3.1)\%$
$\mathcal{B}(D_s \to \tau)$	$(5.48 \pm 0.23)\%$
$\mathcal{B}(b \to D^+)$	$(23.7 \pm 1.8)\%$
$\mathcal{B}(b \to D^-)$	$(0.5 \pm 0.5)\%$
$\mathcal{B}(D \to \tau)$	$(1.2 \pm 0.27) \times 10^{-3}$

Table 4.2: τ production at $\sqrt{s} = 13\,\text{TeV}$

Source	Generation $\sigma^{4\pi}$[µb]	Decay $\mathcal{B}$	$\sigma^{4\pi}(\tau)$[µb]	Contribution in 4π
$b\bar{b} \to \tau$	$2 \times (495 \pm 52)$	$(2.41 \pm 0.23)\%$	23.9 ± 3.4	17.37%
$D_s \to \tau$	1732 ± 373	$(5.48 \pm 0.23)\%$	94.9 ± 20.8	69.09%
$b\bar{b} \to D_s \to \tau$	$2 \times (495 \pm 52)$	$(1.36 \pm 0.21)\%$	13.5 ± 2.5	9.79%
$D \to \tau$	4054 ± 379	$(0.12 \pm 0.027)\%$	4.86 ± 1.2	3.54%
$b\bar{b} \to D \to \tau$	$2 \times (495 \pm 52)$	$(0.029 \pm 0.007)\%$	0.29 ± 0.07	0.21%

As a summary of the five major sources of τ production mentioned above, Tab. 4.2 shows the contribution from each source in 4π.

4.2 Reference channel

While we aim to estimate the branching fraction (or an upper limit of the branching fraction) of the decay $\tau^- \to \phi(K^+K^-)\mu^-$, the quantity we will measure is the yield of the signal channel $N_{\text{sig}} = N(\tau^- \to \phi(K^+K^-)\mu^-)$. The yield is related to the branching fraction and the integrated luminosity by:

$$N_{\text{sig}} = \mathcal{B}(\tau^- \to \phi(K^+K^-)\mu^-)\epsilon_{\text{sig}}\sigma^{4\pi}(\tau)\mathcal{L}_{\text{int}} \tag{4.7}$$

where ϵ_{sig} stands for the efficiency of the signal channel, $\sigma^{4\pi}(\tau)$ stands for the τ production cross section in 4π and $\mathcal{L}_{\text{int}}$ stands for the integrated luminosity.

The number of events of the signal decay $\tau^- \to \phi(K^+K^-)\mu^-$ is evaluated normalising to a well-known reference channel. The uncertainty related to the efficiency can be cancelled partially by using the relative efficiency between the signal and the reference channel. The reference channel chosen is $D_s^- \to \phi(\mu^+\mu^-)\pi^-$. For the

Table 4.3: D_s production at $\sqrt{s} = 13\,\text{TeV}$

Source	Generation $\sigma^{4\pi}[\,\mu\text{b}\,]$	Decay $\mathcal{B}$	$\sigma^{4\pi}(D_s)[\,\mu\text{b}\,]$	Contribution in 4π
$b\bar{b} \to D_s$	$2 \times (495 \pm 52)$	$(24.8 \pm 3.7)\%$	246 ± 45	12.4%
prompt D_s	1732 ± 373	-	1732 ± 373	87.6%

reference channel, there is a similar formula as for the signal channel:

$$N_{\text{norm}} = \mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-)\epsilon_{\text{norm}}\sigma^{4\pi}(D_s)\mathcal{L}_{\text{int}} \tag{4.8}$$

where N_{norm} stands for the yield of the reference channel and ϵ_{norm} stands for the efficiency of the reference channel. We derive:

$$\mathcal{B}(\tau^- \to \phi(K^+K^-)\mu^-) = \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \cdot \frac{N_{\text{sig}}\sigma^{4\pi}(D_s)}{N_{\text{norm}}\sigma^{4\pi}(\tau)}\mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-) \tag{4.9}$$

which will be used to evaluate the signal branching fraction from the observed signal yield N_{sig}.

As described in the previous section, the source of D_s (important for the reference channel) in LHCb can come from direct production at pp collision (prompt D_s) or from the decay of b hadrons. Tab. 4.3 shows the production of D_s mesons from these two sources.

4.3 Crossfeed Background

The main sources of background are D, D_s meson decays, which are incorrectly reconstructed. As described in the previous sections, the production cross sections of D, D_s mesons (including prompt production and b hadron decays) in 4π can be estimated:

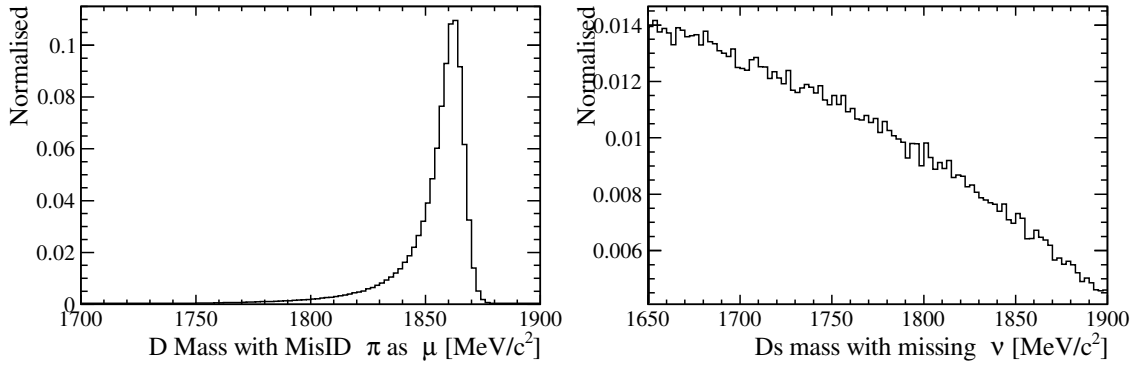
$$\begin{aligned} \sigma^{4\pi}(D_s) &= (1977 \pm 376)\,\mu\text{b} \\ \sigma^{4\pi}(D) &= (4293 \pm 381)\,\mu\text{b} \end{aligned} \tag{4.10}$$

Given the τ mass window and the branching fractions, only some of the D, D_s decay channels need to be considered. Without any selection applied, the 4π contributions for the significant background channels can be roughly estimated. Tab. 4.4 shows several main crossfeed backgrounds, where the branching fractions are taken from PDG [53]. In these background channels, if a charged pion $\pi^\pm$ is misidentified as a muon $\mu^\pm$ and the neutral particle (ν or π^0) is not detected, the decays will have very similar topology as the signal channel. The estimation of the possibility of misidentifying a π as a μ (MisID rate) is complicated and depends on variables such as momentum etc. For a rough estimation of different crossfeed contributions, the misidentification rate is set to an average value of 0.02 [38].

The partially reconstructed backgrounds with a missing neutrino or a π^0 have a broad distribution of the invariant mass of the reconstructed D or D_s, while a fully

Table 4.4: Estimation for some background contributions in 4π

Source	$\sigma^{4\pi}(D_{(s)})$[μb]	$\mathcal{B}$ [%]	MisID rate	4π Contribution[μb]
$D \to \phi(KK)\pi\pi^0$	4293	1.1 ± 0.5	0.02	0.97
$D \to (KK)_{\text{non}\phi}\pi\pi^0$	4293	$1.5^{0.7}_{-0.6}$	0.02	1.3
$D \to KK\pi$	4293	1.1 ± 0.5	0.02	0.85
$D_s \to \phi(KK)\mu\nu_\mu$	1977	0.93 ± 0.25	1	18.5
$D_s \to KK\pi\pi^0$	1977	6.3 ± 0.6	0.02	2.5



(a)$D^- \to K^+K^-\pi^-$ (b)$D_s^+ \to \phi(K^+K^-)\mu^+\nu_\mu$

Figure 4.2: Incorrectly reconstructed $K^+K^-\mu$ mass distributions for two background channels.

reconstructed decay has a peaky distribution. The main source of the broad distribution background comes from $D_s^+ \to \phi(K^+K^-)\mu^+\nu_\mu$ with a missing neutrino. The channel $D^- \to K^+K^-\pi^-$ constitutes a peaky background. The mass distribution of these two channels are explored with RapidSim[1] [56] as shown in Fig. 4.2. The peaky background $D^- \to K^+K^-\pi^-$ distribution is relatively far from the τ mass window and can be easily identified with a fit in the invariant mass distribution. The reduction of the $D_s^+ \to \phi(K^+K^-)\mu^+\nu_\mu$ background requires a multivariate analysis.

[1]RapidSim is a software package for the fast simulation of beauty and charm hadron phase space decays.

5 Data and Monte Carlo samples preparation

In LHCb, before data and Monte Carlo samples are processed and analyzed by analysts, they have already gone through many centralized procedures (e.g. trigger filtering and stripping selections) and have been classified and selected preliminarily. In this chapter, the stripping selections applied to the signal channel and the reference channel will be described.

In addtion, this chapter summarizes some other necessary preparation work for the data and Monte Carlo samples before they can be used in more complicated analysis procedures. The Monte Carlo simulations of the signal are done for different τ lepton production sources, hence a blending procedure with weighted contributions is required. As this analysis aims to search for a very rare decay, a blinding technique is applied to data in order to avoid the possible bias in the analysis. The blinded signal region will not be revealed until the very end of the analysis when the whole analysis procedure is settled. Some preliminary selections (i.e. preselection) and trigger line selections are also described in this chapter.

5.1 Data stripping

As described in section 3.3, the data samples were processed with stripping selections before they are used for analysis. The selection algorithm is included in some stripping lines defined in LHCb. Tab. 5.1 summarizes the stripping selection criteria applied in the lines used for the signal channel and the reference channel. In this table, some variables defined in the LHCb data processing are used. They are explained as follows:

- PV: primary vertex.
- ndf: the number of degrees of freedom associated with the fit. This variable is usually used together with the fit χ^2 for fit quality control.
- χ^2_{vtx}: the χ^2 of a vertex fit. A requirement on this variable selects good quality vertices.
- track χ^2: the χ^2 of a track fit. A requirement on this variable selects good quality tracks.
- χ^2_{IP}: Impact Parameter (IP) χ^2. IP refers to the distance of a track trajectory to a vertex. The χ^2_{IP} here is a measurement of the compatibility of a track to

Table 5.1: Stripping selection [57]

Variables	Signal channel	Reference channel
Final daughters $K^{\pm}\mu^{\pm}\pi^{\pm}$		
track χ^2/ndf	< 3	
track ghost probability	< 0.3	< 0.45
p_{T}	$> 300\,\mathrm{MeV}/c$	
χ^2_{IP} on related PV	> 9	
daughters from ϕ	$K^{\pm}$ PIDK> 0	$\mu^{\pm}$ IsMuon
daughters from ϕ	$K^{\pm}$ ISLONG	-
ϕ		
Mass(ϕ)	$\lvert M - M_{PDG}\rvert < 30\,\mathrm{MeV}/c^2$	$970\,\mathrm{MeV}/c^2 < M < 1070\,\mathrm{MeV}/c^2$
minimal χ^2_{IP} from any PV	> 9	-
χ^2_{vtx}	< 25	-
Mother particle τ/D_s^+		
Mass	$\lvert M - M_{PDG}\rvert < 150\,\mathrm{MeV}/c^2$	$\lvert M - M_{PDG}\rvert < 250\,\mathrm{MeV}/c^2$
$c\tau_0$	$> 50\,\mu\mathrm{m}$	$> 100\,\mu\mathrm{m}$
χ^2_{vtx}	< 25	< 15
χ^2_{IP} on related PV	< 100	< 225

Table 5.2: The requirements to trigger IsMuon [38].

Track momentum	Required muon stations
$3\,\mathrm{GeV}/c < p < 6\,\mathrm{GeV}/c$	M2 and M3
$6\,\mathrm{GeV}/c < p < 10\,\mathrm{GeV}/c$	M2 and M3 and (M4 or M5)
$p > 10\,\mathrm{GeV}/c$	M2 and M3 and M4 and M5

originate from a vertex. It is calculated as the difference of the vertex fit χ^2 with and without the specific track. With a larger χ^2_{IP}, the track is believed less likely to be from the vertex.

- IsMuon: as described in section 3.2.7, it is a binary PID decision based on the number of muon stations where a hit in FOI is found. The specific requirement varies according to the muon track momentum p, which is shown in Tab. 5.2.

- ISLONG: checks if a track is a LONG track i.e. the track goes through Velo, TT, and T1-T3 stations.

- PIDx: a combined delta log likelihood for a given hypothesis x, w.r.t the π hypothesis i.e. $\Delta\mathcal{L} = \log(\mathcal{L}_x/\mathcal{L}_\pi)$. The likelihood $\mathcal{L}$ is derived by simply multiplying the particle identification likelihood produced by RICH, Muon stations and calorimeters.

- M_{PDG}: the nominal mass value of the particle taken from PDG.

- $c\tau_0$: the measured proper lifetime of the particle times the speed of light c.

5.2 Monte Carlo samples

Since the signal τ candidates are produced from different sources, the Monte Carlo samples from different sources need to be blended according to the contribution of each source.

The contribution factors calculated in section 4.1 are in 4π. Considering the LHCb detector acceptance, blending ratios will need to include the acceptance. The acceptance is calculated using generator level simulations. The procedure includes two configurations. Firstly, a normal generator level simulation is performed, with generator level cuts including the LHCb acceptance as well as a requirement of the τ comes from a specific source. The generator level efficiency derived from this procedure is denoted as $\epsilon_{Gen\&Cut}$. The second simulation is performed with the LHCb acceptance cut removed. Only the requirement of the production source is included in the generator level cut for this configuration and the corresponding efficiency is denoted as ϵ_{Gen}. The cut efficiency can be calculated as $\epsilon_{Cut} = \epsilon_{Gen\&Cut}/\epsilon_{Gen}$. The contribution from each source is thus calculated as (contribution in 4π) $\times\, \epsilon_{Cut}$. Normalised contributions are denoted as f_{source} factors which are used for MC blending. We will use the notation $f_{source}(i)$ for the factor of source i. Tab. 5.3 shows the efficiencies and blending factors for the signal channel.

To weight the MC samples to the correct fractions, we evaluate a source weight for each event. If an event belongs to a source i, its source weight (w_{source}) is calculated as:

$$w_{source}(i) = \frac{N_{all} \times f_{source}(i)}{N(i)} \tag{5.1}$$

where $N(i)$ is the number of events of a given source in the MC sample and N_{all} is the total size of the MC sample. As f_{source} factor has already taken the acceptance efficiency into consideration, the number of events is evaluated after the acceptance level.

For the reference channel, the same procedure is applied to the D_s production. Tab. 5.4 shows the fractions of the two sources of D_s production which are used to weight the MC sample.

Table 5.3: MC composition for $\tau^- \to \phi(K^+K^-)\mu^-$ at 13 TeV

Source	Contribution in 4π	$\epsilon_{Gen\&Cut}(\%)$	$\epsilon_{Gen}(\%)$	$\epsilon_{Cut}(\%)$	$f_{source}(\%)$
$b\bar{b} \to \tau$	17.37%	4.67 ± 0.03	31.94 ± 0.08	14.63 ± 0.10	16.42
$D_s \to \tau$	69.09%	14.08 ± 0.04	89.24 ± 0.07	15.78 ± 0.05	70.46
$b\bar{b} \to D_s \to \tau$	9.79%	1.574 ± 0.005	10.73 ± 0.02	14.67 ± 0.06	9.29
$D \to \tau$	3.54%	14.33 ± 0.04	90.30 ± 0.06	15.87 ± 0.05	3.63
$b\bar{b} \to D \to \tau$	0.21%	1.407 ± 0.009	9.64 ± 0.02	14.60 ± 0.10	0.20

Table 5.4: MC composition for $D_s^- \to \phi(\mu^+\mu^-)\pi^-$ at 13 TeV

Source	Contribution in 4π	$\epsilon_{Gen\&Cut}(\%)$	$\epsilon_{Gen}(\%)$	$\epsilon_{Cut}(\%)$	$f_{source}(\%)$
$b\bar{b} \to D_s$	12.4%	1.23 ± 0.03	11.1 ± 0.3	11.2 ± 0.5	11.5
prompt D_s	87.6%	10.8 ± 0.3	88.6 ± 0.9	12.2 ± 0.4	88.5

5.3 Mass distribution and signal region blinding

The invariant mass of the reconstructed signal candidate is one of the most important variables for the analysis. The final extraction of signal will depend on the distribution of the reconstructed mass of the signal candidates.

To avoid the bias of the analysis strategy e.g. event selection optimization, the signal region of the data samples is blinded before the analysis begins. Three regions are defined based on the mass spectrum of the mother particle. They are:

- Signal region: $|M - M_{\mathrm{PDG}}(\tau)| < 20\,\mathrm{MeV}/c^2$. This region is blinded.
- Inner SideBand(SB): $20\,\mathrm{MeV}/c^2 < |M - M_{\mathrm{PDG}}(\tau)| < 30\,\mathrm{MeV}/c^2$. This region is used for multivariate analysis training.
- Outer SideBand(SB): $|M - M_{\mathrm{PDG}}(\tau)| > 30\,\mathrm{MeV}/c^2$

Fig. 5.1 shows the distribution of the reconstructed mass of the mother particle after stripping selection. The signal region is already blinded and the boundaries of the sidebands are marked.

5.4 Sample preselection

After the stripping selection, the data samples are selected further. It is possible to apply a preselection on the ϕ mass to distinguish between the background and the signal. Fig. 5.2 shows the ϕ mass distribution of the signal channel and the reference channel.

Based on the different ϕ mass peak distributions of the signal channel and the reference channel, we require different ϕ mass selections for them. In the signal channel the selection is chosen as $|M(\phi) - M_{\mathrm{PDG}}(\phi)| < 15\,\mathrm{MeV}/c^2$ and in the reference channel the selection is $|M(\phi) - M_{\mathrm{PDG}}(\phi)| < 20\,\mathrm{MeV}/c^2$.

In the signal channel, crossfeed from $D^- \to K^+K^-\pi^-$ mostly falls into the peaky shape of the upper sideband (Fig. 5.1) and can be easily described by a fit to the mass spectrum. However, in order to concentrate on the main broad background distribution during the sample selection, a cut is applied to reduce $D^- \to K^+K^-\pi^-$ crossfeed during the data sample preselection. Under the hypothesis that the muon candidate is a misidentified pion, the invariant mass of the mother particle is recalculated after assigning the π mass to the μ track in data samples. This recalculated mass is marked as variable $\mathrm{M(mother}; \mu \to \pi)$. Fig. 5.3 shows this new mass distribution of the MC sample and the blinded data. In the data preselection, we require: $\mathrm{M(mother}; \mu \to \pi) < 1850\,\mathrm{MeV}/c^2$, which removes a large fraction of the background with a misidentified π. This will facilitate the optimization of further selections since these selections will focus on distinguishing the signal and the continuously distributed background. In the final fit, to avoid biasing the mass distribution, the cutted data (the sample removed by cut $\mathrm{M(mother}; \mu \to \pi) < 1850\,\mathrm{MeV}/c^2$) will be added. This selection is not applied to the MC sample.

