H. Niewodniczański, Institute of Nuclear Physics
Polish Academy of Sciences

Doctoral dissertation

Defended by

Marcin Chrząszcz

Search for Charged Lepton Flavour Violation at LHCb experiment

Thesis Supervisor:
prof. dr hab. Tadeusz Lesiak,
Auxiliary Supervisor:
dr hab. Alberto Lusiani

Krakow, November 2014

Abstract

The thesis describes searches for charged lepton flavour violation in the process $\tau^- \to \mu^-\mu^+\mu^-$. The study is performed at the LHCb experiment data sample corresponding to an integrated luminosity of 3 fb^{-1} of proton-proton collisions at a centre-of-mass energy of 7 TeV (8 TeV), collected in 2011 (2012), respectively. No significantly statistical signal of the decay in question has been found and as a result the upper limit was set: $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 4.6\times10^{-8}$ at 90% confidence level. The limit was then interpreted in terms of an effective field theory approach including beyond the Standard Model operators with different lepton chirality structures. We found that the limits are within the range: $(4.1 - 6.8) \times 10^{-8}$ at 90% confidence level.

The thesis presents also the combination of upper limits for all lepton flavour violating τ lepton decays. For the decay in question, i.e. $\tau^- \to \mu^-\mu^+\mu^-$, the combination leads to the upper limit of $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 1.2 \times 10^{-8}$ at 90% confidence level.

Streszczenie

W rozprawie przedstawiono wyniki poszukiwań rozpadu $\tau^- \to \mu^-\mu^+\mu^-$ łamiącego zachowanie zapachu leptonów. Badania przeprowadzono w ramach współpracy LHCb przy użyciu próbki danych odpowiadającej scałkowanej świetlności 3 fb^{-1}, zebranej w zderzeniach proton-proton przy energii w układzie środka masy wynoszącej 7 TeV (8 TeV) dla okresu zbierania danych, odpowiednio w roku 2011 (2012). Nie zaobserwowano znaczącego statystycznie sygnału pochodzącego od badanego rozpadu i wyznaczono górną granicę częstości jego występowania jako $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 4.6 \times 10^{-8}$ przy poziomie ufności wynoszącym 90%. Wynik ten zinterpretowano następnie w ramach efektywnej teorii pola zawierającej operatory spoza modelu standardowego, uwzględniające różne kombinacje chiralności leptonów. Uzyskano przy tym ograniczenia na poziomie $(4.1 - 6.8) \times 10^{-8}$.

W rozprawie przedstawiono także obecne ograniczenia częstości rozpadów łamiących zachowanie liczby leptonowej, uwzględniające wszystkie obecne wyniki doświadczalne w tej dziedzinie. W szczególności, dla procesu będącego przedmiotem rozprawy, prowadzi to do górnej granicy częstości występowania badanego rozpadu w wysokości $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 1.2 \times 10^{-8}$ przy poziomie ufności 90%.

Acknowledgments

For over seven years I have received support, encouragement and ideal scientific environment to develop my own ideas from prof. Tadesz Lesiak. I would like to thank him for being a mentor, colleague and friend during my scientific journey.

I would also like to thank prof. Alberto Lusiani for passing a small part of his huge knowledge and skills to me. Prof. Lusiani listened to my crazy ideas with stoic calm and allowed me to discover my mistakes on my own.

A big impact on my scientific career had prof. Nicola Serra. Long discussions that took place late at CERN were extremely illuminating and encouraged further scientific ideas. Prof. Serra, who was my supervisor when I was visiting the University of Zurich, gave me a lot of scientific freedom to pursue my own ideas and interests.

Let me express also my deep gratitude to dr Małgorzata Duraj. It is thanks to her support and kind help that I ended up doing science.

I would also like to thank all the founding agencies: Polish Ministry of Science for the Diamond Grand, SCIEX foundation for the scholarship, INFN Pisa for the visiting contracts, KNOW for PhD scholarship and National Science Center for travel grants. My special thanks also go to H. Niewodniczański, Institute of Nuclear Physics, Polish Academy of Sciences for healthy work environment and support for all my work.

Last but not least, I would like to thank my parents for constant support and interest in my research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Standard Model (SM) of particle physics is the current theory describing fundamental particles as well as strong, weak and electromagnetic interactions. It was formulated in the 1960−1970s and since then has passed with flying colours extensive experimental tests, including the recent discovery of the Higgs boson. Despite these facts, the SM has several severe drawbacks and limitations which lead to extensive studies of extended theories, commonly labelled as physics Beyond the Standard Model (BSM).

Most of the studies performed in this thesis are devoted to the search for the phenomenon of Lepton Flavour Violation (LFV), which has been predicted in a vast category of BSM models and has been searched for over the past sixty years. Studies of LFV can shed some light on two particular drawbacks of the SM. Namely, it does not describe the nature of neutrino oscillations and does not provide the quantitative mechanism of baryogenesis. The studies described in this thesis aimed mostly at searching for lepton flavour violation in decay of $\tau^- \to \mu^-\mu^+\mu^-$ and were performed within the LHCb experiment.

Lepton flavour violation has already been observed in the sector of neutral leptons as neutrino oscillations, thus providing clearly the first evidence of a phenomenon which is not described by the SM. However, no evidence for the LFV in the sector of charged leptons has been found so far.

The thesis describes the first search for the decay $\tau^- \to \mu^-\mu^+\mu^-$ performed at the hadron collider. Similar studies have been carried out at e^-e^+ colliders. In comparison with the latter, the environment of hadronic collisions has much harsher conditions. In spite of this fact the results obtained in this study are compatible with the ones originating from e^-e^+ colliders.

The dissertation is organized as follows: Chapter 1 describes the theoretical foundations of the SM as well as the BSM models. Also it provides a model independent description of the decay $\tau^- \to \mu^-\mu^+\mu^-$, which was implemented by the author in the TAUOLA Monte Carlo (MC) generator.

Chapter 2 introduces the experimental apparatus of the LHCb detector. Special attention is put on subdetectors that are of particular importance to this analysis. The results of data analysis in search for the decay $\tau^- \to \mu^-\mu^+\mu^-$ are presented in Chapter 3. The chapter first discusses the preselection requirements. Next, a blending technique is described, which was first introduced into data analysis in particle physics by the author and proved to offer a gain in sensitivity of 6%. In the following parts of this chapter particle identification and calibration issues are discussed. Finally background estimation is performed and in view of the lack of

observation of the decay in question, a limit on the respective branching fraction is computed.

Chapter 4 describes the work performed by the author within the Heavy Flavour Averaging Group, which yielded a combination of limits among all lepton flavour violation τ decays.

Chapter 2

Theoretical formalism of charged lepton flavour violation

2.1 History of lepton flavour violation

The search for Charged Lepton Flavour Violation (CLFV) processes has been ongoing since the discovery of the muon by C. Anderson and S. Neddermeyer in 1936 in Caltech [1], which was confirmed a year later by J. Street and E. Stevenson [2]. Because of the mass value matching the expectations, for a carrier of strong nuclear force, the muon was firstly identified as a Yukawa's meson [3]. However, further experiments led to the discovery of another particle of similar mass — the pion. Contrary to the muon, the pion was proved to participate in strong interactions and was eventually identified as Yukawa's meson. As a result, the muon was recognized as a "heavy electron", which was a big surprise at that time. A Nobel Laureate I. Rabi was supposed to quip: "Who ordered that?", commenting in this way on the present situation in particle physics.

A natural solution to this problem was to treat the muon as an excited electron. In this case one expected to observe the decay $\mu^- \to e^-\gamma$[1] with a branching fraction $\mathcal{B}(\mu^- \to e^-\gamma) \sim 10^{-4}$, unless, as pointed by R. Feinberg [4], there existed a second neutrino.

In 1962 the second muon neutrino was discovered at the Alternating Gradient Synchrotron in Brookhaven [5]. In the experiment led by L. Lederman, M. Schwartz and J. Steinberger the neutrinos from the decays $\pi^\pm \to \mu^\pm(\nu/\bar{\nu})$ were used [6]. Their interactions with matter yielded muons but no single electron, thus implying the existence of the second muon neutrino.

A decade later, a group led by M. Perl at Stanford Linear Accelerator (SLAC) [7] observed the process:

$$e^+ + e^- \to e^\pm\mu^\mp + nX, \tag{2.1}$$

where at least two particles remained undetected [8], as deduced from four-momentum conservation. Thus the process from Eq. 2.1 was interpreted as a decay with the participation of a pair of the third lepton τ and its antiparticle:

$$e^+ + e^- \to \tau^-\tau^+ \to e^\pm\mu^\mp + 4\nu. \tag{2.2}$$

This conjecture was unfortunately hard to prove due to the presence of $\psi(3770)$ resonance, which is very close in mass to $\tau\tau$ threshold. However, further experiments

[1] Charge-conjugate modes are included everywhere, unless otherwise specified.

done at Deutsches Elektronen-Synchrotron (DESY) [9] provided precise measurements of τ spin and mass. The discovery of the τ lepton immediately implied the existence of ν_τ, i.e. a neutrino associated to the third lepton. At the end of 2000 the DONUT collaboration [10] reported the observation of ν_τ [11]. The neutrino beam was formed from 800 GeV protons dumped on a tungsten target. Using a special emulsion, the DONUT collaboration was able to distinguish a separated τ vertex, which proved the existence of the ν_τ.

Finally, it is appropriate to mention the discovery of neutrino oscillation, i.e. a phenomenon in which a neutrino created with a definite lepton flavour can later be measured to carry a different flavour. This kind of neutral flavour violation was predicted by B. Pontecorvo in the 1950s [12]. The evidence for this effect has been building up from the 1960s, first in the experiments by R. Davis [13], who saw a deficit in the flux of solar neutrinos. In 1998 the oscillation was also observed for atmospheric neutrinos in the Super Kamiokande experiment [14]. Later, this phenomenon was also confirmed in experiments using reactor neutrinos as well as neutrino beams (for review see [15]). The observation of the mixing phenomenon proved unequivocally that for neutrinos the flavour eigenstates do not coincide with the mass eigenstates. This is possible only if neutrinos are massive particles. Thus the neutrino mixing constitutes the first clear-cut violation of the SM predictions and leads to a wide scope of its possible extensions.

2.2 Standard Model of particle physics

2.2.1 Basics of the Standard Model

The Standard Model [16] is a Quantum Field Theory (QFT) which describes the strong, weak and electromagnetic interactions via exchange of the respective spin-one bosons between spin-half particles that constitute the building blocks of matter. The symmetries of the SM are in the form of a direct product of three groups: $SU_c(3) \times SU_L(2) \times U_Y(1)$ (the indices are explained below). Eight bosons associated with the group $SU(3)_c$ and responsible for strong interactions are called gluons and denoted as G^α_μ ($\alpha = 1, 2 \ldots 8$ and $\mu = 0, 1, 2, 3$ numerates space-time coordinates). They are massless and carry a quantum number called colour, hence the subscript "c" in the group's name. Four spin-one bosons arising from $SU_L(2) \times U_Y(1)$ group, $W^\pm$, Z^0 and γ mediate the weak and electromagnetic interactions. These two forces will be discussed in detail in Sect. 2.2.2 and 2.2.3.

The fermions, i.e. fundamental matter building blocks, are grouped accordingly to their transformation properties with respect to the gauge group $SU_c(3) \times SU_L(2) \times U_Y(1)$. The pattern of elementary fermions repeats itself three times in the form of the so called generations. Fermions are also divided into two categories: quarks and leptons. The former participate in strong interactions, the latter do not. There are three quarks with fractional electric charge $\frac{2}{3}$ ($-\frac{1}{3}$), denoted as u, c and t (d, s and b), respectively, see Fig. 2.1. The quarks are not observed in nature as free

particles. Instead, they form bound states called hadrons, which are composed either of three quarks (baryons) or a quark-antiquark pair (mesons) [17, 18]. All hadrons are colourless [19].

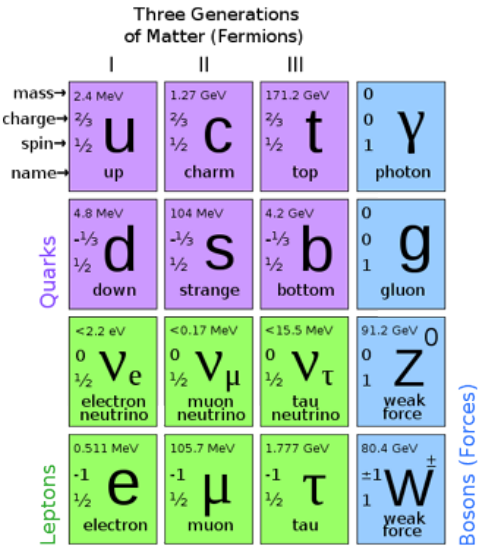


Figure 2.1: Elementary particles and their properties. Particles are grouped in three types: quarks, leptons and intermediate bosons. The first three columns correspond to the generations of fermions.

Leptons are particles that do not interact via strong interactions. To date six leptons have been discovered (three charged and three neutral), as presented in Sect. 2.1. As mentioned above, they are grouped in three generations: $(\mathrm{e}, \nu_\mathrm{e})$, (μ, ν_μ), (τ, ν_τ). The lepton number (flavour) L is defined as a quantum number with value $+1$ for leptons, -1 for anti-leptons and 0 for non-leptonic particles, see Fig. 2.1. Historically, this number was introduced to express the experimental fact that the overall difference between the number of leptons and anti-leptons in a given process is a conserved quantity. Three additional lepton numbers, denoted as L_f, $f = \mathrm{e}, \mu, \tau$ [20], each corresponding to the leptons from a single generation, have also been introduced, which can be seen in Table 2.1.

It is worthwhile to underline here that there is certainly some confusion over the terminology regarding the relation between "lepton number" and "lepton flavour", in particular in the context of violation of the abovementioned quantum numbers. Contrary to many textbooks in particle physics (and also to the definitions from the previous paragraph), in the recent literature related to the matters of CLFV it is widely accepted to treat the "lepton number" as the total number of leptons, where each lepton contributes +1, and each anti-lepton contributes -1. At the same time the "lepton flavour (violation)" is generally associated with the separate lepton number (violation) in each individual lepton family. It is thus almost unavoidable to comply in this thesis with this terminology and treat the decay in question $\tau^- \rightarrow$

Table 2.1: Assignment of lepton numbers L and L_f, $f = \mathrm{e},\ \mu,\ \tau$ to elementary leptons (the quark states assume values of zero for these quantum numbers).

Particle	L	L_e	L_μ	L_τ
e	1	1	0	0
μ	1	0	1	0
τ	1	0	0	1
ν_e	1	1	0	0
ν_μ	1	0	1	0
ν_τ	1	0	0	1

$\mu^-\mu^+\mu^-$ as the one which conserves the lepton number but violates lepton flavour. A similar decay $\tau^- \to \mathrm{p}\mu^-\mu^-$ would then violate both lepton number and lepton flavour.

2.2.2 Weak interactions

The first manifestations of weak interactions corresponded with the observations of beta decays of radioactive atoms. Here the basic process was $\mathrm{n} \to \mathrm{pe}^-\overline{\nu}_\mathrm{e}$. In such decays two particles (p and e^-) were seen in the final state. Based on the postulate of four-momentum conservation, W. Pauli proposed the existence of a spin half particle called the neutrino. In 1933 E. Fermi formulated the theory of weak interactions modifying the well-know current-current couplings of quantum electrodynamics. In the 1950s T. Lee and C. Yang [21] made a bold conjecture that parity transformation that changes the space coordinates: $\vec{x} \to -\vec{x}$; might be violated in the weak interaction. This hypothesis was confirmed in 1956 by C. S. Wu [22] using β decay of ${}^{60}_{27}\mathrm{Co}$:

$$ {}^{60}_{27}\mathrm{Co} \to {}^{60}_{28}\mathrm{Ni}\ \mathrm{e}^-\ \nu_\mathrm{e}\ 2\gamma. \tag{2.3} $$

Here, the two photons originated from ${}^{60}_{28}\mathrm{Ni}$ because it was an excited state. These photons were used as the signature of the decay in question. The results of this experiment indicated that the violation of parity in weak interactions is maximal. Based on that, massless neutrinos were included into Fermi's theory as purely left-handed states (and right-handed for anti-neutrinos). The weak interaction Lagrangian was postulated as the product of two currents:

$$ \mathcal{L}_\mathcal{W} = -\frac{G_F}{2}\sum_{\alpha=0}^{\alpha=3} J^\alpha(x)^\dagger J_\alpha(x), \tag{2.4} $$

where G_F is the Fermi constant [23] and the currents J_α read:

$$ J_\alpha = \overline{\nu}_\mathrm{e}(x)\gamma_\alpha(1-\gamma_5)\mathrm{e}(x) + \overline{\nu}_\mu(x)\gamma_\alpha(1-\gamma_5)\mu(x) + \overline{\nu}_\tau(x)\gamma_\alpha(1-\gamma_5)\tau(x), \tag{2.5} $$

where $x = (t, \vec{x})$ denotes the position of four-vector[2], the fermion fields are labelled after the respective particle names and γ_α are the gamma (Dirac) matrices.

The current J_α changes the electric charge by $\Delta Q = 1$, while the $J^\alpha(x)^\dagger$ modifies the charge by $\Delta Q = -1$. Thus the weak interaction Lagrangian is described as current-current interactions. The current itself can be decomposed into parity conserving vector term:

$$V_\alpha = \sum_{\ell=\mathrm{e},\mu,\tau} \overline{\nu}_\ell(x)\gamma_\alpha\ell(x), \tag{2.6}$$

and parity violating axial vector part:

$$A_\alpha = \sum_{\ell=\mathrm{e},\mu,\tau} \overline{\nu}_\ell(x)\gamma_\alpha\gamma_5\ell(x), \tag{2.7}$$

so the full current can be written as:

$$J_\alpha = V_\alpha - A_\alpha. \tag{2.8}$$

As a result, the $V - A$ combination is the only one which enters $\mathcal{L}_W$, which is the mathematical manifestation of the fact that neutrino field is only left-handed. This construction yields also parity violation as the $V - A$ combination transforms into $V + A$ under the application of the parity operator. Moreover, the operation of charge conjugation does not change the vector term and reverses the sign of the axial vector. As a result, $\mathcal{L}_\mathcal{W}$ is invariant under the CP operation, which is the combination of the previously discussed P and C transformations. Based on the Lagrangian $\mathcal{L}_\mathcal{W}$, the precise description of the decay $\mu^- \rightarrow \mathrm{e}^- \, \overline{\nu}_\mathrm{e} \, \nu_\mu$ was provided and confronted with the respective experimental observables. In consequence, this decay turned out to be the main test of the space-time structure of the Fermi theory of weak interactions.

2.2.3 Electroweak theory

Despite the fact that the Fermi model of the weak interaction successfully passed many experimental tests, it suffered from some problems. The most important ones were:

- divergences in higher-order processes (lack of renormalizability),
- lack of neutral current in $V - A$ couplings,
- discrepancy of neutrino-lepton scattering cross-section.

A generalization of this model, constituting also the unification of weak and electromagnetic interactions, is called the electroweak theory. It was elaborated on by S. Weinberg, A. Salam and S. Glashow in 1967—1968 [24, 25, 26]. The electroweak model assumed the existence of four intermediate bosons W^+, W^-, Z^0 and

[2] the natural system of units is used in this thesis

Table 2.2: Eigenvalues of the third component of weak isospin (T_3), weak hypercharge (Y) and electric charge (Q) for elementary leptons.

Fermion Type			T_3	Y	Q
$\nu_{\mathrm{e}L}$	$\nu_{\mu L}$	$\nu_{\tau L}$	$1/2$	$-1/2$	0
e_L	μ_L	τ_L	$-1/2$	$-1/2$	-1
$\nu_{\mathrm{e}R}$	$\nu_{\mu R}$	$\nu_{\tau R}$	0	0	0
e_R	μ_R	τ_R	0	-1	-1

γ. The states $\mathrm{W}^{\pm}(\mathrm{Z}^0)$ are responsible for charged (neutral) weak currents, respectively, while the photon mediates electromagnetic interactions. In order to fulfill the above assumptions, a theory based on two vector fields was developed. The first (second) one constitutes a weak-isospin triplet (singlet), denoted as $A^{\mu}_{i=1,2,3}$, B^{μ}, respectively. These fields are directly related to physical particles W^+, W^-, Z^0 and γ. Since weak interactions violate the parity, the left-handed and right-handed fields need to behave differently. A standard procedure to include this property is to place the above fields in separate representations of the symmetry group(s) of the model. In the case of the electroweak interactions the left-handed leptons form isospin doublets ($L_{\ell=\mathrm{e},\mu,\tau}$) of the group $SU(2)_L$:

$$L_e = \frac{1-\gamma_5}{2}\begin{pmatrix}\psi_{\nu_\mathrm{e}}\\ \psi_e\end{pmatrix},\ L_\mu = \frac{1-\gamma_5}{2}\begin{pmatrix}\psi_{\nu_\mu}\\ \psi_\mu\end{pmatrix},\ L_\tau = \frac{1-\gamma_5}{2}\begin{pmatrix}\psi_{\nu_\tau}\\ \psi_\tau\end{pmatrix}, \tag{2.9}$$

while the right-handed components of the lepton field are singlets of weak isospin:

$$R_e = \frac{1+\gamma_5}{2}\psi_{\nu_\mathrm{e}},\ R_\mu = \frac{1+\gamma_5}{2}\psi_{\nu_\mu},\ R_\tau = \frac{1+\gamma_5}{2}\psi_{\nu_\tau}. \tag{2.10}$$

Thus the kinetic part of the Lagrangian can be expressed as:

$$\mathcal{L}_{kin} = \sum_{\ell=\mathrm{e},\mu.\tau}\sum_{\mu=0}^{3}\bar{L}_\ell i\gamma^\mu\partial_\mu L_\ell + \bar{R}_\ell i\gamma^\mu\partial_\mu R_\ell, \tag{2.11}$$

and is invariant under transformations of both $SU(2)_L$ and $U(1)_Y$ groups.

Within this framework, the electromagnetic interaction based on $U(1)_Q$ gauge transformations can be expressed as the product of an element belonging to a $U(1)$ subgroup of the weak $SU(2)_L$ and an element attributed to $U(1)_Y$ group. This yields the relation:

$$Q = T_3 + Y, \tag{2.12}$$

where T_3 denotes the third component of the weak isospin (eigenvalues of the third generator of the $SU(2)_L$). The values of quantum numbers T_3, Y and Q for elementary leptons are listed in Table 2.2.

Having defined the symmetries of electroweak interactions, we can formulate the general Lagrangian of electroweak theory:

$$\mathcal{L} = \mathcal{L}_{gauge} + \mathcal{L}_f + \mathcal{L}_\phi, \tag{2.13}$$

which refers to the gauge, fermion and Higgs sectors, respectively. The gauge part is given by:

$$\mathcal{L} = -\frac{1}{4}\sum_{\mu,\nu=0}^{3} F_i^{\mu\nu}F_{\mu\nu}^i - \frac{1}{4}\sum_{\mu,\mu=3}^{3} B_i^{\mu\nu}B_{\mu\nu}^i, \tag{2.14}$$

where the field tensors of $SU(2)_L$ and $U(1)_Y$ groups are defined as:

$$F_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g' \sum_{j,k=0}^{3} \varepsilon_{ijk} W_\mu^j W_\nu^k, \tag{2.15}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.16}$$

where μ, ν are the Lorentz indices, i is the tensor index, g' is the coupling constant for the $SU(2)_L$ gauge field and ε_{ijk} is the Levi-Civita antisymmetric tensor. Both $F_{\mu\nu}^i$ and $B_{\mu\nu}$ tensors include kinetic terms. The last expression in Eq. 2.15 is responsible for the self-interaction of W bosons. The latter is a direct consequence of the fact that the $SU(2)_L$ group is non-abelian.

Interactions with the fermion can be written in terms of covariant derivatives:

$$\mathcal{L}_f = \sum_f i\bar{\psi}_f D_\mu \gamma^\mu \psi^f, \tag{2.17}$$

where

$$D_\mu = \partial_\mu + i\frac{g}{2} Y B_\mu + i g' T_a W_\mu^a, \tag{2.18}$$

and g is the coupling constant of the $U(1)_Y$ gauge field and $a = 1,\ 2,\ 3$. It is worth pointing out that the $SU(2)_L$ and $U(1)_Y$ representations are chiral so no fermion mass terms are allowed. In the SM gauge the invariant mass terms are generated via coupling to a complex doublet spin-zero field:

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \tag{2.19}$$

of spin zero boson, which spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry. This so called Higgs sector of Lagrangian Eq. 2.13 reads:

$$\mathcal{L}_H = D_\mu \phi^\dagger D^\mu \phi - V(\phi) + \sum_f c_f (\bar{\psi}_f^L \phi^\dagger \psi_f^R + \bar{\psi}_f^R \phi \psi_f^L), \tag{2.20}$$

where the potential, containing two parameters λ and μ, is defined as:

$$V(\phi) = \mu \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.21}$$

For $\lambda > 0$ and $\mu^2 < 0$ the Higgs field from Eq. 2.19 can be rotated so that only the lower component remains:

$$\phi = \begin{pmatrix} 0 \\ H + \nu \end{pmatrix}, \tag{2.22}$$

where $\nu = \sqrt{\frac{-\mu^2}{\lambda}}$. The latter choice breaks the symmetry $SU(2)_L \times U(1)_Y$ but conserves the other one, associated with electric charge and the group $U(1)_Q$. This spontaneous symmetry breaking generates masses of $W^{\pm}$ and Z bosons and induces the mixing between the gauge fields:

$$A_\mu = \sin\Theta_W W^3 + \cos\Theta_W B_\mu, \tag{2.23}$$

$$Z_\mu = \cos\Theta_W W^3 - \sin\Theta_W B_\mu. \tag{2.24}$$

Here the mixing angle Θ_W (the so called Weinberg angle) is defined as the ratio of coupling constants which appeared in Eq. 2.18

$$\tan\Theta_W = \frac{g}{g'}. \tag{2.25}$$

Masses of the weak intermediate bosons are then expressed in terms of couplings as:

$$M_{\mathrm{W}} = \frac{\nu}{2} g', \tag{2.26}$$

$$M_{\mathrm{Z}} = \frac{\nu}{2}\sqrt{g^2 + g'^2}. \tag{2.27}$$

2.2.4 Quark mixing

Electroweak couplings of gauge fields to quarks are analogous to the ones for leptons. For a general fermion field which is a representation in $SU(2)_L \times U(1)_Y$, one can write Lagrangian as in Eq. 2.17. At the time when the electroweak theory was formulated, only three quarks u, d and s had been observed experimentally. In such a case the Lagrangian led to the expectation of the so-called flavour changing neutral currents (FCNC) transitions. However, these processes have not been observed experimentally. The apparent suppression of FCNCs was explained by the so-called GIM mechanism (Glashow-Iliopoulos-Maiani [27]), i.e. by postulating the existence of the fourth charm (c) quark forming a left-handed doublet with the strange quark. This expectation was soon confirmed experimentally. Later, this structure was complemented by the third doublet composed of b and t quarks. Thus the charged current for quarks can be written as:

$$\left(\overline{u} \quad \overline{c} \quad \overline{t}\right) \gamma^\mu (1-\gamma_5) V_{CKM} \left(\overline{d} \quad \overline{s} \quad \overline{b}\right)^T, \tag{2.28}$$

where V_{CKM} is Cabbio-Kabayashi-Maskawa (CKM) matrix [28]. V_{CKM} can be regarded as the transformation matrix between flavour eigenstates and mass eigenstates. The V_{ij} element in Eq. 2.31 can be interpreted as the mixing strength between i and j quark. The CKM structure encompasses in a natural way the GIM mechanism of FCNC suppression.

The neutral current in this case reads:

$$J_3^\mu = \frac{1}{2}\left(\overline{u_L} \quad \overline{c_L} \quad \overline{t_L}\right) \gamma^\mu \left(u_L \quad c_L \quad t_L\right)^T \tag{2.29}$$

$$-\frac{1}{2}\left(\overline{d_L} \quad \overline{s_L} \quad \overline{b_L}\right) V_{CKM}^{-1} \gamma^\mu V_{CKM} \left(d_L \quad s_L \quad b_L\right)^T, \tag{2.30}$$

The important impact of this result is that Z^0 decays to hadrons through its coupling to the weak neutral current.