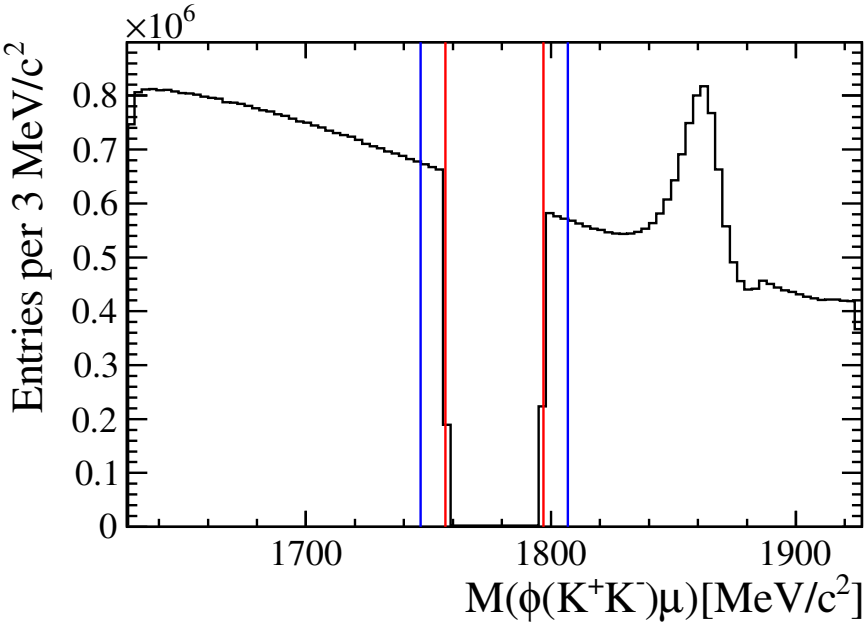
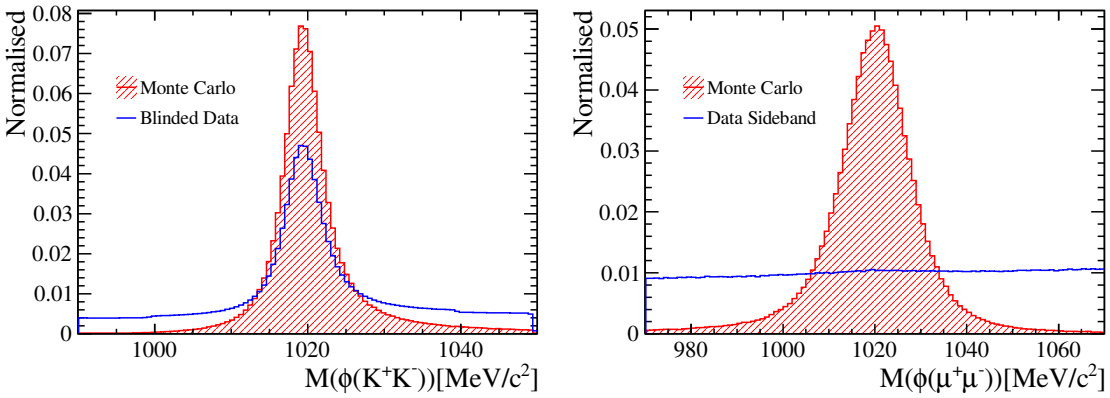


Figure 5.1: Invariant mass spectrum of mother particle in the signal channel. The red lines mark the boundaries of the signal region ($|M - M_{\mathrm{PDG}}| = 20\,\mathrm{MeV}/c^2$), and the blue lines mark the outer boundaries of the inner sideband ($|M - M_{\mathrm{PDG}}| = 30\,\mathrm{MeV}/c^2$).



(a) Signal Channel (b) Reference Channel

Figure 5.2: Invariant mass distribution of ϕ candidates in the signal channel ($\phi \to K^+K^-$) and the reference channel ($\phi \to \mu^+\mu^-$). In the plot of the signal channel, "blinded data" is defined by removing signal region in the $M(\phi(K^+K^-)\mu)$ distribution. In the plot of the reference channel, the data sideband is selected by requiring $20\,\mathrm{MeV}/c^2 < |M(D_s) - M_{PDG}(D_s)| < 50\,\mathrm{MeV}/c^2$.

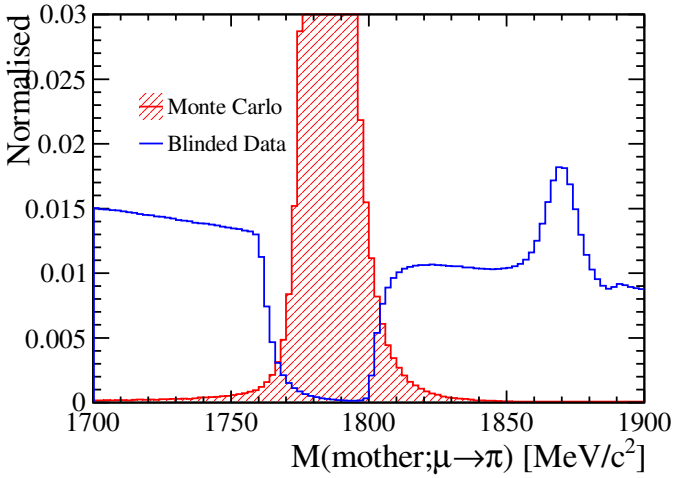


Figure 5.3: In signal channel, the mass of mother particle after assigning π mass to μ track.

Table 5.5: Preselection for the signal channel and the reference channel.

Signal channel	Reference channel
$\lvert M(\phi) - M_{\mathrm{PDG}}(\phi)\rvert < 15\,\mathrm{MeV}/c^2$	$\lvert M(\phi) - M_{\mathrm{PDG}}(\phi)\rvert < 20\,\mathrm{MeV}/c^2$
$M(\mathrm{mother};\mu \to \pi) < 1850\,\mathrm{MeV}/c^2$	$-0.01\mathrm{ns} < \tau_0(D_s) < 0.025\mathrm{ns}$
	D_s DIRA> 0.99
	μ ProbNN$\mu > 0.2$ and π ProbNN$\pi > 0.2$

For the reference channel, additional selections are applied. Tab. 5.5 provides a summary of preselections applied to the signal channel and the reference channel. In the table, some variables and their corresponding notations are used. The variable $\tau_0(D_s)$ stands for the measured decay time of the reconstructed D_s candidates. The DIRA (DIRection Angle) of the D_s candidate stands for the cosine of the angle between the momentum vector of the D_s candidate and the direction vector which points from the primary vertex to the end vertex of the D_s. ProbNN is a response of a neural network which combines the particle identification information from several subdetectors (including RICH, the muon system and the calorimeter system) using a multivariate technique [58]. The variable ProbNNx is normalised in the range $0 \sim 1$, with higher value suggesting higher probability of a particle belonging to a certain species x.

5.5 Trigger selection

As mentioned in section 3.3, the recorded data has been filtered using 3 levels of triggers: L0, HLT1 and HLT2. Each level has many different lines, which apply different requirements. In the data sample, the response for each line is recorded.

Table 5.6: Selected trigger lines for the signal channel.

Trigger Level	Signal channel trigger line
L0	L0Muon (TOS)
	L0Global (TIS)
HLT1	HLT1TrackMuon (TOS)
	HLT1TrackMuonMVA (TOS)
	HLT1TrackMVA (TOS)
HLT2	HLT2PhiIncPhi (TOS)

Table 5.7: Selected trigger lines for the reference channel.

Trigger Level	Reference channel trigger line
L0	L0Muon (TOS)
	L0DiMuon (TOS)
HLT1	HLT1DiMuonLowMass (TOS)
	HLT1TrackMuon (TOS)
HLT2	HLT2DiMuonDetached (TOS)[1]
	HLT2TopoMuMu2Body (TOS)

Each triggered event and the decay candidate of our interest in the event can be categorized for its response of a certain trigger line. Triggered On Signal (TOS) for a given line stands for the events, in which the signal candidate is sufficient to trigger the line. Triggered Independent of Signal (TIS) for a given line stands for the events, in which if the signal candidate is removed, the rest of the event is still enough to trigger the line. The events belonging to neither TOS nor TIS are called Triggered On Both (TOB) [59].

To simplify the calculation of the trigger efficiency and to improve the systematic uncertainty associated with it, several trigger lines are selected to be used in the further analysis. These trigger selections impose the requirement on the sample that for each trigger level, the selected event should at least trigger one of the selected trigger lines. Tab. 5.6 and Tab. 5.7 show the trigger lines selected for the signal channel and the reference channel.

These trigger lines involve complicated requirements. An overview of the L0 trigger lines has been given in section 3.4. Further information on the exact selections contained in these trigger lines are described in the appendix A.1.

[1]This line doesn't exist for the samples of 2016, thus, these samples are required to pass the HLT2TopoMuMu2Body (TOS) line

6 Kinematic correction of Monte Carlo simulation

In order to improve the Monte Carlo sample's consistency with the real data, a correction in kinematics is applied. As the simulation can not perfectly match the kinematic distribution of real data, some correction weights are assigned to the simulated events based on the value of kinematic variables i.e. the momentum of the mother particle $p(D_s/\tau)$, the transverse momentum of the mother particle $p_T(D_s/\tau)$ and the track multiplicity nTracks (meaning the number of tracks for an event).

To get the correction, a comparsion of the kinematic variable distributions between the simulated sample and the signal component of the real data sample is needed. Assuming the discrepancy between the simulation and the real data is the same for the signal channel and the reference channel, the correction can be determined on the reference channel, where it is possible to have an estimation of the real data distribution with only the signal component.

Since the real data sample is a mixture of signal events and background events, we need a method to extract the signal-only distribution from the real data sample. The method is called the *sPlot* technique [60].

The information of the mother particle's invariant mass is used to distinguish the signal from the background. Firstly, a fit of the mass spectrum of the mother particle is performed for the reference channel. The fit describes the signal and the background components of the mass spectrum. It allows to assign a weight to each event depending on its position in the mass spectrum. The weighted distributions of the kinematic variables are taken as the distributions of the signal component. By comparing the weighted data sample and the MC sample, a correction table is calculated as a 3 dimensional histogram ($p(D_s/\tau)$, $p_T(D_s/\tau)$, nTracks). The MC sample of the signal channel is weighted by applying this table.

6.1 *sPlot* technique

Following the description in [60], the *sPlot* technique is summarized in this section.

sPlot is a technique used on a group of events from different sources (in this case, the signal and the background), which are associated with two sets of variables. The distribution models of one set of the variables are known and they are called discriminating variables (in this case, the invariant mass). The other set of variables with unknown distributions are called control variables (in this case, $p(D_s/\tau)$, $p_T(D_s/\tau)$ and nTracks). It is assumed that the discriminating variables and the control variables are independent.

The distribution of the discriminating variable (the invariant mass) is described by a composition of contributions from different sources (signal/background). A maximum likelihood fit of the distribution of the discriminating variable can be performed. The fit will provide an estimation of the expected yield (number of events) for each source. Given the fit result, for each event, a weight (sWeight) can be calculated as a function of its discriminating variable value. The sWeighted distributions of control variables will then provide an estimation of the distributions of the individual sources.

6.2 Reference channel fit

In order to determine the sWeights, the fit of the data of the reference channel is performed. The invariant mass spectrum of the selected candidates is modelled with two parts i.e. the signal distribution and the combinatorial background distribution.

The signal peak is modelled as a Double-Sided Crystall Ball (DSCB) function [61], which can be formulated as:

$$f(x) \propto \begin{cases} e^{-\frac{(x-M)^2}{2\sigma^2}}, \text{ if } \dfrac{x-M}{\sigma} > -\alpha_1 \text{ and } \dfrac{X-M}{\sigma} < \alpha_2 \\ \left(\dfrac{n_1}{|\alpha_1|}\right)^n_1 e^{-\alpha_1^2/2} \left(\dfrac{n_1}{|\alpha_1|} - |\alpha_1| - \dfrac{x-M}{\sigma}\right)^{-n_1}, \text{ if } \dfrac{x-M}{\sigma} < -\alpha_1 \\ \left(\dfrac{n_2}{|\alpha_2|}\right)^n_2 e^{-\alpha_2^2/2} \left(\dfrac{n_2}{|\alpha_2|} - |\alpha_2| + \dfrac{x-M}{\sigma}\right)^{-n_2}, \text{ if } \dfrac{x-M}{\sigma} > \alpha_2 \end{cases} \quad (6.1)$$

where M and σ stand for the mean value and the width of the central gauss distribution, $\alpha_1, \alpha_2, n_1, n_2$ are the tail parameters which define the shape of the two tails. To improve the fit of the data, the tail parameters are obtained by performing a fit of the MC sample. The MC sample is weighted with the source weights calculated in section 5.2 in order to correct the production sources of D_s. The fit result of the MC sample is shown in the left plot of Fig. 6.1.

With the tail of the signal peak fixed, the data of the reference channel is fitted with the DSCB function as the signal model and an exponential function as the combinatorial background model. The data fit is shown in the right plot of Fig. 6.1.

Tab. 6.1 shows the parameters of the fit.

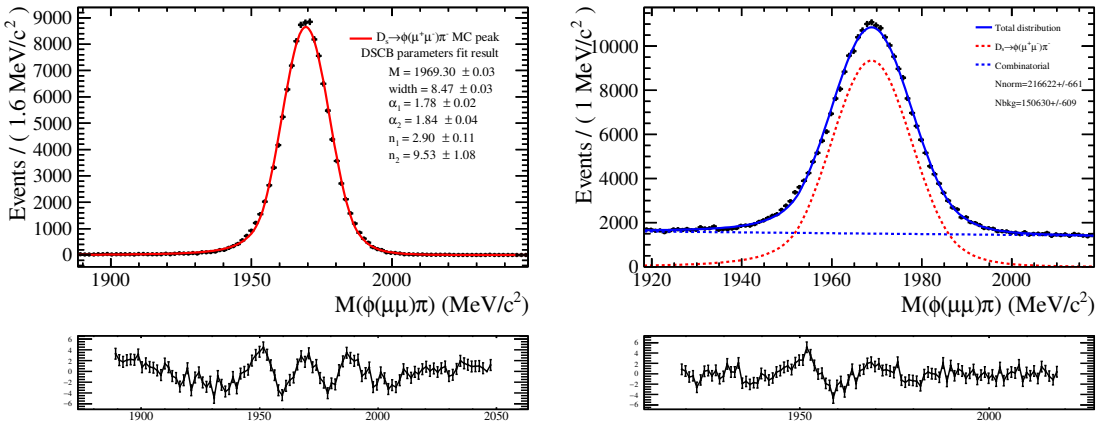


Figure 6.1: Fit of the invariant mass distribution for the reference channel MC(left) and data(right). The MC sample is corrected for different D_s production sources as described in section 5.2.

Table 6.1: Parameters of the invariant mass fit for the reference channel

Parameters	MC	Data
Signal (DSCB)		
$M(D_s)$	1969.30 ± 0.03	1968.85 ± 0.03
width σ	8.47 ± 0.03	8.91 ± 0.03
α_1	1.78 ± 0.02	fixed from MC
α_2	1.84 ± 0.04	fixed from MC
n_1	2.90 ± 0.11	fixed from MC
n_2	9.53 ± 1.08	fixed from MC
Background (exponential)		
exponential constant	-	$(-1.27 \pm 0.10) \times 10^{-3}$

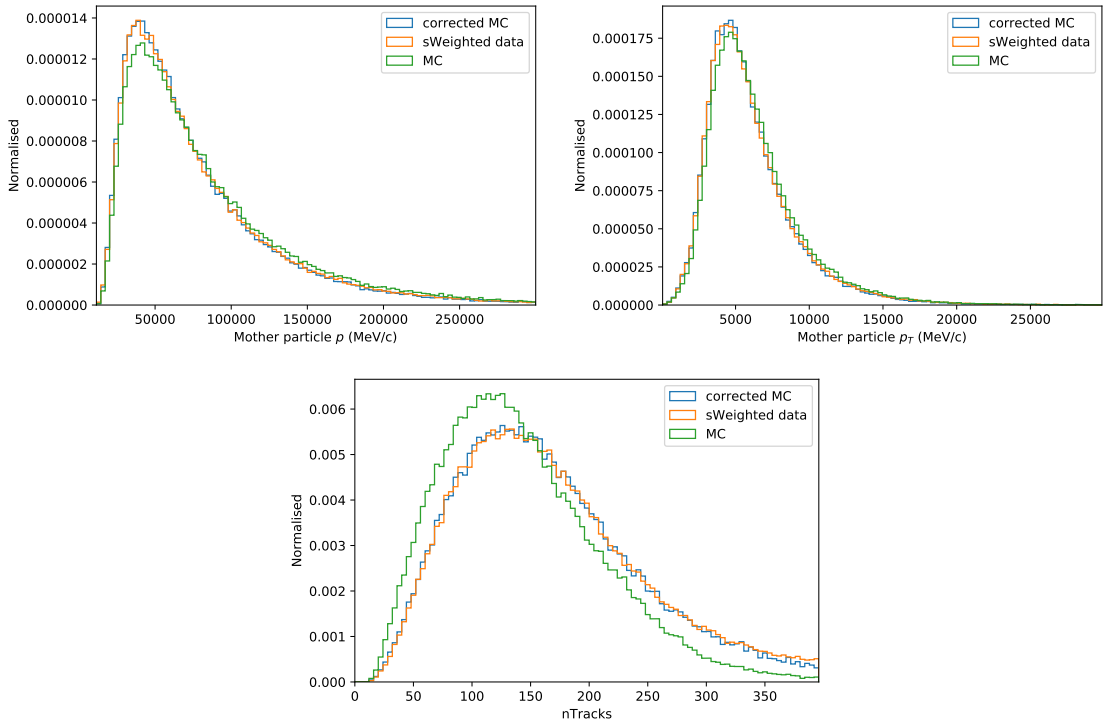


Figure 6.2: Comparison of the distributions of correction variables. The corrected MC shows a better match with the sWeighted data.

6.3 Correction table

Based on the fit result shown in Fig. 6.1(right), the real data signal contribution for different kinematic variables of the reference channel are derived with the *sPlot* technique.

The sWeights are assigned to each event in the data sample. In order to evaluate the correction, the sWeighted data and the MC sample are compared in the 3 dimensional binned distribution ($p(D_s/\tau)$, $p_T(D_s/\tau)$, nTracks). For each bin (bin_x), the correction weight (w_{cor}) is calculated as the ratio of two factors:

$$w_{cor}(\text{bin}_x) = \frac{f_{\text{swdata}}(\text{bin}_x)}{f_{\text{MC}}(\text{bin}_x)} \tag{6.2}$$

where the factor f_{swdata} stands for the fraction of the sWeighted data events belonging to the given bin and the factor f_{MC} stands for the fraction of the MC events belonging to the given bin. In the MC sample, the D_s production weights (section 5.2) have been applied.

As a test, the correction table is applied to the reference channel itself. The corrected distribution of the reference channel MC is compared with the sWeighted data and the non-corrected MC. The result is shown in Fig. 6.2. As expected, Fig. 6.2 shows that the corrected MC matches the sWeighted data rather well.

7 Multivariate analysis

The multivariate (MV) technique is useful when the variables individually do not supply enough separation power to extract signal candidates from a sample with a large amount of background events. By performing the MV analysis, the information from several input variables are combined into a more powerful single variable.