The CKM matrix V_{CKM} in the SM is a unitary 3x3 matrix that can be described by 4 parameters: $\theta_{12},\ \theta_{13},\ \theta_{23},\ \delta$. It can be parameterized as:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & c_{23}c_{13} \\ s_{12}s_{23} - c_{12}s_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}. \tag{2.31}$$

Here the convention that

$$c_{i,j} = \cos\theta_{i,j}, \tag{2.32}$$

$$s_{i,j} = \sin\theta_{i,j}, \tag{2.33}$$

has been applied. The moduli of matrix parameters are measured at the moment to be [23]:

$$\begin{pmatrix} 0.9728 \pm 0.0030 & 0.2252 \pm 0.0009 & (4.15 \pm 0.49) \times 10^{-3} \\ 0.230 \pm 0.011 & 1.006 \pm 0.023 & (40.9 \pm 1.1) \times 10^{-3} \\ (8.4 \pm 0.6) \times 10^{-3} & (42.9 \pm 2.6) \times 10^{-3} & 0.89 \pm 0.07 \end{pmatrix}. \tag{2.34}$$

2.2.5 Neutrino mixing

The discovery of neutrino oscillation, mentioned in Sect. 2.1, implies directly that neutrinos are not massless particles. Thus, similarly to quarks, in the case of neutrinos the mass eigenstates do not coincide with weak interaction eigenstates. As a result the respective mixing matrix called PMNS [29] (after the names of Pontecorvo, Maki, Nakagawa and Sakata) reads:

$$V_{PMNS} = U_{ij} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & c_{23}c_{13} \\ s_{12}s_{23} - c_{12}s_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{2.35}$$

where $s_{mn} = \sin\theta_{mn}$ and $c_{mn} = \cos\theta_{mn}$. The PMNS matrix has a analogous form to the CKM matrix from Eq. 2.31. Over the past decades a big improvement in the experimental accuracy of the moduli of the PMNS matrix elements has been made. In particular the smallest angle θ_{13} is measured up to 7% precision [30]. Currently, the following intervals (three standard deviations from the central value) are measured [31]:

$$V_{PMNS} = \begin{pmatrix} 0.795 \to 0.846 & 0.513 \to 0.585 & 0.126 \to 0.178 \\ 0.205 \to 0.543 & 0.416 \to 0.730 & 0.579 \to 0.808 \\ 0.215 \to 0.548 & 0.409 \to 0.725 & 0.567 \to 0.800 \end{pmatrix}. \tag{2.36}$$

Contrary to the CKM matrix, where there is a distinguished hierarchy in the sizes of the elements, all components of the PMNS matrix are of the same order of magnitude. The origin of this difference between the quark and lepton sectors is unknown and a big effort is made to link the two structures. The observation of neutrino oscillation provided also indisputable evidence that lepton flavour is not conserved. This implies that lepton flavour violation has to occur also in the charged sector in the processes such as: $\ell \to \ell'\gamma$, $\ell \to \ell'\ell''\ell'''$, $X \to \ell\ell'$, etc. which involve the loops like the one shown in Fig. 2.2.

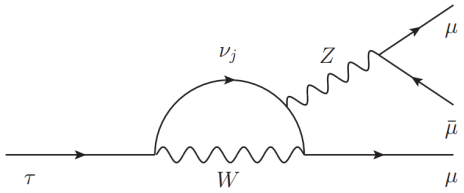


Figure 2.2: Feyman diagram for the decay $\tau^- \to \mu^-\mu^+\mu^-$

However, the charged lepton flavour violation (CLFV) induced by neutrino oscillations was estimated to be negligibly small:

$$\mathcal{B}(\tau^- \to \mu^-\mu^-\mu^+) = \frac{3\alpha}{32\pi}\left|\sum_{i=1,3} U^*_{\tau i}U_{\mu i}\frac{\Delta m^2_{i2}}{M^2_{\mathrm{W}}}\right| < 10^{-40}, \tag{2.37}$$

where α is the fine structure constant, M_{W} is mass of W bosson and Δm^2_{ij} are neutrino square mass differences. This tiny value of the branching fraction for the decay $\tau^- \to \mu^-\mu^+\mu^-$ is due to the GIM mechanism. Obviously, these vanishingly small values of branching fractions that are expected in the SM rule out any possibility of the observation of CLFV in the foreseeable future, unless they are significantly enhanced by phenomena from beyond the SM.

2.2.5.1 Strong interactions

Strong interactions are described by Quantum Chromodynamics (QCD) [32, 33]. This theory is based on the gauge group $SU(3)_{color}$, with eight gluons as particles corresponding to the gauge field. The peculiar feature of the strong force is that, contrary to other interactions, it does not diminish in strength with the increasing distance between the interacting objects. This feature leads to the phenomenon of confinement which states that only colourless states called hadrons can be observed.

Strong interactions are the main source of uncertainty in theoretical predictions of rare decays. The above uncertainty enters in the so-called hadronic matrix elements and its calculation is performed using lattice calculations [34], QCD sum rules [35], heavy quark effective theory [36], chiral perturbation theory [37] and phenomenological quark models [38].

2.3 Beyond the Standard Model theories

The violation of charged lepton flavour is predicted in many extensions of the SM (generically named as BSM theories). The inclusion of the CLFV is usually straightforward and follows directly from the model's assumptions (cf example for neutrinos in Section 2.2.5). Among these theories are supersymmetry (SUSY) [39], seesaw models [40], little Higgs scenarios [41] and models with four generations of fermions [42]. Below, three BSM models with the biggest impact on the studies described in this thesis will be briefly described. Next, an effective field approach including BSM operators with different lepton chrality structures will be briefly discussed.

2.3.1 Charged lepton flavour violation in supersymmetry

One of the most natural extensions of the SM is the so-called supersymmetry (SUSY). It is a theory that introduces an additional symmetry between integer spin particles (bosons) and spin-half particles (fermions). For each particle from the SM the existence of its supersymmetric partner is predicted. The particles in such pairs would differ in spin by one half. For instance, the SUSY partner of leptons (which are fermions) are sleptons (which are bosons). The counterparts of bosons are obtained by adding the suffix "ino" to the name of the boson, e.g. gluino, zino etc. Moreover, the supersymmertic fields of wino, bino and higgsino mix together to form four eigenstates called neutralinos. Each SM particle is associated with its supersymmetric partner that falls into the opposite category; for example a lepton which is a fermion has the associated SUSY particle called slepton, which is a boson. The names of other susy particles are constructed in the above mentioned way.

Since we do not observe SUSY particles at the same mass scale as SM particles, SUSY must be broken. As a result, the supersymmetric particles can be sufficiently heavy to escape experimental detection. In the general case of so-called Minimal Supersymmetry, there are 124 free parameters that describe the model [23]. Below the discussion of the charged lepton flavour violation will be provided in terms of the minimal extension to the SM that is realized in SUSY. It is called the Constrained Minimal Supersymmetric Standard Model (CMSSM) and depends only on five parameters: m_0 scalar mass at the Grand Unification scale $\Lambda \sim 10^{16}$ GeV, $m_{1/2}$ mass of gauginos and higgsinos at the so called Grand Unification scale, A_0—the common trilinear coupling, $\tan\beta$ - trilinear coupling, sign(μ) - sign of the higgsino mass parameter. In the MSSM lepton flavour violation arises naturally in soft supersymmetry breaking Lagrangian [43] involving non-diagonal slepton masses and anti tri-linear couplings. Leading effects arise by sneutrino-chargino and slepton-neutralino loops, where LFV is caused by mass differences between leptons and sleptons. Using mass insertion approximation [44], one can parametrize the mass matrix as [45]:

$$(\ell_L^\dagger \ell_R^\dagger) \begin{pmatrix} m_L(1+\delta_{LL}) & (A-\mu\tan\beta)m_l + m_L m_R \delta_{LR} \\ (A-\mu\tan\beta)m_l + m_L m_R \delta_{LR}^\dagger & m_R^2(1+\delta_{RR}) \end{pmatrix} \begin{pmatrix} \widetilde{\ell}_L \\ \widetilde{\ell}_R \end{pmatrix}, \quad (2.38)$$

where m_L and m_R are the averaged masses of sleptons, A is proportional to m_0, μ is mass eigenstate of higgsino, δ_{ij} are the mass insertion and $\widetilde{\ell}_{L,R}$ are the bispinor components. Another way to impose LFV is to introduce non-holomorphic couplings of Higgs doublets, which are then enhanced for large values of $\tan\beta$ [45]. Generally, such a mass insertion approximation is more useful and easy to incorporate. The small off-diagonal masses δ_{ij} insertion can be generated in various ways depending on the SUSY model, for instance by introducing additional flavour symmetries [46] or even using see-saw mechanism [47]. A good example of this kind of model is [46], where a group: $A_4 \times Z_3 \times U(1)$ to enforce nearly tri-bimaximal lepton mixing was introduced. The common feature of this kind of models is that $\mathcal{B}(\mu \to e\gamma)$ is more sensitive experimentally than $\mathcal{B}(\tau \to \mu\gamma)$, because $R_{\mu e} \approx R_{\tau\mu} \approx R_{\tau e}$, where $R_{xy} = \dfrac{\mathcal{B}(x \to y\gamma)}{\mathcal{B}(x \to h\nu\bar{\nu})}$, $x = \mathrm{e},\ \mu$.

2.3.2 Littlest Higgs model with T-parity

An alternative method of solving the hierarchy problem [48] is the Little Higgs model. In this model the Higgs particle is considered as a pseudo-Goldstone boson that arises from breaking a new symmetry at $\mathcal{O}(10$ TeV). The variation of this approach, called the Littlest Higgs [49, 50] included additionally the violation of time reversal (T). Here, new leptons with masses $\mathcal{O}(1$ TeV) are postulated in addition to new heavy gauge bosons. The free parameters in this kind of models are the so-called mirror leptons masses, mirror lepton mixing angles and Dirac phases. The drawback of this class of models is that they require a fine tuning of the parameters. On the other hand, the Littlest Higgs theory provides clear-cut experimental expectations, postulating in particular that $\mathcal{B}(\ell \to \ell'\gamma) \approx \mathcal{B}(\ell \to \ell'\ell'\ell')$, as illustrated in Fig. 2.3.

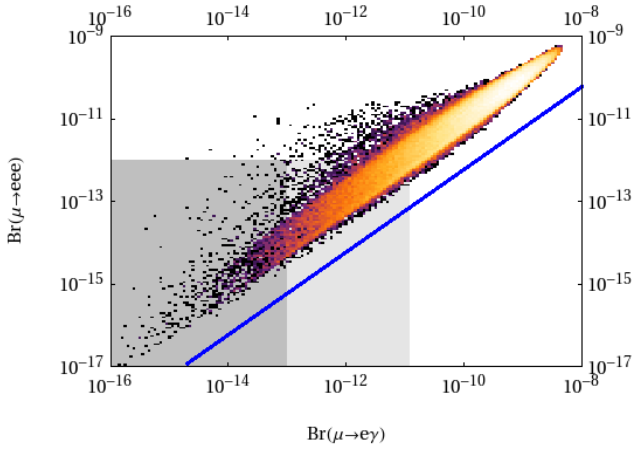


Figure 2.3: Correlation between branching fractions of $\mu^- \to e^-e^+e^-$ and $\mu^- \to e^-\gamma$ decays in the littlest Higgs Model with the violation of T-parity. Blue line represents results with only dipole contribution to $\mu^- \to e^-\gamma$. Figure taken from [49].

2.3.3 Models with four generations of leptons

The next BSM theory discussed in this thesis is based on the introduction of the 4^{th} generation of leptons and quarks, with a new heavy lepton ℓ_4 and neutrino ν_4. In this case the PMNS matrix introduced in Eq. 2.35 requires an extension by one dimension. This matrix is dependent on six mixing angles, three Dirac phases and three Majorana phases. As showed by A. Buras [51], one can easily relate the branching fractions of LFV decays:

$$\frac{\mathcal{B}(\tau \to \mu\gamma)}{\mathcal{B}(\mu \to \mathrm{e}\gamma)} = \left|\frac{U_{\mu 4}}{U_{\mathrm{e}4}}\right|^2 \frac{\mathcal{B}(\tau \to \mu\nu\nu)}{\mathcal{B}(\tau \to \mathrm{e}\nu\nu)} \approx \left|\frac{U_{4\mu}}{U_{\mathrm{e}4}}\right|^2, \tag{2.39}$$

where $U_{4\mu}$ are the elements of the "extended" PMNS matrix [51]. The expected branching fractions in this model are just below the current experimental limits, which makes them very interesting, as illustrated in Fig. 2.4.

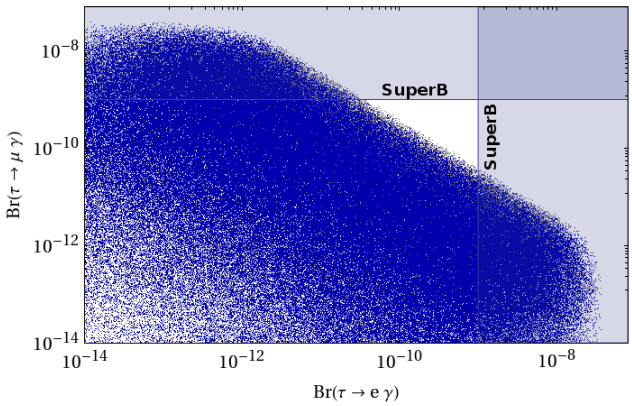


Figure 2.4: Correlation between branching fractions of $\tau \to \mu\gamma$ and $\tau \to \mathrm{e}\gamma$ decays in the 4^{th} generation model [51]. Here the sensitivity of the proposed SuperB experiment [52] was marked as blue bands.

2.3.4 Effective field theory approach as applied to $\tau^- \to \mu^-\mu^+\mu^-$ decay

CLFV processes resulting from BSM theories can be described in a model independent way in terms of new operators. If new physics exists at a mass scale Λ, it can manifest itself at an electroweak scale in the form of higher order operators which, however, do not spoil the $SU(2)_L \times U(1)_Y$ symmetry. As mentioned in Sect. 2.2.3, the left-handed leptons from isospin doublets cf. Eq. 2.9 and the right-handed partners belong to isospin singlets. In the EFT approach the right-handed singlets are written as the following isospin doublets [53]:

$$R_e = \frac{1-\gamma_5}{2}\begin{pmatrix}0\\ \psi_{\mathrm{e}}\end{pmatrix},\ R_\mu = \frac{1-\gamma_5}{2}\begin{pmatrix}0\\ \psi_{\mu}\end{pmatrix},\ R_\tau = \frac{1-\gamma_5}{2}\begin{pmatrix}0\\ \psi_{\tau}\end{pmatrix}. \tag{2.40}$$

Taking into account Eq. 2.9 and 2.40 and the matrix of Higgs fields from [53], one can derive the following relevant dimension six operators:

$$O_1 = (\bar{L}\gamma_\mu L)(\bar{L}\gamma^\mu L), \tag{2.41}$$
$$O_2 = (\bar{L}\tau^a\gamma_\mu L)(\bar{L}\tau^a\gamma^\mu L), \tag{2.42}$$
$$O_3 = (\bar{R}\gamma_\mu R)(\bar{R}\gamma^\mu R), \tag{2.43}$$
$$O_4 = (\bar{R}\gamma_\mu R)(\bar{L}\gamma^\mu L), \tag{2.44}$$
$$R_1 = g'(\bar{L}H\sigma_{\mu\nu}R)B^{\mu\nu}, \tag{2.45}$$
$$R_2 = g(\bar{L}\tau^a H\sigma_{\mu\nu}R)W^{\mu\nu}, \tag{2.46}$$

as defined above, $B_{\mu\nu}$ and $W_{\mu\nu,a}$ are the electroweak gauge fields, g and g' are the coupling constants of $SU(2)_L$ and $U(1)_Y$, H denotes the matrix of Higgs fields, $L(R)$ are the left(right)-handed fields and $\sigma^{\mu,\nu} = \frac{i}{4}[\gamma^\mu, \gamma^\nu]$. According to S. Turczyk et. al. [54], higher order operators are suppressed by small lepton Yukawa couplings, so we will not consider them in this thesis. In the effective field theory the most general Hamiltonian that describes the discussed process is formed as the sum of the operators from Eq. 2.41 - 2.46. For the studied process $\tau^- \to \mu^-\mu^+\mu^-$ the operators O_1 and O_2 are identical after projecting them on charged leptons. The O_3 corresponds to a purelly right-handed current and is completely analogous to O_1. For radiative operators R_1 and R_2 the latter is suppressed by small Yukawa coupling of τ, so only the photonic operator R_1 is relevant.

The analysis performed in this dissertation was also interpreted in terms of the BSM operators, as described in Sect. 4.10. The respective decay widths can be presented in the form of Dalitz distributions [55], which were derived in the following five cases, corresponding to different lepton chirality structures:

- Four left-handed leptons (O_1 operator):

$$\frac{d^2\Gamma_V^{(LL)(LL)}}{d^2m_{23}d^2m_{12}} = \frac{\left|g_V^{(L_\mu L^\tau)(L_\mu L^\mu)}\right|^2}{\Lambda^4}\frac{(m_\tau^2 - m_\mu^2)^2 - (2m_{12}^2 - m_\tau^2 - 3m_\mu^2)^2}{256\pi^3 m_\tau^3}. \tag{2.47}$$

- Two left-handed, two right-handed leptons (O_4 operator):

$$\frac{d^2\Gamma_V^{(LL)(RR)}}{d^2m_{23}d^2m_{12}} = \frac{\left|g_V^{(L_\mu L^\tau)(L_\mu L^\mu)}\right|^2}{\Lambda^4}\left[\frac{(m_\tau^2 - m_\mu^2)^2 - 4m_\mu^2(m_\tau^2 + m_\mu^2 - m_{12}^2)}{512\pi^3 m_\tau^3}\right.$$
$$\left. - \frac{(2m_{12}^2 - m_\tau^2 - 3m_\mu^2)^2 + (2m_{23}^2 - m_\tau^2 - 3m_\mu^2)^2}{1024\pi^3 m_\tau^3}\right]. \tag{2.48}$$

- Radiative right-handed τ leptons (R_1 operator):

$$\frac{d^2\Gamma^{(LR)}_{rad}}{d^2m_{23}d^2m_{12}} = \alpha^2_{em}\frac{\left|g^{(L_\mu R^\tau)}_{rad}\right|^2\nu^2}{\Lambda^4}\left[\frac{4m^2_\mu(m^2_\tau+m^2_\mu-m^2_{12})}{128\pi^3m^3_\tau}(\frac{1}{m^4_{13}}+\frac{1}{m^4_{23}})\right.$$
$$+\frac{m_\mu(m^4_\tau(-3m^2_\tau m^2_\mu+2m^2_\mu)}{128\pi^3m^3_\tau m^2_{23}m^2_{12}}+\frac{2m^2_{12}-3m^2_\mu}{128\pi^3m^3_\tau}$$
$$\left.+\frac{(m^2_{13}+m^2_{23})(m^4_{12}+m^4_{13}+m^4_{23}-6m^2_\mu(m^2_\mu+m^2_\tau))}{256\pi^3m^3_\tau m^2_{23}m^2_{12}}\right]. \quad (2.49)$$

- Interference between O_1 and R_1:

$$\frac{d^2\Gamma^{(LL)(RR)}_{mix}}{d^2m_{23}d^2m_{12}} = \alpha^2_{em}\frac{2\nu Re\left[g^{(L_\mu L^\tau)(L_\mu L^\mu)}_V g^{*LR}_{rad}\right]}{\Lambda^4}\left[\frac{m^2_{12}-3m^2_\mu}{64\pi^3m^2_\tau}+\right.$$
$$\left.\frac{m^2_\mu(m^2_\tau-m_\mu)^2(m^2_{13}+m^2_{23})}{128\pi^3m^3_\tau m^2_{23}m^2_{12}}\right]. \quad (2.50)$$

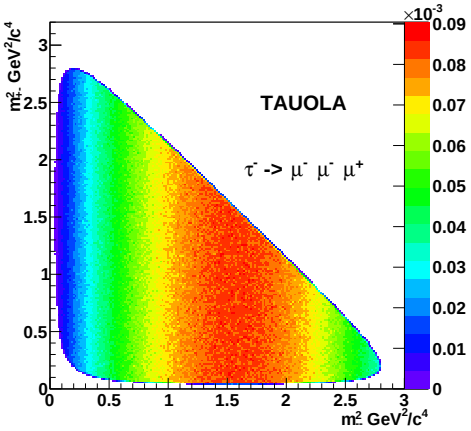
- Interference between O_4 and R_1:

$$\frac{d^2\Gamma^{(LL)(RR)}_{rad}}{d^2m_{23}d^2m_{12}} = \alpha_{em}\frac{2\nu Re\left[g^{(L_\mu L^\tau)(R_\mu R^\mu)}_V g^{*LR}_{rad}\right]}{\Lambda^4}\left[\frac{m^2_\tau-m^2_{12}-3m^2_\mu}{256\pi^3m^3_\tau}+\right.$$
$$\left.\frac{m_\mu(m^2_\tau-m^2_\mu)(m^2_{13}+m^2_{23})}{256\pi^3m^3_\tau m^2_{23}m^2_{12}}\right]. \quad (2.51)$$

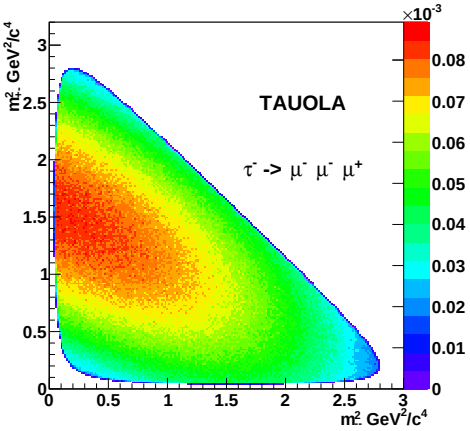
In the Eq. 2.47 - 2.51 the following dimuon masses are defined:

$$m^2_{--} = m_{12} = (p_{\mu^-}+p'_{\mu^-})^2, \qquad m^2_{+-} = m_{23} = (p_{\mu^-}+p_{\mu^+})^2, \quad (2.52)$$

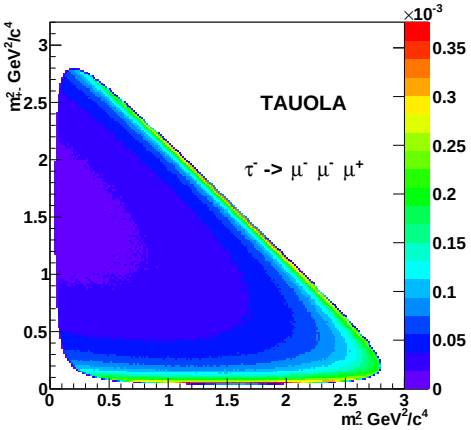
and m_ℓ are the masses of corresponding leptons, g_V are the coupling constants and ν is the element from the Higgs matrix. All the above models and several others were implemented by the author in the TAUOLA library [56] of Monte Carlo programs dedicated to τ physics (the relevant publication by M. Chrząszcz and Z. Wąs is in preparation). The Dalitz distributions corresponding to the simulation of each of the above mentioned models can be found in Fig. 2.5.



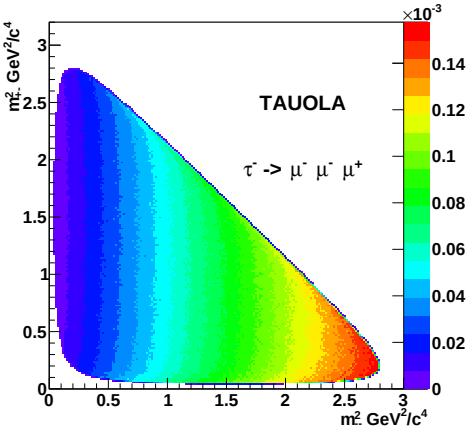
(a) Simulated Dalitz distribution for Eq. 2.47.



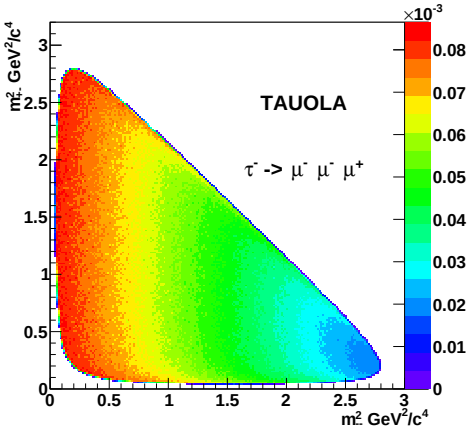
(b) Simulated Dalitz distribution for Eq. 2.48.



(c) Simulated Dalitz distribution for Eq. 2.49.



(d) Simulated Dalitz distribution for Eq. 2.50.



(e) Simulated Dalitz distribution for Eq. 2.51.

Figure 2.5: Dalitz distributions simulated in the effective field approach for the five different BSM operators corresponding to different lepton chirality structures [53]. The distributions were implemented in the TAUOLA package and normalized to unit area.

Chapter 3

Experimental apparatus

This study has been carried out using the data collected over the years 2011 and 2012 at the Large Hadron Collider beauty (LHCb) experiment, located at the Large Hadron Collider (LHC) at CERN (Organisation Européenne pour la Recherche Nucléaire) near Geneva. Below we will briefly describe the LHC accelerator and provide the detailed description of the LHCb subsystems which are relevant to the topic of this study.

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) [57] is the world's largest accelerator located at CERN. It accelerates and collides protons with energies of 4 TeV (as of 2012 running period). The LHC started functioning in 2008 and constitutes the final stage of acceleration in the CERN complex (cf. Fig. 3.1). The boosting process starts with Linear Accelerator 2 (LINAC2) [58], in which protons are accelerated to 50 MeV. The protons are then injected to the Proton Synchrotron Booster [59], which further increases their energy to 1.4 GeV. In order to surge the intensity of the beam, the booster is made up of four superimposed synchrotron rings. Having left the booster, the protons are injected into the Proton Synchrotron (PS) [60], which is a synchrotron of 628 m in circumference made of 277 electromagnets. It accelerates the protons up to 25 GeV. In the last stage before entering the LHC, the protons are boosted in the Super Proton Synchrotron (SPS) [61], composed of over 1300 room-temperature electromagnets and able to accelerate the protons up to 450 GeV.

Finally the protons are injected into the LHC, which is located in the 26.7 km tunnel previously occupied by the Large Electron-Positron Collider [63]. The tunnel comprises eight arcs (exploited for bending) and eight straight sections that host the experiments and are used for injection.

In 2011 (2012) the LHC accelerated the protons up to the energy of 7 (8) TeV, respectively, as measured in the centre-of-mass frame. Each beam is composed of 1380 proton bunches with 50 ns time spacing. There exist four interaction points where seven experiments (ATLAS [64], CMS [65], LHCb [66], ALICE [67], TOTEM [68], LHCf [69] and MoEDAL [70]) are located. The protons are accelerated using 400 MHz radio frequency (RF) cavities, which are situated at Point 4. The whole chain of CERN accelerators is illustrated in Fig. 3.1. The LHC machine is composed of $1,232$ dipole magnets that are responsible for bending the beam and 392 quadrupole magnets used to focus the protons bunches. Liquid helium is

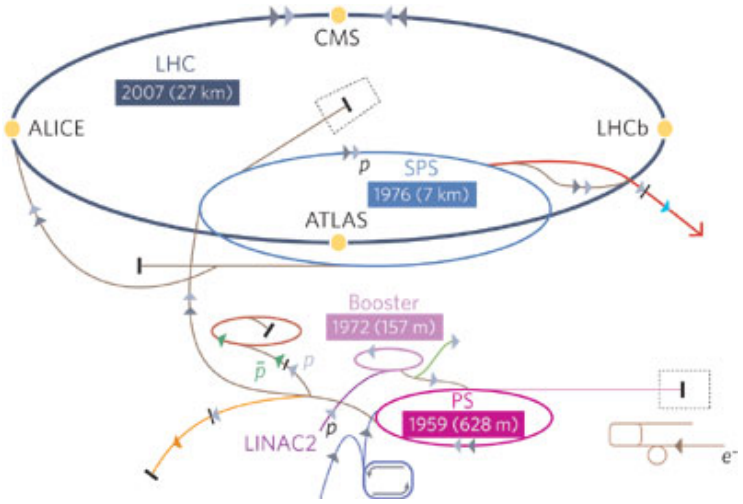


Figure 3.1: The layout of CERN accelerator system [62].

used to cool down the copper-clad niobium-titanium magnets to a super-conducting temperature of 1.9 K.

In the LHCb experiment the beams are purposely defocused to a number of collisions per bunch crossing (the so-called pile-up), thus reducing both the risk of radiation damage of the vertex detector as well as combinatorial background. Because of the defocusing of the beams the instantaneous luminosity of the LHCb amounts to $\mathcal{L} = 4 \times 10^{32}$ cm^{-2} s^{-1} (Fig. 3.2 and Table 3.1), which is substantially lower when compared with the values at ATLAS and CMS ($\mathcal{L} = 1 \times 10^{34}$ cm^{-2} s^{-1}).

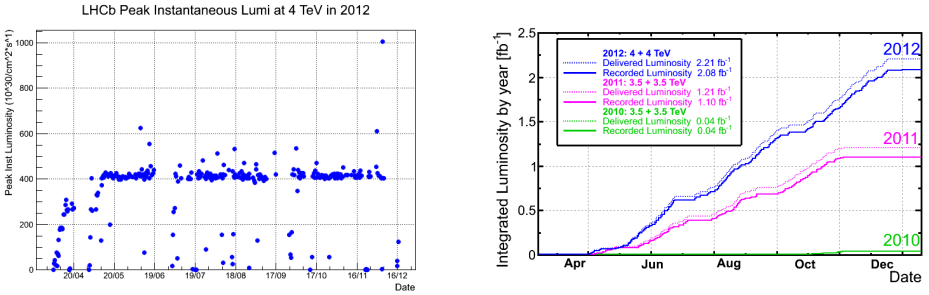


Figure 3.2: Instantaneous (a) and integrated (b) luminosity collected in LHCb detector.

3.2 LHCb detector

LHCb is a single arm forward spectrometer located at Point 8 in the LHC ring. The detector covers the pseudorapidity (η)[1] range $2 < \eta < 5$, which corresponds to the

[1] $\eta = -\ln\tan\frac{\theta}{2}$, where θ is the polar angle w.r.t. the beam axis.

Table 3.1: Beam parameters at LHCb beam crossing point.

Quantity	Achieved	Designed
No. protons in bunch	1.49×10^{11}	1.15×10^{11}
Luminosity(at LHCb)	4×10^{32} cm^{-2} s^{-1}	2×10^{32} cm^{-2} s^{-1}
Energy	4 TeV	7 TeV
No. bunches	1380	2808
Stored Energy	115 MJ	362 MJ
Time steep of bunches	50 ns	25 ns

geometric acceptance $(10 - 250)$ mrad $(10 - 300)$ mrad in the vertical (horizontal) plane, respectively. Such a choice is dictated by the fact that $b\bar{b}$ pairs are boosted in the direction of the higher energy parton in the laboratory frame. This results in the highly correlated directions of flight of the b and $\bar{b}$ quark (cf Fig. 3.3). The above angular acceptance makes it possible to capture every second pair of the produced $b\bar{b}$ pairs.

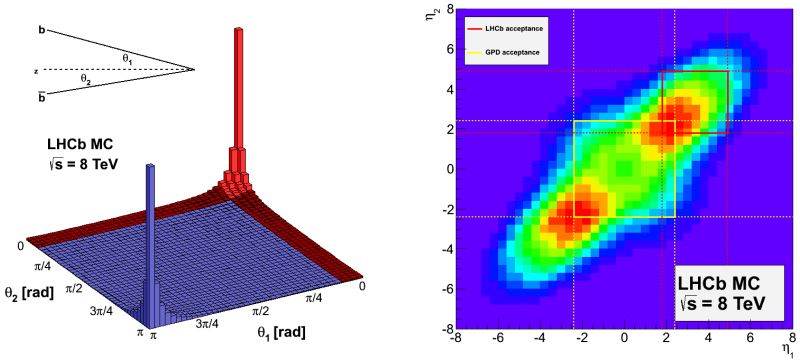


Figure 3.3: Simulated $b\bar{b}$ production angles at LHCb. Left: forward-backward productions fractions as functions of b angles to the beam axis. LHCb acceptance is marked in red. Right: Distribution of pseudorapidities for $b\bar{b}$ pairs with LHCb acceptance marked in red.

The LHCb detector sees on average 1.4 interactions per bunch crossing occuring every 50 ns. Usually each pp interaction creates around 80 tracks in the LHCb detector that originate from the so-called primary vertex (PV) [71].

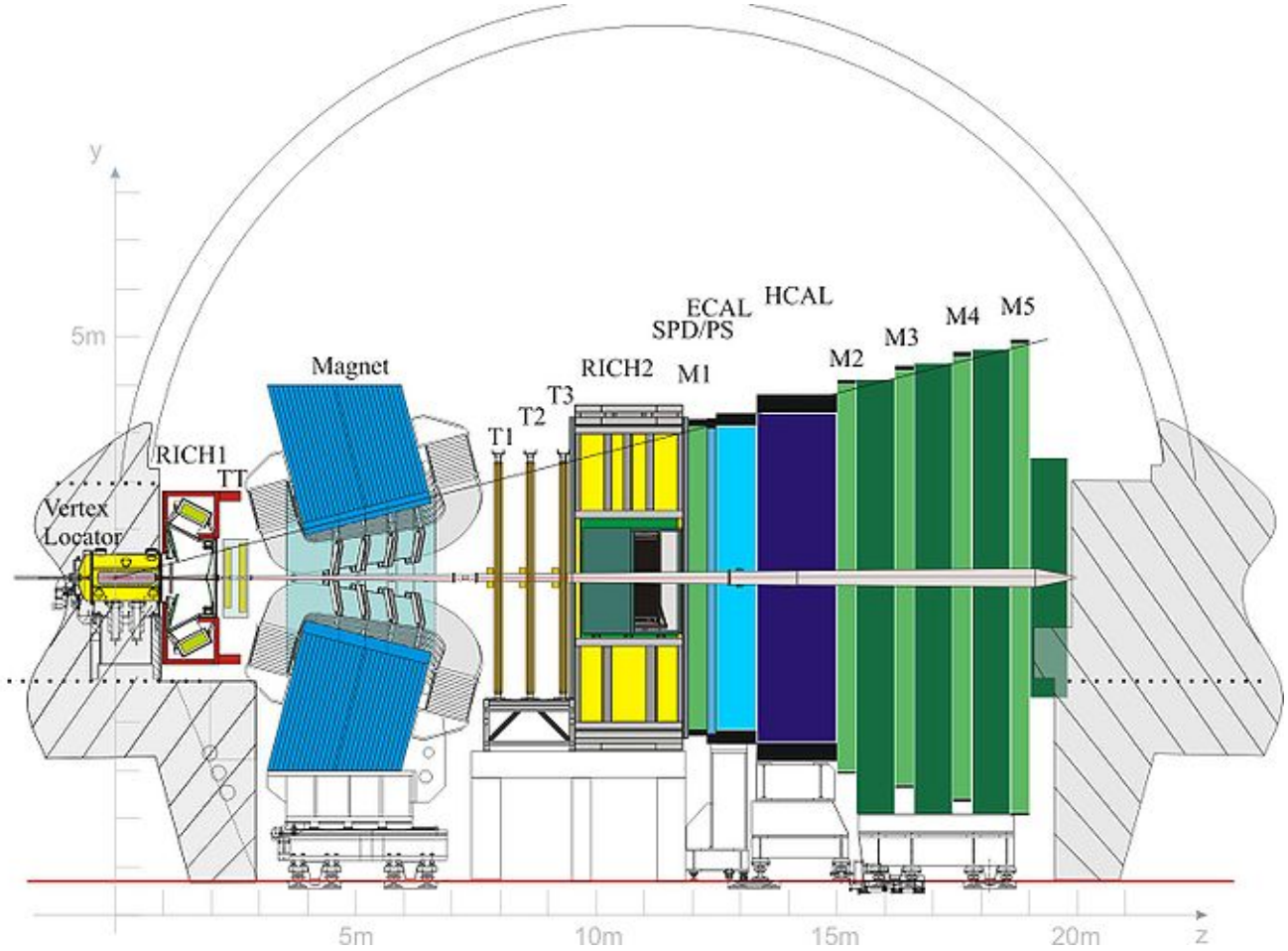


Figure 3.4: Side view of the LHCb detector [72]. The z axis coincides with the beam direction and y axis points vertically towards the Earth's surface. The x axis is oriented towards from the centre of the LHC ring.

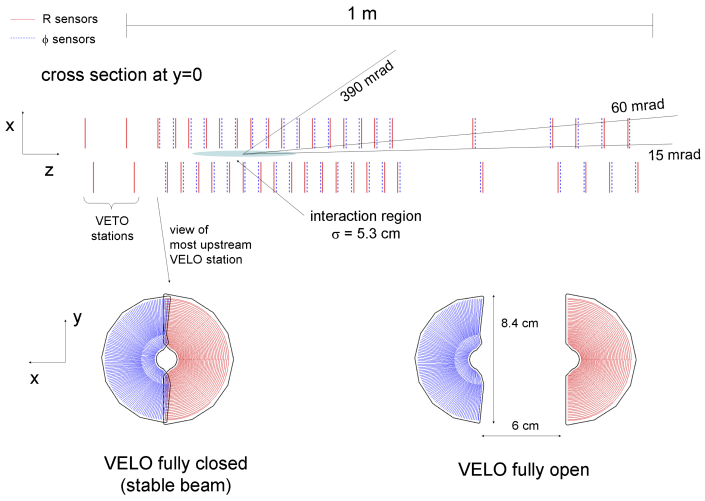


Figure 3.5: The layout of the VELO detector. The lower drawings show the closed (left) and open (right) configuration.

B hadrons are reconstructed in the LHCb detector as the displaced secondary vertices (SV) of their decays. The separation between the PV and SV is typically of the order of 5 mm, which is due to the substantial lifetime of B hadrons ($\sim$ 1 ps) and their large momentum. Moreover, the decay products of B hadrons are characterized by a large value of the impact parameter (IP) w.r.t., the primary vertex. In addition, the SV tracks have on average higher values of transverse momenta in comparison with PV tracks. The IP is later used in the trigger. The LHCb spectrometer is made of several sub-detectors, each serving different purposes. The full layout of the LHCb detector is presented in Fig. 3.4. The charged track reconstruction is performed in LHCb with the Vertex Locator (VELO), Tracker Turicensis (TT) and the magnet. Particle identification is performed by two Ring Imaging Cherenkov detectors (RICH) and the muon system (M1-M5). The Scintillating Pad Detector (SPD), Pre-Shover (PS), Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL) measure the energy of neutral particles.

3.2.1 Vertex Locator

The Vertex Locator (VELO) system provides very precise information about charged track positions close to the interaction point and plays a crucial role in the reconstruction of the secondary vertices (Fig. 3.5). The VELO is composed of two 1 m long semi-cylinders, each containing 21 modules of silicon microstrip sensors (Fig. 3.5). In addition, two silicon modules are placed before the interaction point. The microstrip detectors of the VELO system are either oriented in such a way that they emanate radially from the beam axis (phi-sensors) or they form concentric circles around the beam axis (r-sensors). The above geometry allows two-dimensional precise measurement of a particle position to be performed in both radial and azimuthal

coordinates (Fig. 3.5). The detector is located only 8 mm away from the beam axis. Due to the fact that the injector to the LHC is placed just before the LHCb cavern, the VELO is a movable detector. During the injection the VELO semi-cylinders retract 35 mm outside the beam axis. After the beam is ramped up and squeezed, the VELO modules are pushed back to their original position. The LHCb vertex detector offers excellent spatial resolution: 10 µm in x and y direction and 60 µm in z. The average resolutions for the determination of the IP (decay length) are of the order of 200 µm (300 µm), respectively. Such a precision in the location of the track's origin is crucial for reducing the so-called combinatorial background in which one of the tracks gets assigned to the wrong decay vertex. This is the dominant background for most of rare decay analyses.

3.2.2 Tracker and magnet

The Tracker is composed of the Tracker Turicensis located upstream of the magnet and the T1, T2 and T3 tracking stations placed downstream of the magnet (Fig. 3.4). Each of the T1-T3 stations is splitted into the Inner Tracker (IT) and Outer Tracker (OT) (Fig. 3.6 and Table 3.2).

Both the TT and each of T1-T3 station are composed of four layers of detectors. The second and third layer are tilted by $\pm 5°$ with respect to the first and fourth one in order to obtain a stereo view of the particle trajectory. They are made of silicon strip detectors with the sensor area of 11×7.6 cm^2. The strips have a pitch of 200 µm and their thickness spans over the range of $(320 - 410)$ µm. The geometry of this setup was optimized to limit the occupancy to a few percent and it reaches a single hit resolution of around 50 µm.

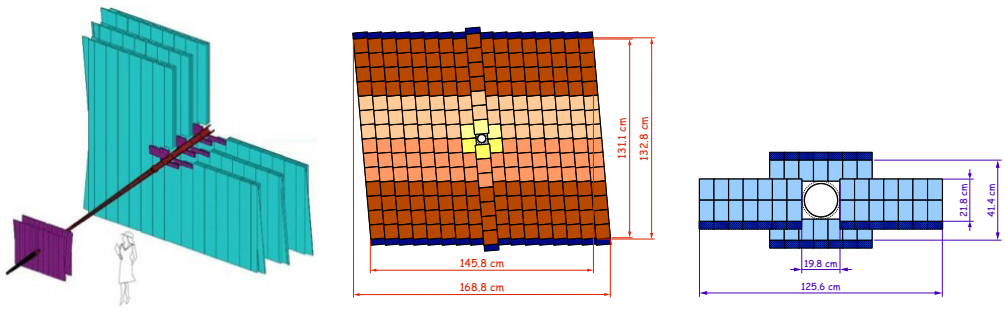


Figure 3.6: Left: schematic representation of TT (purple), IT (purple) and OT (cyan). Middle: the layout of TT last layer. Right: the layout of IT last layer.

The OT constitutes a drift time gas detector based on straw tube modules. Each station is composed of four layers of modules and each module contains two staggered layers of drift tubes. The layers are tilted in the same way as described above for the IT. The drift-coordinate resolution is 200 µm and the drift time does not exceed 50 ns.

In order to determine the charge and the momentum of the charged particles, the curvature of the track trajectories inside the magnetic field is measured. For

Table 3.2: Basic properties of LHCb tracking system.

Quantity / Detector	TT	IT	OT
Size [cm^2]	130×150	120×40	up to 290×240
Spatial resolution [µm]	50	50	200

that purpose LHCb uses a room temperature electromagnet with 4 Tm integrated field strength.

3.2.3 Ring Imaging Cherenkov detectors

The LHCb spectrometer is instrumented with two Ring Imaging Cherenkov (RICH) sub-detectors, marked as RICH1 and RICH2 (Fig. 3.7). RICH1 is located upstream of the magnet. It performs particle identification (PID) with a momentum range of $(2-60)$ GeV/c and uses C_4F_{10} and aerogel radiators as Cherenkow light emitters. This light is then focused using a set of spherical mirrors onto optical elements that are located outside the LHCb acceptance, thus reducing the material budget of the spectrometer. The angular acceptance of RICH1 amounts to 250 mrad (300 mrad) in the vertical (horizontal) direction, respectively.

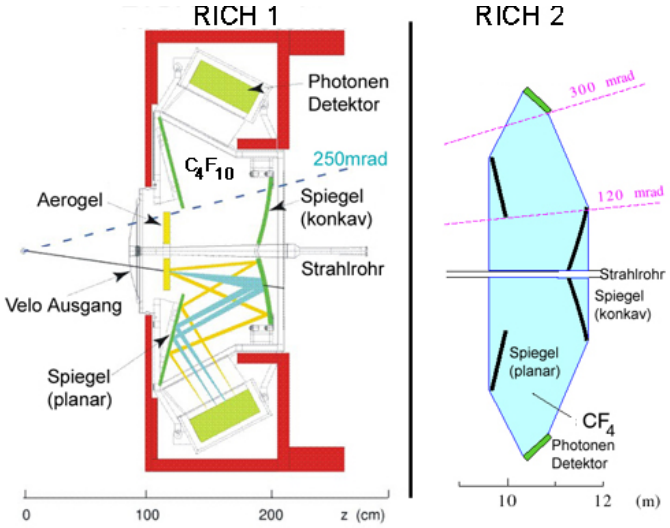


Figure 3.7: The layout of RICH detectors

RICH2 is located further downstream between T3 and ECAL. It uses CF_4 as a radiator and is optimized for identifying charge particles in the momentum range of $(15-100)$ GeV/c. The acceptance of RICH2 is 100 mrad (120 mrad) in the vertical (horizontal) direction, respectively.

The readout of both RICH detectors is based on Hybrid Photon Detectors (HPDs) [73], operating in the wavelength range $(200-600)$ µm.

3.2.4 The calorimeters

The LHCb spectrometer is composed of four calorimetric subsystems: Scintillating Pad Detector (SPD), Pre-Shover (PS), Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL). All of them follow the same "sandwich" structure of alternating layers of active and absorbing material. The ECAL is located dowstream of RICH2 and HCAL is situated as the next component dowstream of the LHCb spectrometer.

The Scintillating Pad Detection and Pre-Shower systems were designed to improve the spatial and energy resolution of electromagnetic showers. They consist of 15 mm lead converters which are sandwiched between two identical planes of rectangular high-granularity scintillation pads. The main task of the SPD is to detect hits from charged particles in order to distinguish electrons from photons and π^0 decays. The PS is aimed at the longitudinal segmentation of the ECAL energy deposits in order to reduce the background from charged pions.

The ECAL is composed of alternating layers of 2 mm thick lead plates and 4 mm scintillating plates. Its thickness corresponds to 25 radiation lengths. Similarly, HCAL is formed from layers of iron and scintillating tiles with a total of 0.59 interaction lengths. The thickness of the iron plates (scintillators) is 16 (4) mm, respectively.

3.2.5 The muon system

The efficient and accurate detection and identification of muons is one of the most important features of LHCb. A lot of rare decays and tagging channels (like $B^0 \to J/\psi(\mu^+\mu^-)K^0_S$) have a muon in the final state. Muon identification plays also a crucial role in the analysis performed in this dissertation.

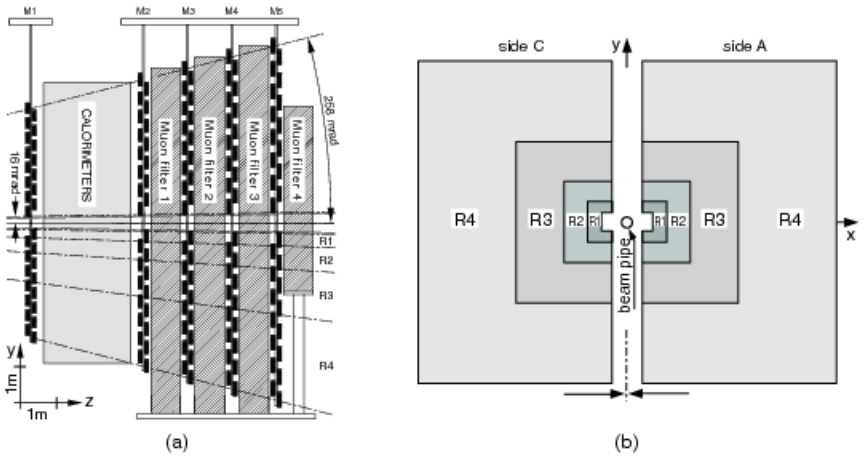


Figure 3.8: The Muon system of LHCb detector: (a) Side view, (b) station layout with four regions R1-R4 described in the text.

The muon system consists of five stations (M1-M5), four of which (M2-M5) are located downstream of the calorimeters (Fig. 3.8). They are interleaved with 80 cm thick iron absorbers. The M1 station is placed before the calorimeters to measure high p_T muons. The data from M1 is also used by the trigger system. The M2-M5 modules are built from multi-wire proportional chambers (MWPCs), with $3-4$ ns time resolution [74]. Because of the higher occupancy of the M1 module, the MWPCs were substituted with Gas Electron Multipliers (GEMs) with 3 ns time resolution. Each station is divided into four regions: $R1-R4$, which are shown in Fig. 3.8. The geometry of those regions is based on granularity requirements.

Table 3.3: Resolution (σ) along x and y coordinates of the distance between the muon track and the muon cluster in each region of the muon detector. The muon track is reconstructed skipping the station whose resolution must be evaluated.

		M1	M2	M3	M4	M5
R1	$\sigma_x \times \sigma_y$ (mm^2)	4×10	15×30	10×12	15×16	33×40
R2	$\sigma_x \times \sigma_y$ (mm^2)	8×18	25×50	15×24	27×32	50×60
R3	$\sigma_x \times \sigma_y$ (mm^2)	16×40	35×70	25×48	48×64	100×110
R4	$\sigma_x \times \sigma_y$ (mm^2)	32×80	60×100	40×96	97×128	150×180

The LHCb muon system offers outstanding resolutions in x and y coordinates for tracks which are reported in Table 3.3. The typical efficiency of muon stations is above 95% [74]. This excellent performance of the muon system makes the LHCb experiment a perfect environment for studying decays with muons in the final state.

3.2.6 The trigger

The LHCb trigger was designed to sieve efficiently the events containing heavy flavour particles from other pp interactions. The trigger is composed of two parts: First Level Trigger (L0) and High Level Trigger (HLT). The L0 component is a hardware trigger that uses the information from the muon system and calorimeters. Its main purpose is to reduce the event rate from 11 MHz to 870 kHz [71]. The HLT is a software-based trigger that uses a computing farm located in the LHCb pit. The HLT consists of two triggers: HLT1 and HLT2. HLT1 uses partly reconstructed events and reduces the rate from 870 kHz to 43 kHz. HLT2 has access to fully reconstructed events and further reduces the bandwidth to 3 kHz, which corresponds to the data volume stored on tape. In the hadron environment the muon rates are considerably smaller than those of hadrons, allowing the trigger to have loose requirements for the candidates while keeping a reasonable rate.

3.2.6.1 First Level Trigger

The L0 consists of three independent triggers: `L0Muon`, `L0PileUp` and `L0Calorimeter`. Each of these triggers uses a different subset of detectors, which allows for making a very fast decision at the L0 level. The analysis described in this dissertation uses only the `L0Muon` trigger, which makes a decision based on information from the muon stations. In this case the `L0Muon` trigger is fired if there exists a track pointing to the interaction point that possesses a $p_T > 0.5$ GeV/c in $x - z$ plane. The trigger threshold is applied to the track with the highest p_T.

3.2.6.2 High Level Trigger

The HLT trigger runs on the Event Filter Farm which employs the same software as the one used by offline analysis. This guarantees the homogeneity of further off-line selection. Each so-called trigger line consists of a sequence of reconstruction algorithms and selection requirements. The descriptions of those requirements and parameters are associated to Trigger Configuration Key (TCK). During the data taking the trigger is reoptimized, e.g. in 2012 eight different TCKs were used. For each event in the data the TCK number (32 bit) is stored together with the information about the trigger lines that accepted the event.

The first task of HLT1 is to perform the reconstruction of the VELO tracks. If the event fired `L0Muon` trigger, then the VELO tracks are extrapolated with the straight line to the M3 station in the vertical plane (the magnet does not bend the tracks in the vertical plane). The muon candidate is required to have a minimum momentum of 6 GeV/c. Each of its hits found in M3 is combined with the VELO track and the search is performed in the M2, M4 and M5 stations. If additional hits are found, a linear χ^2 fit is performed in the horizontal plane. The event is accepted if χ^2/ndf of the corresponding fit χ^2 divided by the number of degrees of freedom is smaller than 25 and then the trigger accepts this event.

The HLT2 trigger uses more advanced reconstruction algorithms that include the removal of clone tracks[2] and select the events based on inclusive and exclusive criteria. Finally, topological triggers target on selecting the inclusive production of b hadrons. There exist several trigger lines that aim to select either prompt charm production or hadronic decays which can be triggered with high efficiency, e.g. processes involving the decay $J/\psi \to \mu\mu$.

[2] A single track reconstructed as two tracks

Chapter 4

Search for the decay $\tau^- \to \mu^-\mu^+\mu^-$

This chapter describes the search for the charged lepton flavour violating decay $\tau^- \to \mu^-\mu^+\mu^-$, using the data collected by the LHCb experiment. The studies have been carried out within the Rare Decays Working Group [75] of the LHCb collaboration. The author of this thesis conducted the main part of each step of the analysis described below. The results of these studies have been published in [76, 77] and presented by the author at the following conferences: CKM 2012 conference [78], Rencontres de Blois 2014 [79], Heavy Quark and Leptons 2013 [80] and International Workshop on Tau Lepton Physics 2012 and 2014 [81, 82].

Similar analyses were carried out by the so-called B factories [83, 84], i.e. high-luminosity e^+e^- colliders running at the centre-of-mass (CMS) energy corresponding to the $\Upsilon(4S)$ resonance. However, the analysis technique was completely different, as at e^+e^- colliders the τ leptons were produced in a clean experimental environment in back-to-back pairs, thus allowing for efficient selection of the decay in question by tagging the decay of the other τ lepton. In contrary in LHCb single τ particles are produced from heavy meson decays and the abovementioned tagging is not available.

4.1 Analysis strategy

The search for the decay $\tau^- \to \mu^-\mu^+\mu^-$ was performed using 2011 and 2012 data collected by the LHCb collaboration, corresponding to the integrated luminosity of $3\,\text{fb}^{-1}$. Two similar selections were implemented: the first one for the $\tau^- \to \mu^-\mu^+\mu^-$ signal mode and the second selection for the normalization channel, which was the decay $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$. The latter was chosen as a process with a known branching fraction (and thus attainable experimentally at LHCb) as well as with the final state similar to the one being searched for. This approach allows for the evaluation of the ratio $\dfrac{\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-)}{\mathcal{B}(D_s^+ \to \phi(\mu^-\mu^+)\pi^+)}$ in which the production cross-sections as well as the most systematic uncertainties cancel. The discrimination between signal and background events is achieved using a binned three-dimensional distribution in three variables: a likelihood based on the 3-body kinematics of the event ($\mathcal{M}_{3body}$), a likelihood based on muon particle identification ($\mathcal{M}_{PID}$) and the invariant mass of the signal candidates ($m_{\mu\mu\mu}$).

The range of $\mu\mu\mu$ invariant mass is divided into three regions:

- The signal region: $|m_{\mu\mu\mu} - m_{\tau,PDG}| < 20\,\mathrm{MeV}/c^2$, where the average experimental value of tau lepton mass $m_{\tau,PDG}$ reads $1776.8\,\mathrm{MeV}/c^2$, as given by the Particle Data Group (PDG) [23]. This part is blinded on data until the analysis strategy is fully developed and checked. Here the background expectation will be compared to the actual event count.

- The middle-sidebands region: $|m_{\mu\mu\mu} - m_{\tau,PDG}| > 20\,\mathrm{MeV}/c^2$ while $m_{\mu\mu\mu} > 1747.18\,\mathrm{MeV}/c^2$ and $m_{\mu\mu\mu} < 1807.18\,\mathrm{MeV}/c^2$. This region of the $\mu\mu\mu$ invariant mass is used for analysis optimization concerning in particular the trigger use, the choice of the multivariate classifiers, their performance and the classifier's binning.

- The sidebands region: $m_{\mu\mu\mu} < 1747.18\,\mathrm{MeV}/c^2$ or $m_{\mu\mu\mu} > 1807.18\,\mathrm{MeV}/c^2$. This mass range is used to fit the background spectra and obtain an expectation of the background yield in the signal regions.

4.2 Data set description

The results described in this dissertation are obtained using the data collected at a centre-of-mass energy of $\sqrt{s} = 7$ TeV in 2011, corresponding to $1.01\,\mathrm{fb}^{-1}$ of integrated luminosity, and at $\sqrt{s} = 8$ TeV in 2012, corresponding to $1.99\,\mathrm{fb}^{-1}$ of integrated luminosity.

The data have been reconstructed using an official reconstruction software of the LHCb collaboration called Brunel [85]. For this analysis the v43r2p2 version of Brunel code was used. Stripping (described in Sect. 4.3) has been performed using a DaVinci software package [86] (the v32r1p1 version was applied in this study), officially dedicated to that purpose by the LHCb collaboration. The data are taken from the the `Stripping20` campaign. Reconstruction is a process in which hits in the detector are grouped in the so-called clusters and assigned to individual tracks by performing a χ^2 fit. Its main purpose is to take tracks reconstructed by Brunel and perform vertex fits. In this study the v32r1p1 version of DaVinci was used.

4.3 Selection criteria in search for the $\tau^- \to \mu^-\mu^+\mu^-$ decay

The overall strategy of the physics studies at LHCb is in a large sense dictated by the large bandwidth of the trigger, which amounts to 3 kHz. The full data sample, corresponding to this trigger rate, is stored on tapes, which are characterized by a relatively slow access time.