Several variables with potential power to distinguish signal from background are selected as input variables. A classifier is able to assign an event to the category of signal or background based on the value of its input variables. The classifier can be trained to do this by adjusting its structure parameters while learning from a training sample, which contains a group of events with input variables and known labels to be either signal or background. In this analysis, the MC sample is used as the signal training sample and the inner sideband of the data sample is used as the background training sample. The trained classifier is then applied to the full data and MC samples. The classifier evaluates an output value for each candidate suggesting an extent to which the candidate is believed to be a signal event.

In this analysis, the multivariate analysis training is performed with the machine learning library scikit [62] and the CatBoost [63] toolkit.

7.1 Introduction to classifiers

A classifier is essentially a mapping from an input variable space to an output space. Before it can be used, it is trained on a training sample which has a known target value in the output space. During the training of a classifier, its structure parameters are adjusted trying to perform the best classification. More specifically, this adjusting process can be done by minimizing a function which takes its input from the classifier output and the known target value of the training sample. This function is called the loss function.

Several constraints can be set prior to the training in order to control it. The parameters concerning these constraints are called hyperparameters.

The classifier used in this analysis is CatBoost. Based on the content of [63] and [62], an introduction of CatBoost is given below.

CatBoost is an improved version of the algorithm Gradient Boosting Decision Tree (GBDT). GBDT is an ensemble method classifier, which means it is applied by combining the results from a group of primitive estimators. GBDT uses Decision Trees (DT) as primitive weak learners and add the functions of these weak learners with certain weights. A decision tree is a basic classifier. It consists of a set of nodes, in which the samples are recursively partitioned in a way that benefits the classification of test samples to match their known labels. The space of input variables are

finally partitioned into disjoint regions R_i. The final output of a decision tree can be generally written as

$$h(\vec{x}) = \sum_i b_i \mathbb{I}_{\{\vec{x} \in R_i\}} \tag{7.1}$$

where $\vec{x}$ stands for a point in the input variable space and $\mathbb{I}_{\{\vec{x} \in R_i\}}$ is the characteristic function which only contributes when $\vec{x}$ belongs to the region R_i. The values b_i are assigned as the predicted class labels for this region. There is a hyperparameter to limit the maximum depth of a DT. In GBDT, these basic decision trees are determined one by one so that every basic decision tree added to the ensemble is used to try to minimize the loss function of the previous ensemble in a greedy fashion. Step by step, the GBDT classifier is constructed as a sequence F^t:

$$F^t = F^{t-1} + \alpha_{lr} h^t \tag{7.2}$$

where h^t is the base estimator of the step t and α_{lr} is called the learning rate, which is introduced as a regularization strategy. The estimator h^t is determined to minimize the loss function $\mathcal{L}$:

$$h^t = \arg\min \mathcal{L}(F^{t-1} + h) \tag{7.3}$$

As proposed in [64], by training each base classifier on a randomly selected subsample of the full sample, it helps to improve the performance of the classifier. A hyperparameter is used to constrain the fraction of samples used for training each base classifier.

GBDT is a powerful classifier and has been used widely for many years. However, it has certain prediction shift problems which can be considered as a special kind of target leakage. CatBoost improves GBDT by introducing an ordered boosting algorithm and a new algorithm for processing categorical features [63].

7.2 Training input

The training use the MC sample as the signal sample and the inner sideband of the data sample as the background sample. The full sample is divided into two parts i.e. 3/4 used for developing the classifier and 1/4 used for the final evaluation of the classifier. In order to correct the τ lepton production fractions from different sources and the kinematic variable distributions (i.e. $p(\tau), p_T(\tau)$ and track multiplicity), we use the product of the source weight and the kinematic correction from the reference channel to calculate the final training weight. In addition, the sizes of the signal sample and the background sample are different, and thus they need to be balanced. The final training weight (w_{train}) for events in the signal category is calculated as:

$$w_{train}(i) = N_{\text{Bkg.}} \cdot \frac{w_{cor}(i) \cdot w_{source}(i)}{\sum_{k \in \text{sig.}} w_{cor}(k) \cdot w_{source}(k)} \tag{7.4}$$

where i and k are labels for events. The symbol $N_{Bkg.}$ stands for the size of the background sample. For events in the background category, the weights are simply set to one.

The input variables for the classifier are listed below with the explanation of the notations. The distribution of these input variables are shown in Fig. 7.1. Due to the similarity between the signal and the background, only very few variables provide a seperation power.

- τ IP: impact parameter of the τ candidate w.r.t PV.
- τ χ^2_{IP}: IP χ^2 of the τ candidate w.r.t PV.
- ϕ p_T: p_T of the $\phi(KK)$ candidate .
- μ p_T: p_T of the μ candidate.
- μ χ^2_{IP}: IP χ^2 of the muon candidate w.r.t PV.
- minimal $p_T(K)$: the minimum of the p_T of the two kaons.
- minimal $\chi^2_{\text{IP}}(K)$: the minimum of the IP χ^2 of the two kaons w.r.t PV.
- sinDIRA: sine value of the angle between the τ momentum and the vector which points from PV to its decay vertex.
- Pointing: calculated as $\frac{p(\tau)\cdot\text{sinDIRA}}{p(\tau)\cdot\text{sinDIRA}+p_T(\phi)+p_T(\mu)}$
- TrkIsoBDT of K^+, K^- and μ: the maximum value of a track isolation variable evaluated between a certain final state particle ($K^\pm, \mu^\pm$) and another track in the remaining part of the event. In the signal decay, the final state particles $K^\pm, \mu^\pm$ are the only particles coming from τ. For the τ production from $D_s^+ \to \tau^+\nu_\tau$ and $D^+ \to \tau^+\nu_\tau$, there are no other charged particles produced. Thus the daughter tracks are expected to be relatively isolated. For each daughter track and any other track in the event, an isolation variable is evaluated with BDT multivariate analysis technique. This variable was developed during previous studies on $B_s^0/B \to \mu^+\mu^-$ at LHCb [65, 66].
- Mother particle lifetime : measured life time of τ candidates

7.3 Hyperparameter optimization

Before the training of the final classifier, an optimization of some hyperparameters of the classifier is performed. Several combinations of hyperparameters are tested with a grid search function provided by the CatBoost toolkit:

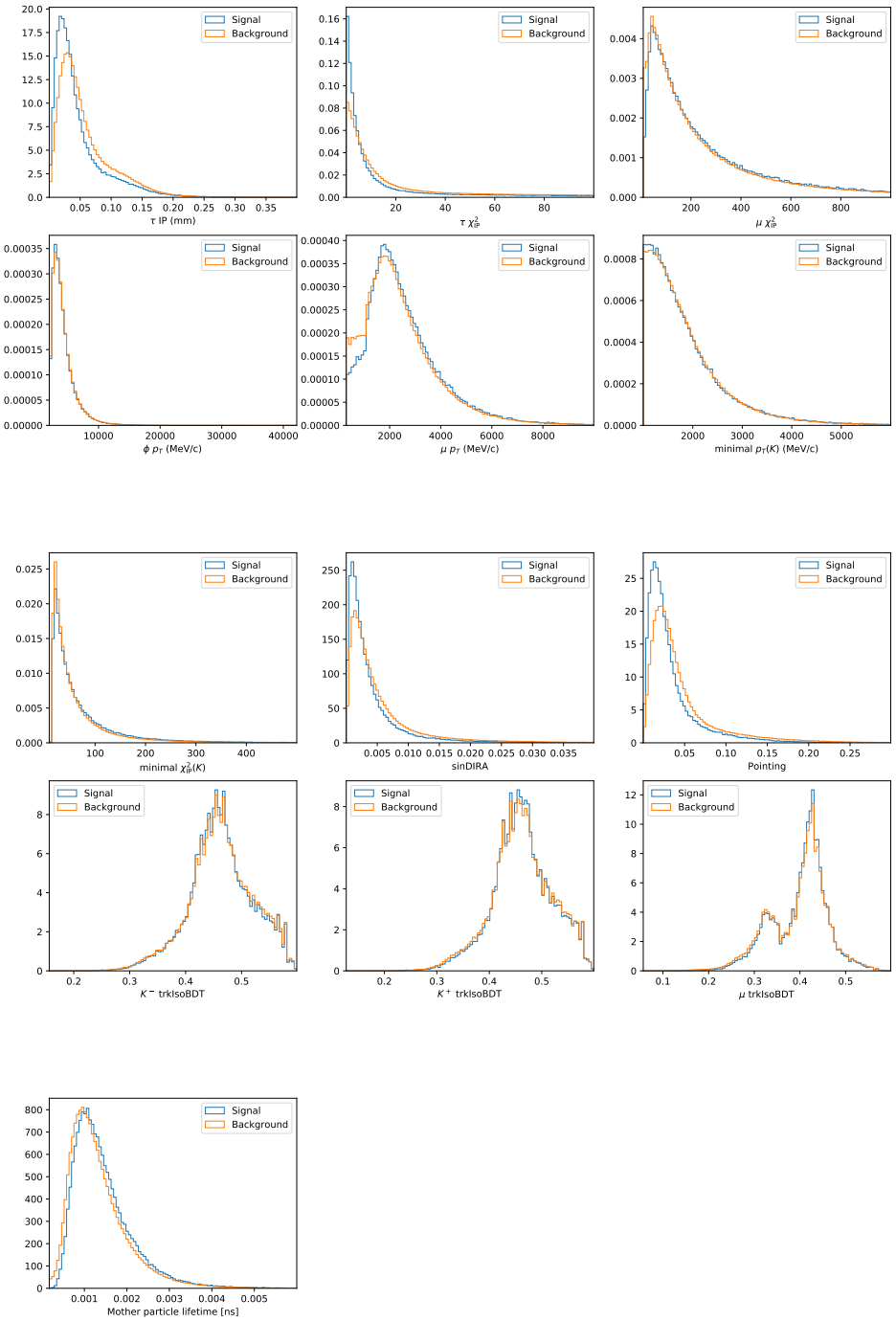


Figure 7.1: Multivariate analysis training input distribution for the background (data inner sideband) and the signal (MC). The distributions are normalised.

- Learning rate: α_{lr} = 0.03, 0.1, 0.3. As described in Eq. 7.2, the learning rate works as a step size of the training iteration.

- The maximum depth of the trees: depth = 4, 5, 6, 10. This hyperparameter sets a limit on the number of allowed node layers for each basic decision tree.

- Sample rate for bagging: subsample = 0.3, 0.5, 0.6, 0.66, 0.7, 0.8. As introduced in section 7.1, a randomly selected subset of the total training sample is used at each iteration. This hyperparameter determines the fraction of the subset within the total sample.

- The maximum number of basic trees that can be built: iterations = 400, 1000. When the number of basic trees reaches this number, the training will stop. If this number is too large, there is a risk of overtraining.

The tests are conducted with a 3-fold cross-validation technique. This technique randomly splits the total developing sample into 3 parts. Each time the classifier is trained with 2 of them and the result is evaluated on the third part. This technique is intended to prevent over-training during hyperparameter tuning.

The grid search picks out the best hyperparameters combination: α_{lr}=0.1, depth=6, subsample=0.5, iterations=1000. However, the training result of this combination shows a considerable extent of over-training. The output is shown in appendix A.2. To suppress the overtraining, the outputs of several combinations with iterations = 400 are compared with the result in appendix A.2. With a 2nd set of hyperparameters: α_{lr}=0.1, depth=6, subsample=0.6, iterations=400, the output result (a description of this output can be found in the next section) is comparable with that of the best hyperparameters combination. We do not lose too much ability while no obvious overtraining is observed. The following analysis will use this combination of hyperparameters.

7.4 Training output

To evaluate the performance of the classifier, the trained classifier is applied to the evaluation sample. For each event, the trained CatBoost classifier evaluates an output based on the value of input variables. The left plot of Fig. 7.2 shows the normalised distribution of the trained classifier. Both the training sample(i.e. the developing sample) and the evalution sample are plotted as histograms. The output of CatBoost will be used as a selection for the signal: CatBoost $> x$. By scanning the cut value x, a group of background rejection rates and corresponding signal retention rates are evaluated based on the output distribution of evaluation sample. The result is plotted as Receiver Operating Characteristic (ROC) curve as shown in the right plot of Fig. 7.2

The distribution of the classifier output shows a large overlap between the signal and the background. The ROC curve is an important criterion of the performance.

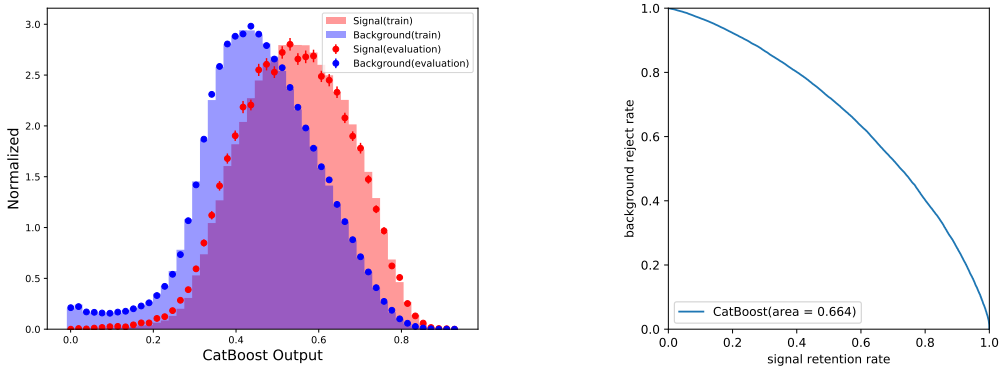


Figure 7.2: The performance of the trained classifier. The left plot shows normalised the CatBoost output distribution for signal/background training/evaluation samples respectively. The right plot shows the ROC curve (the background reject rate against the signal retention rate) and the area under the curve.

With a better performance, the area under the ROC curve will be larger. Fig. 7.2 indicates that the performance of the classifier is limited. The reason is the similarity between the signal and the background. The main background in this case is the $D_s^+ \to \phi(K^+K^-)\mu^+\nu_\mu$ decay. Comparing with the signal decay $\tau^- \to \phi(K^+K^-)\mu^-$, the final state particles only differ by a missing neutrino. The topological structure of these two decays are very similar and to disentangle them is a difficult task. Thus we consider the result of this MV analysis acceptable and we believe further improvements are difficult to realize.

Because the final extraction of signal will rely on the invariant mass spectrum, it is important that the MV selection is uncorrelated to the invariant mass i.e. the selection on CatBoost output itself does not cause any potential peaky structure in the signal region. Several two dimensional distribution histograms are produced for MC, data inner sideband and data outer sideband samples to test the correlation between the CatBoost output and the variable $|M(\phi(KK)\mu) - M_{\mathrm{PDG}}(\tau)|$. The results are shown in Fig. 7.3. No obvious correlation between these two variables is found.

Finally, the trained CatBoost classifier is applied to all data and MC samples. To suppress the background while preserving the potential signal candidates, the selection CatBoost $> x$ will be applied with x remained to be optimized later.

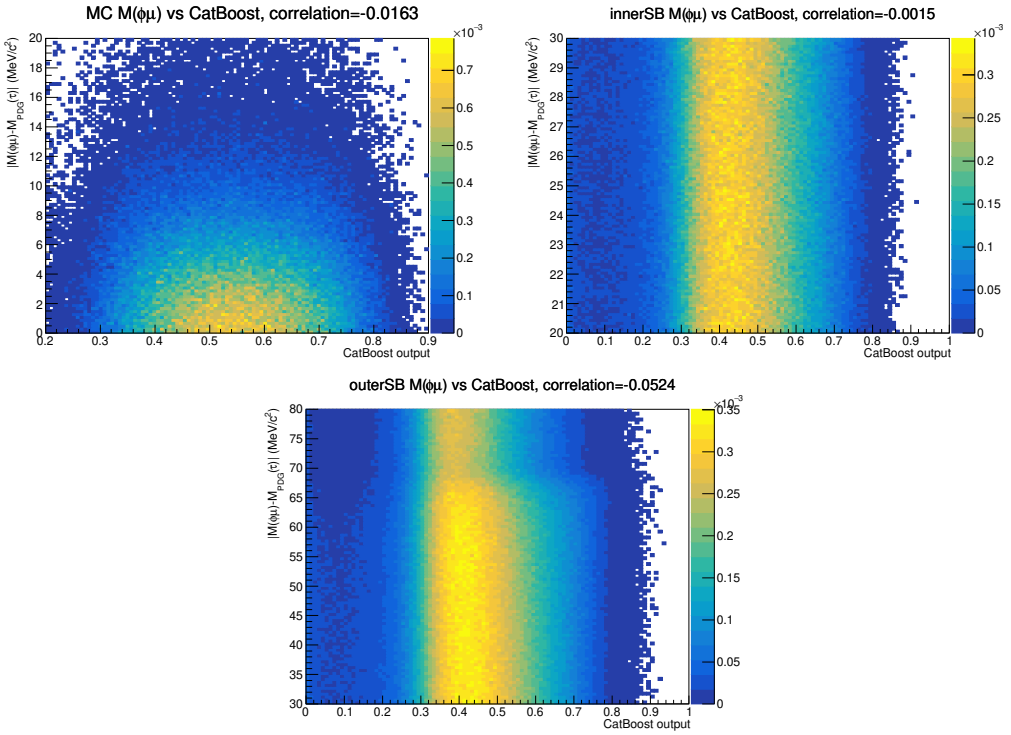


Figure 7.3: Two dimensional distribution of the CatBoost output and the invariant mass variable $|M(\phi(KK)\mu) - M_{\mathrm{PDG}}(\tau)|$ for MC, data inner sideband (innerSB) and data outer sideband (outerSB) samples. The colored z axis is in normalised scale. The linear correlation factors are shown in the sub-titles of the plots.

8 Multivariate and PID selection optimization

In the previous chapter, the output of the multivariate classifier provides an useful variable for the signal/background classification. In addition, the particle identification information of the final state particles in the signal decay $\tau^- \to \phi(K^+K^-)\mu^-$ has not been fully exploited so far. This chapter describes a selection optimization procedure which combines the MV output and the PID information. During the optimization, the evaluation of the PID selection efficiency is needed. A PID efficiency calibration method is introduced, which will contribute to a more accurate optimization.

8.1 PID selection strategy

8.1.1 Kaon PID selection

In the signal decay $\tau^- \to \phi(K^+K^-)\mu^-$, the $K^\pm$ candidates could be misidentified muons or some other particles. Thus, two selections are applied to both kaon candidates: ProbNNk > 0.4 and IsMuon is set to false.

The cut point of ProbNNk is determined based on a rough estimation from the sideband/MC comparsion. Fig. 8.1 shows the distribution of the kaon ProbNNk variable.

As introduced in section 3.2.7 and section 5.1, the variable IsMuon is a very loose binary selection of muon candidates. By requiring its value to be false, we use it to veto the potential false kaon candidates.