As mentioned in Chapter 3, to speed up the process of analysing the data, the LHCb computing services provide the centrally made selection of data subsamples containing preselected events, which corresponds to about 0.5% of the total data volume. This process is called "stripping" and each individual LHCb user can provide his/her set of loose selection criteria for his/her decay in question, which is called the

Table 4.1: Stripping selections criteria as applied in the line `StrippingTau23MuLinesTau23MuLine` for the `Stripping20` campaign as well as all additional offline cuts. The muon candidates are required to leave signals in at least three muon stations (so-called isMuon criteria).

	$\tau^- \to \mu^-\mu^+\mu^-$	$D_s^+ \to \phi(\mu^-\mu^+)\pi^+$
Criteria for $\mu^\pm$ and $\pi^\pm$ candidates		
p_T [MeV/c]	> 300	
Track χ^2/ndf	< 3	
IP χ^2/ndf	> 9	
track ghost probability	< 0.3	
Criteria for μ pair candidates		
$(m_{\mu^-\mu^+} - m_\phi)$ [MeV/c^2]	> 20	< 20
$m_{\mu^+\mu^-}$ [MeV/c^2]	> 450	
$m_{\mu^+\mu^+}$ [MeV/c^2]	> 250	
Criteria for τ and D_s candidates		
Δm [MeV/c^2]	< 400	< 50
Vertex χ^2	< 15	
IP χ^2	< 225	
$\cos\alpha$	> 0.99	
$c\tau$ [µm]	> 100	
lifetime [ns]	< 0.025	

stripping line. These stripped data are stored on hard drives and made available to all LHCb users for performing further studies. Each central production of stripping lines has its unique number. In this analysis the most up-to-date `Stripping20` was used. The particular stripping line used in this study was officially labelled in the LHCb collaboration as `StrippingTau23MuLinesTau23MuLine` and was intended to select events with at least two muons in the final state and with selection criteria reported in Table 4.1. The following variables have been used in the stripping procedure: $m_{\mu^+\mu^-}$ - the invariant mass of the opposite-sign dimuon pair, $m_{\phi,PDG}$ - the mass of ϕ meson as reported in PDG, Δm - the difference of the reconstructed mass of a given decay and the PDG value for τ and D_s, respectively. The variables "Track χ^2/ndf" ("Vertex χ^2") correspond to the chi-square of the fit to the given track (vertex), respectively (the acronym "ndf" denotes the number of degrees of freedom). The "IP χ^2/ndf" is defined as an increase in the χ^2 of the PV vertex when the track is added to the PV. The so-called pointing angle α is the angle between the momentum vector of the $\tau(D_s)$ candidate, respectively, and a line connecting the primary vertex with the decay vertex. The $c\tau$ is the distance between PV and the SV. "Track ghost probability" denotes the probability that a given track is reconstructed as a clone of another track.

Clone tracks are removed by placing a requirement on the same-sign dimuon mass

to be greater than 250 MeV/c^2. Moreover, the events with an opposite dimuon mass within 20 MeV/c^2 of the $\phi(1020)$ nominal mass are rejected as the $\tau^- \to \mu^-\mu^+\mu^-$ candidates. In addition, the same-sign dimuon mass should exceed 450 MeV/c^2 to remove the background from $\mathrm{D_s^-} \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$ decays. This background is discussed in detail in Sect. 4.7.2.

The efficiencies of all these selection criteria are listed in Sect. 4.8. The requirements reported in Table 4.1 include additional (offline) cuts on the angle α and on lifetime that are not part of the stripping, but are applied afterwards. These supplementary selection criteria were found to have a combined efficiency of the value greater than 99.9%, as measured on the signal MC samples.

4.3.1 Monte Carlo simulation samples

The Monte Carlo (MC) samples used in this analysis can be broadly divided into four categories: signal, normalization, physics backgrounds and inclusive backgrounds, all of which are detailed in Table 4.2 and 4.3. The signal (normalization) events contain the $\tau^- \to \mu^-\mu^+\mu^-$ ($\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+$) decay, respectively. Physics background samples encompass the events with decays that can mimic the signal in the selection procedure. The detailed list of these kinds of processes is presented in Sect. 4.7. The inclusive background is a general background sample with special generator selection requirements that will be discussed in detail in Sect. 4.3.2.

Table 4.2: MC simulation samples used for the analysis of data collected in 2011. The MC production version and the number of events generated are also given. All samples are produced with an approximately equal amount of both magnet polarities. Event type denotes a unique label for each type of decay produced by LHCb software called Gauss [87].

Process (event type)	Monte Carlo production	Number of events
Signal:		
$c\bar{c} \to \mathrm{D_s}^- \to \tau^- \to \mu^-\mu^+\mu^-$ (23513001)	"Official" Pythia 6	1,319,998
$b\bar{b} \to \mathrm{D_s}^- \to \tau^- \to \mu^-\mu^+\mu^-$ (23513000)	"Official" Pythia 6	198,099
$c\bar{c} \to \mathrm{D}^- \to \tau^- \to \mu^-\mu^+\mu^-$ (21513001)	"Official" Pythia 6	75,000
$b\bar{b} \to \mathrm{D}^- \to \tau^- \to \mu^-\mu^+\mu^-$ (21513000)	"Official" Pythia 6	5,150
$b\bar{b} \to \tau^- \to \mu^-\mu^+\mu^-$ (31113002)	"Official" Pythia 6	195,300
Control/normalization decay channels:		
$c\bar{c} \to \mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+$ (23173003)	"Official" Pythia 6	1,385,245
$b\bar{b} \to \mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+$ (23173002)	"Official" Pythia 6	167,249

It has been found that the simulated inclusive $b\bar{b} \to \mu^+\mu^- X$ and $c\bar{c} \to \mu^+\mu^- X$ events accurately describe the overall background to the decay in question. Hence, these events have been produced as background samples.

The Monte Carlo samples relevant for this study were created as part of the so-called Sim08 campaign, which provides the latest description of the detector conditions throughout 2011 and 2012 data taking. As this analysis involves data samples collected in 2011 and 2012 data sets, the simulation sets were also divided into two groups (cf Table 4.2 and 4.3), to take into account the differences in beam energy, trigger, and other effects between these two run periods.

Table 4.3: MC simulation samples used for the analysis of data collected in 2012. The MC production version and the number of events generated are also given. All samples are produced with an approximately equal amount of both magnet polarities. Event type denotes a unique label for each type of decay produced by LHCb software called Gauss [87].

Process (event type)	Monte Carlo production	Number of events
Signal:		
$c \to D_s \to \tau^- \to \mu^-\mu^+\mu^-$ (23513001)	"Official" Pythia 6	875,499
$b \to D_s \to \tau^- \to \mu^-\mu^+\mu^-$ (23513000)	"Official" Pythia 6	120,250
$c \to D \to \tau^- \to \mu^-\mu^+\mu^-$ (21513001)	"Official" Pythia 6	71,250
$b \to D \to \tau^- \to \mu^-\mu^+\mu^-$ (21513000)	"Official" Pythia 6	5,850
$b \to \tau^- \to \mu^-\mu^+\mu^-$ (31113002)	"Official" Pythia 6	163,999
$c \to D_s^- \to \tau^- \to \mu^-\mu^+\mu^-$ (23513001)	"Krakow" Pythia 6	706,198
$b \to D_s^- \to \tau^- \to \mu^-\mu^+\mu^-$ (23513000)	"Krakow" Pythia 6	74,700
$c \to D^- \to \tau^- \to \mu^-\mu^+\mu^-$ (21513001)	"Krakow" Pythia 6	41,100
$b \to D^- \to \tau^- \to \mu^-\mu^+\mu^-$ (21513000)	"Krakow" Pythia 6	219,298
$b \to \tau^- \to \mu^-\mu^+\mu^-$ (31113002)	"Krakow" Pythia 6	101,925
Control/normalization decay channels:		
$c \to D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ (23173003)	"Krakow" Pythia 6	1,270,893
$b \to D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ (23173002)	"Krakow" Pythia 6	237,450
$D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$(23513002)	"Official" Pythia 6	5,087,241
$D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$(23513002)	"Official" Pythia 8	5,095,982
Background:		
Inclusive $c\bar{c}$ (20072002)	"Official" Pythia 6	5,094,980
Inclusive $c\bar{c}$ (20072002)	"Official" Pythia 8	5,101,241
Inclusive $c\bar{c}$ (20072002)	"Krakow" Pythia 6+8	4,157,649
Inclusive $b\bar{b}$ (10012013)	"Official" Pythia 6	2,586,990
Inclusive $b\bar{b}$ (10012013)	"Official" Pythia 8	2,544,493
Inclusive $b\bar{b}$ (10012013)	"Krakow" Pythia 6+8	10,196,221

The LHCb system of Monte Carlo samples production is currently in a transitional period between the use of Pythia 6 [88] (written in FORTRAN) and Pythia 8 [89] (coded in C++). While the productions using Pythia 8 are encouraged,

this configuration has not been yet correctly tuned to the LHCb event parameters in data. Therefore some properties, such as track multiplicity, pseudorapidity etc., show up some discrepancies between the two Pythia versions. To avoid making corrections for this effect we use Pythia 6 exclusively for the signal and normalization channels, where generator efficiencies need to be precisely determined. For inclusive and for the background from $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$ decays, where the above difference in Pythia versions is not so important, we use a mixture of 50% Pythia 6 and 50% Pythia 8 simulations, as indicated in Table 4.2 and 4.3 (denoted below as Pythia 6+8). This equal share is suggested by the respective Monte Carlo simulation working group of the LHCb collaboration.

A substantial part of the Monte Carlo samples was prepared "centrally" by the respective team of the LHCb collaboration (denoted below as the "official" setup). The remaining MC samples ($\sim 50\%$) were produced by the author of this dissertation using the computing resources of CYFRONET [90] and IFJ PAN [91] (labelled as the "Krakow" configuration). In the latter case an identical setup, in comparison with the central production, was used so that these two categories of produced MC samples were perfectly compatible.

To save computing time and to increase the number of generated events, a set of selection criteria was applied already at the generation level, as described in Sect. 4.3.2. In the MC production the trigger conditions were emulated using the TCK 0x40760037 for 2011 and TCK 0x409f0045 for 2012, which were chosen to describe the most common trigger conditions throughout the respective year. Since significant changes are expected between TCKs in several trigger lines of interest to this analysis, a number of additional TCKs were also studied, as detailed in Sect. 4.4 and 4.8. To assess the trigger efficiency for the prescaled trigger lines correctly, the `TCKPrescaleEmulator` [92] code has also been applied to the MC samples.

For the signal channels the events were simulated using the so-called "Flat Phase Space" model, based on the phase-space decay of the τ lepton. To produce a correctly normalized mixture of production channels, the signal and normalization MC samples have been splitted into five and two sub-channels respectively. This procedure is detailed in Sect. 4.3.3.

4.3.2 Selection criteria at the generator level

A set of selection criteria (labelled below as "generator cuts") was introduced in order to reduce the number of simulated events. They were aimed at removing those generated events that would be rejected either way at further stages of the analysis. In particular the muon tracks can reach the muon chambers and thus can be reconstructed in the LHCb spectrometer only when their momentum exceeds 3 GeV/c and the transverse momentum (p_T) is above 0.3 GeV/c. Therefore, these two requirements can be imposed already at the generator level without any reduction in selection efficiency. As seen in Table 4.4, slightly lower cut values have been chosen for these two variables in order to provide a safety margin for the momentum smearing which occurs in the process of detector simulation. The selection criteria

Table 4.4: The selection criteria applied to Monte Carlo samples at the generator level. For the signal, normalization and $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$ samples, these requirements are applied to all final state particles, whilst for inclusive backgrounds samples these requirements are imposed to muons only.

Variable	Signal, normalization and $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$ Samples	Inclusive backgrounds Samples
p_T [MeV/c]	> 250	> 280
p [MeV/c]	> 2500	> 2900
DOCA [mm]	-	< 0.35
mass($\mu\mu$) [MeV/c^2]	-	< 4500

which were applied at the generator level to all charged final state particles in the signal, normalization and $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$ samples are summarized in the second column of Table 4.4. For the inclusive background samples the cuts are slightly tighter and are supplemented with a requirement on the distance of the closest approach (denoted as DOCA) of the two muons, as listed in the third column of Table 4.4. In addition, all simulated muons are additionally required to be inside the LHCb detector acceptance (fiducial volume of the LHCb detector). The efficiencies for generated events to pass these generator level cuts are described in Sect. 4.3.3.

4.3.3 Correction of τ and D_s production rates

In the experimental conditions of the LHCb apparatus, τ leptons are coming predominantly from D_s, D and B mesons. It has been observed that Pythia generator produces these mesons in wrong proportions. To calculate the correct fractions of τ leptons from different sources, we used the PDG values of appropriate branching fractions as well as the following cross-section measurements performed by LHCb: [93, 94] for 7 TeV CMS energy data and [95] for 8 TeV CMS energy. Since a measurement of charm cross-section ($\sigma_{c\bar{c}}$) is yet to be performed for 8 TeV, the respective value of the cross-section measured at 7 TeV was taken and scaled by a factor indicated by Pythia. Different parton distributions, taken from [96], were tested and gave consistent results. These cross-sections are listed in Table 4.5. The τ production from Z^0 and $W^\pm$ decays and Drell-Yan processes were neglected as the rates of these contributions were found several orders of magnitude below coming those from heavy quark sources.

The information about five major processes contributing to the cross-section of $\sigma(p\overline{p} \to \tau X)$ is collected in Table 4.6. In the second column the fraction of each process denoted as $\mathrm{Calc}_{4\pi}$, as given in the 4π solid angle at 8 TeV CMS energy, is presented. As previously, we have found that prompt D_s decays are the source of

Table 4.5: Cross-section measurements used in the calculation of contributions of various Monte Carlo simulated processes (called subsequently MC mixing).

Cross-section	7 TeV CMS energy	8 TeV CMS energy
$\sigma(pp \to \mathrm{D}^+X)$ [μb]	3163 ± 363 [94]	3606 ± 415
$\sigma(pp \to \mathrm{D_s}^+X)$ [μb]	976 ± 153 [94]	1113 ± 175
$\sigma(pp \to b\bar{b})$ [μb]	$2 \times (288 \pm 48)$ [93]	$2 \times (298 \pm 36)$ [95]

the majority of τ leptons produced at LHCb. The third column of Table 4.6 gives the generator level efficiency, denoted as $\varepsilon_{GEN|CUT}$, including the requirements that a generated τ lepton was produced in the specific sub-channel of interest[1] and that it decayed within the LHCb acceptance and passed the generator level cuts. The purpose of the mixing method is to provide the correct fractional contribution of each sub-channel, which is known to be incorrect in the simulation procedure of the LHCb experiment. Therefore the "efficiency" for starting with the particular sub-channel[2], ε_{GEN}, must be calculated and divided out to give the actual cut efficiency, ε_{CUT}. This is achieved by preparing a generator-level-only sample of each sub-channel with no cuts applied. As kinematic variations in the different sub-channels cause ε_{CUT} to vary slightly, the fraction of each process that makes up the mixed MC sample should be corrected from its cut-independent value listed in the second column of Table 4.6. The corrected value, f_{Gauss}, is determined by summing the product of ε_{CUT} and $\mathrm{Calc}_{4\pi}$ for each sub-channel and dividing the individual product for the sub-channel of interest by this sum over all decays in question. Then the two most right columns of the Table 4.6 indicate the correct number of events of each sub-channel in the mixed sample, N_{mix}, if a given population of each sub-channel, N_{prod}, is produced. This is determined by first finding the decay with limiting statistics (the minimum value of N_{prod}/f_{Gauss}) and scaling all other sub-channels relative to this.

The above procedure is repeated for the $\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+$ normalization channel, but with two sub-channels only, and is detailed in Table 4.7. The difference in the beam energy between the 2011 and 2012 datasets can potentially cause variations in all of the generator level cut efficiencies. Therefore the mixing calculations were repeated for all channels, using 7 TeV CMS energy Monte Carlo samples, as shown in Table 4.8-4.9. The systematic uncertainty resulting from this method is considered in Sect. 4.8.

[1] This means that τ leptons from all sources are generated. For each of the MC samples, we cut away at generator level the ones that come from the "wrong" parent hadron. However, we are interested in the requirements efficiency of all other generator level cuts, but Gauss (simulation software of LHCb) only provides the full cut efficiency corresponding to the production mechanism, acceptance, and the physics requirements from Table 4.4.

[2] In other words ε_{GEN} is the probability that a τ leptons produced by Gauss simulation comes from the production channel of interest.

Decay chain	Calc$_{4\pi}$ 8 TeV (%)	$\varepsilon_{GEN\|CUT}$ (%)	ε_{GEN} (%)	ε_{CUT} (%)	f_{Gauss} (%)	N$_{prod}$	N$_{mix}$
$D_s \to \tau$	70.2 ± 4.0	5.978 ± 0.016	63.111 ± 0.077	9.472 ± 0.028	72.4 ± 2.7	875,499	807,723
$B_x \to D_s \to \tau$	9.30 ± 2.00	0.8445 ± 0.0024	10.063 ± 0.019	8.392 ± 0.017	8.5 ± 1.7	120,250	94,829
$D^+ \to \tau$	4.10 ± 0.75	6.194 ± 0.016	65.202 ± 0.077	9.500 ± 0.027	4.24 ± 0.77	71,250	47,303
$B_x \to D^+ \to \tau$	0.20 ± 0.04	0.6613 ± 0.0019	7.948 ± 0.015	8.320 ± 0.015	0.172 ± 0.036	5,850	1,919
$B_x \to \tau$	16.2 ± 2.8	2.479 ± 0.012	29.691 ± 0.079	8.349 ± 0.046	14.7 ± 2.3	163,999*	163,999

Table 4.6: The parameters (defined in the text) relevant for the MC mixing method for the decay $\tau^- \to \mu^-\mu^+\mu^-$ at 8 TeV CMS energy. The asterisk mark (*) indicates the MC sample which statistically limits the mixing procedure.

Decay chain	Calc$_{4\pi}$ 8 TeV (%)	$\varepsilon_{GEN\|CUT}$ (%)	ε_{GEN} (%)	ε_{CUT} (%)	f_{Gauss} (%)	N$_{prod}$	N$_{mix}$
D_s	88.3 ± 8.5	10.066 ± 0.088	86.13 ± 0.22	11.69 ± 0.11	89.7 ± 2.0	573,159*	573,159
$B_x \to D_s$	11.7 ± 2.5	1.394 ± 0.014	13.776 ± 0.089	10.12 ± 0.12	10.3 ± 2.0	118,482	65,814

Table 4.7: The parameters (defined in the text) relevant for the MC mixing method for the decay $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ at 8 TeV CMS energy. The asterisk mark (*) indicates the MC sample which statistically limits the mixing procedure.

Decay chain	Calc$_{4\pi}$ 7 TeV (%)	$\varepsilon_{GEN\|CUT}$ (%)	ε_{GEN} (%)	ε_{CUT} (%)	f_{Gauss} (%)	N$_{prod}$	N$_{mix}$
$D_s \to \tau$	68.4 ± 4.4	5.885 ± 0.016	63.49 ± 0.12	9.301 ± 0.033	71.1 ± 3.0	1,319,998	895,860
$B_x \to D_s \to \tau$	10.0 ± 2.4	0.7818 ± 0.0039	9.666 ± 0.029	8.088 ± 0.047	9.0 ± 2.0	198,099	113,400
$D^+ \to \tau$	4.00 ± 0.72	6.076 ± 0.029	65.64 ± 0.12	9.257 ± 0.047	4.14 ± 0.76	75,000	52,164
$B_x \to D^+ \to \tau$	0.204 ± 0.048	0.613 ± 0.011	7.599 ± 0.023	8.07 ± 0.15	0.184 ± 0.044	5,150	2,318
$B_x \to \tau$	17.3 ± 3.4	2.321 ± 0.011	28.927 ± 0.084	8.024 ± 0.045	15.5 ± 2.7	195,300*	195,300

Table 4.8: The parameters (defined in the text) relevant for the MC mixing method for the decay $\tau^- \to \mu^-\mu^+\mu^-$ at 7 TeV CMS energy. The asterisk mark (*) indicates the MC sample which statistically limits the mixing procedure.

Decay chain	Calc$_{4\pi}$ 7 TeV (%)	$\varepsilon_{GEN\|CUT}$ (%)	ε_{GEN} (%)	ε_{CUT} (%)	f_{Gauss} (%)	N$_{prod}$	N$_{mix}$
D_s	88.4 ± 8.5	9.885 ± 0.027	86.897 ± 0.099	11.376 ± 0.034	88.9 ± 2.2	1,385,245	1,339,499
$B_x \to D_s$	11.6 ± 2.5	1.2697 ± 0.0046	13.090 ± 0.039	9.700 ± 0.045	11.1 ± 2.2	167,249*	167,249

Table 4.9: The parameters (defined in the text) relevant for the MC mixing method for the decay $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ at 7 TeV CMS energy. The asterisk mark (*) indicates the MC sample which statistically limits the mixing procedure.

4.3.4 Luminosities of background Monte Carlo samples

For the inclusive background samples it is crucial to know the corresponding integrated luminosities. Firstly, this allows for a comparison of the simulated event yield to the observed event yield. In particular, if the simulated backgrounds contain fewer events than the same luminosity in data, this would indicate that a significant background source would be missing in the background simulation. Secondly, the two inclusive backgrounds have to be mixed together in the right proportions to correctly describe the background seen in data.

The integrated luminosities for inclusive background samples are calculated from the relation:

$$\mathcal{L} = \frac{N_{MC}}{\varepsilon_{ACC} \times \varepsilon_{CUT} \times \sigma_{LHCb}}, \tag{4.1}$$

where $\mathcal{L}$ is the integrated luminosity of the sample, N_{MC} is the number of generated MC events, ε_{ACC} is the generator level efficiency for the decay products to be in the LHCb acceptance, ε_{CUT} is the generator level efficiency for the final state that contains two muons and passes the requirements collected in Table 4.4 and σ_{LHCb} is the cross-section for charm or beauty production at 8 TeV CMS energy. For the latter the values discussed in Sect. 4.3.3 have been used. The values of N_{MC} for different production samples are given in Table 4.3, while the values of other input variables are provided in Table 4.10. Finally, the resulting effective integrated luminosities are listed in Table 4.11.

Table 4.10: Inputs to the calculation of the luminosity of the background samples. Although the samples used in the analysis are a combination of Pythia 6 and 8, only Pythia 6 numbers are given as the resulting effective luminosities are identical in both versions within the uncertainties of the method.

	Inclusive $c\bar{c}$	Inclusive $b\bar{b}$
ε_{ACC}	0.4554 ± 0.001	0.437 ± 0.003
ε_{CUT}	$(1.29 \pm 0.01) \times 10^{-3}$	$(6.38 \pm 0.05) \times 10^{-3}$
σ_{LHCb}	(6.95 ± 1.07) mb	(298 ± 36) μb

Table 4.11: Integrated luminosities of the inclusive background simulation.

Process (event type)	luminosity per 1M events	"official" MC	"Krakow" MC
Inclusive $c\bar{c}$ (20072002)	$(0.25 \pm 0.04)\,\mathrm{pb}^{-1}$	$(2.5 \pm 0.4)\,\mathrm{pb}^{-1}$	$(1.0 \pm 0.2)\,\mathrm{pb}^{-1}$
Inclusive $b\bar{b}$ (10012013)	$(1.2 \pm 0.2)\,\mathrm{pb}^{-1}$	$(6.2 \pm 1.0)\,\mathrm{pb}^{-1}$	$(12 \pm 2)\,\mathrm{pb}^{-1}$

4.3.5 Differences between conditions of Monte Carlo simulations in 2011 and 2012

The MC samples listed above share a common reconstruction for both 2011 and 2012 setups. However, apart from the difference in CMS energy (7 TeV vs 8 TeV)

these two sets of Monte Carlo samples are also characterized by several other individual properties. In particular the beam conditions were very different for the samples corresponding to 2011 and 2012 conditions, as summarized in Table 4.12. The change in the $z-$position of the primary vertex causes the tracks of all particles coming from the decays of interest to traverse a different region of the VELO detector, and as a result the reconstruction efficiency between the 2011 and 2012 signal MC samples can be expected to vary by a small amount. This effect is quantified in Sect. 4.8, where all efficiencies required to compute the normalization factor are given for both 2011 and 2012 MC. The above discussed selection criteria have been used in all studies described in the next chapters.

Year	Energy [TeV]	z_{PV}[mm]	crossing angle in x coordinate	crossing angle in y coordinate
2011	3.5	0.5	-0.520	0
2012	4	25.7	0.236	0.100

Table 4.12: Beam conditions for the Monte Carlo samples used in this analysis.

4.4 Trigger requirements

As described in Sect. 3.2.6.2, the HLT2 trigger is composed of many "lines", each designed to select different kind of processes. In this analysis the choice of the trigger lines was optimized to ensure the best signal to background ratio. In a given event a trigger can be fired by a signal candidate (in our case $\tau^- \to \mu^-\mu^+\mu^-$ or $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$) or some other tracks that are not part of the signal candidate. The first case is called Trigger On Signal (TOS), while the second one is labelled as Trigger Independent of Signal (TIS). Since Monte Carlo simulation does not describe properly the features of the underlying event, the efficiency of TIS triggers is not correctly simulated in MC, and analyses are done predominantly using TOS triggers. In this study we considered the use of both scenarios. It was found that the performance of TOS trigger is better in terms of signal to background ratio.

For trigger optimization we defined the following quantities:

$$\varepsilon_{\text{TOS,line}} = \frac{N(\tau \text{ MC candidates triggering line})}{N(\tau \text{ MC candidates triggered by any line})}, \tag{4.2}$$

where $N(\tau$ MC candidates triggering line) is the number of signal MC events that pass the selection criteria of a given trigger line with TOS requirement and $N(\tau$ MC candidates triggered by any line) is the number of events that were triggered by any line (TIS or TOS).

For background middle-sidebands in data (defined in Sect. 4.1) were used and, in analogy with Eq. 4.2, we define:

$$\beta_{\text{TOS,line}} = \frac{N(\text{ middle-sideband candidates TOS triggering line})}{N(\text{ middle-sideband candidates TOS triggered by any line})}, \tag{4.3}$$

where $N($ middle-sideband candidates TOS triggering line$)$ is the number of candidates passing the given trigger line from middle-sidebands and $N($ middle-sideband candidates TOS triggered by any line$)$ is the number of events from middle-sidebands that were triggered by any line (TIS and TOS) with the same selection as for the nominator.

The order in which the trigger stages are optimized is from HLT2 to L0. The following trigger optimization was applied:

- In the optimization of the HLT2 trigger choice, events are required to pass any trigger in L0 and HLT1.
- In the optimization of the HLT1 trigger choice, events are required to pass any physics trigger in L0 and to fulfil the HLT2 requirement which was found optimal in the previous step.
- In the optimization of the L0 trigger choice, events are required to fulfil the HLT2 and HLT1 requirements which were found optimal in the previous two steps.

For each line, the Punzi figure of merit [97] is computed as:

$$p_{TOS,\text{line}} = \frac{\varepsilon_{TOS,\text{line}}}{1 + \sqrt{\beta_{TOS,\text{line}}}}. \tag{4.4}$$

The quantity $p_{TOS,\text{line}}$ will be used for ranking the individual trigger lines.

4.4.1 Optimization of HLT2 trigger in 2012 data taking conditions

As mentioned in Sect. 4.4, trigger lines are sorted by $p_{\text{TOS,line}}$ and according to their order for each line it is determined how much efficiency and retention is gained by adding TOS candidates from this line to the analysis:

$$\varepsilon'_{\text{TOS,line}} = \frac{N(\tau \text{ MC candidates triggering line, but not any better line})}{N(\tau \text{ MC events triggered by any line})}, \tag{4.5}$$

$$\beta'_{\text{TOS,line}} = \frac{N(\text{ middle-sideband candidates triggering line, but not any better line})}{N(\text{ middle-sideband events triggered by any line})}, \tag{4.6}$$

where $N(\tau$ MC candidates triggering line, but not any better line$)$ is the number of events from MC that are selected by this line as TOS and not selected by any line that has a higher $p_{\text{TOS,line}}$ rank. Therefore $\varepsilon'_{\text{TOS,line}}$ quantity can be considered as a measure of efficiency gain. In analogy $N($middle-sideband candidates triggering line, but not any better line$)$ is the number of background events from middle-sidebands that are selected by a given line as TOS and are not selected by any better ranked line. We would like to point out that $\sum_{\text{all lines}} \varepsilon'_{\text{line}} < 1$ and $\sum_{\text{all lines}} \beta'_{\text{line}} < 1$ because the nominator has only TOS requirements and the denominator is either TIS or TOS. The reason for the different

treatment of the nominator and denominator is that the same method was used to select TIS lines in the nominator. As these lines were found to be much worse then the TOS ones, we will focus on TOS below.

The best trigger lines will be used in the following analysis. Further trigger lines are added depending on the *cumulative trigger figure of merit* (CTFM) defined as:

$$\mathrm{CTFM} = \frac{\sqrt{\sum_{\text{trigger lines}} \beta'_{\text{evt,line}}}}{\sum_{\text{trigger lines}} \varepsilon'_{\text{evt,line}}}. \tag{4.7}$$

The results of the above considerations are given in Table 4.13. It is concluded that only candidates which are selected by Hlt2TriMuonTau as TOS will be used in futher steps.

Table 4.13: Results of the trigger optimization based on the HLT2 trigger decision for 2012 data taking conditions. Trigger lines not contributing any additional events are omitted. The parameters $\varepsilon'_{\text{TOS,line}}$, $\beta'_{\text{TOS,line}}$ and CTFM are described in the text.

trigger line name	$\varepsilon'_{\text{TOS,line}}$	$\beta'_{\text{TOS,line}}$	CTFM
Hlt2TriMuonTauDecision	0.877306	0.712216	0.961955
Hlt2DiMuonDetachedDecision	0.0672294	0.171496	0.995261
Hlt2CharmSemilep3bodyD2KMuMuDecision	0.0207291	0.0189898	0.984296
Hlt2CharmHadD2HHHDecision	0.00410846	0.0013704	0.980868
Hlt2CharmSemilep3bodyD2KMuMuSSDecision	0.00141929	0.00195771	0.980494
Hlt2CharmSemilep3bodyD2PiMuMuDecision	0.00209158	0.00509005	0.98113
Hlt2CharmHadD02HH_D02KKDecision	0.00145664	0.00156617	0.980505
Hlt2TopoMu2BodyBBDTDecision	0.00201688	0.00469851	0.980995
Hlt2SingleMuonDecision	0.00257713	0.00587314	0.981539
Hlt2CharmSemilep3bodyD2PiMuMuSSDecision	0.000112049	0	0.981427
Hlt2TriMuonDetachedDecision	7.46993e-05	0	0.981352
Hlt2CharmHadD02HH_D02KPiDecision	3.73497e-05	0	0.981315
Hlt2TopoMu3BodyBBDTDecision	0.000112049	0.000195771	0.981306
Hlt2CharmSemilepD02HMuNu_D02KMuNuTightDecision	0.000410846	0.000587314	0.981207
Hlt2CharmSemilepD02HMuNu_D02KMuNuDecision	0.000298797	0.000391543	0.981115
Hlt2CharmSemilepD02HMuNu_D02PiMuNuDecision	0.000261448	0	0.980854
Hlt2CharmHadD02HH_D02KKWideMassDecision	3.73497e-05	0	0.980816
Hlt2Dst2PiD02MuMuDecision	0.000224098	0.000978857	0.981111
Hlt2CharmSemilepD02PiPiMuMuDecision	0.000784343	0.00156617	0.981156
Hlt2CharmSemilepD02KPiMuMuDecision	0.00138194	0.00176194	0.980707
Hlt2CharmSemilepD02KKMuMuDecision	0.000859042	0.000783085	0.980263
Hlt2SingleMuonHighPTDecision	7.46993e-05	0.000783085	0.980602
Hlt2Dst2PiD02KMuDecision	3.73497e-05	0	0.980564
Hlt2CharmHadD02HHXDst_hhXDecision	7.46993e-05	0	0.98049
Hlt2CharmSemilepD02HMuNu_D02PiMuNuWSDecision	7.46993e-05	0.000195771	0.980519
Hlt2CharmSemilepD02HMuNu_D02KMuNuWSDecision	7.46993e-05	0.000195771	0.980547

4.4.2 HLT1 trigger in 2012 data taking conditions

Given the findings of the previous Sect. 4.4.1, the optimization of HLT1 trigger is done using events which are TOS in the Hlt2TriMuonTau line. The optimization for HLT1 trigger is presented in Table 4.14. We conclude that Hlt1TrackMuonDecision is the optimum solution for HLT1 lines selection.

Table 4.14: Results of the trigger optimization based on the HLT1 trigger decision for 2012 data taking conditions. Trigger lines not contributing any additional events are omitted. The parameters $\varepsilon'_{\text{TOS,line}}$, $\beta'_{\text{TOS,line}}$and CTFM are described in the text.

trigger line name	$\varepsilon'_{\text{TOS,line}}$	$\beta'_{\text{TOS,line}}$	CTFM
Hlt1TrackMuonDecision	0.882839	0.717427	0.959417
Hlt1DiMuonLowMassDecision	0.0770573	0.192139	0.993557
Hlt1TrackAllL0Decision	0.017455	0.0195162	0.986226
Hlt1SingleMuonHighPTDecision	0.000936609	0.000549753	0.985573
Hlt1TrackPhotonDecision	0.00170292	0.00192413	0.984878

4.4.3 L0 trigger in 2012 data taking conditions

Given the findings in the previous sections, the optimization of L0 trigger is done using events which are selected by the Hlt2TriMuonTau and Hlt1TrackMuon lines as TOS candidates. The optimization for L0 is presented in Table 4.15.

Table 4.15: Results of the trigger optimization based on the L0 trigger decision for 2012 data taking conditions. Trigger lines not contributing any additional events are omitted. The parameters $\varepsilon'_{\text{TOS,line}}$, $\beta'_{\text{TOS,line}}$ and CTFM are described in the text.

trigger line name	$\varepsilon'_{\text{TOS,line}}$	$\beta'_{\text{TOS,line}}$	CTFM
L0Muon	0.953417	0.850192	0.96711
L0DiMuon	0.0140811	0.0693487	0.991141
L0Hadron	0.000530453	0.00153257	0.991423
L0Photon	4.8223e-05	0	0.991374

4.4.4 Optimization of trigger lines in 2011 data taking conditions.

The above described procedure is applied also for the conditions of 2011 data taking. The optimization for HLT2 is shown in Table 4.16. We conclude that events that are selected by Hlt2TriMuonTau or Hlt2CharmSemilepD2HMuMu as TOS will be used in futher considerations. We would like to point out that the line Hlt2CharmSemilepD2HMuMu was not present in 2012 data taking conditions, so the previous optimization was obviously different.

The same procedure of HLT1 and L0 triggers optimizations as described in Sect. 4.4.2 and 4.4.3 was repeated for 2011 data taking conditions and is reported in Tables 4.16, 4.17 and 4.18.

Table 4.16: Results of the trigger optimization based on HLT2 TOS candidates in 2011 data taking conditions. Trigger lines not contributing any additional events are omitted. The parameters $\varepsilon'_{\text{TOS,line}}$, $\beta'_{\text{TOS,line}}$ and CTFM are described in the text.

trigger line name	$\varepsilon'_{\text{TOS,line}}$	$\beta'_{\text{TOS,line}}$	CTFM
Hlt2CharmSemilepD2HMuMuDecision	0.684252	0.466071	0.997722
Hlt2TriMuonTauDecision	0.1054399	0.118452	0.968151
Hlt2DiMuonDetachedDecision	0.060081	0.133929	0.997461
Hlt2CharmHadD2HHHDecision	0.010061	0.00238095	0.987421
Hlt2CharmSemilepD02HMuNu_D02KMuNuTightDecision	0.00340579	0.000595238	0.983932
Hlt2TopoMu2BodyBBDTDecision	0.00472641	0.00595238	0.982602
Hlt2SingleMuonDecision	0.00808007	0.0142857	0.983053
Hlt2CharmSemilepD2HMuMuWideMassDecision	0.00250222	0.00297619	0.982218
Hlt2TopoMu3BodyBBDTDecision	0.000330153	0.000595238	0.982242
Hlt2TriMuonDetachedDecision	6.9506e-05	0.00178571	0.983340
Hlt2CharmHadD02HH_D02KKDecision	1.73765e-05	0	0.983320
Hlt2CharmSemilepD02HHMuMuHardHadronsAndMuonsDecision	0.00542147	0.00833333	0.982728
Hlt2CharmSemilepD02HHMuMuDecision	0.000243271	0.00178571	0.983619
Hlt2CharmSemilepD02HMuNu_D02PiMuNuDecision	0.000660307	0.00119048	0.983658
Hlt2CharmSemilepD02HMuNu_D02KMuNuDecision	0.00069506	0.00119048	0.983657
Hlt2Dst2PiD02MuMuDecision	0.000417036	0.00119048	0.983965
Hlt2CharmSemilepD02HHMuMuHardHadronsAndMuonsWideMassDecision	0.000903578	0.00119048	0.983731
Hlt2CharmSemilepD02HHMuMuWideMassDecision	0.00099046	0.000595238	0.983018
Hlt2CharmHadD02HH_D02KKWideMassDecision	5.21295e-05	0	0.982961
Hlt2TopoMu4BodyBBDTDecision	3.4753e-05	0	0.982922
Hlt2SingleMuonHighPTDecision	0.00034753	0	0.982538
Hlt2Dst2PiD02KMuDecision	3.4753e-05	0	0.982499
Hlt2CharmSemilepD02HMuNu_D02KMuNuWSDecision	0.000191141	0	0.982288
Hlt2CharmSemilepD02HMuNu_D02PiMuNuWSDecision	0.000121635	0.000595238	0.982537
Hlt2CharmHadLambdaC2KPPiDecision	1.73765e-05	0.000595238	0.982901
Hlt2CharmSemilepD02HHMuMuHardHadronsSoftMuonsWideMassDecision	0.000121635	0	0.982766
Hlt2diPhotonDiMuonDecision	1.73765e-05	0	0.982747

Table 4.17: Results of the trigger optimization based on HLT1 TOS candidates in 2011 data taking conditions. Trigger lines not contributing any additional events are omitted. The parameters $\varepsilon'_{\text{TOS,line}}$, $\beta'_{\text{TOS,line}}$ and CTFM are described in the text.

trigger line name	$\varepsilon'_{\text{TOS,line}}$	$\beta'_{\text{TOS,line}}$	CTFM
Hlt1TrackMuonDecision	0.953065	0.862375	0.974374
Hlt1DiMuonLowMassDecision	0.0269415	0.0754717	0.988182
Hlt1TrackAllL0Decision	0.0089414	0.0266371	0.993057
Hlt1SingleMuonHighPTDecision	0.00111401	0.00554939	0.994789
Hlt1TrackPhotonDecision	0.00111401	0.00110988	0.994239
Hlt1SingleMuonNoIPDecision	2.93161e-05	0	0.99421

Table 4.18: Results of the trigger optimization based on L0 TOS candidates in 2011 data taking conditions. Trigger lines not contributing any additional events are omitted. The parameters $\varepsilon'_{\text{TOS,line}}$, $\beta'_{\text{TOS,line}}$ and CTFM are described in the text.

trigger line name	$\varepsilon'_{\text{TOS,line}}$	$\beta'_{\text{TOS,line}}$	CTFM
L0Muon	0.971824	0.876448	0.963331
L0DiMuon	0.012673	0.0810811	0.993943
L0Hadron	0.000276838	0	0.993664

4.4.4.1 Summary of triggered requirements

We performed a trigger line selection optimization based on MC events and the middle-sidebands. The resulting trigger lines chosen for each year of data taking are listed in Table 4.19.

Table 4.19: Triggers lines that were selected in the optimization procedure.

trigger line name	2011 data taking conditions	2012 data taking conditions
L0	L0Muon	L0Muon
HLT1	Hlt1TrackMuonDecision	Hlt1TrackMuonDecision
HLT2	Hlt2TriMuonTauDecision or Hlt2CharmSemilepD2HMuMuDecision	Hlt2TriMuonTauDecision

4.4.5 Cross-check of trigger requirements with further selection

The choice of trigger lines was optimized for "$\tau^- \to \mu^-\mu^+\mu^-$ sample" obtained after the preselection requirements described in Sect. 4.3. These criteria tend to be loose and will be preceded by the requirements described in Sect. 4.5. We performed a final test by applying stronger requirements on the classifiers ($\mathcal{M}_{3body}, \mathcal{M}_{PID}$) described in the next section. For a cross-check we considered the following three requirements:

(a) $\mathcal{M}_{3body} > 0.1, \mathcal{M}_{PID} > 0.5$,

(b) $\mathcal{M}_{3body} > 0.5, \mathcal{M}_{PID} > 0.1$,

(c) $\mathcal{M}_{3body} > 0.5, \mathcal{M}_{PID} > 0.5$.

After applying each of these conditions in sequence, the procedure of trigger optimization was repeated. The ordering of trigger lines was found to be stable against the above algorithm, yielding in this case the same selection of trigger lines.

4.4.6 Optimization of trigger lines used for the normalization decay channel

The usage of the Hlt2TriMuonTauLine for the normalization decay channel $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ was not found to be the optimal choice. As a result, it was replaced by the Hlt2DiMuonDetachedLine (TOS as well), which was checked to provide the better performance against the procedure underlined above. For simplicity only this line is used for both 2011 and 2012 data. In HLT1 and L0 trigger the normalization candidates have to fulfil the same requirements as the signal candidates.

4.5 Signal and background discrimination

To discriminate between signal and background events in the sample that passed the above selection criteria, we attribute to each event the probability of being signal- or background-like. This assignment is given according to the values of the following three variables:

- The geometric likelihood ($\mathcal{M}_{3body}$): a multivariate classifier which uses the geometry of the reconstructed τ decay to distinguish displaced three-body decays from N-body decays (with $N > 3$) and combinations of tracks originating from different vertices.

- The PID likelihood ($\mathcal{M}_{PID}$): a multivariate classifier which quantifies the compatibility of each of the three decay particles with the muon hypothesis. This classifier is provided by the particle identification working group of the LHCb experiment [98].

- The invariant mass of the τ candidate $m_{\mu\mu\mu}$.

The data is splitted into several bins in each of the above three variables. The details of the partition in the bins will be given in Sect. 4.5.4. Generally, the bins with a low likelihood (for any of the three variables) are expected to contain mainly background events and only a small portion of the signal $p.d.f$. The high-likelihood bins (in all three variables) are expected to contain only a small number of background events but a significant part of the signal $p.d.f$. The latter are most suited for searching for the decay in question.

4.5.1 Blending technique

The blending (or ensemble technique) [99] is a procedure that transforms multiple likelihood estimates of the same variable into a single one, with a higher efficiency/purity characteristics. Since in the analysis described in this thesis the τ leptons are produced in five different processes, we proposed a method which combines not only different classifiers but also different production channels. We found that this method was the most effective in terms of signal efficiency vs. background rejection (as will be justified below).

The blending procedure is composed of two steps. In the first one, denoted as "Stage I", classifiers for individual sources of τ leptons were constructed. In the following "Stage II", the classifiers prepared and trained in the previous step were combined to a single one with an optimal performance.

4.5.1.1 Variables used in the training of the classifiers

The multi-variate classifiers were trained using the following variables:

- **DOCA:** the minimum of distances of the closest approach (as defined in Sect. 4.3.2) of two muons in each of three possible two muon pairings,

- **$\tau(D_s)$ Vertex χ^2:** the quality of the vertex parametrized as the chi square of the τ secondary vertex fit (as defined in Sect. 4.3),

- **$c\tau$:** The measured decay length of the τ lepton, assuming its production at the primary vertex. To smooth out the distribution, the decay time is transformed according to the formula $T = \exp(-1000 \cdot \tau)$,

- **IP χ^2 (τ):** τ lepton impact parameter χ^2/ndf as defined in Sect. 4.3,

- **Min. IP χ^2 (μ):** the minimum value of the three μ impact parameter (χ^2/ndf)s, as defined in Sect. 4.3,

- **Track χ^2/ndf:** maximum of track's (χ^2)s of the three muons as defined in Sect. 4.3.
- **Pointing angle α:** the angle between the direction of τ momentum and a straight line from the τ decay vertex to the primary vertex as defined in Sect. 4.3,
- $\mathbf{p_T}$: the τ transverse momentum,
- **Track isolation:** the sum of three track isolations variables, each parametrising how far in space is an individual muon candidate w.r.t. the rest of event. The detailed explanation of this variable is given in Appendix A.1,
- **BDT (Boosted Decision Tree) isolation:** the response of multivariate analysis (MVA) working at the charged track level and aimed at discriminating between isolated and non-isolated tracks. The full construction of this variable is given in Appendix A.2,
- **Cone isolation:** the fraction of the τ candidate transverse momentum among the sum of all transverse momenta within a certain cone around the τ candidate. The full description of this variable is given in Appendix A.3.

4.5.1.2 The first stage of blending technique

For each production channel of the τ lepton ($D_s \to \tau X$, $D \to \tau X$, $B \to D_s X \to \tau Y$, $B \to DX \to \tau Y$, $B \to \tau X$) each individual mutivariate classifier is trained separately on one third of the signal MC sample and half of the simulated backgrounds (encompassing a mixture of $b\bar{b}$ and $c\bar{c}$ MC samples), using eleven variables listed in Sect. 4.5.1.1. This process is illustrated in Fig. 4.1 as Stage I. The samples used in the training do not take part in further steps of the blending procedure. At this stage the signals are not mixed so each classifier corresponds to a "clean" source of τ leptons from a single production channel.

The following ten multi-variate classifiers [100] were trained in the "Stage I" of blending technique:

- BDTG - Gradient Boosted Decision Tree,
- FDA - Function Discriminant Analysis,
- Fisher - Fisher (linear) discriminant,
- GFisher Fisher discriminant with Gaussian-transformed input variables,
- LD - Linear Discriminant,
- MLPC - Multi Layer Perception (Artificial Neural Network) with tanh as the activation function, and Cross-Entropy estimator (Bernoulli Likelihood),

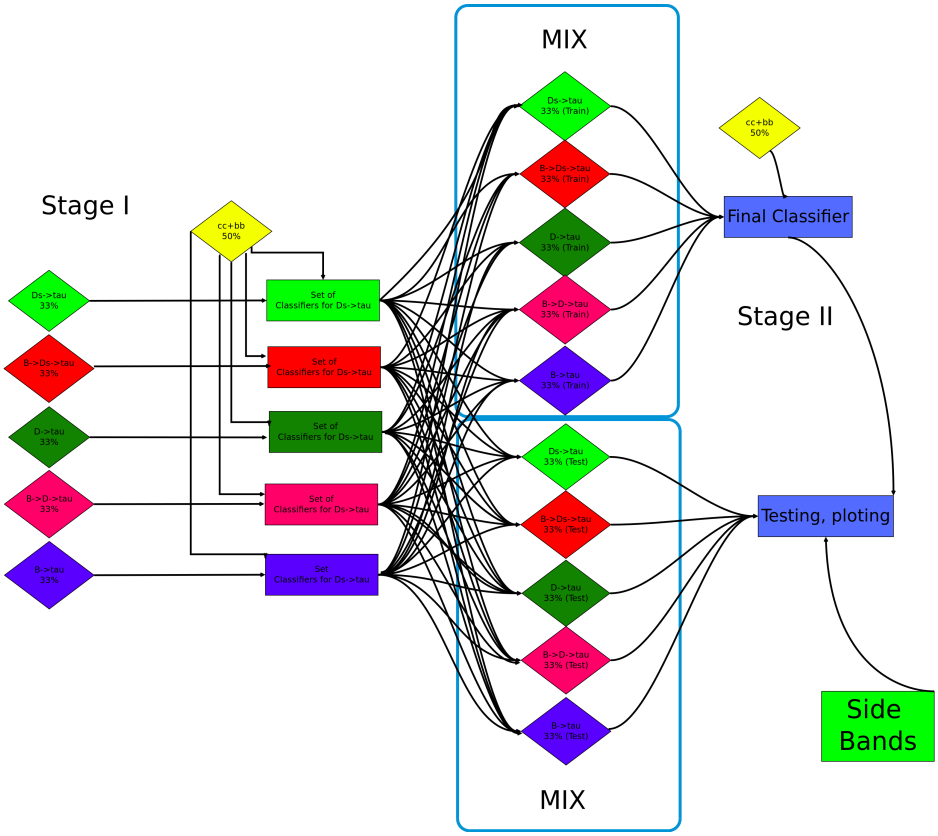


Figure 4.1: Schematic representation of the blending technique.

- MLPCJ_SIG - Multi Layer Perception (Artificial Neural Network) with sigmoid[3] as the activation function and Cross-Entropy estimator (Bernoulli Likelihood),
- MLP - Multi Layer Perception (Artificial Neural Network) with tanh as the activation function and Mean Square Estimator (Gaussian Likelihood),
- MLP_SIG - Multi Layer Perception (Artificial Neural Network) with sigmoid as the activation function and Mean Square Estimator (Gaussian Likelihood),
- MN - MatrixNet discriminant.

The selection of the above set of classifiers was dictated by the fact that they are implemented in the TMVA [101] and MatrixNet [102, 103] packages and thus directly accessible. Moreover, they provide a high efficiency/purity characteristics and are relatively easy to train.

[3] $S(x) = \frac{1}{1 - e^{-x}}$

4.5.1.3 The second stage of blending technique

In this stage the responses of ten classifiers listed above (each using the set of eleven variables from Sect. 4.5.1.2) are used as input to the final classifier which is trained on the remaining half of the simulated background events and the second third of signal MC events (using the correctly mixed signal MC in this step). This procedure is labelled as "Stage II" in Fig. 4.1. The classifiers were tested on the remaining third part of signal MC and middle-sidebands defined in Sect. 4.1 of data. For the evaluation of the final classifiers together with the resulting distribution we use the remaining, independent subsamples which remove potential biases which might be caused by the procedure.

As the final classifier we have chosen the one exploiting the MatrixNet approach [102, 103]. Since it belongs to the category of gradient boosted decision trees, the response distribution is characterized by the presence of narrow maxima close to 0 (1) for background (signal) events, respectively. This feature is, however, cumbersome to handle in further optimization, so both distributions have been transformed in such a way, that they have flat signal distribution. Fig. 4.2 shows the classifier distribution for signal and background.

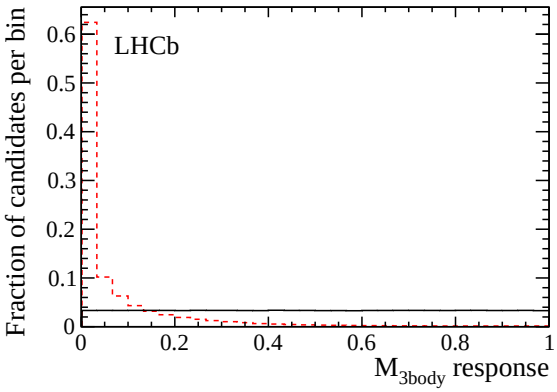


Figure 4.2: Distribution of signal (black solid line) and background (red dashed histogram) events for the final ("Stage II") $\mathcal{M}_{3body}$ classifier. The distributions have been transformed accordingly to show up a flat signal spectrum.

In Fig. 4.3 we compare the performance of the $\mathcal{M}_{3body}$ variable with other commonly used classifier techniques. Firstly, $\mathcal{M}_{3body}$ is compared to various TMVA classifiers with different optimization strategies (the best of which was BDT). Over a hundred classifiers were trained. Furthermore we compared our $\mathcal{M}_{3body}$ classifier to the output of MatrixNet approach without blending. Finally we present also the performance of the approach known as Multinomial classification [104], which is less known in high Energy Physics. The performance of the abovementioned methods was compared by presenting Receiver Operating Characteristics (ROC). The relatively small errors characterising the distributions in Fig. 4.3 resulted from the use of a large statistics sample of sidebands data. As can be seen in Fig. 4.3, the

performance of the $\mathcal{M}_{3body}$ classifier evaluated with blending technique is optimal.

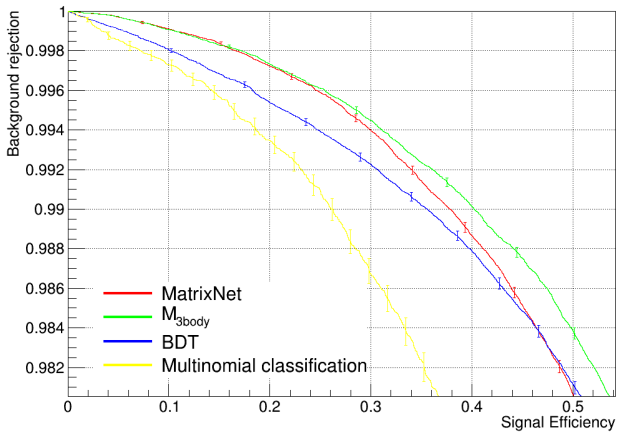


Figure 4.3: Receiver Operating Characteristics for trained classifiers. The names of the classifiers are explained in the text.

4.5.1.4 Calibration of the blending classifier response

The final $\mathcal{M}_{3body}$ classifier response is calibrated using the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ normalization samples to correct the differences in response between data and simulation. The correction factors are then applied to the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ response from data. It is assumed that $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ and $\tau^- \to \mu^-\mu^+\mu^-$ differ only in their response distributions because of physics effects which are correctly described in Monte Carlo simulation. These are kinematic and geometric differences, e.g. $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ candidates point on average a bit better to the PV than $\tau^- \to \mu^-\mu^+\mu^-$ candidates, $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ decays through an intermediate resonance, etc. The correction factors are therefore the ones which must be applied in MC sample to the distribution of $\mathcal{M}_{3body}$ cut efficiency corresponding to the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ decay to obtain the $\mathcal{M}_{3body}$ cut efficiency relevant for $\tau^- \to \mu^-\mu^+\mu^-$ decay. Subsequently, these corrections are applied to the $\mathcal{M}_{3body}$ distribution of $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ data sample, yielding the final response for the $\tau^- \to \mu^-\mu^+\mu^-$ data sample.

The systematic uncertainty of the above method has been investigated by comparing smeared (in vertex and IP resolutions) with unsmeared Monte Carlo samples. We did not observe any statistically significant difference and assigned a conservative 1.3% error, which was taken as the statistical error on a single bin. The response distribution on $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ data sample is evaluated using the *sPlot* technique [105]. The systematics uncertainty from the *sPlot* method were also taken into account because of the use of different fit models as well as because of statistical errors of the calibration procedure.

We attributed also an additional systematic uncertainty because of the dependence of $\mathcal{M}_{3body}$ variable on the τ lepton production mechanism. For each of the

five production channels a 1σ change upwards and downwards in cross-sections is assumed to give a different signal *p.d.f.* The respective five nuisance parameters resulting from this procedure were used in the limit computation to address the uncertainty on the τ lepton production. The correlation between signal *p.d.f.* change and normalization factor variation was also taken into account.

4.5.2 Particle identification classifier

The multivariate particle identification classifier $\mathcal{M}_{PID}$ is an artificial neural net combining inputs from the PID subdetectors and tracking information. This classifier is provided by DaVinci package and is trained by the respective PID working group of the LHCb collaboration [98]. Such a classifier provides response for each particle hypothesis. It was found that the rejection of background due to $\mu\mu X$ combinations is optimized by applying the requirement on the minimum value of the muon PID classifier for the three muon track candidates selected in the search for $\tau^- \to \mu^-\mu^+\mu^-$ decay: $\mathcal{M}_{PID} = \min(\mathcal{M}_{PID,1}, \mathcal{M}_{PID,2}, \mathcal{M}_{PID,3})$.

4.5.2.1 Calibration of $\mathcal{M}_{PID}$ variable

The $\mathcal{M}_{PID}$ variable is calibrated using the `MultiTrack` tool, a part of the `PIDCalib` package [98], as provided by the particle identification group of the LHCb collaboration. The calibration is performed on a dimuon data and Monte Carlo samples containing the decays $J/\psi \to \mu^-\mu^+$ which were selected without any requirements on particle identification. The respective correction factors are provided in bins of track momentum, pseudorapidity and track multiplicity. They were applied to the Monte Carlo samples used in the analysis.

For each track in the signal MC sample the efficiency to pass a $\mathcal{M}_{PID}$ requirement is evaluated on the calibration sample. Since this is done on an event-by-event basis, the kinematics of one track affect the kinematics of another. Thus the kinematic correlations between all final state tracks are included when calculating the PID efficiency for that event and are averaged over all phase space. The probability to retain a τ candidate within a certain $\mathcal{M}_{PID}$ bin ($x_1 < \mathcal{M}_{PID} < x_2$) is the difference of selection efficiencies for requirements at the boundaries $p_{\mathcal{M}_{PID}>x_1}(\tau) - p_{\mathcal{M}_{PID}>x_2}(\tau)$.