8.1.2 Muon PID selection

As described in section 4.3, possible crossfeed backgrounds e.g. $D_s \to KK\pi(\pi^0)$ or $D \to KK\pi(\pi^0)$ may contribute due to a misidentified muon. To suppress these crossfeed backgrounds, a selection on the muon is applied based on a combined muon PID variable:

$$\text{MuPIDvar} = \text{ProbNN}\mu \cdot (1 - \text{ProbNNk}) \cdot (1 - \text{ProbNNp}) \tag{8.1}$$

Fig. 8.2 shows the distribution of this variable as a comparsion between the MC (signal) sample and the data inner sideband (background) sample. This combination of the ProbNN variables was tested in a previous study on $B^0 \to \mu^+\mu^-$ conducted

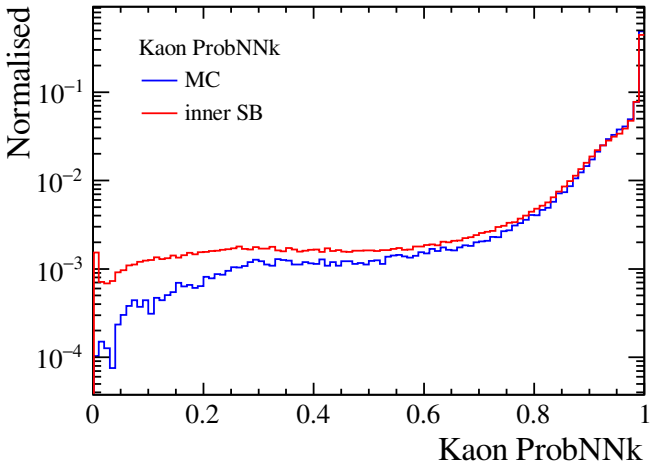


Figure 8.1: Comparison of the kaon ProbNNk distribution between the MC sample (signal) and the data inner sideband sample (background). Based on the plot, the cut point of the variable ProbNNk is chosen to be 0.4.

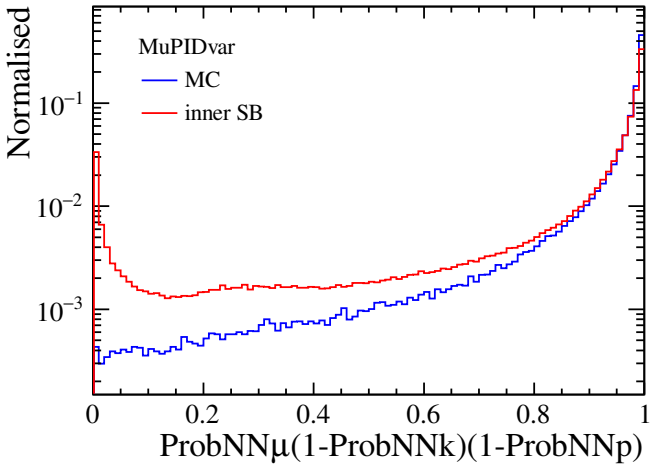


Figure 8.2: Normalised distribution of a combined muon PID variable for the MC sample (signal) and the data inner sideband sample (background). The kaon PID selection has been applied.

at LHCb, and was found to have good rejection performance for wrongly identified muons [66].

The selection will be applied as the requirement MuPIDvar $> y$ with the cut value y remaining to be optimized.

8.2 PID Calibration method

It is difficult to simulate the particle identification response of the detectors accurately. Conditions such as the temperature, gas pressure and alignments etc. can have an influence on the PID response [67]. MC simulation is not able to accurately describe all the interactions at low energy which happen in RICH and in the calorimeters. In order to optimize the muon PID selection and to evaluate the PID selection efficiency more precisely, a data-driven method PIDCalib is used [68]. The basic idea of the PIDCalib tool package is to use a calibration data sample to evaluate an efficiency. It is assumed that the PID efficiency can be parameterized by a set of known variables e.g. momentum, transverse momentum of the particle of interest and the event track multiplicity. According to the value of these variables, samples can be divided into different subsets (binning). Then the efficiency of a certain PID requirement for a particle in a certain subset can be corrected by the corresponding efficiency of the calibration sample [67].

Calibration samples are usually based on decay modes which are clean and have large branching fractions compared with the channel under study. Final states with only charged particles are preferred. Their selection strategy is carefully designed to try to avoid introducing a bias in PID variables. For the particles of our interest, the muon calibration samples are based on the decay modes $J/\psi \to \mu^+\mu^-$ and $B^+ \to J/\psi(\mu^+\mu^-)K^+$, and the kaon calibration samples are based on $D_s^+ \to \phi(K^+K^-)\pi^+$ and $D^{*+} \to D^0(K^-\pi^+)\pi^+$ decays [68].

8.3 MV and PID selection working point

To fully exploit the selections of the multivariate analysis and the muon PID variable, the optimization of the cut values (CatBoost $> x$ and MuPIDvar $> y$) is performed in a two dimensional manner. The goal is to maximize the Punzi Figure of Merit (FoM) [69]:

$$\text{Punzi FoM} = \frac{\epsilon_{\text{sig}}}{a/2 + \sqrt{B}} \tag{8.2}$$

where ϵ_{sig} stands for the efficiency of the selection (CatBoost $> x$ and MuPIDvar $> y$ with given value x, y) and B stands for the expected background yield in the signal region after the cuts. The parameter a stands for the number of sigmas at the desired significance. A detailed discussion can be found in [69]. The significance is chosen at $3\sigma (a = 3)$.

As described in section 8.2, for each selection value (x, y), the efficiency ϵ_{sig} is evaluated with the PIDCalib tool. For a certain selection MuPIDvar $> y$, PIDCalib will generate an efficiency table which is parameterized as a three dimensional histogram in the muon momentum, the muon transverse momentum and the event track multiplicity. Each event in the MC sample is assigned an efficiency MuPIDEff(y) according to this table.

In the efficiency calculation, the τ production source weights w_{source} (section 5.2) and the kinematic correction weights w_{cor} (section 6.3) are used. For a given pair (x, y) with CatBoost $> x$ and MuPIDvar $> y$, the selection efficiency is calculated with the weighted MC sample:

$$\epsilon_{\text{sig}} = \frac{\sum_{\text{CatBoost}(i)>x} w_i \cdot \text{MuPIDEff}_i(y)}{\sum_{\text{all}} w_i} \tag{8.3}$$

where the weights w_i are the product of the source weights and the kinematic correction weights. The variable $\text{MuPIDEff}_\text{i}(\text{y})$ is the PID selection efficiency derived from the PIDCalib table. In the numerator, the sum includes those candidates accepted by the CatBoost cut, while in the denominator the sum includes all candidates.

The expected background yield B is extrapolated from the invariant mass fit of the outer sideband data sample. Within the mass window $1720\,\text{MeV}/c^2 < M < 1746.86\,\text{MeV}/c^2$ (left outer sideband) and $1806.86\,\text{MeV}/c^2 < M < 1820\,\text{MeV}/c^2$ (right outer sideband), a 3rd-order Chebychev polynomial fit is performed for each x and y value. The expected yield B is extrapolated from the number of events in the fit region and the distribution model. The extrapolation can be written as:

$$B = N_{\text{fit}} \times \frac{\int_{|M-M_{\text{PDG}}(\tau)|<20\,\text{MeV}/c^2} f(x)\text{d}x}{\int_{\text{fit}} f(x)\text{d}x} \tag{8.4}$$

where $f(x)$ stands for the probablility density function of the mass distribution and the variable N_{fit} stands for the number of events in the fit region.

A scan of the x and y value in a two dimensional grid is performed and the above procedure to calculate the Punzi FoM is repeated for each point. The result is shown in Fig. 8.3. The relation between the Punzi FoM and the selections is further displayed in appendix A.3, which shows the slices of Fig. 8.3.

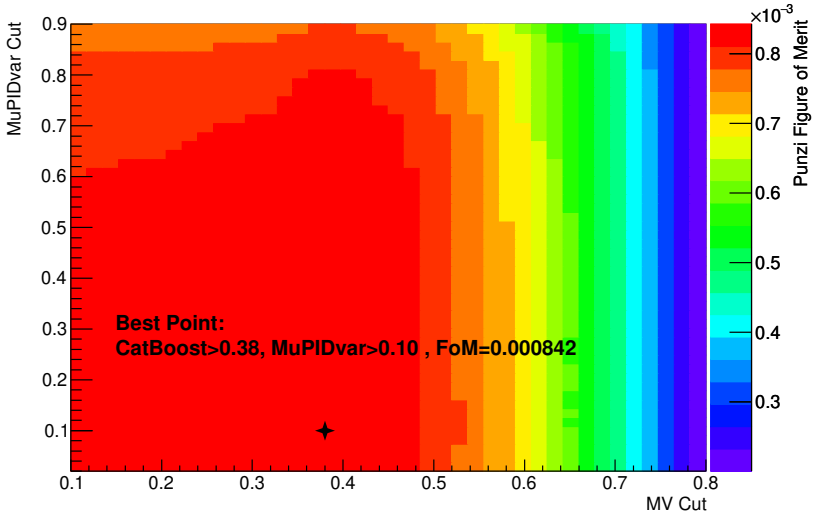


Figure 8.3: Punzi Figure of Merit as a function of CatBoost cut value x and the MuPIDvar cut value y. The optimized result is CatBoost>0.38 and MuPIDvar>0.10, which is shown in the plot as a black mark.

9 Signal and background models

After the event selection, the samples are ready to be used in the final evaluation of the upper limit of the signal branching fraction, which will be discussed in the subsequent parts of the thesis. The corresponding procedure requires models to describe the signal shape and the background shape in the invariant mass ($M(\phi(K^+K^-)\mu)$) distribution. In this chapter, an evaluation of the signal shape and the background shape without unblinding the data signal region is performed. The signal shape model will be extracted based on a fit of the MC signal sample. The background model will be estimated based on a fit of the data sidebands.

9.1 Signal shape

To derive the signal shape in the $M(\phi(K^+K^-)\mu)$ distribution, an unbinned maximum likelihood fit is performed with the signal channel Monte Carlo sample. To correct the kinematic variables and the τ production sources, the product of the kinematic corrections and the source weights are applied as weights. The fit model is chosen to be a Double-Sided Crystal Ball function [61]. Fig. 9.1 shows the fit result and Tab. 9.1 shows the parameters of the fit.

9.2 Background shape

The mass distribution shape of background events is extracted with a binned extended maximum likelihood fit of the data sidebands. As described in section 5.4, the $D^- \to K^+K^-\pi^-$ peak is included by adding the sample which was removed earlier by the requirement M(mother;$\mu \to \pi$)$< 1850\,\mathrm{MeV}/c^2$.

The sidebands are modelled with two distribution functions. The contribution from the decay $D^- \to K^+K^-\pi^-$ with a pion misidentified as a muon is mod-

Table 9.1: Shape parameters of the signal MC fit.

Parameters	Fit result
$M(\tau)$	1777.44 ± 0.02
width σ	4.88 ± 0.02
α_1	1.83 ± 0.02
α_2	1.58 ± 0.02
n_1	2.89 ± 0.10
n_2	11.64 ± 0.99

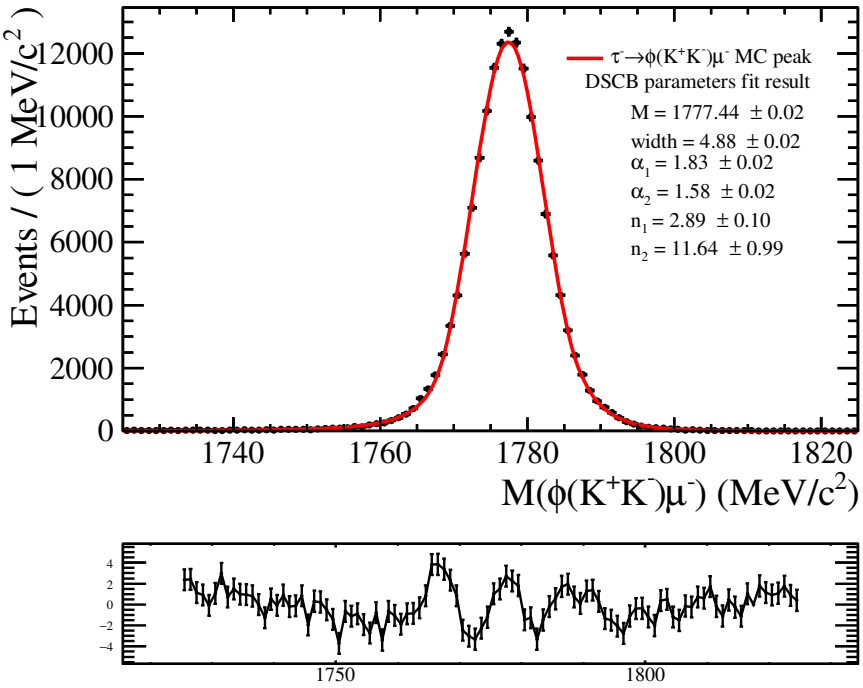


Figure 9.1: Fit of the signal MC sample.

elled as a single Crystal Ball function [61]. The combinatorial background, the $D_s^+ \to \phi(K^+K^-)\mu^+\nu_\mu$ crossfeed background and other small contributions (e.g. $D \to KK\pi\pi^0, D_s \to KK\pi\pi^0$ as discussed in section 4.3) are modelled as a 3rd-order Chebychev polynomial. The fit of the data sidebands is shown in Fig. 9.2.

The shape parameters of the fit are shown in Tab. 9.2.

The expected yields of backgrounds in the full mass window (as shown in Fig. 9.2) are extracted from the fit: $N(\text{Cheby.}) = (6.099 \pm 0.008) \times 10^6$ and $N(D \to KK\pi) = (4.30 \pm 0.06) \times 10^5$.

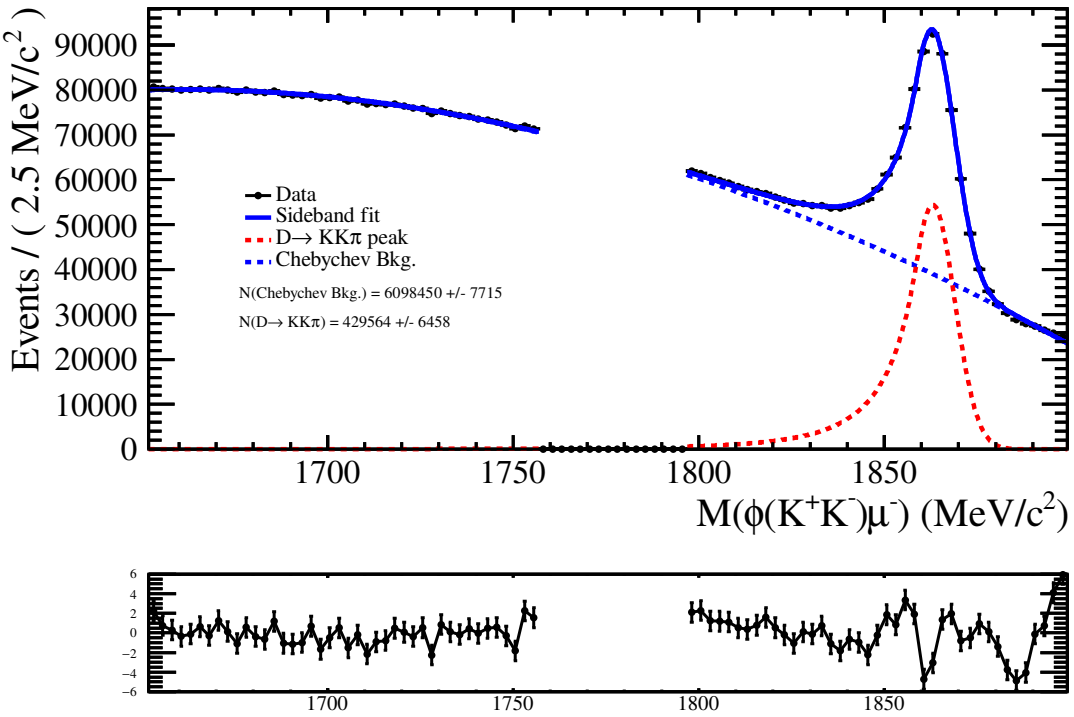


Figure 9.2: Fit of the data sidebands in the signal channel.

Table 9.2: Shape parameters of the data sidebands fit in the signal channel.

Parameters	Fit result
3rd-order Chebychev polynomial	
c_0	-0.4795 ± 0.0011
c_1	-0.1246 ± 0.0011
c_2	0.0022 ± 0.0013
$D^- \to K^+K^-\pi^-$ Crystal Ball peak	
mean	1863.06 ± 0.02
width σ	5.96 ± 0.03
α	0.748 ± 0.009
n	3.8 ± 0.4

10 Efficiency and normalisation

To translate a signal yield or a limit on the signal yield into the information concerning the branching fraction, one needs to normalise the signal channel to the reference channel which has a known branching fraction. The normalisation method gives a normalisation factor which will be used as an input for the evaluation of the branching fraction upper limit.

Following the derivation of Eq. 4.9 while considering that the signal channel contains τ leptons from 5 different sources and the reference channel contains D_s mesons from 2 different sources, the final normalisation equation can be written as:

$$\begin{aligned}\mathcal{B}(\tau^- \to \phi(K^+K^-)\mu^-) &= \mathcal{B}(\tau \to \phi\mu)\mathcal{B}(\phi \to K^+K^-) \\ &= \frac{\mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-)}{N_{\text{norm}}} \frac{\sum_{j=1}^{2} \epsilon_{j,\text{norm}} \sigma_j^{4\pi}(D_s)}{\sum_{i=1}^{5} \epsilon_{i,\text{sig}} \sigma_i^{4\pi}(\tau)} N_{\text{sig}}\end{aligned} \tag{10.1}$$

And hence:

$$\begin{aligned}\mathcal{B}(\tau \to \phi\mu) &= \frac{\mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-)}{N_{\text{norm}}\mathcal{B}(\phi \to K^+K^-)} \frac{\sum_{j=1}^{2} \epsilon_{j,\text{norm}} \sigma_j^{4\pi}(D_s)}{\sum_{i=1}^{5} \epsilon_{i,\text{sig}} \sigma_i^{4\pi}(\tau)} N_{\text{sig}} \\ &= \alpha N_{\text{sig}}\end{aligned} \tag{10.2}$$

where α stands for the normalisation factor, the sum over i stands for the different τ lepton sources and the sum over j stands for the D_s meson sources.

For the evaluation of α, the efficiencies are important inputs. The following part of this chapter will describe the details of the efficiency evaluation for each selection step. Due to the different trigger configurations for different data-taking years, the efficiencies can vary for different years. Thus the efficiencies and the reference channel yield N_{norm} will be evaluated for the years 2016, 2017 and 2018 separately. Based on the efficiencies and the reference channel yields, the normalisation factors for the three years can be evaluated. The normalisation factors of the three years are then combined into one single normalisation factor. The uncertainty of the normalisation factor will be evaluated with a toy MC which takes the errors of the cross sections of the different sources into consideration.

10.1 Overview of efficiency evaluation

The efficiencies are evaluated based on the simulation with corrections using data-driven methods. The total efficiency is factorized into different steps, with each step taking the previous step as the prerequisite. The total efficiency can be written as:

$$\epsilon_{tot} = \epsilon_{acc} \cdot \epsilon_{trig|acc} \cdot \epsilon_{rec|trig} \cdot \epsilon_{sel|rec} \tag{10.3}$$

where the total efficiency is calculated as the product of the acceptance of the detector (ϵ_{acc}), the trigger efficiency ($\epsilon_{trig|acc}$), the reconstruction efficiency ($\epsilon_{rec|trig}$) and the selection efficiency ($\epsilon_{sel|rec}$). The notation $\epsilon_{x|y}$ is used for the efficiency of x under the prerequisite y. The particle tracks for a potential candidate are required firstly to be in the LHCb detector acceptance and then pass the trigger requirements. Then the candidate should be reconstructed and pass certain selections. In the analysis, the decisions of the trigger lines are recorded in the simulation. The sequence of efficiencies can be rearranged:

$$\begin{aligned}\epsilon_{trig|acc} \cdot \epsilon_{rec|trig} \cdot \epsilon_{sel|rec} &= \frac{N_{sel\&rec\&trig\&acc}}{N_{acc}} \\ = \frac{N_{rec\&acc}}{N_{acc}} \frac{N_{sel\&rec\&acc}}{N_{rec\&acc}} \frac{N_{sel\&rec\&trig\&acc}}{N_{sel\&rec\&acc}} &= \epsilon_{rec|acc} \cdot \epsilon_{sel|rec} \cdot \epsilon_{trig|sel}\end{aligned} \tag{10.4}$$

where the notation N_x stands for the number of events after the requirement x. In this way, the trigger efficiency is evaluated using a sample after passing the selections [59]. For the signal channel of this analysis, efficiencies for the stripping selection and the preselection are evaluated before the trigger selection while the multivariate and kaon/muon PID selections are placed after it.