The `MultiTrack` tool provides also statistical uncertainties associated with the $p_{\mathcal{M}_{PID}>x_1}(\tau)$ requirement arising from the $J/\psi \to \mu^-\mu^+$ statistics which is added in quadrature to a systematic error of 1% per track. Since the $\mathcal{M}_{PID}$ cut efficiency at $\mathcal{M}_{PID} > x_1$ enters the two *p.d.f.* bins $x_1 < \mathcal{M}_{PID} < x_2$ and $x_3 < \mathcal{M}_{PID} < x_1$, the errors on the *p.d.f.* bin contents due to the error on the $\mathcal{M}_{PID}$ cut efficiency at $\mathcal{M}_{PID} > x_1$ are considered as fully anticorrelated.

4.5.2.2 Correction to calibration of $\mathcal{M}_{PID}$ variable

The correctness of the $\mathcal{M}_{PID}$ calibration from `PIDCalib` package is checked using $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ samples. For each of the $\mathcal{M}_{PID}$ bin boundaries a requirement is

applied on the $\mathcal{M}_{PID}$ of both muons from the final state. Because of this requirement the reduction of the signal yield is compared with the PIDCalib predicted cut efficiency, using the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ MC as a reference sample. Moreover, we provided, not only the efficiencies for a $\mathcal{M}_{PID}$ requirement on both muons but also in the case where a requirement was applied at random to one of the two muons. In the following these efficiencies are called ε_1 for a cut on one random muon, ε_2 for a cut on two random muons, and ε_3 for a cut on three muons (only applicable to $\tau^- \to \mu^-\mu^+\mu^-$).

Table 4.20: Efficiencies of the requirements on the $\mathcal{M}_{PID}$ classifier response (the parameters ε_1 and ε_2 are defined in the text) obtained on $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ normalization sample in data collected in 2011.

$\mathcal{M}_{PID}$ cut	0.4	0.45	0.54	0.63	0.75
ε_1 [%]	87.2	84.9	79.1	71.8	55.9
ε_2 [%]	76.5	72.5	62.4	51.0	32.1
ε_1^2 [%]	76.0	72.1	62.6	51.6	31.2

Table 4.21: Efficiencies of the requirements on the $\mathcal{M}_{PID}$ classifier response (the parameters ε_1 and ε_2 are defined in the text) obtained on $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ normalization sample in data collected in 2012.

$\mathcal{M}_{PID}$ cut	0.4	0.54	0.61	0.71	0.8
ε_1 [%]	86.7	77.0	71.1	60.5	47.3
ε_2 [%]	75.2	60.3	52.0	37.5	22.8
ε_1^2 [%]	76.2	59.3	50.6	36.6	22.3

Tables 4.20 and 4.21 list the requirement efficiencies measured on data (from PIDCalib package respectively), while Tables 4.22 and 4.23 give the efficiencies from PIDCalib package. It can be clearly seen that the numbers from PIDCalib are different when compared with the correct values evaluated from data. Furthermore, it is visible that the correlations between the two muons are negligibly small (i.e. $\varepsilon_1^2 \approx \varepsilon_2$). The same tendency is observed in the $\tau^- \to \mu^-\mu^+\mu^-$ simulation (Table 4.24 and 4.25). Therefore, a correction is applied to the PIDCalib package efficiencies for the $\tau^- \to \mu^-\mu^+\mu^-$ sample in the following way: for each cut value and year of data taking, the single particle efficiency ratio c is determined as the ratio of ε_1 from the cut-and-fit method over ε_1 from PIDCalib package. No dependence on momentum (p), pseudorapidity (η) or the occupancy of c is assumed. However, for each $\mathcal{M}_{PID}$ cut value an individual c factor is determined (it is visible that c is not constant as a function of $\mathcal{M}_{PID}$ variable).

Table 4.22: Efficiencies of the requirements on the $\mathcal{M}_{PID}$ classifier response (the parameters ε_1 and ε_2 are defined in the text) obtained from `PIDCalib` package for 2011 data taking conditions.

$\mathcal{M}_{PID}$ cut	0.4	0.45	0.54	0.63	0.75
ε_1 [%]	90.9	88.2	82.5	75.1	60.7
ε_2 [%]	82.3	77.8	68.2	56.6	37.2
ε_1^2 [%]	82.6	77.8	68.1	56.4	36.8
$c = \varepsilon_1(DATA)/\varepsilon_1(PIDCalib)$	0.959	0.963	0.959	0.956	0.921

Table 4.23: Efficiencies of the requirements on the $\mathcal{M}_{PID}$ classifier response (the parameters ε_1 and ε_2 are defined in the text) obtained from `PIDCalib` package for 2012 data taking conditions.

$\mathcal{M}_{PID}$ cut	0.4	0.54	0.61	0.71	0.8
ε_1 [%]	90.6	82.3	76.8	66.4	52.9
ε_2 [%]	82.2	67.8	59.0	44.2	28.2
ε_1^2 [%]	82.2	67.8	59.0	44.1	28.0
$c = \varepsilon_1(DATA)/\varepsilon_1(PIDCalib)$	0.957	0.936	0.926	0.911	0.894

Table 4.24: Efficiencies of the requirements on the $\mathcal{M}_{PID}$ classifier response for $\tau^- \to \mu^-\mu^+\mu^-$ simulated events in 2011 data taking conditions.

$\mathcal{M}_{PID}$ cut	0.4	0.45	0.54	0.63	0.75
ε_1 [%]	91.7	89.7	85.6	79.9	68.4
ε_2 [%]	83.9	80.3	72.9	63.8	46.9
ε_1^2 [%]	84.1	80.5	73.3	63.8	46.8
ε_3 [%]	76.9	72.2	62.5	51.3	32.8
ε_1^3 [%]	77.1	72.2	62.7	51.0	32.0

Table 4.25: Efficiencies of the requirements on the $\mathcal{M}_{PID}$ classifier response for $\tau^- \to \mu^-\mu^+\mu^-$ simulated events in 2012 data taking conditions.

$\mathcal{M}_{PID}$ cut	0.4	0.54	0.61	0.71	0.8
ε_1 [%]	89.5	82.9	78.2	69.7	57.4
ε_2 [%]	80.7	68.3	60.8	48.5	33.9
ε_1^2 [%]	80.1	68.7	61.2	48.6	33.0
ε_3 [%]	72.3	57.1	48.4	34.5	20.2
ε_1^3 [%]	71.6	56.9	47.9	33.8	19.0

The third power of c is used as a correction to calibration of $\mathcal{M}_{PID}$ variable. The resulting $\mathcal{M}_{PID}$ efficiencies are listed in Tables 4.26 and 4.27. Since `PIDCalib` package estimates an error of 0.5% for $\varepsilon_2(\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+)$, the cut-and-fit method

estimates an error of about 1.5% for the requirement: $\mathcal{M}_{PID} > 0.4$, and since the deviation of ε_n from ε_1^n is smaller than 1%, the conservative overall uncertainty of 0.02 was attributed to the factor c^3.

Table 4.26: Efficiencies for $\mathcal{M}_{PID}$ classifier response cuts for $\tau^- \to \mu^-\mu^+\mu^-$ simulated events in 2011 data taking conditions, after applying the correction described in the text.

$\mathcal{M}_{PID}$ cut	0.4	0.45	0.54	0.63	0.75
ε_1 [%]	89.7	86.9	80.9	73.2	58.6
ε_2 [%]	80.5	75.5	65.4	53.7	34.6
ε_1^2 [%]	80.5	75.5	65.4	53.6	34.3
ε_3 [%]	72.3	65.7	53.2	39.7	20.7
ε_1^3 [%]	72.2	65.6	52.9	39.2	20.1
$\varepsilon_1 \times c$ [%]	86.0	83.6	77.6	70.0	54.0
$\varepsilon_2 \times c^2$ [%]	74.0	70.0	60.1	49.1	29.3
$\varepsilon_1^2 \times c^2$ [%]	74.0	70.0	60.1	49.0	29.1
$\varepsilon_3 \times c^3$ [%]	63.8	58.6	46.9	34.7	16.2
$\varepsilon_1^3 \times c^3$ [%]	63.7	58.5	46.7	34.2	15.7

Table 4.27: Efficiencies for $\mathcal{M}_{PID}$ classifier response cuts for $\tau^- \to \mu^-\mu^+\mu^-$ simulated events in 2012 data taking conditions, after applying the correction described in the text.

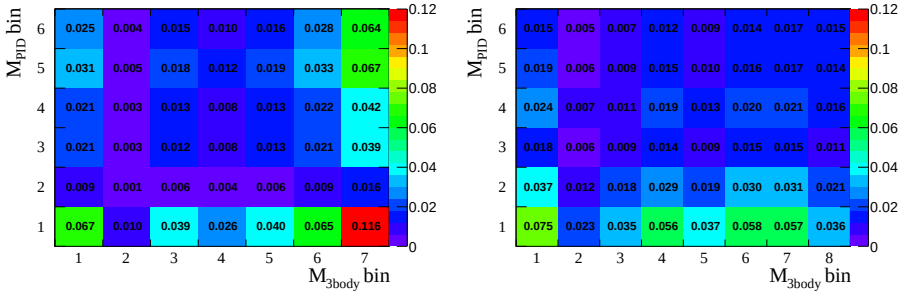
$\mathcal{M}_{PID}$ cut	0.4	0.54	0.61	0.71	0.8
ε_1 [%]	89.5	80.5	74.8	64.1	50.5
ε_2 [%]	80.2	65.2	56.4	41.6	26.0
ε_1^2 [%]	80.0	64.8	56.0	41.1	25.5
ε_3 [%]	71.8	52.7	42.4	27.1	13.6
ε_1^3 [%]	71.7	52.2	41.8	26.4	12.9
$\varepsilon_1 \times c$ [%]	85.6	75.4	69.2	58.4	45.1
$\varepsilon_2 \times c^2$ [%]	73.4	57.1	48.3	34.5	20.8
$\varepsilon_1^2 \times c^2$ [%]	73.3	56.8	48.0	34.1	20.3
$\varepsilon_3 \times c^3$ [%]	63.0	43.2	33.6	20.5	9.7
$\varepsilon_1^3 \times c^3$ [%]	62.8	42.8	33.2	19.9	9.2

The discrepancies observed between the efficiencies evaluated from data and those obtained from the `PIDCalib` package are due to the fact that the muon tracks used for calibration in the `PIDCalib` procedure originated from $B^+ \to J/\psi K^+$ decay, which covers a different fiducial region in the detector when compared with muons from $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ decay. When restricting ourselves to the common region of the two samples, we found `PIDCalib` in agreement with our fit-and-count method. The detector regions with less statistics of $B^+ \to J/\psi K^+$ give on average lower

efficiencies when compared to the ones determined from $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ data.

4.5.2.3 Calibration of the correlation between $\mathcal{M}_{3body}$ and $\mathcal{M}_{PID}$

Finally the amount of correlation between the $\mathcal{M}_{3body}$ and $\mathcal{M}_{PID}$ variables was evaluated in the following way. The calibration of $\mathcal{M}_{PID}$ variable obtained from the `MultiTrack` package is performed in the restricted range of $\mathcal{M}_{3body}$ classifier's response, corresponding to a single bin of this distribution for the input MC samples. Next this procedure is repeated for each individual bin of $\mathcal{M}_{3body}$ variable. This calibration is then applied to data yielding correlations in the two-dimensional signal $p.d.f.$ in $(\mathcal{M}_{3body}, \mathcal{M}_{PID})$ plane, as shown in Fig. 4.4.



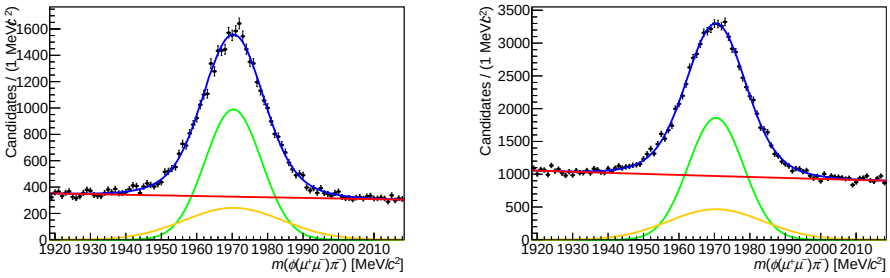
(a) 2011 running conditions (b) 2012 running conditions

Figure 4.4: The distribution of the $\mathcal{M}_{PID}$ variable vs the value of the classifier $\mathcal{M}_{3body}$ for 2011 (left) and 2012 (right) MC signal distribution after applying the calibration procedure (as described in the text).

4.5.3 Invariant mass of the $\mu\mu\mu$ system.

The shape of the invariant mass spectrum for the $\tau^- \to \mu^-\mu^+\mu^-$ signal is obtained from a fit to the respective $\mu^+\mu^-\pi^-$ mass distribution of the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ control sample, using the selection outlined in Sect. 4.3. The signal distribution is modelled with the sum of two Gaussian functions with a common central value, where the first one is constrained to be narrower than the second and the yield of the first Gaussian is fixed to be 70% of the total signal yield. The above restrictions were proved to enable a common fit model across all samples. If the fractional contribution of the Gaussians is also allowed to float, the variation in the signal efficiency is considered as a systematic uncertainty on the fit model. The combinatorial background is modelled with an exponential function.

The distributions of invariant mass $\phi(\mu^+\mu^-)\pi^-$ obtained from the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ normalization channel sample in 2011 and 2012 data together with the above described fits are shown in Fig. 4.5(a) and 4.5(b) respectively, and are summarized in Table 4.28. The values of N_{cal} (number of $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ events) include the systematic error due to the fit model, determined by treating the fraction of the first Gaussian distribution as a free fit parameter.



(a) 2011 data sample (b) 2012 data sample

Figure 4.5: Distribution of the $\mu^+\mu^-\pi^-$ invariant mass corresponding to the sample of normalization channel $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ after applying the trigger requirements and selection criteria described in the text. The left (right) plot corresponds to the 2011 (2012) data, respectively. The blue curve represents the results of the fit described in the text. The green and orange lines correspond to the respective Gaussian signals while the red line represents the combinatorial background

Table 4.28: Results of the fits to the invariant mass obtained from the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ normalization channel sample. See Table 4.42 for the corresponding $\tau^- \to \mu^-\mu^+\mu^-$ mass shape parameters.

Data Sample	2011	2012
N_{cal}	$28,207 \pm 417 \pm 138$	$52,131 \pm 619 \pm 317$
Mean (MeV/c^2)	1970.33 ± 0.09	1970.42 ± 0.07
σ_1 (MeV/c^2)	8.0 ± 0.1	7.82 ± 0.08
σ_2 (MeV/c^2)	13.9 ± 0.9	13.3 ± 0.6

The values of the D_s mass determined from the fit to data samples in 2011 and 2012 (given in Table 4.28) differ slightly from the respective masses determined by the fits to Monte Carlo distributions. Therefore, in the following considerations for the mass *p.d.f.* used in the final result evaluation (c.f. Sect. 4.9) we applied the central values of the D_s meson mass, as given by the fit to the distribution of the in $\tau^- \to \mu^-\mu^+\mu^-$ signal MC corrected by the ratio of data and MC central values obtained from the abovementioned studies of the normalization channel $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$. This procedure yielded (1779.1 ± 0.1) MeV/c^2 and (1779.0 ± 0.1) MeV/c^2 for 2011 and 2012 respectively. It is appropriate to mention here that for the fits to the $\tau^- \to \mu^-\mu^+\mu^-$ signal sample we used the τ lepton mass value as given by the PDG [23]. The $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ mass resolution in data is also found to be considerably different from that in MC and is corrected in the same way as the mean.

4.5.4 Optimization of binning in the variables $\mathcal{M}_{3body}$, $\mathcal{M}_{PID}$ and $m_{\mu\mu\mu}$

To evaluate the levels of consistency of the data sample with background b and signal plus background $s+b$ hypotheses, the samples of the selected signal candidates are analyzed in three dimensions: $\mathcal{M}_{3body}$, $\mathcal{M}_{PID}$ and $m_{\mu\mu\mu}$. The data are binned in the values of these variables; the choice of the binning is discussed in this section.

The optimization method encompasses both the number of bins and the position of the bin boundaries by maximizing the following figure of merit variable:

$$\Delta LQ = 2\ln(\mathrm{Q_{s+b}}) - 2\ln(\mathrm{Q_b})$$

where,

$$Q_{s+b} = \prod_i \frac{P(s_i+b_i, s_i+b_i)}{P(s_i+b_i, b_i)}, \tag{4.8}$$

$$Q_b = \prod_i \frac{P(b_i, s_i+b_i)}{P(b_i, b_i)}. \tag{4.9}$$

In the above, the symbol $P(a,e)$ denotes the probability to observe a events for a Poisson distribution with expectation value e, and the index i indicates that the probabilities are computed for each bin i. The s_i denote the numbers of signal events in bin i, while b_i refers to respective background events.

The binning is optimized in a two-dimensional plane ($\mathcal{M}_{3body}$, $\mathcal{M}_{PID}$) simultaneously. The bin boundaries of the lowest likelihood bins are not constrained to be zero for $\mathcal{M}_{3body}$ variable. Hence there is also a non-sensitive range from 0 to the lower bin boundary of the lowest sensitive $\mathcal{M}_{3body}$ bin called "trash bin", which is not counted as a bin by the optimization method. For $\mathcal{M}_{PID}$ variable we have put a constraint on the lowest bin at 0.4. This removes contributions from $D_s \to 3h$ background where two hadrons were misidentified as muons, which will be discussed in Sect. 4.7.

It was found that the optimization procedure yields the same outcome, irrespective of the particular choice of the binning in the $m_{\mu\mu\mu}$ mass. Thus, for simplicity, the binning w.r.t. the mass variable was chosen in such a way that the $\pm 20\,\mathrm{MeV}/c^2$ signal mass window has been centered around the PDG τ mass ($1776.82\,\mathrm{MeV}/c^2$) and was composed of eight bins of $5\,\mathrm{MeV}/c^2$ width.

The bin partitionings resulting from the above procedure are given in Table 4.29 and 4.30, for 2011 and 2012 data, respectively. The number of expected background events in each bin is provided in Table 4.36 and 4.37, for 2011 and 2012 data respectively.

4.6 Comparison of data and Monte Carlo distributions for the normalization decay channel

In this section we check the overall agreement betwen the relevant distributions of data and Monte Carlo samples corresponding to the samples selected in the search for

the normalization decay channel $D_s \to \phi(\mu\mu)\pi$. An unbinned maximum likelihood fit is performed to the $\phi(\mu^+\mu^-)\pi^-$ mass distribution in data, where the fit model is the same as described Sect. 4.5.3 and shown in Fig. 4.5.

Based on these fits, we performed comparisons of data and MC distributions for several variables which were used in the selection and were involved as input variables in the blending approach (cf Sect. 4.5.1.1 for definition of those variables). The signal-only distributions are obtained using the *sPlot* technique [105]. The distributions of those variables, presented in Fig. 4.6 and 4.7, show a reasonable agreement between data and MC samples. Some differences are observed in the cone isolation and BDT isolation variables, but this disagreement has no effect on the preselection efficiency as these variables are not used in the preselection.

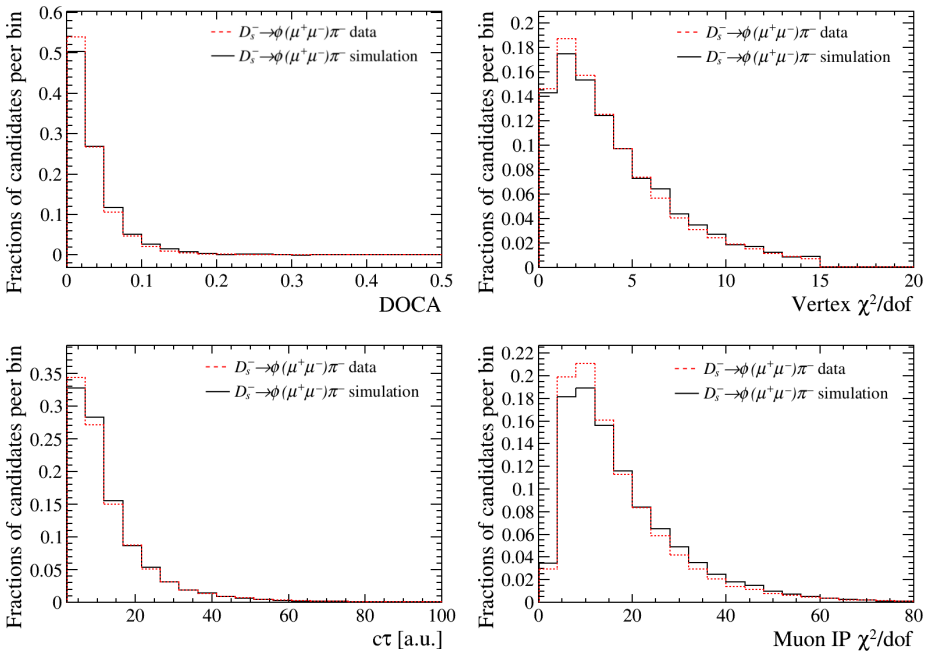


Figure 4.6: Distributions of the first four input variables relevant for the calculation of $\mathcal{M}_{3body}$ classifier for $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ signal events obtained with the *sPlot* technique and normalized to unit area. The variables are defined in Sect. 4.5.1.1. The dashed (red) line corresponds to data while the solid (black curve represents the Monte Carlo distribution.

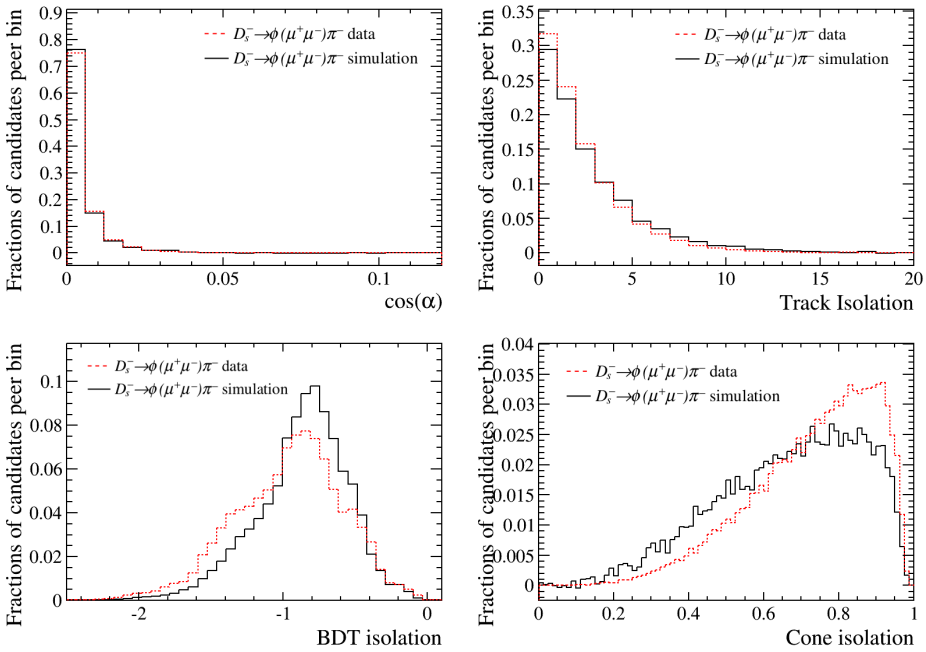


Figure 4.7: Distributions of the remaining four input variables relevant for the calculation of $\mathcal{M}_{3body}$ classifier for $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ signal events obtained with the *sPlot* technique and normalized to unit area. The variables are defined in Sect. 4.5.1.1. The dashed (red) line corresponds to data while the solid (black curve represents the Monte Carlo distribution.

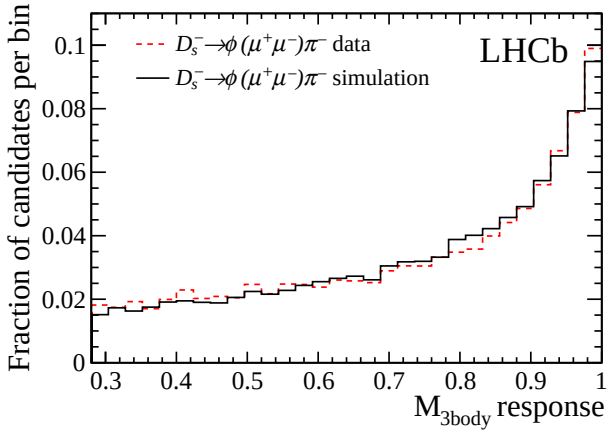


Figure 4.8: Distributions of $\mathcal{M}_{3body}$ classifier for $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ signal events obtained with the *sPlot* technique. The dashed (red) line corresponds to data while the solid (black curve represents the Monte Carlo distribution.

These particular variables have been studied here because they were used in the construction of $\mathcal{M}_{3body}$ classifier, where the data-MC discrepancy is corrected by

the calibration of $\mathcal{M}_{3body}$ variable on data. To conclude, it was checked that the variable $\mathcal{M}_{3body}$ (c.f. Fig. 4.8) exhibits a good overall agreement between data and Monte Carlo samples containing events from the decay $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ i.e. the normalization channel.

4.7 Background characterisation

The background processes for the rare decay $\tau^- \to \mu^-\mu^+\mu^-$ comprise long decay chains of heavy mesons with three muons in the final state as well as one or two muons in the combination with one or two mis-identified hadrons. Below these two categories of background contributions will be discussed in more detail.

4.7.1 Background processes with three muons in the final state

The main source of events with three muons of total charge ± 1 in the final state are semileptonic decays of D_s, D^+ and B mesons into $\mu\bar{\nu}_\mu$, in combination with a light unflavoured meson decaying into $\mu^+\mu^-$ or $\mu^+\mu^-\gamma$. Because of the significantly smaller production rate of B mesons in comparison to D_s and D^+ mesons, and two to three orders of magnitude smaller semileptonic decay rates of B mesons when compared with D_s and D^+ states (e.g.: $\mathcal{B}(D_s^+ \to \eta\mu^+\nu) = 2.67 \times 10^{-2}$ and $\mathcal{B}(B^+ \to \eta\mu^+\nu) = 3.7 \times 10^{-5}$), only decays of the D_s and D^+ mesons are taken into account when evaluating the backgrounds.

The branching fractions for the processes in which light unflavoured mesons decay into a pair of muons plus an additional photon or π^0 are between 10^{-4} and 10^{-6}. The total cross-sections of possible $D_s \to 3\mu + X$ decays are listed in Table 4.31. Additional contributions from $D_s^+ \to \omega\mu^+\nu_\mu, \omega \to \mu^+\mu^-\pi^0$ decays are expected to be small, since the branching ratio $D_s^+ \to \omega\mu^+\nu_\mu$ is expected to be of the order of 10^{-3} (it was not measured yet) as well as $\mathcal{B}(\omega \to \mu^+\mu^-\pi^0) = 1.3 \times 10^{-4}$ [23]. The total cross-sections of possible $D^+ \to 3\mu + X$ decays are collected in Table 4.32. Given the relative small size of the expected cross-sections when compared to those from D_s, and the shift in mass of approximately 100 MeV/c^2 between the D_s and D mesons, the backgrounds originating from these decays are assumed to be negligible.

Table 4.29: Bin partitioning of variables $m_{\mu\mu\mu}$, $\mathcal{M}_{3body}$ and $\mathcal{M}_{PID}$ used for 2011 data sample. The lowest likelihood bins in both $\mathcal{M}_{3body}$ and $\mathcal{M}_{PID}$ are not considered in the evaluation of the final result.

Bin number	Bin range	
	Lower limit	Upper limit
$m_{\mu\mu\mu}$ (MeV/c^2)		
1	1756.82	1761.82
2	1761.82	1766.82
3	1766.82	1771.82
4	1771.82	1776.82
5	1776.82	1781.82
6	1781.82	1786.82
7	1786.82	1791.82
8	1791.82	1796.82
$\mathcal{M}_{3body}$		
1	0	0.28
2	0.28	0.32
3	0.32	0.46
4	0.46	0.54
5	0.54	0.65
6	0.65	0.80
7	0.80	1.00
$\mathcal{M}_{PID}$		
1	0.00	0.40
2	0.40	0.45
3	0.45	0.54
4	0.54	0.63
5	0.63	0.75
6	0.75	1.00

Table 4.30: Bin partitioning of variables $m_{\mu\mu\mu}$, $\mathcal{M}_{3body}$ and $\mathcal{M}_{PID}$ used for 2012 data sample. The lowest bin in both $\mathcal{M}_{3body}$ and $\mathcal{M}_{PID}$ are not considered in the evaluation of the final result.