10.2 Acceptance

As described in section 4.1, the LHCb detector does not cover the full 4π solid angle. Only those particles which enter the acceptance range can interact with the detector.

In section 5.2, the acceptance factors for the different τ/D_s production sources have been calculated with generator level simulations. The value of ϵ_{Cut} from section 5.2 is quoted as the acceptance:

$$\epsilon_{acc} = \epsilon_{Cut} = \epsilon_{Gen\&Cut}/\epsilon_{Gen} \tag{10.5}$$

The acceptance factors for the signal channel and the reference channel with different sources are listed in Tab. 10.1. These values are used for all the three data-taking years. The different values for the different τ/D_s production sources are related to the different topology and kinematics of the production decays.

10.3 Reconstruction, stripping and preselection

After the particles of the signal candidates enter the LHCb detector, they interact with the detector. The detector output is used for reconstruction. However, not all signal candidates can be correctly reconstructed. Thus a reconstruction efficiency needs to be taken into account. As described in section 3.3, the reconstructed data recorded by the LHCb detector are selected with stripping lines. In this analysis, after the stripping selection, the samples are further selected by the preselection described in section 5.4.

Table 10.1: Acceptance of the LHCb detector for the signal channel and the reference channel.

Source	Acceptance [%]
Signal Channel	
$b\bar{b} \to \tau$	14.63 ± 0.10
$D_s \to \tau$	15.78 ± 0.05
$b\bar{b} \to D_s \to \tau$	14.67 ± 0.06
$D \to \tau$	15.87 ± 0.05
$b\bar{b} \to D \to \tau$	14.60 ± 0.10
Reference Channel	
$b\bar{b} \to D_s$	11.15 ± 0.46
prompt D_s	12.17 ± 0.36

Table 10.2: Combined efficiencies for reconstrution and stripping selection ($\epsilon_{rec\&strip}$).

Source	$\epsilon_{rec\&strip}(2016)[\%]$	$\epsilon_{rec\&strip}(2017)[\%]$	$\epsilon_{rec\&strip}(2018)[\%]$
Signal Channel			
$b\bar{b} \to \tau$	6.841 ± 0.022	6.878 ± 0.023	6.830 ± 0.024
$D_s \to \tau$	5.963 ± 0.010	5.958 ± 0.010	5.946 ± 0.010
$b\bar{b} \to D_s \to \tau$	9.898 ± 0.033	9.928 ± 0.034	9.892 ± 0.032
$D \to \tau$	10.210 ± 0.052	10.297 ± 0.049	10.278 ± 0.044
$b\bar{b} \to D \to \tau$	12.336 ± 0.120	12.116 ± 0.118	12.061 ± 0.102
Reference Channel			
$b\bar{b} \to D_s$	19.260 ± 0.076	19.419 ± 0.074	19.514 ± 0.077
prompt D_s	6.595 ± 0.016	6.653 ± 0.016	6.634 ± 0.018

Based on the MC samples, the reconstruction and stripping efficiency ($\epsilon_{rec.\&strip.}$) is evaluated as the ratio between the number of events after the stripping selection and the number of accepted events. Tab. 10.2 shows the reconstruction and stripping efficiency for the different years. The preselection efficiency ($\epsilon_{presel.}$) is evaluated as the ratio between the number of events in the MC samples before and after the preselection. The result is shown in Tab. 10.3.

The reconstruction and stripping selection efficiencies for different τ/D_s production sources are very different. The reason is that the selections include the criteria concerning the decay topology and kinematics, which are very different for the different τ/D_s production decays.

The reconstruction, stripping and preselection efficiencies for the different years agree within their statistical uncertainties.

Table 10.3: Preselection efficiencies (ϵ_{presel}).

Source	$\epsilon_{presel}(2016)[\%]$	$\epsilon_{presel}(2017)[\%]$	$\epsilon_{presel}(2018)[\%]$
Signal Channel			
$b\bar{b} \to \tau$	94.979 ± 0.076	94.860 ± 0.078	94.834 ± 0.081
$D_s \to \tau$	95.034 ± 0.039	95.208 ± 0.038	95.121 ± 0.038
$b\bar{b} \to D_s \to \tau$	95.029 ± 0.078	95.029 ± 0.080	95.071 ± 0.076
$D \to \tau$	95.116 ± 0.116	95.031 ± 0.110	95.231 ± 0.097
$b\bar{b} \to D \to \tau$	94.424 ± 0.241	95.026 ± 0.230	94.888 ± 0.203
Reference Channel			
$b\bar{b} \to D_s$	85.062 ± 0.175	85.452 ± 0.168	85.149 ± 0.176
prompt D_s	86.316 ± 0.091	86.437 ± 0.086	86.410 ± 0.097

Table 10.4: Trigger efficiencies (ϵ_{trig}) for different data-taking years.

Source	$\epsilon_{trig}(2016)[\%]$	$\epsilon_{trig}(2017)[\%]$	$\epsilon_{trig}(2018)[\%]$
Signal Channel			
$b\bar{b} \to \tau$	12.24 ± 0.12	13.70 ± 0.13	11.99 ± 0.12
$D_s \to \tau$	13.95 ± 0.06	15.58 ± 0.07	13.75 ± 0.06
$b\bar{b} \to D_s \to \tau$	14.07 ± 0.13	15.90 ± 0.14	13.94 ± 0.12
$D \to \tau$	14.06 ± 0.19	15.85 ± 0.19	13.68 ± 0.16
$b\bar{b} \to D \to \tau$	14.68 ± 0.38	15.51 ± 0.39	14.14 ± 0.33
Reference Channel			
$b\bar{b} \to D_s$	16.54 ± 0.20	37.46 ± 0.25	32.01 ± 0.25
prompt D_s	12.27 ± 0.09	30.61 ± 0.12	26.71 ± 0.13

10.4 Trigger Efficiency

As an estimation, the trigger efficiencies are evaluated using the MC samples. Tab. 10.4 shows the trigger efficiencies for the different sources of the signal channel and the reference channel in the different years. The trigger efficiency varies largely for different data-taking years. This is related to the different trigger configurations during these years. For example, a lower threshold setting of the trigger lines will result in an increase in the efficiency.

10.5 Kaon PID selection

The MC simulation usually does not describe the PID variables precisely. Therefore, the data-driven PIDCalib tool is used to evaluate the kaon PID selection efficiency.

The kaon PID selection efficiency is parameterized as a 3 dimensional table, which is a function of the variables $p(K)$, $p_T(K)$ and nTracks. The efficiency for a single kaon is derived from the PIDCalib table according to its p, p_T and the track multiplicity. Assuming that the PID selections of the two kaons are independent,

the efficiency for the whole event is calculated as the product of the PID selection efficiencies for the two kaons from the ϕ decay.

The final efficiency is calculated as a weighted average:

$$\epsilon_{KPID} = \frac{\sum_i w_{cor}(i) \cdot \mathrm{KPIDEff}_i(K^+) \cdot \mathrm{KPIDEff}_i(K^-)}{\sum_i w_{cor}(i)} \tag{10.6}$$

where i loops over all the MC events, $\mathrm{KPIDEff}_i$ are the efficiency values derived from the kaon PIDCalib table and the weights $w_{cor}(i)$ are the kinematic corrections calculated from the reference channel (section 6.3).

In principle, the kinematic correction itself has an uncertainty. As described in section 6, the kinematic correction is derived with the *sPlot* method. The evaluation of its uncertainty will have to involve the *sPlot* method and will be very complicated. For the purpose of studying the feasibility, not all uncertainties are taken into account in this analysis. The uncertainty of the kinematic correction is not considered and only the uncertainty of the PIDCalib table is included. The uncertainty in the PIDCalib table can be related to the limited size of the calibration sample [67]. The uncertainty contributed by the PIDCalib table is estimated with a toy MC simulation. The toy MC simulation is done as follows:

- Step 1: smearing the PIDCalib table. For each bin in the 3 dimensional PIDCalib table, the original efficiency is substituted by a random number generated from a gaussian distribution, which takes the original efficiency of the given bin as the mean value and the error of the given bin as the sigma value.

- Step 2: calculating a toy MC efficiency. The kaon selection efficiency is recalculated using Eq. 10.6 with the smeared PIDCalib table.

The above procedure is repeated for the production of the toy MC sample. For each of the 5 τ lepton sources, a toy MC sample is generated. The mean value of the toy MC efficiency sample is used as the efficiency for the corresponding τ lepton source. The standard deviation of the sample is used as the uncertainty of the efficiency.

The result of the kaon PID efficiency evaluation is shown in Tab. 10.5. Since these efficiencies are only related to the particle identification ability of the kaon candidates, the results are very similar between the different τ production sources and the different data-taking years. The slight difference can be caused by the uncertainties, which include the errors shown in Tab. 10.5 (originating from the PIDCalib table) and the uncertainties from the other possible sources (e.g. the kinematic corrections and the binning scheme of the PIDCalib table).

Table 10.5: Kaon PID selection efficiency (ϵ_{KPID}) of the signal channel

Source	$\epsilon_{KPID}(2016)[\%]$	$\epsilon_{KPID}(2017)[\%]$	$\epsilon_{KPID}(2018)[\%]$
$bb \to \tau$	90.943 ± 0.009	90.886 ± 0.010	91.080 ± 0.011
$D_s \to \tau$	90.786 ± 0.007	90.841 ± 0.008	90.804 ± 0.008
$b\bar{b} \to D_s \to \tau$	90.734 ± 0.008	90.954 ± 0.010	90.731 ± 0.012
$D \to \tau$	90.261 ± 0.011	90.594 ± 0.011	90.508 ± 0.014
$b\bar{b} \to D \to \tau$	90.874 ± 0.015	90.944 ± 0.022	90.237 ± 0.019

10.6 Multivariate and muon PID selection

In the selection described in section 8.3, the multivariate (MV) classifier output and the muon PID variables selections are optimized in a two dimensional manner:

- MV selection: CatBoost > x
- Muon PID selection: MuPIDvar > y

The efficiency of the selections is separated into two parts i.e. the MV selection and the muon PID selection with the MV cut as its prerequisite. The efficiency of the MV selection is simply the events which have passed the MV selection (CatBoost$> x$) divided by the total events with the kinematic corrections as weights:

$$\epsilon_{MV} = \frac{\sum_{\text{CatBoost}>x} w_{cor}(i)}{\sum_{tot} w_{cor}(i)} \tag{10.7}$$

where the index i labels the events in the MC sample. The index "tot" means all the events which passed the previous level selection. The uncertainty of this weighted efficiency is calculated as:

$$\Delta\epsilon_{MV} = \frac{\sqrt{\sum_{tot}(w_{cor}(i))^2}}{\sum_{tot} w_{cor}(i)} \cdot \sqrt{\epsilon_{MV}(1-\epsilon_{MV})} \tag{10.8}$$

This uncertainty is due to the limited size of the MC sample. Again, the uncertainties asscociated with the the kinematic corrections are neglected due to the complexity of the *sPlot* method.

The calculation of the muon PID selection efficiency is based on the samples that passed the MV selection. The PIDCalib tool is applied to correct the MC simulation. The calculation can be formulated as:

$$\epsilon_{\mu PID|MV} = \frac{\sum_{\text{CatBoost}>x} w_{cor}(i) \cdot \text{MuPIDEff}_i(y)}{\sum_{\text{CatBoost}>x} w_{cor}(i)} \tag{10.9}$$

where MuPIDEff_i are the values from the muon PIDCalib table. Same as in the previous procedures, the uncertainty contributed by the kinematic corrections is not considered and the PIDCalib table is considered as the only error source. As

Table 10.6: MV selection efficiency (ϵ_{MV}) of the signal channel

Source	$\epsilon_{MV}(2016)[\%]$	$\epsilon_{MV}(2017)[\%]$	$\epsilon_{MV}(2018)[\%]$
$b\bar{b} \to \tau$	75.868 ± 0.545	75.510 ± 0.514	75.684 ± 0.556
$D_s \to \tau$	92.853 ± 0.152	92.957 ± 0.143	92.734 ± 0.150
$b\bar{b} \to D_s \to \tau$	84.681 ± 0.446	84.771 ± 0.415	85.039 ± 0.426
$D \to \tau$	97.747 ± 0.264	97.517 ± 0.247	97.274 ± 0.249
$b\bar{b} \to D \to \tau$	91.166 ± 0.963	87.127 ± 1.091	87.871 ± 0.990

Table 10.7: Muon PID selection efficiency ($\epsilon_{\mu PID|MV}$) of the signal channel, with the prerequisite of the MV selection.

Source	$\epsilon_{\mu PID\|MV}(2016)[\%]$	$\epsilon_{\mu PID\|MV}(2017)[\%]$	$\epsilon_{\mu PID\|MV}(2018)[\%]$
$b\bar{b} \to \tau$	97.403 ± 0.025	97.261 ± 0.049	97.600 ± 0.015
$D_s \to \tau$	97.454 ± 0.014	97.341 ± 0.014	97.630 ± 0.011
$b\bar{b} \to D_s \to \tau$	97.340 ± 0.026	97.293 ± 0.023	97.537 ± 0.016
$D \to \tau$	97.278 ± 0.030	97.208 ± 0.029	97.568 ± 0.020
$b\bar{b} \to D \to \tau$	97.312 ± 0.075	97.040 ± 0.059	97.410 ± 0.031

the PIDCalib table is derived with the kinematic binning, the uncertainty of the efficiency $\epsilon_{\mu PID|MV}$ is also calculated with binning:

$$\Delta\epsilon_{\mu PID|MV} = \sqrt{\sum_{\text{bin}} \left(r_{\text{bin}} \cdot \Delta\text{MuPIDEff}_{\text{bin}}(y)\right)^2} \tag{10.10}$$

where the uncertainty term $\Delta\text{MuPIDEff}_{\text{bin}}(y)$ is the error of the corresponding bin in the muon PIDCalib table. The parameter r_{bin} stands for the ratio of the MC sample in a certain bin, weighted with the kinematic corrections:

$$r_{\text{bin}} = \frac{\sum_{i \in \text{bin}} w_{cor}(i)}{\sum_{all} w_{cor}(i)} \tag{10.11}$$

As a summary, Tab. 10.6 and Tab. 10.7 show the efficiencies of these two steps. The efficiencies for the different data-taking years are very similar. Since the multivariate analysis includes the information of the decay topology and kinematics, the MV efficiencies vary for the different τ lepton production sources.

Table 10.8: Total efficiency (ϵ_{tot}) for different sources in the signal channel and the reference channel.

Source	$\epsilon_{tot}(2016)[\%]$	$\epsilon_{tot}(2017)[\%]$	$\epsilon_{tot}(2018)[\%]$
Signal Channel			
$b\bar{b} \to \tau$	0.078 ± 0.001	0.087 ± 0.001	0.076 ± 0.001
$D_s \to \tau$	0.102 ± 0.001	0.115 ± 0.001	0.101 ± 0.001
$b\bar{b} \to D_s \to \tau$	0.145 ± 0.002	0.165 ± 0.002	0.145 ± 0.002
$D \to \tau$	0.186 ± 0.003	0.211 ± 0.003	0.182 ± 0.002
$b\bar{b} \to D \to \tau$	0.201 ± 0.006	0.201 ± 0.006	0.182 ± 0.005
Reference Channel			
$b\bar{b} \to D_s$	0.302 ± 0.013	0.693 ± 0.029	0.593 ± 0.025
prompt D_s	0.085 ± 0.003	0.214 ± 0.006	0.186 ± 0.006

10.7 Normalisation factor

The normalisation factor introduced in Eq. 10.2 can be written as:

$$\alpha = \frac{\mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-)}{N_{\text{norm}}\mathcal{B}(\phi \to K^+K^-)} \frac{\sum_{j=1}^{2} \epsilon_{j,\text{norm}}\sigma_j^{4\pi}(D_s)}{\sum_{i=1}^{5} \epsilon_{i,\text{sig}}\sigma_i^{4\pi}(\tau)} \tag{10.12}$$

The calculation of this normalisation factor requires the following input: the number of events for the reference channel N_{norm}, the branching fractions $\mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-)$, $\mathcal{B}(\phi \to K^+K^-)$, the efficiencies and the cross sections of the τ/D_s production sources (τ production for the signal channel and D_s production for the reference channel).

The branching fraction $\mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-))$ can be obtained from the branching fraction $\mathcal{B}(D_s^- \to \phi(K^+K^-)\pi^-)$ using the ϕ decay branching fraction. The branching fractions of the relevant decays, taken from PDG [53], are $\mathcal{B}(D_s^- \to \phi(K^+K^-)\pi^-) = (2.24 \pm 0.08)\%$, $\mathcal{B}(\phi \to K^+K^-) = (49.2 \pm 0.5)\%$ and $\mathcal{B}(\phi \to \mu^+\mu^-) = (2.86 \pm 0.19) \times 10^{-4}$.

The branching fraction ratio in Eq. 10.12 can be calculated as:

$$\begin{aligned} \frac{\mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-)}{\mathcal{B}(\phi \to K^+K^-)} &= \frac{\mathcal{B}(D_s^- \to \phi(K^+K^-)\pi^-)}{(\mathcal{B}(\phi \to K^+K^-))^2} \cdot \mathcal{B}(\phi \to \mu^+\mu^-) \\ &= (2.647 \pm 0.207) \times 10^{-5} \end{aligned} \tag{10.13}$$

The total efficiencies, calculated as the product of the efficiencies of each step described in the previous sections, are shown in Tab. 10.8.

Following a similar procedure as in section 6.2, the yields of the reference channel for the three years are extracted using a maximum likelihood fit. During the fit, the tail parameters of the signal peak are taken from the MC fit in section 6.2. The corresponding fit result is shown in Fig. 10.1. The expected event yields are listed in Tab. 10.9.

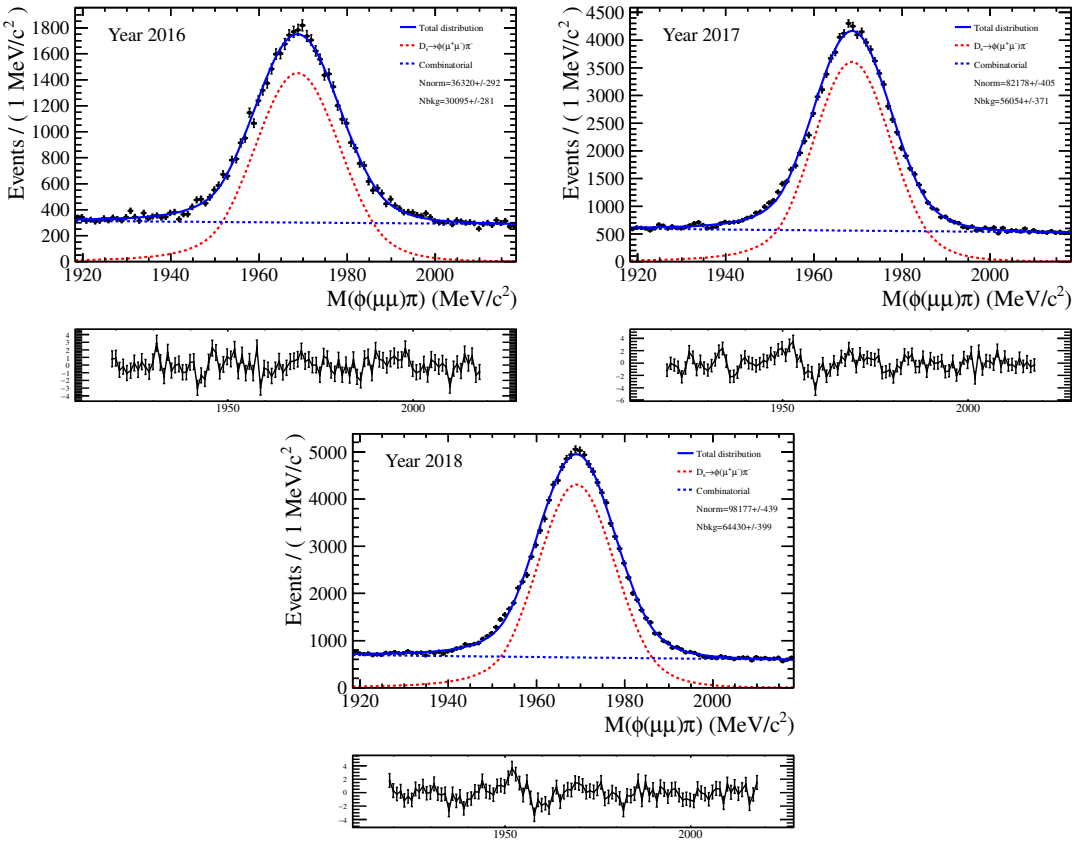


Figure 10.1: Reference channel fit for the three years.

Table 10.9: Reference channel yields (N_{norm}) for the different years.

Year	2016	2017	2018
N_{norm}	36320 ± 292	82178 ± 405	98177 ± 439

The production cross sections for the different D_s and τ sources have already been calculated in section 4.1 and section 4.2. As shown in Eq. 4.6 and Tab. 4.1, these cross sections of interest can be expressed by a set of variables (Tab. 4.1), which in this analysis are considered uncorrelated to simplify the estimation of the uncertainty.

With the efficiencies and the reference channel yields for the three years, we can calculate the normalisation factors for the three years with Eq. 10.12. The three normalisation factors are then combined into one single combined normalisation factor. For each year (denoted with the index y), the normalisation factor (α_y) associates the signal yield (N^y_{sig}) to the branching fraction of interest:

$$\alpha_y N^y_{sig} = \mathcal{B}(\tau \to \phi\mu) \Rightarrow N^y_{sig} = \frac{\mathcal{B}(\tau \to \phi\mu)}{\alpha_y} \tag{10.14}$$

The combined sample for all years gives a total signal yield, which can be related to the signal branching fraction with the combined normalisation factor (α_{comb}).

By summing up N^y_{sig} with Eq. 10.14, we derive:

$$\sum_y N^y_{sig} = \left(\sum_y \frac{1}{\alpha_y}\right) \cdot \mathcal{B}(\tau \to \phi\mu) = \frac{\mathcal{B}(\tau \to \phi\mu)}{\alpha_{comb}} \tag{10.15}$$

with:

$$\alpha_{comb} = \left(\sum_y \frac{1}{\alpha_y}\right)^{-1} \tag{10.16}$$

The uncertainty of the normalisation factor can be evaluated with the toy MC method. By smearing the input parameters in Tab. 4.1, Tab. 10.8, Tab. 10.9 and the branching fraction ratio in Eq. 10.13 with gaussian distributions [1], a toy MC of 10^6 entries is produced. Each toy MC entry uses the smeared input parameters to calculate the three normalisation factors corresponding to the three years: $(\alpha_{2016}, \alpha_{2017}, \alpha_{2018})$. The three normalisation factors are then used to calculate a α_{comb} using Eq. 10.16. These toy MC values of α_{comb} constitute the toy MC sample of the combined normalisation factor. The distribution of these α_{comb} entries is displayed in Fig. 10.2. The mean value of the toy MC sample of the normalisation factor can be considered as the final value of the normalisation factor and the standard deviation of the sample is considered as its uncertainty. The combined normalisation factor takes the value:

$$\alpha_{comb} = (3.37 \pm 0.32) \times 10^{-9} \tag{10.17}$$

[1] This means that every input parameter is substituted by a random number generated with a gaussian distribution, which uses the nominal value of the given parameter as the mean value and the uncertainty of the given parameter as the sigma value.

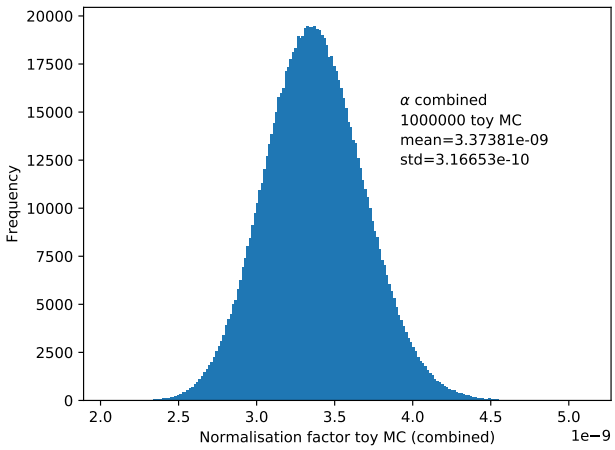


Figure 10.2: Combined normalisation factor toy MC distribution.

11 Upper limit

This analysis aims to search for a very rare signal decay $\tau^- \to \phi(K^+K^-)\mu^-$, and the search result should be expressed as an upper limit of the branching fraction $\mathcal{B}(\tau \to \phi\mu)$. The search for a potential signal is eventually a test of a hypothesis on the branching fraction $\mathcal{B}(\tau \to \phi\mu)$. In the situation of no signal observed, an upper limit of the parameter of interest $\mathcal{B}(\tau \to \phi\mu)$ can be set. This can be done by comparing the test hypothesis with the background only hypothesis.

In order to perform such a comparison, the so-called CL_s technique [70] is used. The CL_s technique is usually used to distinguish the test hypothesis (in our case, the background plus a signal corresponding to a certain $\mathcal{B}(\tau \to \phi\mu)$) against the alternative hypothesis (in our case, the background only hypothesis). In this analysis, the toolkit RooStats [71] is used to implement this technique.

In the following chapter, the method used for the upper limit setting will be introduced and the upper limit on the branching fraction $\mathcal{B}(\tau \to \phi\mu)$ will be estimated as the result of this search.

11.1 Evaluation method

The CL_s method for the upper limit determination is introduced in this section based on the reference [70] and [72].

11.1.1 Test statistics

To distinguish the test hypothesis from the alternative hypothesis, certain criteria need to be set. A test statistic is constructed as a variable to set up the criteria. The distribution of the test statistic will be used to separate different hypotheses. We use the symbol q_μ as the notation for the test statistic, with the index μ denoting the test value of the parameter of interest. There are several test statistics available for different purposes. In this analysis, the one side profile likelihood ratio (q_μ) test is used, which is defined as follows:

$$q_\mu = \begin{cases} -2\ln(L(\mu,\hat{\hat{\nu}})/L(\hat{\mu},\hat{\nu})) & \text{if } \hat{\mu} \leq \mu \\ 0 \quad \text{if } \hat{\mu} > \mu \end{cases} \tag{11.1}$$

In this equation, μ and $\hat{\mu}$ stand for the parameter of interest i.e. the branching fraction $\mathcal{B}(\tau \to \phi\mu)$. The notations $\hat{\hat{\nu}}$ and $\hat{\nu}$ stand for the nuisance parameters (e.g. the normalisation factor α in our case, the detailed description and treatment of the nuisance parameters can be found in section 11.1.2). The numerator and the

denominator $L(\mu, \hat{\hat{\nu}}), L(\hat{\mu}, \hat{\nu})$ stand for two likelihoods obtained from the maximum likelihood fits under two different assumptions. The parameter μ is fixed to the test value of the parameter of interest, and $\hat{\hat{\nu}}$ stands for the fitted value of the nuisance parameter under the condition that μ is fixed. The parameters $\hat{\mu}$ and $\hat{\nu}$ stand for the fit result without additional constraints (the best fit values).

The test statistic q_μ can be considered as a measurement of the compatibility between the data sample (information contained in $\hat{\mu}$) and the test hypothesis (information contained in μ). A high value of q_μ indicates the incompatibility between the data sample and the test hypothesis. Since we are testing against the background only hypothesis (no signal is expected in the Standard Model), only the upper limit of the branching fraction is of interest (we are not interested in a potential lower limit which can be caused by fluctuation). In this case, the situation $\hat{\mu} > \mu$ is not considered as a sign of incompatibility between the data sample and the test hypothesis. Thus, for $\hat{\mu} > \mu$, the test statistic is set to zero.

For a certain data sample with an observable $\vec{x}$ (in this case, the invariant mass), the likelihoods in the above equation are calculated as follows:

$$L(\mu, \nu) = f_m(\vec{x}|\mu, \nu_i) \times \prod_i G(\nu_{i,obs}|\nu_i, \nu_{i,err}) \tag{11.2}$$

where the different nuisance parameters are denoted by ν_i and the function f_m stands for the distribution function of $\vec{x}$. The distribution of $\vec{x}$ depends on the parameters μ, ν_i. Function $G(\xi|\nu_i, \nu_{i,err})$ stands for a gaussian distribution function for the variable ξ with the mean value ν_i and the standard deviation $\nu_{i,err}$. During the maximum likelihood fit, the parameters ν_i are allowed to float, while the parameters $\nu_{i,obs}$ and $\nu_{i,err}$ are fixed to constant. In this way, the nuisance parameters are considered as global observables and constrained with the gaussian distributions. A global observable is a variable which has been measured before the CL_s method is applied. It has an observed value $\nu_{i,obs}$ and an uncertainty $\nu_{i,err}$, which can be determined by an initial fit prior to the procedure of the CL_s method.

In this analysis, due to the large statistics of the data sample, a binned likelihood for the invariant mass spectrum is used.

With the form of the likelihood given, the distribution of the test statistic q_μ is then explored under the two hypotheses: the signal-plus-background model (S+B model) and the background only model (B only model). The two hypotheses lead to two different distributions of $\vec{x}$. For a fixed test value μ, the test statistic q_μ is a function of the observable $\vec{x}$. With the given distribution of $\vec{x}$, the distribution of q_μ can be derived. When $\vec{x}$ is set to the observed value $\vec{x}_{obs}$ corresponding to the observed data, the observed value of test statistic ($q_{\mu,obs}$) is calculated. The distribution of q_μ can be evaluated with a toy MC simulation. In addition, there is an asymptotic theory for the distribution of q_μ, which requires less computation resources and time [72].

Based on the distributions of q_μ under the two hypotheses, the p-value of the two

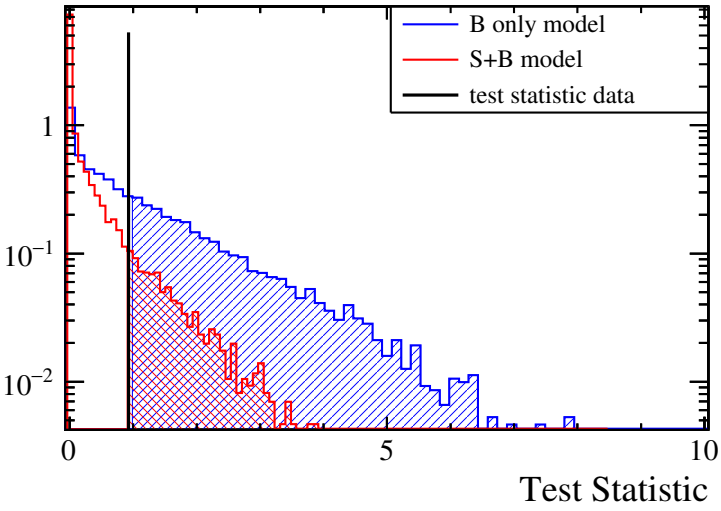


Figure 11.1: An example for the test statistic distribution (normalised). The black line stands for the observed value derived from the data. The red shaded part represents the p-value of S+B model and the blue shaded part represents the p-value of B only model.

models are calculated as follows:

$$\begin{cases} p_{S+B} = \int_{q_{\mu,obs}}^{+\infty} f(q_\mu|S+B)\mathrm{d}q_\mu \\ p_B = \int_{q_{\mu,obs}}^{+\infty} f(q_\mu|B)\mathrm{d}q_\mu \end{cases} \tag{11.3}$$

where $f(q_\mu|H)$ stands for probability density function of the test statistic q_μ under the hypothesis H. Fig. 11.1 shows an example of the distribution of the test statistic q_μ in case of the background only hypothesis and signal-plus-background hypothesis.

The observed CL_s value is then defined:

$$CL_s = \frac{p_{S+B}}{p_B} \tag{11.4}$$

For each test branching fraction value, a CL_s value is evaluated. A larger test branching fraction tends to result in a larger discrepancy between the B-only model and the S+B model, and hence a smaller CL_s value. When a scan of the test branching fraction is performed, the CL_s values constitutes a CL_s curve. The CL_s curve is considered as a criterion for the upper limit evaluation, e.g. the 90% confidence level upper limit is set with requirement $1 - CL_s = 90\%$. Fig. 11.2 shows basic idea of setting upper limit of a certain confidence level with the CL_s curve.

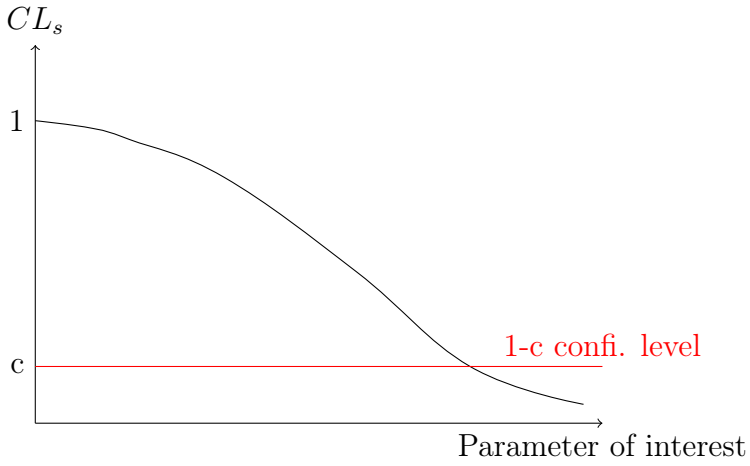


Figure 11.2: A sketch of CL_s curve. The red confidence level line intercepts with the CL_s curve. The intercept point is taken to set the limit on the parameter of interest corresponding to the confidence level c.

11.1.2 Nuisance parameters

In addition to the parameter of interest ($\mathcal{B}(\tau \to \phi\mu)$), the distribution function for the observable $\vec{x}$ also depends on other parameters i.e. the nuisance parameters.

In this analysis, the parameters used in the description of the invariant mass distribution include the branching fraction of the signal decay, the normalisation factor α, the signal shape parameters, the background shape parameters and the expected background yields. The branching fraction of the signal decay is the parameter of interest. The other parameters can be treated in two ways. Before performing the CL_s evaluation, an initial fit is performed. The parameters other than the parameter of interest can be fixed to the result of this fit or they can be constrained with gaussian distributions as described in Eq. 11.2. In the case of a gaussian constraint, the value derived from the initial fit can be used as the observed value of the nuisance parameters ($\nu_{i,obs}$) and the corresponding uncertainty derived from the fit can be used as the uncertainty of the nuisance parameters ($\nu_{i,err}$).

During the evaluation of the test statistic distribution using the toy MC method, when a toy MC sample is generated, a set of pseudo-observed value of nuisance parameters can be generated. In this way, the uncertainty of nuisance parameters is considered in the final result.

11.1.3 Expected limit

Under the background only hypothesis, an expected limit can be evaluated without involving the observed data. Given the information of the signal and background models, the distribution of the test statistic q_μ can be derived. As in Eq. 11.3, the median value of the test statistic ($q_{\mu,med}$) under the background only model can substitute the observed value $q_{\mu,obs}$ as the integral boundary. This substitution gives

the expected CL_s curve and the corresponding expected upper limit. Similarly, the CL_s curve can also be evaluated for the median value minus or plus some deviations (e.g. $\pm\sigma, \pm 2\sigma$ with σ being the standard deviation of the test statistic distribution under the background only hypothesis). When comparing with the observed CL_s curve (derived using $q_{\mu,obs}$ as the integral boundary), the deviation can help to evaluate to what extent the observed data is compatible with the background only hypothesis.

11.2 Extrapolated result from sidebands fit

Before unblinding the signal region, it is possible to evaluate an expected upper limit under the background only hypothesis. The background model can be extracted from the sidebands and extrapolate to the signal region. Based on the sidebands fit in section 9.2, the expected number of events per background source (i.e. the Chebychev-shaped background and the $D^- \to K^+K^-\pi^-$ crossfeed) and the shape parameters for these background models are derived. Together with the signal model derived in section 9.1, they are used as the input for the CL_s method. All the background shape parameters and the tail parameters of the signal peak are fixed to the fit result shown in section 9.1 and section 9.2. The mean parameter of the signal peak, the width parameter of the signal peak, the expected yields of the background sources and the normalisation factor are constrained with gaussian distributions.

Based on the q_μ distribution under the background only model, the CL_s curves corresponding to median value, median plus/minus 1σ and median plus/minus 2σ are calculated. Fig. 11.3 shows the evaluation result of these expected CL_s curves.

With the extrapolated background model from the sidebands fit, the median value of the expected upper limit at 90% confidence level takes the value:

$$\mathcal{B}(\tau \to \phi\mu) < 4.2 \times 10^{-6} \tag{11.5}$$

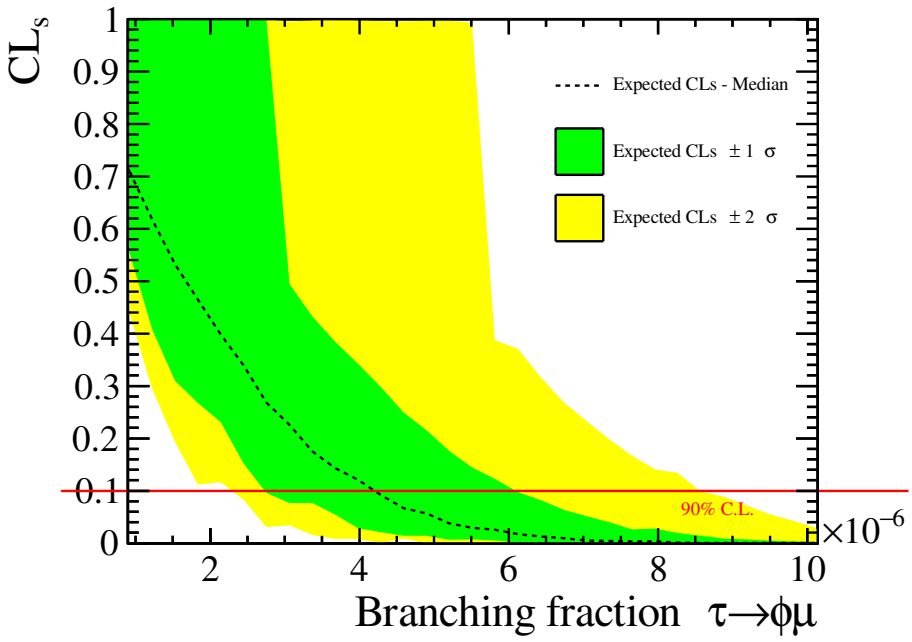


Figure 11.3: The expected CL_s curve and its corresponding upper limit, calculated with toy MC method. The background models are extrapolated from the data sidebands. The red line marks the 90% confidence level.