Bin number	Bin range	
	Lower limit	Upper limit
$m_{\mu\mu\mu}$ (MeV/c^2)		
1	1756.82	1761.82
2	1761.82	1766.82
3	1766.82	1771.82
4	1771.82	1776.82
5	1776.82	1781.82
6	1781.82	1786.82
7	1786.82	1791.82
8	1791.82	1796.82
$\mathcal{M}_{3body}$		
1	0.00	0.26
2	0.26	0.34
3	0.34	0.45
4	0.45	0.61
5	0.61	0.70
6	0.70	0.83
7	0.83	0.94
8	0.94	1.00
$\mathcal{M}_{PID}$		
1	0.00	0.40
2	0.40	0.54
3	0.54	0.61
4	0.61	0.71
5	0.71	0.80
6	0.80	1.00

Table 4.31: Branching fractions of trimuon final states from D_s^+ decays [23]. The expected production cross-sections are calculated by multiplying the measured, 4π solid angle production cross-section $\mathcal{B}(pp \to D_s + X) = 976\,\mu b$ with the corresponding detector acceptance efficiencies (determined from MC), which amounts to $\varepsilon_\eta = 0.183$ for decays into η or η' and $\varepsilon_\phi = 0.19$ for decays including a ϕ meson as intermediate state.

D_s^+ decay	$\mathcal{B}_1^{(*)}$	Secondary decays	$\mathcal{B}_2$	$\mathcal{B}_{tot} = \mathcal{B}_1 \times \mathcal{B}_2$	$\sigma(D_s \to 3\mu X)$
$\eta\mu^+\nu_\mu$	2.67×10^{-2}	$\eta \to \mu^+\mu^-$	5.8×10^{-6}	1.5×10^{-7}	0.03 nb
		$\eta \to \mu^+\mu^-\gamma$	3.1×10^{-4}	8.2×10^{-6}	1.5 nb
		$\eta \to \pi^0\mu^+\mu^-\gamma$	$< 3 \times 10^{-6}$	$< 8.0 \times 10^{-8}$	< 0.02 nb
$\eta'\mu^+\nu_\mu$	9.9×10^{-3}	$\eta' \to \mu^+\mu^-\gamma$	1.09×10^{-4}	1.1×10^{-6}	0.20 nb
$\phi\mu^+\nu_\mu$	2.49×10^{-2}	$\phi \to \mu^+\mu^-$	2.87×10^{-4}	7.1×10^{-6}	1.3 nb
		$\phi \to \mu^+\mu^-\gamma$	1.4×10^{-5}	3.5×10^{-7}	0.06 nb
		$\phi \to \mu^+\mu^-\pi^0$	1.12×10^{-5}(†)	2.8×10^{-7}	0.05 nb

(∗) : given branching ratios are assumed to be of the same value as the corresponding $e\nu_e$ decays.
(†) : given branching ratio are assumed to be of the same value as the corresponding $\phi \to e^+e^-\pi^0$ decay.

Table 4.32: Branching fractions of trimuon final states from D^+ decays [23]. The expected production cross-sections are calculated by multiplying the measured, 4π production cross-section $\mathcal{B}(pp \to D^+ + X) = 3.16\,mb$ with the corresponding detector acceptance efficiencies (determined from MC) $\varepsilon_\eta = 0.183$ for decays into η or η', $\varepsilon_\phi = 0.19$ for the decays including a ϕ meson as intermediate state, $\varepsilon_\omega = 0.20$ for decays into ω and $\varepsilon_\rho = 0.19$ for decays into ρ^0.

D^+ decay	$\mathcal{B}_1^{(*)}$	Secondary decays	$\mathcal{B}_2$	$\mathcal{B}_{tot} = \mathcal{B}_1 \times \mathcal{B}_2$	$\sigma(D^+ \to 3\mu X)$
$\eta\mu^+\nu_\mu$	1.14×10^{-3}	$\eta \to \mu^+\mu^-$	5.8×10^{-6}	6.6×10^{-9}	< 0.01 nb
		$\eta \to \mu^+\mu^-\gamma$	3.1×10^{-4}	3.5×10^{-7}	0.20 nb
		$\eta \to \pi^0\mu^+\mu^-\gamma$	$< 3 \times 10^{-6}$	$< 3.4 \times 10^{-9}$	< 0.01 nb
$\eta'\mu^+\nu_\mu$	2.2×10^{-4}	$\eta' \to \mu^+\mu^-\gamma$	1.09×10^{-4}	2.4×10^{-8}	0.01 nb
$\omega\mu^+\nu_\mu$	1.6×10^{-3}	$\omega \to \mu^+\mu^-$	9.0×10^{-5}	1.4×10^{-7}	0.09 nb
		$\omega \to \mu^+\mu^-\pi^0$	1.3×10^{-4}	2.1×10^{-7}	0.13 nb
$\rho^0\mu^+\nu_\mu$	2.4×10^{-3}	$\rho^0 \to \mu^+\mu^-$	4.55×10^{-5}	1.1×10^{-7}	0.07 nb
$\phi\mu^+\nu_\mu$	$< 9 \times 10^{-5}$	$\phi \to \mu^+\mu^-$	2.87×10^{-4}	2.6×10^{-8}	0.02 nb

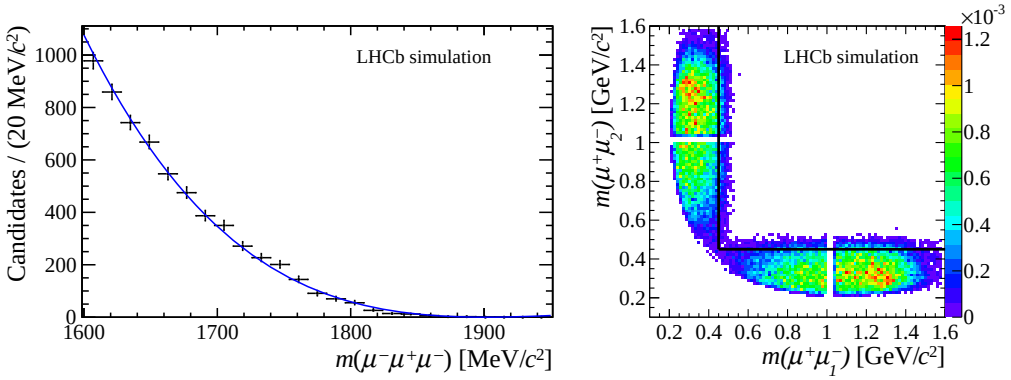
(∗) : given branching ratios are assumed to be of the same value as the corresponding $e\nu_e$ decays.

4.7.2 Background from $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decays

Special attention has been paid to the study of the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ background, which provides the largest contribution to the studies in question according to Table 4.31.

A study was performed using a specific simulation, which implemented the results of dimuon mass spectra distribution for the η radiative decay from [106, 107]. Using this model, over 10 million events have been generated. The simulated sample corresponds to over $40\,\text{fb}^{-1}$ of integrated luminosity.

The rejection of this background was optimized by varying the allowed range of the $\mu^+\mu^-$ invariant mass and evaluating the ratio $S/\sqrt{B}$ (S - signal, B - background). The optimal value of this selection criterium was found to be: $m_{\mu^+\mu^-} > 450$ MeV/c^2, which rejects $\sim 93\%$ of the $\mathrm{D_s^-} \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$, background while preserving 83% of the $\tau^- \to \mu^-\mu^+\mu^-$ signal (cf. Fig. 4.9).



(a) (b)

Figure 4.9: Relevant distributions of invariant masses obtained for the Monte Carlo sample containing the decay $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$. Left: trimuon invariant mass distribution. Right: Scatter plot of dimuon invariant masses $\mu^+\mu_2^-$ vs $\mu^+\mu_1^-$. The applied selection criteria are marked as black solid lines.

4.7.3 Backgrounds due to reflections

In this study, backgrounds due to reflections arise as a result of mis-identification of final state hadrons as muons. In the LHCb experiment pions and kaons can be mis-identified as muons with a mis-identification rate of the order of 2%, which depends on the hadron's momentum. Hadronic three-body charm mesons decays, which might yield a reflection to the decay studies, are summarized in Table 4.33. It is evident that the decays $D^+ \to K^-\pi^+\pi^+$ and $D^{*+} \to D^0(K^-\pi^+\pi^0)\pi^+$ are the dominant sources of the background in this category. However, $D^{*+} \to D^0(K^-\pi^+\pi^0)\pi^+$ decays are expected to fall out of range close to the τ lepton mass in the trimuon invariant mass distribution. This is motivated by the fact that the π^0 meson is not reconstructed in the above decay which causes a shift in the mass distribution. A clear peak of $D^+ \to K^-\pi^+\pi^+$ events is observed in the $\tau^- \to \mu^-\mu^+\mu^-$ data

sample (cf. Fig. 4.10). Using an unbinned likelihood fit to the $m_{K\pi\pi}$ hypothesis distribution, the corresponding yield of this background in 2011 data was evaluated as 4138 ± 310 (16052 ± 1003) in 2011 (2012) data, respectively.

Table 4.33: Charm decay modes which can potentially contribute to the background due to reflections. The number of events produced per 1 fb^{-1} of integrated luminosity in the LHCb acceptance, presented in the fourth column, is estimated using the LHCb measured inclusive D^+, D_s^+ and D^{*+} cross-sections in 4π solid angle (summed over charm and beauty contributions) and the respective PDG branching fraction and a 20% acceptance efficiency.

Decay channel	$\sigma(pp \to c\bar{c} + b\bar{b} \to ...)$[mb]	$\mathcal{B}$ (10^{-2})	$\mathcal{N}$ (10^9)	comments
$D^+ \to K^-\pi^+\pi^+$	3.30 ± 0.36	9.13 ± 0.19	60	main contribution
$D^+ \to K^-K^+\pi^+$	3.30 ± 0.36	0.954 ± 0.026	6.3	small peak visible
$D^+ \to \pi^+\pi^+\pi^-$	3.30 ± 0.36	0.318 ± 0.018	2.1	small peak visible
$D^+ \to K^-\pi^+\pi^+\pi^0$	3.30 ± 0.36	5.99 ± 0.18	40	peak at to low mass
$D^+ \to \pi^+\pi^+\pi^-\pi^0$	3.30 ± 0.36	1.13 ± 0.08	7.4	peak at to low mass
$D^+ \to K^-\pi^+\mu^+\nu^+$	3.30 ± 0.36	3.8 ± 0.4	25	peak at to low mass
$D_s^+ \to K^-K^+\pi^+$	1.11 ± 0.16	5.49 ± 0.27	12	small peak visible
$D_s^+ \to K^-\pi^+\pi^-$	1.11 ± 0.16	0.69 ± 0.05	1.5	negligible
$D_s^+ \to \pi^-\pi^+\pi^+$	1.11 ± 0.16	1.10 ± 0.06	2.4	small peak visible
$D_s^+ \to K^-K^+\pi^+\pi^0$	1.11 ± 0.16	5.6 ± 0.5	12	peak at to low mass
$D^{*+} \to D^0(K^-\pi^+\pi^0)\pi^+$	3.48 ± 0.41	9.41 ± 0.34	>1	peak at to low mass

In order to estimate the yield of this background in the $(\mathcal{M}_{3body}, \mathcal{M}_{PID})$ bins, we follow a data-driven approach. By exploiting the parametrization of the $D^+ \to K^-\pi^+\pi^+$ signal as shown in Fig. 4.10 and using *sPlot* [105] technique, we evaluated the yields of $D^+ \to K^-\pi^+\pi^+$ events in data in ($\mathcal{M}_{PID}$, $\mathcal{M}_{3body}$) bins. The results of this calculation are presented in Table 4.34 and 4.35.

We conclude that the yields of the background due to reflections are small and statistically insignificant outside of the lowest bin of $\mathcal{M}_{PID}$ variable. Therefore we decided to neglect the lowest bins when performing the fit to the sidebands of $\mu^-\mu^+\mu^-$ invariant mass.

4.7.4 Background estimate in the τ lepton mass region

The yield of background events in the τ lepton mass region is obtained from a set of extended unbinned likelihood fits to the $m_{\mu\mu\mu}$ distribution, corresponding to the sidebands of the τ lepton mass: $(1600, 1756.82) \cup (1806.82, 1950)\,\mathrm{MeV}/c^2$. The 2011 data sample is binned in $\mathcal{M}_{PID}$ and $\mathcal{M}_{3body}$ bins with 42 bins, as detailed in Table 4.29, while 2012 data sample is binned similarly but with 48 bins, as detailed in Table 4.30. We blind the region of ± 20 MeV/c^2 around the τ lepton mass, as given by the PDG [23] i.e. 1776.8 MeV/c^2, and remove the additional range in mass of ± 10 MeV/c^2 to test the classifiers and perform the binning optimization. Altogether the range of ± 30 MeV/c^2 around τ lepton nominal mass was excluded for the fit. As mentioned in Sect. 4.7.3, the reflection background is concentrated in the lowest bins of $\mathcal{M}_{PID}$ variable.

Table 4.34: The yields of $D^+ \to K^-\pi^+\pi^+$ decays in two-dimensional bins of the variables ($\mathcal{M}_{3body}$, $\mathcal{M}_{PID}$), obtained using *sPlot* technique for the data sample selected in search for the decay $\tau^- \to \mu^-\mu^+\mu^-$ in 2011. The quoted errors are those provided by the *sPlot* method.

N	$\mathcal{M}_{PID}$	$\mathcal{M}_{3body}$	Yield
1	(0 , 0.4)	(0 , 0.28)	648.36 ± 95.81
2	(0 , 0.4)	(0.28 , 0.32)	97.42 ± 18.58
3	(0 , 0.4)	(0.32 , 0.46)	396.68 ± 31.47
4	(0 , 0.4)	(0.46 , 0.54)	323.33 ± 24.94
5	(0 , 0.4)	(0.54 , 0.65)	619.87 ± 30.98
6	(0 , 0.4)	(0.65 , 0.8)	993.12 ± 38.18
7	(0 , 0.4)	(0.8 , 1.00)	1009.03 ± 38.64
8	(0.4 , 0.45)	(0 , 0.28)	4.64 ± 7.95
9	(0.4 , 0.45)	(0.28 , 0.32)	1.19 ± 1.33
10	(0.4 , 0.45)	(0.32 , 0.46)	0.06 ± 1.96
11	(0.4 , 0.45)	(0.46 , 0.54)	-0.72 ± 0.56
12	(0.4 , 0.45)	(0.54 , 0.65)	2.72 ± 2.30
13	(0.4 , 0.45)	(0.65 , 0.8)	-2.74 ± 1.38
14	(0.4 , 0.45)	(0.8 , 1.00)	3.30 ± 3.31
15	(0.45 , 0.54)	(0 , 0.28)	12.59 ± 10.43
16	(0.45 , 0.54)	(0.28 , 0.32)	-0.20 ± 1.11
17	(0.45 , 0.54)	(0.32 , 0.46)	3.19 ± 2.55
18	(0.45 , 0.54)	(0.46 , 0.54)	1.41 ± 1.41
19	(0.45 , 0.54)	(0.54 , 0.65)	-0.48 ± 1.67
20	(0.45 , 0.54)	(0.65 , 0.8)	1.85 ± 2.08
21	(0.45 , 0.54)	(0.8 , 1.00)	1.81 ± 2.86
22	(0.54 , 0.63)	(0 , 0.28)	16.76 ± 7.80
23	(0.54 , 0.63)	(0.28 , 0.32)	-1.51 ± 1.11
24	(0.54 , 0.63)	(0.32 , 0.46)	-2.77 ± 1.99
25	(0.54 , 0.63)	(0.46 , 0.54)	1.99 ± 1.87
26	(0.54 , 0.63)	(0.54 , 0.65)	1.13 ± 2.09
27	(0.54 , 0.63)	(0.65 , 0.8)	-0.43 ± 0.98
28	(0.54 , 0.63)	(0.8 , 1.00)	3.95 ± 2.58
29	(0.63 , 0.75)	(0 , 0.28)	4.57 ± 7.34
30	(0.63 , 0.75)	(0.28 , 0.32)	1.41 ± 1.35
31	(0.63 , 0.75)	(0.32 , 0.46)	-0.06 ± 2.26
32	(0.63 , 0.75)	(0.46 , 0.54)	-2.07 ± 0.96
33	(0.63 , 0.75)	(0.54 , 0.65)	0.43 ± 0.33
34	(0.63 , 0.75)	(0.65 , 0.8)	-1.06 ± 0.75
35	(0.63 , 0.75)	(0.8 , 1.00)	-1.83 ± 1.61
36	(0.75 , 1.00)	(0 , 0.28)	-5.38 ± 5.18
37	(0.75 , 1.00)	(0.28 , 0.32)	-0.45 ± 0.45
38	(0.75 , 1.00)	(0.32 , 0.46)	1.88 ± 2.47
39	(0.75 , 1.00)	(0.46 , 0.54)	3.75 ± 1.97
40	(0.75 , 1.00)	(0.54 , 0.65)	0.30 ± 0.94
41	(0.75 , 1.00)	(0.65 , 0.8)	0.94 ± 1.64
42	(0.75 , 1.00)	(0.8 , 1.00)	0.78 ± 2.29

Table 4.35: The yields of $D^+ \to K^-\pi^+\pi^+$ decays in two-dimensional bins of the variables ($\mathcal{M}_{3body}$, $\mathcal{M}_{PID}$), obtained using *sPlot* technique for the data sample selected in search for the decay $\tau^- \to \mu^-\mu^+\mu^-$ in 2012. The quoted errors are those provided by the *sPlot* method.

N	$\mathcal{M}_{PID}$	$\mathcal{M}_{3body}$	Yield
1	(0 , 0.4)	(0 , 0.26)	3397.72 ± 202.73
2	(0 , 0.4)	(0.26 , 0.34)	756.26 ± 47.98
3	(0 , 0.4)	(0.34 , 0.45)	1647.61 ± 55.90
4	(0 , 0.4)	(0.45 , 0.61)	2992.73 ± 68.85
5	(0 , 0.4)	(0.61 , 0.7)	2031.90 ± 54.38
6	(0 , 0.4)	(0.7 , 0.83)	2802.14 ± 63.73
7	(0 , 0.4)	(0.83 , 0.94)	1845.58 ± 51.30
8	(0 , 0.4)	(0.94 , 1.00)	591.10 ± 29.35
9	(0.4 , 0.54)	(0 , 0.26)	-15.82 ± 21.21
10	(0.4 , 0.54)	(0.26 , 0.34)	-1.06 ± 4.92
11	(0.4 , 0.54)	(0.34 , 0.45)	-7.37 ± 3.95
12	(0.4 , 0.54)	(0.45 , 0.61)	-4.72 ± 4.07
13	(0.4 , 0.54)	(0.61 , 0.7)	-0.69 ± 3.28
14	(0.4 , 0.54)	(0.7 , 0.83)	5.94 ± 3.78
15	(0.4 , 0.54)	(0.83 , 0.94)	8.02 ± 4.45
16	(0.4 , 0.54)	(0.94 , 1.00)	0.69 ± 2.42
17	(0.54 , 0.61)	(0 , 0.26)	18.19 ± 12.36
18	(0.54 , 0.61)	(0.26 , 0.34)	-2.28 ± 2.43
19	(0.54 , 0.61)	(0.34 , 0.45)	-0.93 ± 2.76
20	(0.54 , 0.61)	(0.45 , 0.61)	-1.33 ± 2.31
21	(0.54 , 0.61)	(0.61 , 0.7)	1.03 ± 1.64
22	(0.54 , 0.61)	(0.7 , 0.83)	0.59 ± 0.84
23	(0.54 , 0.61)	(0.83 , 0.94)	-3.54 ± 1.93
24	(0.54 , 0.61)	(0.94 , 1.00)	1.90 ± 1.81
25	(0.61 , 0.71)	(0 , 0.26)	-1.60 ± 11.56
26	(0.61 , 0.71)	(0.26 , 0.34)	-1.93 ± 2.48
27	(0.61 , 0.71)	(0.34 , 0.45)	-1.26 ± 2.70
28	(0.61 , 0.71)	(0.45 , 0.61)	-3.81 ± 2.35
29	(0.61 , 0.71)	(0.61 , 0.7)	2.20 ± 2.30
30	(0.61 , 0.71)	(0.7 , 0.83)	3.86 ± 3.10
31	(0.61 , 0.71)	(0.83 , 0.94)	1.00 ± 2.40
32	(0.61 , 0.71)	(0.94 , 1.00)	5.37 ± 2.96
33	(0.71 , 0.8)	(0 , 0.26)	-11.08 ± 8.20
34	(0.71 , 0.8)	(0.26 , 0.34)	2.94 ± 2.46
35	(0.71 , 0.8)	(0.34 , 0.45)	0.65 ± 0.65
36	(0.71 , 0.8)	(0.45 , 0.61)	0.51 ± 2.29
37	(0.71 , 0.8)	(0.61 , 0.7)	0.71 ± 2.31
38	(0.71 , 0.8)	(0.7 , 0.83)	0.80 ± 2.12
39	(0.71 , 0.8)	(0.83 , 0.94)	-0.40 ± 1.56
40	(0.71 , 0.8)	0.94 , 1.00)	1.58 ± 2.35
41	(0.8 , 1.00)	(0 , 0.26)	-4.52 ± 6.80
42	(0.8 , 1.00)	(0.26 , 0.34)	0.53 ± 0.44
43	(0.8 , 1.00)	(0.34 , 0.45)	-1.96 ± 0.98
44	(0.8 , 1.00)	(0.45 , 0.61)	-3.98 ± 1.54
45	(0.8 , 1.00)	(0.61 , 0.7)	1.48 ± 2.62
46	(0.8 , 1.00)	(0.7 , 0.83)	-3.17 ± 1.45
47	(0.8 , 1.00)	(0.83 , 0.94)	0.32 ± 1.72
48	(0.8 , 1.00)	(0.94 , 1.00)	-2.36 ± 1.07

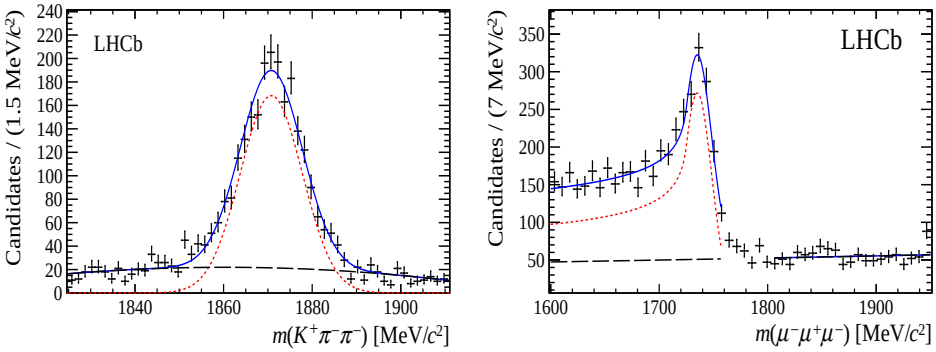
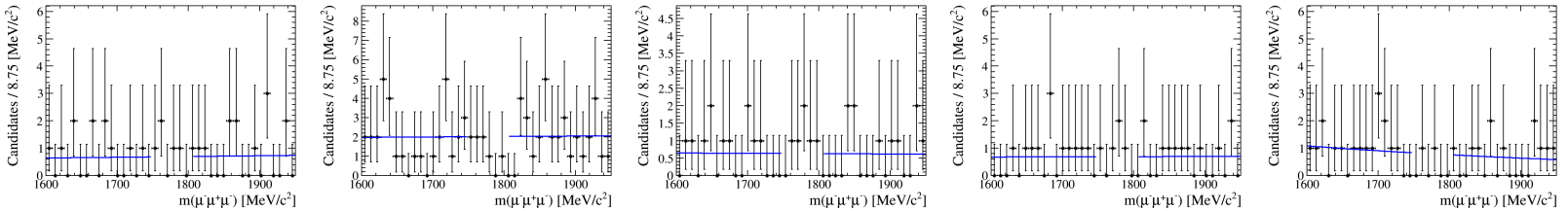


Figure 4.10: The invariant mass distribution of the $K^+\pi^-\pi^+$ system (left) and the respective distribution of reflection, assuming $\mu^-\mu^+\mu^-$ mass assignment (right). The signal distribution (red-dotted curve) is fitted with a Gaussian (Crystal Ball) function on the left (right) plot, respectively. The background is shown as dark black-bashed lines while the overall fit parametrization is marked with blue-solid line.

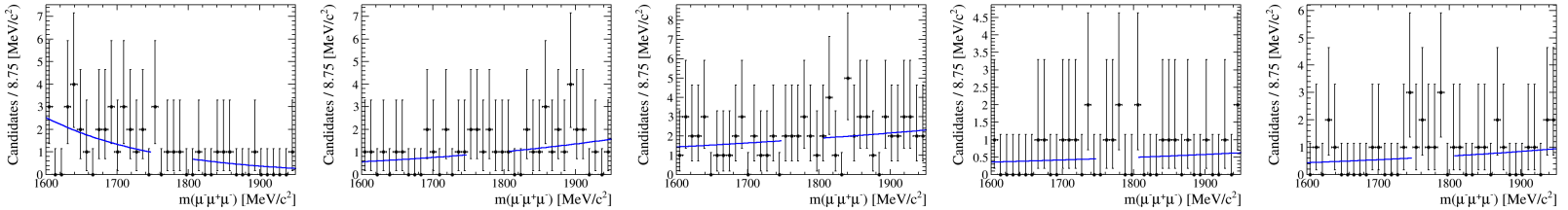
4.7.5 Background estimate with $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$ veto

As described in Sect. 4.7.2, one can veto the $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$with the requirement on dimuon invariant mass. This enabled us to apply a simple fit strategy relying on just one exponential $p.d.f.$ to account for the combinatorial background in the sensitive bins of $\mathcal{M}_{PID}$ and $\mathcal{M}_{3body}$ variables, as detailed in Sect. 4.5.4. These fits were performed in the sidebands of the trimuon mass spectrum and their results are presented in Fig. 4.11–4.14.

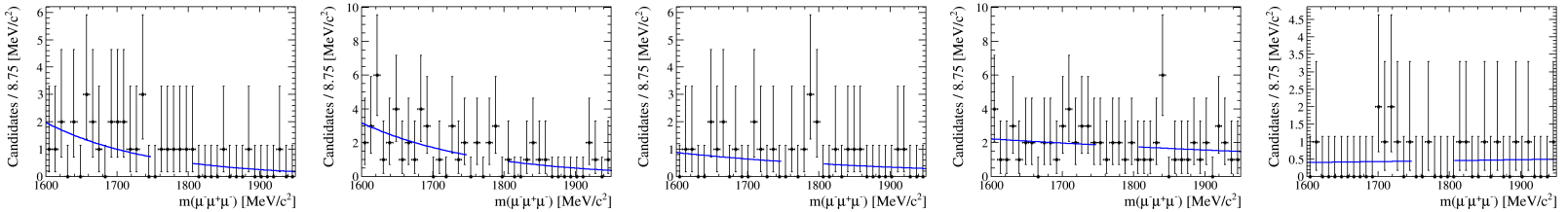
The results of the above procedure were used in the determination of the expected number of background events in each of the eight mass bins in signal window, i.e. $(1756.82 - 1796.82)$ MeV/c^2. The total numbers of the expected background events for all bins are given, with their uncertainties in Tables 4.36 and 4.37. The observed number of events is also included in these tables and will be described in more detail in Sect. 4.9.



$\mathcal{M}_{3body}$: 0.28 – 0.32 $\mathcal{M}_{PID}$: 0.4 – 0.45 | $\mathcal{M}_{3body}$: 0.32 – 0.46 $\mathcal{M}_{PID}$: 0.4 – 0.45 | $\mathcal{M}_{3body}$: 0.46 – 0.54 $\mathcal{M}_{PID}$: 0.4 – 0.45 | $\mathcal{M}_{3body}$: 0.54 – 0.65 $\mathcal{M}_{PID}$: 0.4 – 0.45 | $\mathcal{M}_{3body}$: 0.65 – 0.8 $\mathcal{M}_{PID}$: 0.4 – 0.45



$\mathcal{M}_{3body}$: 0.8 – 1.0 $\mathcal{M}_{PID}$: 0.4 – 0.45 | $\mathcal{M}_{3body}$: 0.28 – 0.32 $\mathcal{M}_{PID}$: 0.45 – 0.54 | $\mathcal{M}_{3body}$: 0.32 – 0.46 $\mathcal{M}_{PID}$: 0.45 – 0.54 | $\mathcal{M}_{3body}$: 0.46 – 0.54 $\mathcal{M}_{PID}$: 0.45 – 0.54 | $\mathcal{M}_{3body}$: 0.54 – 0.65 $\mathcal{M}_{PID}$: 0.45 – 0.54



$\mathcal{M}_{3body}$: 0.65 – 0.8 $\mathcal{M}_{PID}$: 0.45 – 0.54 | $\mathcal{M}_{3body}$: 0.8 – 1.0 $\mathcal{M}_{PID}$: 0.45 – 0.54 | $\mathcal{M}_{3body}$: 0.28 – 0.32 $\mathcal{M}_{PID}$: 0.54 – 0.63 | $\mathcal{M}_{3body}$: 0.32 – 0.46 $\mathcal{M}_{PID}$: 0.54 – 0.63 | $\mathcal{M}_{3body}$: 0.46 – 0.54 $\mathcal{M}_{PID}$: 0.54 – 0.63

Figure 4.11: Distributions of tri-muon invariant mass (data points) in bins of $(\mathcal{M}_{3body}, \mathcal{M}_{PID})$ variables for the $\tau^- \to \mu^-\mu^+\mu^-$ data in 2011 after $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$veto. The blue-solid line shows the results of the fit to an exponential *p.d.f.* performed in the sidebands of the tri-muon invariant mass. Continued in Fig. 4.12.