11.3 Unblinded result

Evaluated in a wide mass window, the aforementioned background model does not perfectly describe the mass distribution. Especially, it does not include the information in the blinded region. With the signal region unblinded, the background can be evaluated more accurately using a smaller sidebands region. In the mass window $M_{PDG}(\tau) \pm 30\,\mathrm{MeV}/c^2$, a fit is performed. The background is modelled with a 3rd-order Chebychev polynominal. The signal shape is fixed to the result of the signal MC fit described in section 9.1. The fit of the unblinded data sample is shown in Fig. 11.4 and the parameters of the fit are shown in Tab. 11.1.

With the fit of the unblinded data and the fit of the signal peak (section 9.1), the models are used as the input for the CL_s method. The three background shape parameters (c_0, c_1, c_2), the mean parameter of the signal peak, the width parameter of the signal peak, the expected background yield (N_{bkg}) and the normalisation factor are treated as floating nuisance parameters with gaussian constraints. The tail parameters of the signal peak are fixed to the fit result in Tab. 9.1.

Using the unblinded data sample, the CL_s curve evaluation is shown in Fig. 11.5. At 90% confidence level, the expected upper limit under the background only model takes the value:

$$\mathcal{B}(\tau \to \phi\mu) < 5.2 \times 10^{-6} \tag{11.6}$$

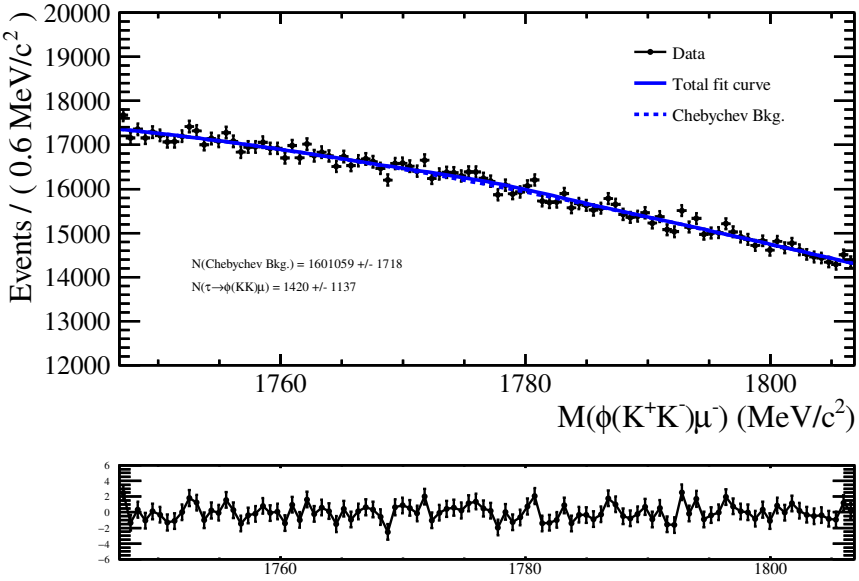


Figure 11.4: Fit on the unblinded data sample.

Table 11.1: Parameters of the fit of the unblinded data sample

Parameters	Fit result
Background shape (3rd-order Chebychev polynomial)	
c_0	-0.0965 ± 0.0016
c_1	-0.0084 ± 0.0018
c_2	0.0010 ± 0.0013
Expected yields	
N_{sig}	$(1.4 \pm 1.1) \times 10^3$
N_{bkg}	$(1.6011 \pm 0.0017) \times 10^6$

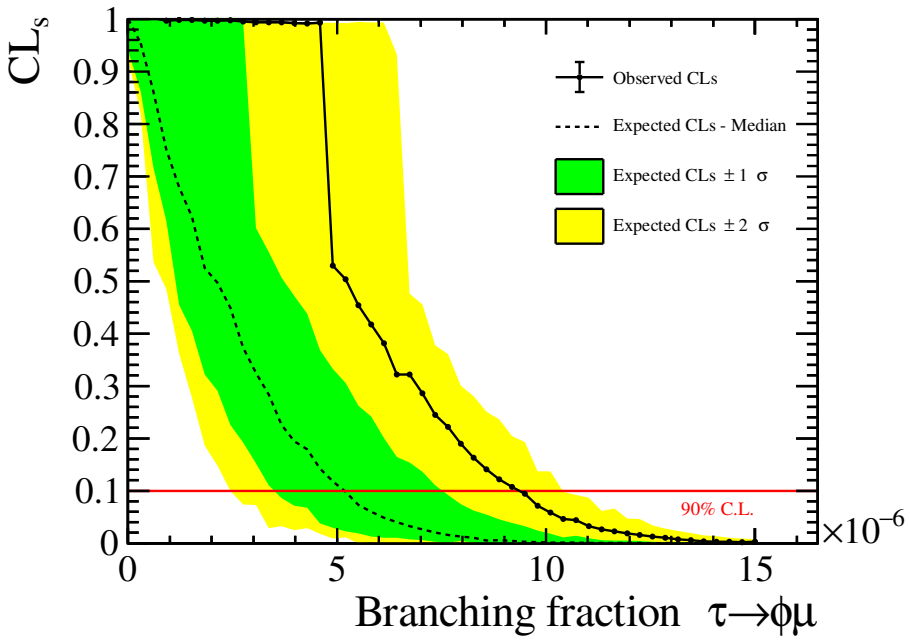


Figure 11.5: Evaluation of the CL_s curve on the unblinded data sample. The distribution of the test statistics is calculated with the toy MC method.

and the observed upper limit takes the value:

$$\mathcal{B}(\tau \to \phi\mu) < 9.4 \times 10^{-6} \tag{11.7}$$

Comparing with the extrapolated result from the sidebands fit (section. 11.2), the expected upper limit has shifted. The reason is that the background model extrapolated from the sidebands fit does not describe the actual background well. A systematic uncertainty asscoiated with the fit model should be considered.

Comparing the observed CL_s curve with the expected CL_s curve, there is an upwards fluctuation. The observed curve is compatible with the background only hypothesis within 2σ. This does not include a complete evaluation of the systematic uncertainty.

12 Conclusion

In the analysis described in this thesis, the feasibility of a search for the LFV decay $\tau^- \to \phi(K^+K^-)\mu^-$ at the LHCb experiment is studied. The data samples used include the data collected by LHCb during the years 2016-2018, which corresponds to an integrated luminosity of 5.6 fb^{-1}. No signal has been observed. The search result is presented as an upper limit of the branching fraction $\mathcal{B}(\tau \to \phi\mu)$. Without fully evaluating the systematic uncertainty, estimation on the branching fraction $\mathcal{B}(\tau \to \phi\mu)$ is made. The expected upper limit of $\mathcal{B}(\tau \to \phi\mu)$ under the background only hypothesis is:

$$\mathcal{B}(\tau \to \phi\mu) < 5.2 \times 10^{-6} \tag{12.1}$$

at 90% confidence level. The observed data in the signal region gives the observed upper limit:

$$\mathcal{B}(\tau \to \phi\mu) < 9.4 \times 10^{-6} \tag{12.2}$$

at 90% confidence level. Considering the fluctuation and the uncertainty, the observed CL_s curve is compatible with the background only hypothesis. Comparing with the world's best result set by Belle [2], the upper limit of this analysis is about 2 orders worse.

The search result is dominated by the background decay $D_s^+ \to \phi(K^+K^-)\mu^+\nu_\mu$, which has the same final state particles except for the missing neutrino. As neutrinos are undetectable with LHCb, the ability to distinguish the signal decay from this background decay is very limited.

If no significant background reduction can be achieved in the future, the LHCb experiment is not competitive with respect to the B-factories. A better signal to background separation is restricted by the nature of the background.

In the data sample used in this analysis, only the signal candidate part of the event is stored in order to save disk space. If in the future other tracks of the event (underlying event) are available, a better background reduction might be achievable.

Under the background only hypothesis, the upper limit is expected to scale as $\alpha\sqrt{N_{bkg}}$, where α is the normalisation factor and N_{bkg} is the background yield. The normalisation factor is related to the integrated luminosity ($\mathcal{L}$) of the data sample and the signal efficiency (ϵ_{sig}):

$$\alpha \propto 1/(\mathcal{L}\epsilon_{sig}) \tag{12.3}$$

which leads to the expected luminosity scaling of the upper limit (U.L.):

$$\text{U.L.} \propto \frac{\sqrt{N_{bkg}}}{\mathcal{L}\epsilon_{sig}} \propto \frac{\sqrt{\mathcal{L}\epsilon_{bkg}}}{\mathcal{L}\epsilon_{sig}} = \frac{\sqrt{\epsilon_{bkg}}}{\epsilon_{sig}} \cdot \frac{1}{\sqrt{\mathcal{L}}} \tag{12.4}$$

If the ability to distinguish the signal against the background (i.e. $\sqrt{\epsilon_{bkg}}/\epsilon_{sig}$) stays the same, the upper limit will improve as $(\mathcal{L})^{-1/2}$. As described in chapter 3, the LHCb experiment is going through an upgrade and will operate at a higher luminosity in the future. In the future Run 3 and Run 4, LHCb experiment is expected to collect data samples with an integrated luminosity of about 50fb^{-1} [73, 74]. However, if no significant improvement in the signal/background separation is achieved, this increase in luminosity will only be able to improve the upper limit of $\mathcal{B}(\tau \to \phi\mu)$ by a factor about 3, which is still not a competitive result.

Appendix

A Supplements

A.1 Trigger line requirements

Two of the L0 trigger lines are of particular interest for this analysis i.e. the L0Muon line and the L0DiMuon line. A description of them can be found in section 3.4. The thresholds of these two lines are shown in Tab. A.1. The thresholds of the L0 trigger lines can change with different setups during the different time of the data taking. The thresholds in the table are shown as an example.

Table A.1: The thresholds of the L0Muon and L0DiMuon trigger lines during the years 2016, 2017 and 2018. The values are the setup used for the majority of the data [42].

L0 trigger line	$p_T(2016)$	$p_T(2017)$	$p_T(2018)$	SPD threshold
L0Muon	$> 1.8\,\text{GeV}$	$> 1.35\,\text{GeV}$	$> 1.75\,\text{GeV}$	< 450
L0DiMuon	$> 2.25\,\text{GeV}^2$	$> 1.69\,\text{GeV}^2$	$> 3.24\,\text{GeV}^2$	< 900

For the HLT trigger, same as the L0 trigger, the exact configurations also change during the operation of the experiment. As an example, the configurations of the lines of interest are listed as follows.

The HLT1 trigger lines:

- HLT1DiMuonLowMass line:
  - previous level requirement: L0Muon or L0DiMuon.
  - track requirement: $p_T > 0\,\text{MeV}/c$ and $p > 3\,\text{GeV}/c$ and $\chi^2/\text{ndf} < 4$ and $\chi^2_{\text{IP}} > 4$ and track ghost probability < 0.2.
  - IsMuon
  - two tracks DOCA < 0.2mm. DOCA (Distance Of Closest Approach) means the closest distance of two tracks.
  - two tracks vertex $\chi^2_{\text{vtx}} < 25$.
  - $m(\mu\mu) > 220\,\text{MeV}/c^2$ or the two muons have opposite charges.
- HLT1TrackMuon line:
  - previous level requirement: L0Muon or L0DiMuon.
  - track requirement before fit: $p_T > 1045\,\text{MeV}/c$ and $p > 5.7\,\text{GeV}/c$.

- track requirement after fit: $p_T > 1.1\,\text{GeV}/c$ and $p > 6\,\text{GeV}/c$ and $\chi^2/\text{ndf} < 3$ and $\chi^2_{\text{IP}} > 35$ and track ghost probability < 0.2.
- IsMuon

- HLT1TrackMVA line:
  - track requirement before fit: $p_T > 570\,\text{MeV}/c$ and $p > 4.75\,\text{GeV}/c$.
  - track requirement after fit: $p_T > 600\,\text{MeV}/c^2$ and $p > 5\,\text{GeV}/c$ and $\chi^2/\text{ndf} < 2.5$ and track ghost probability < 0.2.
  - further track requirement:

    one of them being true:
    * $p_T > 25\,\text{GeV}/c$ and $\chi^2_{\text{IP}} > 7.4$
    * $1\,\text{GeV}/c < p_T < 25\,\text{GeV}/c$ and

$$log(\chi^2_{\text{IP}}) > \left(\frac{1}{(p_T/(\,\text{GeV}/c) - 1)^2} + 1.1 \times (1 - \frac{p_T}{25\,\text{GeV}/c}) + log(7.4)\right)$$

- HLT1TrackMuonMVA line:
  - previous level requirement: L0Muon or L0DiMuon
  - IsMuon
  - track requirement after fit: $p_T > 600\,\text{MeV}/c$ and $p > 5\,\text{GeV}/c$ and $\chi^2/\text{ndf} < 2.5$ and track ghost probability < 0.2.
  - further track requirement:

    one of them being true:
    * $p_T > 25\,\text{GeV}/c$ and $\chi^2_{\text{IP}} > 7.4$
    * $1\,\text{GeV}/c < p_T < 25\,\text{GeV}/c$ and

$$log(\chi^2_{\text{IP}}) > \left(\frac{1}{(p_T/(\,\text{GeV}/c) - 1)^2} + 1.1 \times (1 - \frac{p_T}{25\,\text{GeV}/c}) + log(7.4)\right)$$

The HLT2 trigger lines:

- HLT2PhiIncPhi line:
  - ϕ candidate requirement: $p_T(\phi) > 2\,\text{GeV}/c$ and $|M(\phi) - M_{PDG}(\phi(1020))| < 20\,\text{MeV}/c^2$ and $\chi^2_{\text{vtx}}/\text{ndf} < 20$
  - kaon daughters requirement: $p_T > 1\,\text{GeV}/c$ and $\chi^2_{\text{IP}} > 9$ and PIDK > 0 and track $\chi^2/\text{ndf} < 5$

- HLT2DiMuonDetached line
  - all muon tracks: $p_T > 300\,\text{MeV}/c$ and $\chi^2_{\text{IP}} > 25$ and track $\chi^2/\text{ndf} < 5$

- $p_T(\mu\mu) > 600\,\text{MeV}/c$ and $\chi^2_{\text{vtx}}/\text{ndf} < 9$
- decay length significance > 9

- HLT2TopoMuMu2Body line
  - require 2 muon daughter candidates by criteria: ProbNN$\mu > 0.2$ and IsMuon and ISLONG
  - requirement on the output of a BDT classifier [75], which is evaluated mainly based on the topological structure of the candidates.
  - corrected mass (M_{cor}): $1\,\text{GeV}/c^2 < M_{cor} < 10\,\text{GeV}/c$. The corrected mass is defined by:

$$M_{cor} = \sqrt{M^2 + |p_{T,\text{missing}}|^2} + |p_{T,\text{missing}}|$$

    where $p_{T,\text{missing}}$ stands for the missing transverse momentum with respect to the particle flight direction [75, 76].

A.2 Multivariate hyperparameters tuning

With the hyperparameters combination: learning rate α_{lr}=0.1, depth=6, subsample=0.5, iterations=1000, the training is performed. Fig. A.1 shows its output. Comparing with the output shown in section 7.4, the ROC curve does not show too much improvement. However, as shown in the left plot of Fig. A.1, there is a considerable discrepancy of the CatBoost output distributions between the training sample and the evaluation sample. The classifier is considered overtrained. The overtraining can introduce a bias in later analysis e.g. the evaluation of the efficiency for the selection based on the MV output.

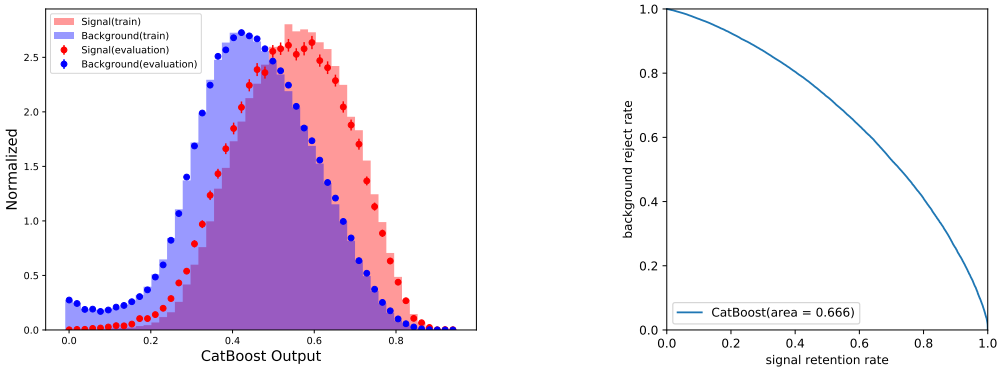


Figure A.1: The multivariate output with the hyperparameters combination: learning rate α_{lr}=0.1, depth=6, subsample=0.5, iterations=1000

A.3 Punzi Figure of Merit slices

Fig. A.2,A.3 andA.4 show the slices of the 2-dimensional Punzi FoM optimization (Fig. 8.3). Each plot shows a curve of FoM against the muon PID variable cut with the MV selection fixed.

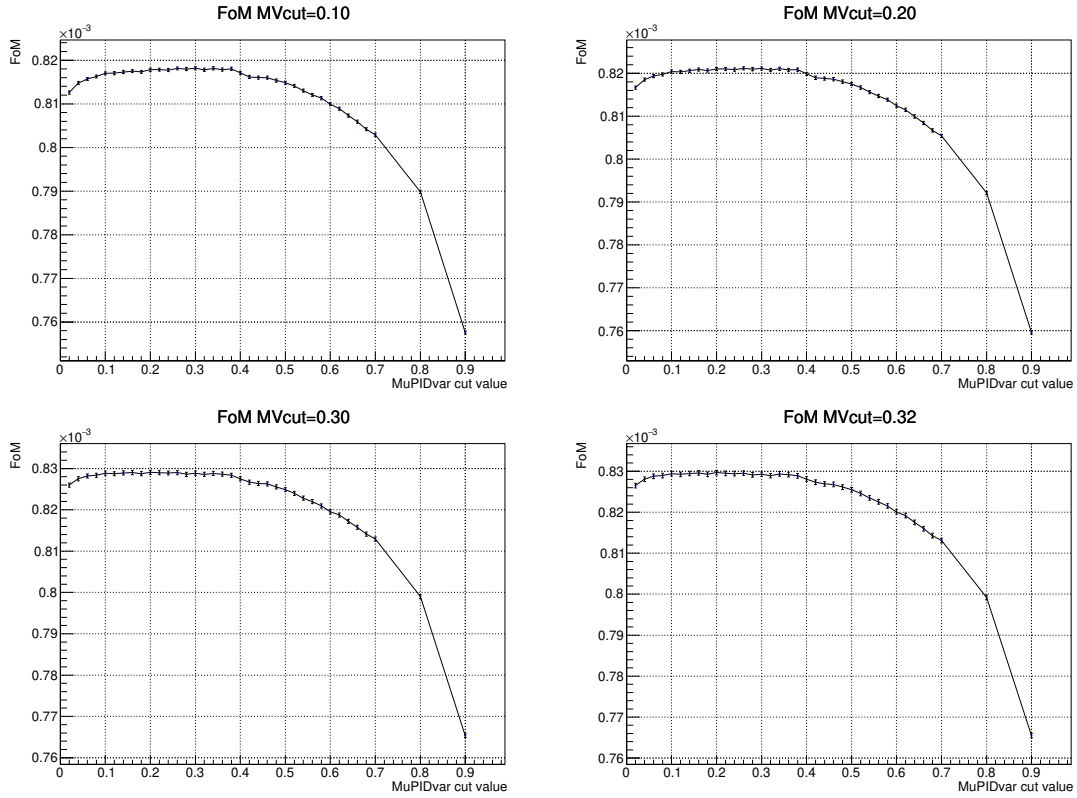


Figure A.2: Punzi Figure of Merit slices (1)

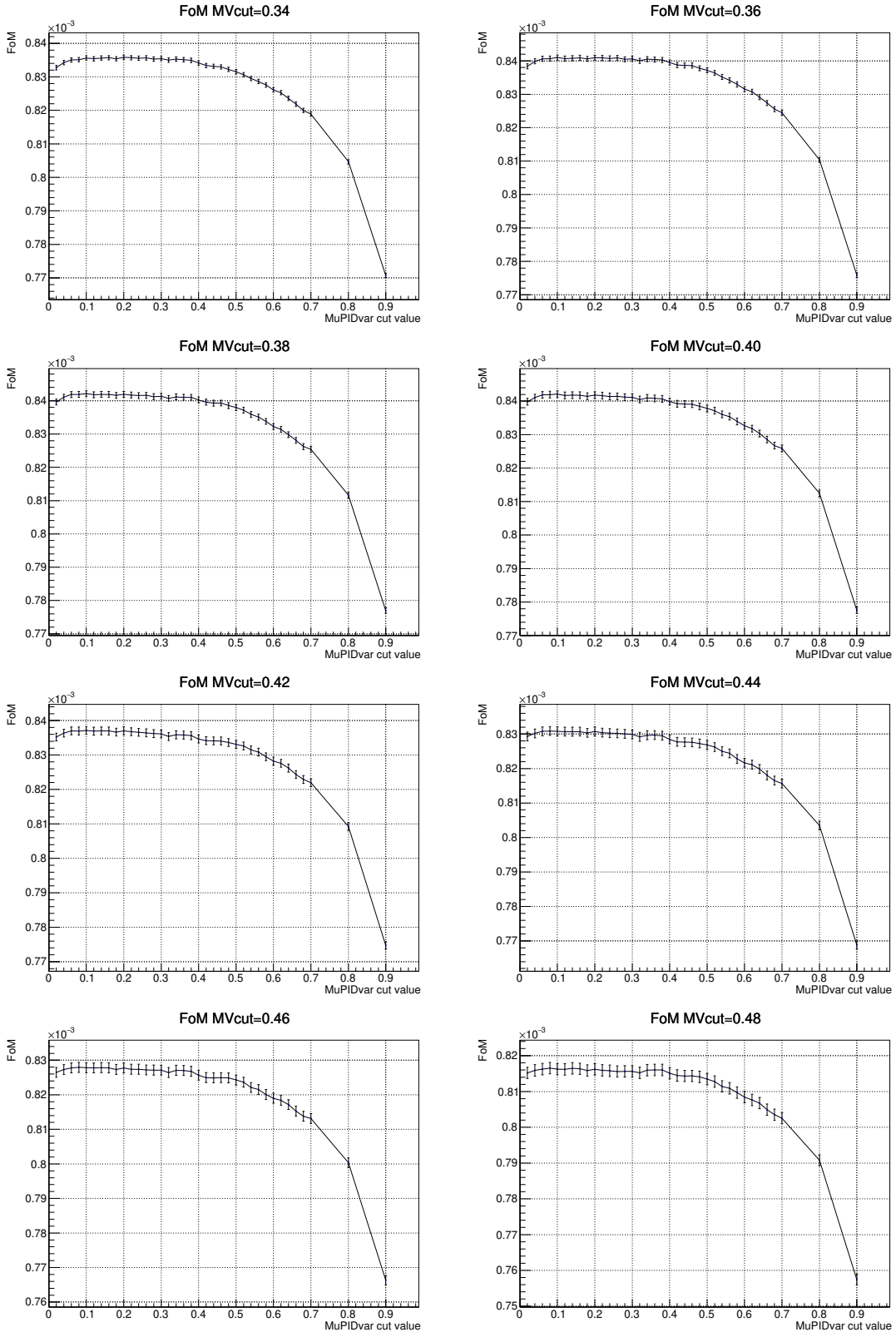


Figure A.3: Punzi Figure of Merit slices (2)

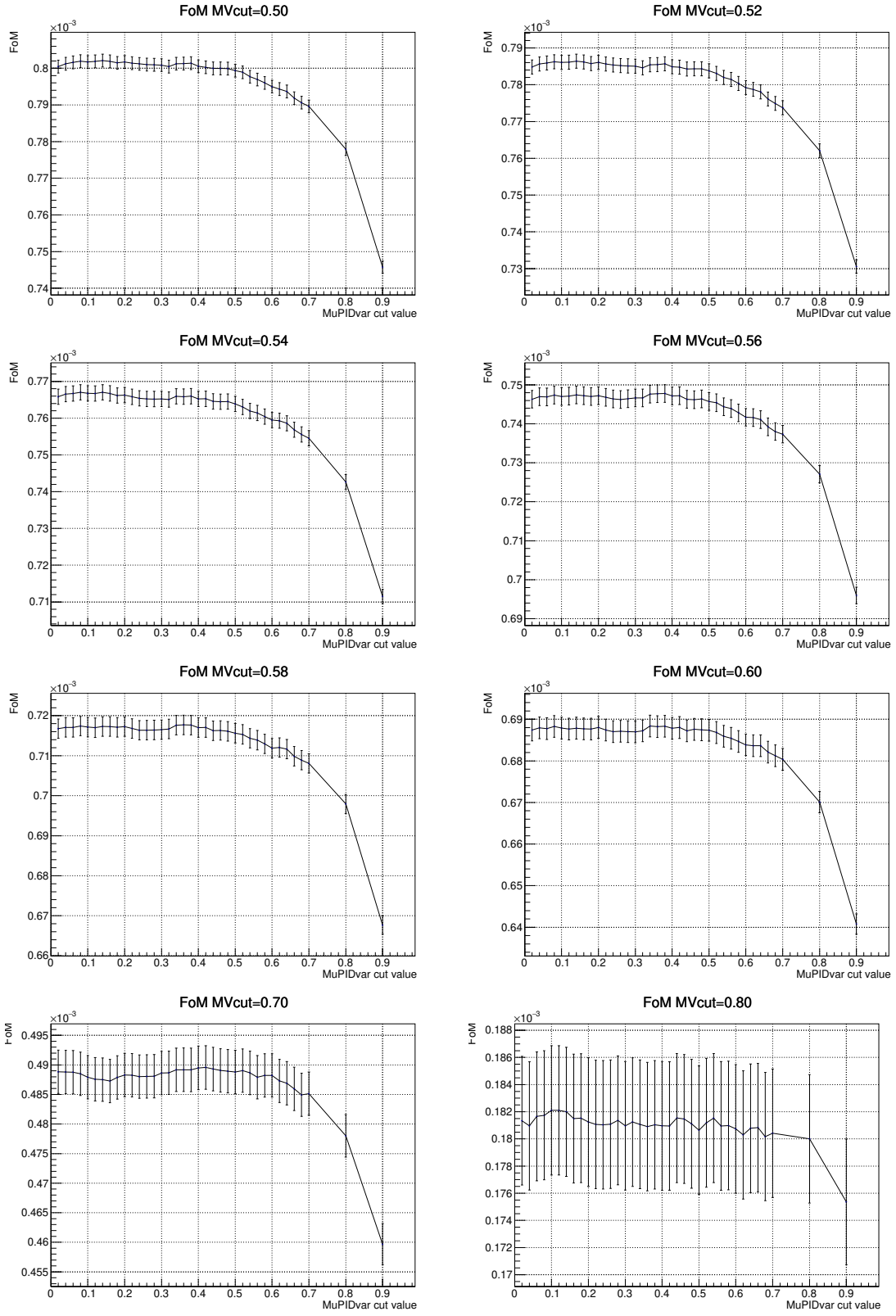


Figure A.4: Punzi Figure of Merit slices (3)

B Bibliography

[1] B. Aubert, Y. Karyotakis, J. P. Lees, V. Poireau, E. Prencipe, X. Prudent, V. Tisserand, J. Garra Tico, E. Grauges, M. Martinelli, and et al. Improved limits on lepton-flavor-violating τ decays to $l\phi, l\rho, lk^*$ and $l\bar{K}^*$. *Physical Review Letters*, 103(2), Jul 2009.

[2] Y. Miyazaki, H. Aihara, K. Arinstein, V. Aulchenko, A.M. Bakich, and et al. Search for lepton-flavor-violating τ decays into a lepton and a vector meson. *Physics Letters B*, 699(4):251 – 257, 2011.

[3] R. Aaij et al. Search for the lepton flavour violating decay $\tau^- \to \mu^-\mu^+\mu^-$. *JHEP*, 02:121, 2015.

[4] R. Aaij et al. Searches for violation of lepton flavour and baryon number in tau lepton decays at lhcb. *Physics Letters B*, 724(1):36 – 45, 2013.

[5] The Wikipedia community. Standard model, from wikipedia, the free encyclopedia, https://en.wikipedia.org/wiki/standard_model.

[6] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[7] Paul Seyfert. *The search for $\tau \to \mu\mu\mu$ at LHCb*. PhD thesis, Heidelberg U., 1 2015.

[8] Roel Aaij et al. Search for the lepton flavour violating decay $\tau^- \to \mu^-\mu^+\mu^-$. *JHEP*, 02:121, 2015.

[9] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[10] Jogesh C. Pati and Abdus Salam. Lepton number as the fourth "color". *Phys. Rev. D*, 10:275–289, Jul 1974.

[11] Howard Georgi and S. L. Glashow. Unity of all elementary-particle forces. *Phys. Rev. Lett.*, 32:438–441, Feb 1974.

[12] Harald Fritzsch and Peter Minkowski. Unified interactions of leptons and hadrons. *Annals of Physics*, 93(1):193 – 266, 1975.

[13] Roel Aaij et al. Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 122(19):191801, 2019.

[14] R. Aaij et al. Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays. *JHEP*, 08:055, 2017.

[15] R. Aaij et al. Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)/\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$. *Phys. Rev. Lett.*, 120:121801, 2018.

[16] R. Aaij et al. Measurement of the ratio of branching fractions $\mathcal{B}(\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+}\mu^-\overline{\nu}_\mu)$. *Phys. Rev. Lett.*, 115:111803, 2015.

[17] R. Aaij et al. Test of lepton flavor universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays. *Phys. Rev.*, D97:072013, 2018.

[18] J.P. Lees et al. Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons. *Phys. Rev. D*, 88(7):072012, 2013.

[19] Y. Sato et al. Measurement of the branching ratio of $\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method. *Phys. Rev. D*, 94(7):072007, 2016.

[20] R. Aaij et al. Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using $3\,\text{fb}^{-1}$ of integrated luminosity. *JHEP*, 02:104, 2016.

[21] Roel Aaij et al. Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay. *Phys. Rev. Lett.*, 125(1):011802, 2020.

[22] Wolfgang Altmannshofer, Stefania Gori, Maxim Pospelov, and Itay Yavin. Quark flavor transitions in $L_\mu - L_\tau$ models. *Phys. Rev. D*, 89:095033, 2014.

[23] Andreas Crivellin, Lars Hofer, Joaquim Matias, Ulrich Nierste, Stefan Pokorski, and Janusz Rosiek. Lepton-flavour violating B decays in generic Z' models. *Phys. Rev. D*, 92(5):054013, 2015.

[24] E. Arganda, M.J. Herrero, and J. Portoles. Lepton flavour violating semileptonic tau decays in constrained MSSM-seesaw scenarios. *JHEP*, 06:079, 2008.

[25] Andrea Brignole and Anna Rossi. Anatomy and phenomenology of mu-tau lepton flavor violation in the MSSM. *Nucl. Phys. B*, 701:3–53, 2004.

[26] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001–S08001, aug 2008.

[27] S. Amato et al. LHCb technical proposal. 1998.

[28] A. A. Alves Jr. et al. The LHCb detector at the LHC. *JINST*, 3:S08005, 2008.

[29] LHCb collaboration. LHCb- large hadron collider beauty experiment. online at https://lhcb-public.web.cern.ch/.

[30] R. Aaij et al. Performance of the LHCb Vertex Locator. *JINST*, 9:P09007, 2014.

[31] LHCb collaboration. *LHCb VELO (VErtex LOcator): Technical Design Report.* Number CERN-LHCC-2001-011 in Technical design report. LHCb. CERN, Geneva, 2001.

[32] LHCb collaboration. *LHCb magnet: Technical Design Report.* Number CERN-LHCC-2000-007 in Technical design report. LHCb. CERN, Geneva, 2000.

[33] C. Abellan Beteta et al. Monitoring radiation damage in the LHCb Tracker Turicensis. 2018.

[34] LHCb collaboration. *LHCb inner tracker: Technical Design Report.* Number CERN-LHCC-2002-029 in Technical design report. LHCb. CERN, Geneva, 2002. revised version number 1 submitted on 2002-11-13 14:14:34.

[35] Ph. d'Argent et al. Improved performance of the LHCb Outer Tracker in LHC Run 2. *JINST*, 9:P11016, 2017.

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

[37] R. Aaij et al. The LHCb trigger and its performance in 2011. *JINST*, 8:P04022, 2013.

[38] F. Archilli et al. Performance of the muon identification at LHCb. *JINST*, 8:P10020, 2013.

[39] LHCb collaboration. The LHCb starterkit lessons. online at https://lhcb.github.io/starterkit-lessons/first-analysis-steps/run-2-data-flow.html.

[40] LHCb collaboration. The MOORE project. online at http://lhcbdoc.web.cern.ch/lhcbdoc/moore/.

[41] LHCb collaboration. The BRUNEL project. online at http://lhcbdoc.web.cern.ch/lhcbdoc/brunel/.

[42] Roel Aaij et al. Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. (LHCb-DP-2019-001), 2018.

[43] R. Aaij et al. Tesla: an application for real-time data analysis in High Energy Physics. *Comput. Phys. Commun.*, 208:35–42, 2016.

[44] LHCb collaboration. The DAVINCI project. online at https://lhcbdoc.web.cern.ch/lhcbdoc/davinci/.

[45] ROOT data analysis framework. online at https://root.cern/.

[46] LHCb collaboration. The GAUSS project. online at http://lhcbdoc.web.cern.ch/lhcbdoc/gauss/.

[47] Torbjörn Sjöstrand, Stephen Mrenna, and Peter" Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[48] Torbjörn Sjöstrand, Stephen Mrenna, and Peter" Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[49] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[50] John Allison, K. Amako, J. Apostolakis, H. Araujo, P.A. Dubois, et al. Geant4 developments and applications. *IEEE Trans.Nucl.Sci.*, 53:270, 2006.

[51] S. Agostinelli et al. Geant4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250, 2003.

[52] LHCb collaboration. The BOOLE project. online at http://lhcbdoc.web.cern.ch/lhcbdoc/boole/.

[53] M. Tanabashi et al. Review of particle physics. *Phys. Rev.*, D98:030001, 2018. and 2019 update.

[54] R. Aaij et al. Erratum to: Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 05:074, 2017.

[55] R. Aaij and et al. Erratum to: Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13TeV$. *JHEP*, 05, 2017.

[56] G. A. Cowan, D. C. Craik, and M. D. Needham. RapidSim: an application for the fast simulation of heavy-quark hadron decays. *Comput. Phys. Commun.*, 214:239–246, 2017.

[57] LHCb collaboration. The stripping project. online at https://lhcbdoc.web.cern.ch/lhcbdoc/stripping/.

[58] Lucio Anderlini, Sean Benson, Olli Lupton, Barbara Sciascia, and Vladimir 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.

[59] S Tolk, J Albrecht, F Dettori, and A Pellegrino. Data driven trigger efficiency determination at LHCb, 2014.

[60] Muriel Pivk and Francois R. Le Diberder. sPlot: A statistical tool to unfold data distributions. *Nucl. Instrum. Meth.*, A555:356–369, 2005.

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

[62] Fabian Pedregosa et al. Scikit-learn: Machine learning in Python. *J. Machine Learning Res.*, 12:2825, 2011. and online at http://scikit-learn.org/stable/.

[63] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features, 2017.

[64] Jerome H. Friedmann. Stochastic gradient boosting. 1999.

[65] G Mancinelli and J Serrano. Study of Muon Isolation in the $B_s^0 \to \mu^+\mu^-$ Channel. Technical Report LHCb-INT-2010-011. CERN-LHCb-INT-2010-011, CERN, Geneva, Mar 2010.

[66] Flavio Archilli, Matteo Palutan, Matteo Rama, Alessio Sarti, Barbara Sciascia, and Ricardo Vazquez Gomez. Background studies for $B^0 \to \mu^+\mu^-$ analysis optimization. Technical Report LHCb-INT-2014-047. CERN-LHCb-INT-2014-047, CERN, Geneva, Nov 2014.

[67] Lucio Anderlini, Andrea Contu, Christopher Rob Jones, Sneha Sirirshkumar Malde, Dominik Muller, Stephen Ogilvy, Juan Martin Otalora Goicochea, Alex Pearce, Ivan Polyakov, Wenbin Qian, Barbara Sciascia, Ricardo Vazquez Gomez, and Yanxi Zhang. The PIDCalib package, 2016.

[68] R. Aaij et al. Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2. 2018.

[69] Giovanni Punzi. Sensitivity of searches for new signals and its optimization. *eConf*, C030908:MODT002, 2003.

[70] A L Read. Presentation of search results: The CL_s technique. *J. Phys.*, G28:2693, 2002.

[71] Roostats package. online at https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome.

[72] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb 2011.

[73] I. Bediaga et al. Framework TDR for the LHCb Upgrade: Technical Design Report. Technical report, Apr 2012.

[74] Physics case for an LHCb Upgrade II. Technical Report LHCb-PUB-2018-009. CERN-LHCb-PUB-2018-009, CERN, Geneva, May 2018.

[75] R Aaij et al. The LHCb Trigger and its Performance in 2011. *JINST*, 8:P04022, 2013.

[76] Albert Puig. The LHCb trigger in 2011 and 2012. Technical Report LHCb-PUB-2014-046. CERN-LHCb-PUB-2014-046, CERN, Geneva, Nov 2014.

Erklärung:

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

Heidelberg, den (Datum) .....