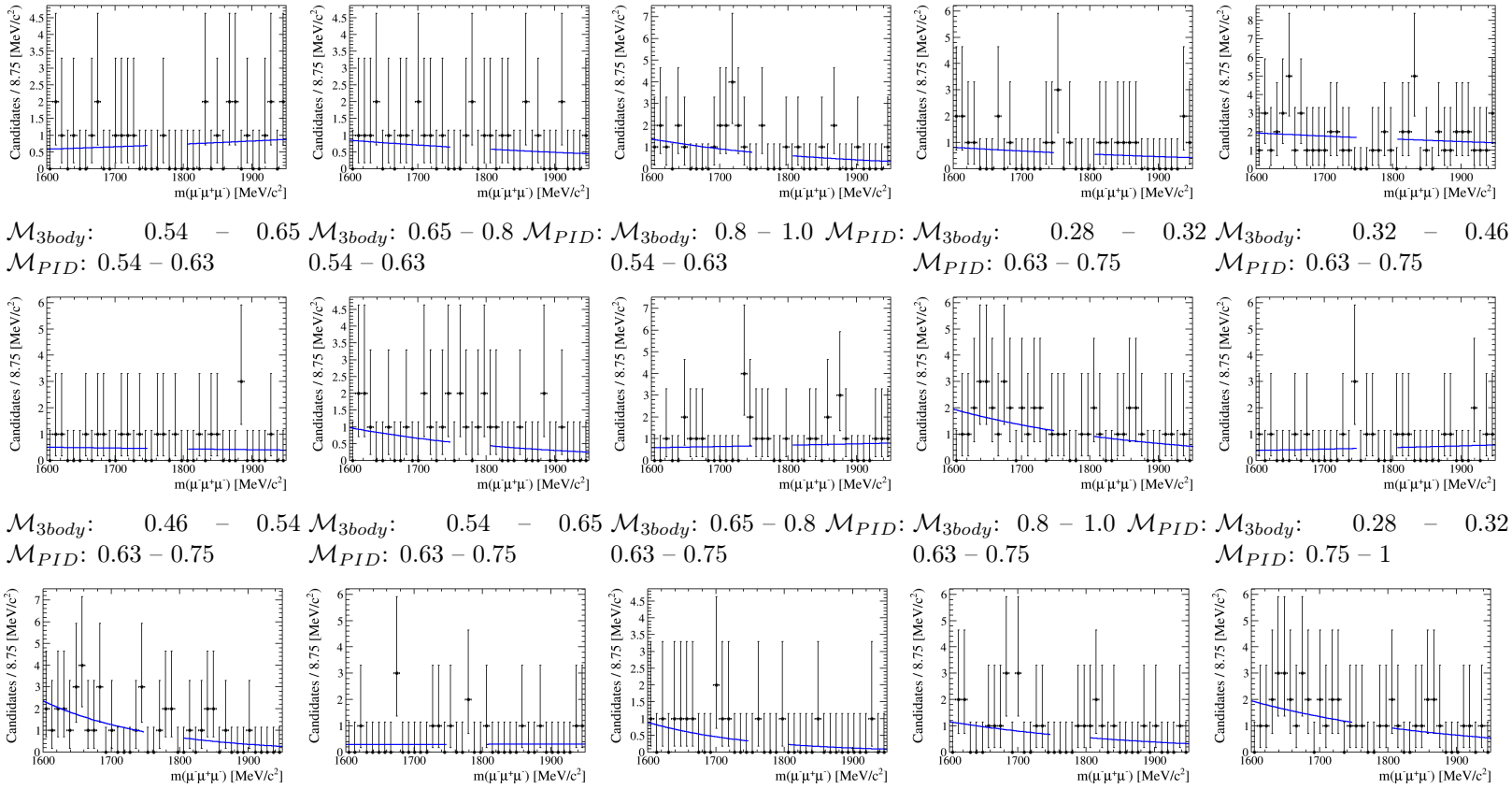
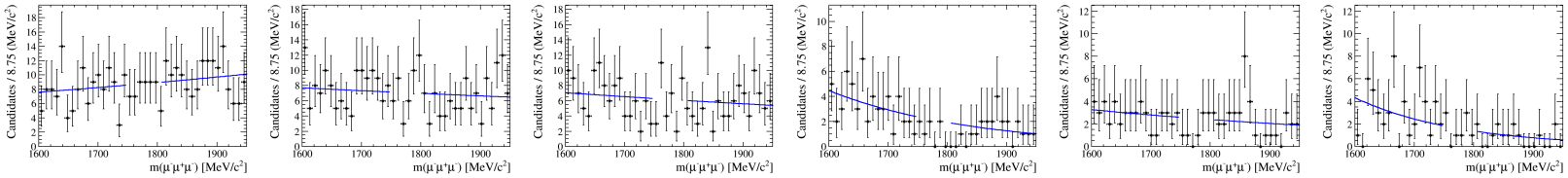
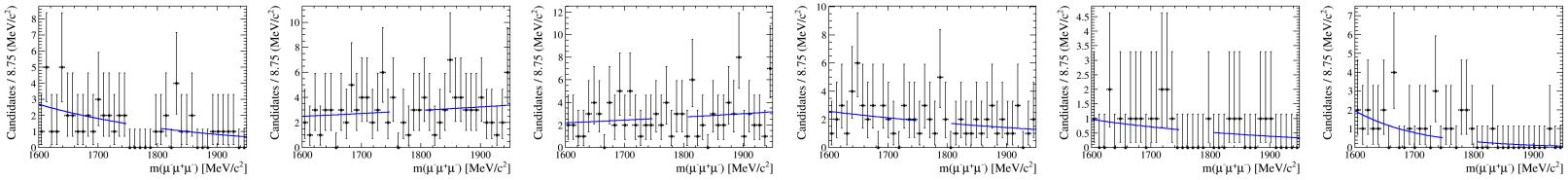


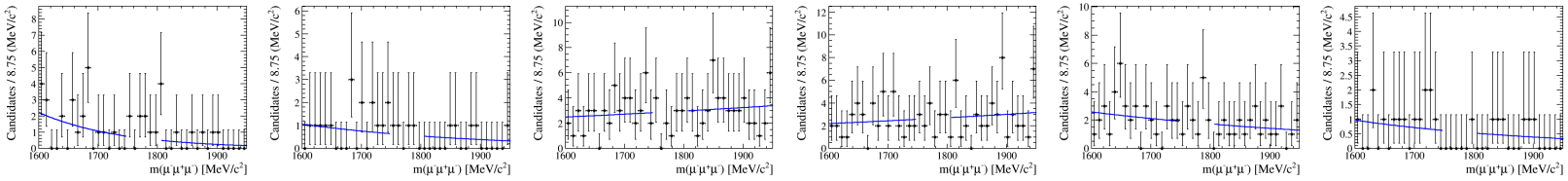
Figure 4.12: Distributions of tri-muon invariant mass (data points) in bins of $(\mathcal{M}_{3body}, \mathcal{M}_{PID})$ variables for the $\tau^- \to \mu^-\mu^+\mu^-$ data in 2011 after $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$veto. The blue-solid line shows the results of the fit to an exponential *p.d.f.* performed in the sidebands of the tri-muon invariant mass.



$\mathcal{M}_{3body}$: 0.26 – 0.34 $\mathcal{M}_{3body}$: 0.34 – 0.45 $\mathcal{M}_{3body}$: 0.45 – 0.61 $\mathcal{M}_{3body}$: 0.61 – 0.71 $\mathcal{M}_{3body}$: 0.71 – 0.8 $\mathcal{M}_{3body}$: 0.83 – 0.94
$\mathcal{M}_{PID}$: 0.4 – 0.45 $\mathcal{M}_{PID}$: 0.4 – 0.45 $\mathcal{M}_{PID}$: 0.4 – 0.45 $\mathcal{M}_{PID}$: 0.4 – 0.45 $\mathcal{M}_{PID}$: 0.4 – 0.45 $\mathcal{M}_{PID}$: 0.4 – 0.45



$\mathcal{M}_{3body}$: 0.94 – 1.00 $\mathcal{M}_{3body}$: 0.26 – 0.34 $\mathcal{M}_{3body}$: 0.34 – 0.45 $\mathcal{M}_{3body}$: 0.45 – 0.61 $\mathcal{M}_{3body}$: 0.61 – 0.71 $\mathcal{M}_{3body}$: 0.71 – 0.8
$\mathcal{M}_{PID}$: 0.4 – 0.45 $\mathcal{M}_{PID}$: 0.45 – 0.54 $\mathcal{M}_{PID}$: 0.45 – 0.54 $\mathcal{M}_{PID}$: 0.45 – 0.54 $\mathcal{M}_{PID}$: 0.45 – 0.54 $\mathcal{M}_{PID}$: 0.45 – 0.54



$\mathcal{M}_{3body}$: 0.83 – 0.94 $\mathcal{M}_{3body}$: 0.94 – 1.00 $\mathcal{M}_{3body}$: 0.26 – 0.34 $\mathcal{M}_{3body}$: 0.34 – 0.45 $\mathcal{M}_{3body}$: 0.45 – 0.61 $\mathcal{M}_{3body}$: 0.61 – 0.71
$\mathcal{M}_{PID}$: 0.45 – 0.54 $\mathcal{M}_{PID}$: 0.45 – 0.54 $\mathcal{M}_{PID}$: 0.54 – 0.61 $\mathcal{M}_{PID}$: 0.54 – 0.61 $\mathcal{M}_{PID}$: 0.54 – 0.61 $\mathcal{M}_{PID}$: 0.54 – 0.61

Figure 4.13: Distributions of tri-muon invariant mass (data points) in bins of $(\mathcal{M}_{3body}, \mathcal{M}_{PID})$ variables for the $\tau^- \to \mu^-\mu^+\mu^-$ data in 2012 after $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$veto. The blue-solid line shows the results of the fit to an exponential $p.d.f.$ performed in the sidebands of the tri-muon invariant mass Continued in Fig. 4.14.

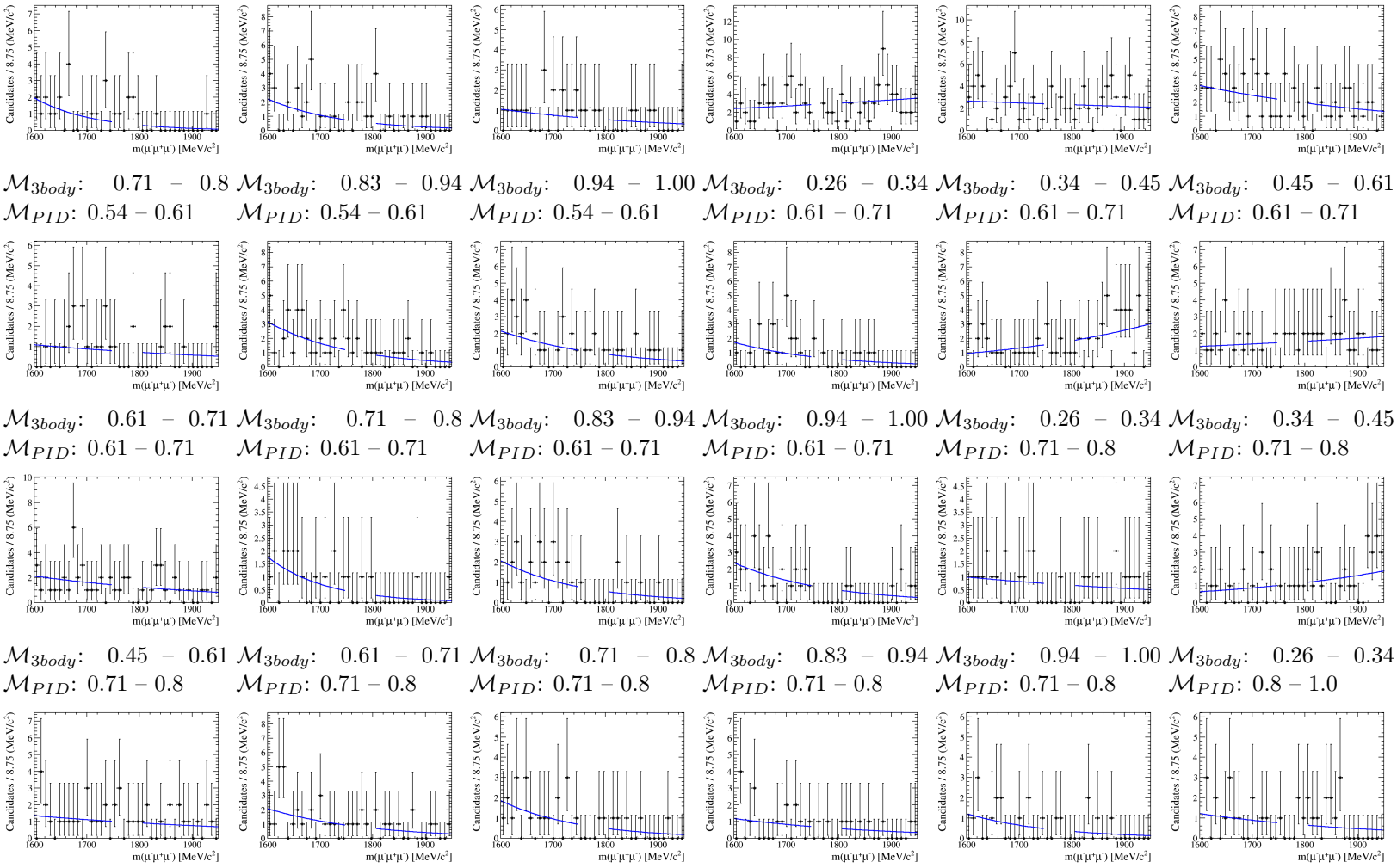


Figure 4.14: Distributions of tri-muon invariant mass (data points) in bins of $(\mathcal{M}_{3body}, \mathcal{M}_{PID})$ variables for the $\tau^- \to \mu^-\mu^+\mu^-$ data in 2012 after $\mathrm{D}_\mathrm{s}^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$veto. The blue-solid line shows the results of the fit to an exponential $p.d.f.$ performed in the sidebands of the tri-muon invariant mass

Table 4.36: Numbers of estimated background events and numbers of observed events in bins of $\mathcal{M}_{PID}$ and $\mathcal{M}_{3body}$ variables obtained from the fits to the sidebands of the τ lepton mass in 2011 data.

$\mathcal{M}_{PID}$	$\mathcal{M}_{3body}$	Estimated	Observed
0.4, 0.45	0.28, 0.32	3.172 ± 0.661	4
0.4, 0.45	0.32, 0.46	9.242 ± 1.129	6
0.4, 0.45	0.46, 0.54	2.894 ± 0.632	6
0.4, 0.45	0.54, 0.65	3.173 ± 0.661	4
0.4, 0.45	0.65, 0.80	3.637 ± 0.716	2
0.4, 0.45	0.80, 1.0	3.787 ± 0.802	3
0.45, 0.54	0.28, 0.32	4.223 ± 0.779	6
0.45, 0.54	0.32, 0.46	8.345 ± 1.077	10
0.45, 0.54	0.46, 0.54	2.317 ± 0.568	4
0.45, 0.54	0.54, 0.65	2.828 ± 0.632	8
0.45, 0.54	0.65, 0.80	2.718 ± 0.688	5
0.45, 0.54	0.80, 1.00	4.825 ± 0.900	7
0.54, 0.63	0.28, 0.32	2.327 ± 0.584	6
0.54, 0.63	0.32, 0.46	8.324 ± 1.077	8
0.54, 0.63	0.46, 0.54	2.068 ± 0.534	1
0.54, 0.63	0.54, 0.65	3.291 ± 0.675	1
0.54, 0.63	0.65, 0.80	2.962 ± 0.646	4
0.54, 0.63	0.80, 1.00	3.114 ± 0.687	3
0.63, 0.75	0.28, 0.32	2.688 ± 0.616	1
0.63, 0.75	0.32, 0.46	7.541 ± 1.023	5
0.63, 0.75	0.46, 0.54	2.059 ± 0.534	3
0.63, 0.75	0.54, 0.65	1.996 ± 0.549	5
0.63, 0.75	0.65, 0.80	3.164 ± 0.661	2
0.63, 0.75	0.80, 1.00	4.674 ± 0.836	2
0.75, 1.00	0.28, 0.32	2.192 ± 0.551	2
0.75, 1.00	0.32, 0.46	3.384 ± 0.755	5
0.75, 1.00	0.46, 0.54	1.517 ± 0.457	3
0.75, 1.00	0.54, 0.65	1.280 ± 0.469	1
0.75, 1.00	0.65, 0.80	2.780 ± 0.645	1
0.75, 1.00	0.80, 1.00	4.421 ± 0.833	7

Table 4.37: Numbers of estimated background events and numbers of observed events in bins of $\mathcal{M}_{PID}$ and $\mathcal{M}_{3body}$ variables obtained from the fits to the sidebands of the τ lepton mass in 2012 data.

$\mathcal{M}_{PID}$	$\mathcal{M}_{3body}$	Estimated	Observed
0.4, 0.54	0.26, 0.34	39.6 ± 2.3	39
0.4, 0.54	0.34, 0.45	32.2 ± 2.1	34
0.4, 0.54	0.45, 0.61	28.7 ± 2.0	28
0.4, 0.54	0.61, 0.7	9.72 ± 1.22	5
0.4, 0.54	0.7, 0.83	11.38 ± 1.26	7
0.4, 0.54	0.83, 0.94	7.34 ± 1.10	6
0.4, 0.54	0.94, 1.00	5.98 ± 0.95	0
0.54, 0.61	0.26, 0.34	13.6 ± 1.37	8
0.54, 0.61	0.34, 0.45	12.1 ± 1.29	12
0.54, 0.61	0.45, 0.61	8.32 ± 1.086	13
0.54, 0.61	0.61, 0.7	2.595 ± 0.616	1
0.54, 0.61	0.7, 0.83	1.833 ± 0.601	5
0.54, 0.61	0.83, 0.94	2.929 ± 0.724	6
0.54, 0.61	0.94, 1.00	2.693 ± 0.632	3
0.61, 0.71	0.26, 0.34	13.457 ± 1.366	7
0.61, 0.71	0.34, 0.45	10.852 ± 1.23	11
0.61, 0.71	0.45, 0.61	9.661 ± 1.18	12
0.61, 0.71	0.61, 0.7	3.346 ± 0.69	2
0.61, 0.71	0.7, 0.83	4.600 ± 0.888	5
0.61, 0.71	0.83, 0.94	4.091 ± 0.809	4
0.61, 0.71	0.94, 1.00	2.780 ± 0.680	1
0.71, 0.8	0.26, 0.34	7.808 ± 1.067	6
0.71, 0.8	0.34, 0.45	7.001 ± 0.985	8
0.71, 0.8	0.45, 0.61	6.170 ± 0.945	6
0.71, 0.8	0.61, 0.7	1.570 ± 0.556	2
0.71, 0.8	0.7, 0.83	2.987 ± 0.717	0
0.71, 0.8	0.83, 0.94	3.929 ± 0.806	0
0.71, 0.8	0.94, 1.00	3.222 ± 0.676	1
0.8, 1.00	0.26, 0.34	5.123 ± 0.861	3
0.8, 1.00	0.34, 0.45	4.435 ± 0.792	6
0.8, 1.00	0.45, 0.61	3.802 ± 0.784	5
0.8, 1.00	0.61, 0.7	2.649 ± 0.676	2
0.8, 1.00	0.7, 0.83	3.053 ± 0.674	2
0.8, 1.00	0.83, 0.94	1.740 ± 0.543	2
0.8, 1.00	0.94, 1.00	3.361 ± 0.702	3

4.8 Normalization of signal yield

To estimate the signal branching fraction for $\tau^- \to \mu^-\mu^+\mu^-$ we normalize the number of observed signal events N_{sig} to the number of events in the $\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+$

calibration channel. This procedure is performed according to the following formula:

$$\begin{aligned}
&\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) \\
&\quad = \mathcal{B}(\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+) \times \frac{f^\tau_{D_s}}{\mathcal{B}(\mathrm{D_s}^- \to \tau^-\overline{\nu}_\tau)} \times \frac{\varepsilon_{\mathrm{cal}}{}^{\mathrm{REC\&SEL}}\varepsilon_{\mathrm{cal}}{}^{\mathrm{TRIG|SEL}}}{\varepsilon_{\mathrm{sig}}{}^{\mathrm{REC\&SEL}}\varepsilon_{\mathrm{sig}}{}^{\mathrm{TRIG|SEL}}} \times \frac{N_{\mathrm{sig}}}{N_{\mathrm{cal}}} \\
&\quad = \alpha \times N_{\mathrm{sig}}\,, \qquad (4.10)
\end{aligned}$$

where α is the overall normalization factor. The value $\mathcal{B}(\mathrm{D_s}^- \to \tau^-\overline{\nu}_\tau)$ is the branching fraction for $\mathrm{D_s}^- \to \tau^-\overline{\nu}_\tau$ decay taken from [108, 109]. The quantity $f^\tau_{D_s}$ is the fraction of τ leptons originating from $\mathrm{D_s}$ decays and it was calculated using the $b\bar{b}$ and $c\bar{c}$ production cross-sections as measured by LHCb [93, 94]. Here we employed the values of the inclusive $\mathrm{b} \to \tau$ and $\mathrm{c} \to \tau$ branching fractions as measured by the LEP experiments [23]. The remaining factors presented in Eq. 4.10 (efficiencies) will be described below in Sect. 4.8.1.

Table 4.38: Branching fractions of decays relevant for the normalization of the signal yield [23]. The value of $\mathcal{B}(\mathrm{D_s}^- \to \mathrm{K^+K^-}\pi^-)$ is taken from a weighted average of the measurements from CLEO [110], BaBar [111] and Belle [112]. It is then multiplied by the branching fraction of the decay $\mathrm{D_s}^+ \to \phi\pi^+$ as measured by BaBar [113] to give $\mathcal{B}(\mathrm{D_s}^- \to \phi(\mathrm{K^+K^-})\pi^-)$. The value of $\mathcal{B}(\mathrm{D_s}^- \to \tau^-\overline{\nu}_\tau)$ is taken from [108] (c.f. Eq. 4.11).

Decay	Branching fraction
$\mathrm{D_s}^- \to \phi(\mathrm{K^+K^-})\pi^-$	$(2.244 \pm 0.076) \times 10^{-2}$
$\phi \to \mathrm{K^+K^-}$	$(48.9 \pm 0.5) \times 10^{-2}$
$\phi \to \mu^+\mu^-$	$(2.87 \pm 0.19) \times 10^{-4}$
$\mathrm{D_s}^- \to \tau^-\overline{\nu}_\tau$	$(5.61 \pm 0.24) \times 10^{-2}$

The branching fractions used in the normalization of the signal yield are listed in Table 4.38. To calculate the $\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+$ branching fraction we use the equation:

$$\mathcal{B}(\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+) = \frac{\mathcal{B}(D_s^+ \to \phi\pi^+, \phi \to K^+K^-)}{\mathcal{B}(\phi \to \mathrm{K^+K^-})}\mathcal{B}(\phi \to \mu^+\mu^-), \qquad (4.11)$$

where only the $\phi \to \mathrm{K^+K^-}$ resonant part of the $\mathrm{D_s}$ decay is considered, motivated by the negligible contribution of non-resonant $\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+$ events in data.

The fraction of τ leptons produced in $\mathrm{D_s}$ decays at the CMS energy of 8 TeV was calculated from the inputs given in Table 4.5 and 4.39, yielding the value $f^\tau_{\mathrm{D_s}} = (71 \pm 10)\ \mu\mathrm{b}/(89 \pm 11)\ \mu\mathrm{b} = 0.80 \pm 0.03$. While the uncertainties in the cross-sections of the individual production channels have been considered uncorrelated, the ratio of the cross-sections in $f^\tau_{\mathrm{D_s}}$ takes into account the correlation between the cross-section from D_s decays and the total cross-section. At the CMS energy of 7 TeV the value of $f^\tau_{D_s}$ changes to 0.78 ± 0.04, due to the difference in the relative increase of the charm and beauty cross-sections.

Table 4.39: Branching fractions from [23] and [108] used in the evaluation of the parameter $f_{D_s}^{\tau}$.

$\mathcal{B}(b \to \tau^- X)$	0.0241 ± 0.0023
$\mathcal{B}(b \to \mathrm{D}^- X)$	0.233 ± 0.017
$\mathcal{B}(b \to \mathrm{D_s} X)$	0.248 ± 0.037
$\mathcal{B}(\mathrm{D_s}^- \to \tau^- \overline{\nu}_\tau)$	0.0561 ± 0.0024

4.8.1 Generation, reconstruction and selection efficiencies

The generator level efficiencies for the mixed MC samples, $\varepsilon_{\mathrm{sig,mix}}{}^{\mathrm{GEN}}$ and $\varepsilon_{\mathrm{cal,mix}}{}^{\mathrm{GEN}}$, are the averages of the individual sub-channel efficiencies, ε_{CUT}, as detailed in Sect. 4.3.3. Their values are given in Table 4.40, for all signal and normalization samples at both 7 and 8 TeV CMS energy. The uncertainty includes the statistical uncertainty, which is due to the finite size of the MC samples and the systematic uncertainty on the branching fractions values used in the calculation of τ lepton production cross-section.

Table 4.40: Generator level efficiencies for the mixed MC samples.

	CMS Energy	
Parameter	7 TeV	8 TeV
$\varepsilon_{\tau^- \to \mu^-\mu^+\mu^-,\mathrm{mix}}{}^{\mathrm{GEN}}$ [%]	8.99 ± 0.40	9.21 ± 0.35
$\varepsilon_{\mathrm{D_s^+} \to \phi(\mu^-\mu^+)\pi^+,\mathrm{mix}}{}^{\mathrm{GEN}}$ [%]	11.19 ± 0.34	11.53 ± 0.34

Subsequently, the reconstruction, isMuon and selection efficiencies, $\varepsilon_{\mathrm{sig}}{}^{\mathrm{REC,isMuon,SEL}}$ and $\varepsilon_{\mathrm{cal}}{}^{\mathrm{REC,isMuon,SEL}}$, are determined after the selection criteria detailed in Sect. 4.3 have been applied. Their values are given in Table 4.41. The motivation for the veto on $\mathrm{D_s^-} \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$ is described in more detail in Sect. 4.7.2.

Table 4.41: Reconstruction, isMuon and selection efficiencies for the mixed MC samples.

	CMS Energy	
Parameter	7 TeV	8 TeV
$\varepsilon_{\tau^- \to \mu^-\mu^+\mu^-,\mathrm{mix}}{}^{\mathrm{REC,isMuon,SEL}}$ [%]	9.927 ± 0.028	9.261 ± 0.023
$\varepsilon_{\mathrm{D_s^+} \to \phi(\mu^-\mu^+)\pi^+,\mathrm{mix}}{}^{\mathrm{REC,isMuon,SEL}}$ [%]	7.187 ± 0.022	6.690 ± 0.022

The ratio of efficiencies is then corrected by four factors to account for the differences between data and MC corresponding to track reconstruction efficiency, muon identification efficiency, $\phi(1020)$ mass cut efficiency and τ mass cut efficiency. These corrections will be described in the following sections. The additional requirement aimed at the removal of trash bins is also applied.

4.8.1.1 Correction to track reconstruction

We determine the track reconstruction efficiency from the simulation and correct it using the tracking efficiency maps as provided by the tracking group of the LHCb experiment [114, 115]. The correction factors, c^{track}, depend on the pseudo-rapidity and the momentum of the tracks and are shown in Table 4.45, where the first uncertainty corresponds to the statistical uncertainty from the tracking efficiency map [114, 115]. The second uncertainty accounts for the combined systematics on tracks (0.6%) and for the uncertainty resulting from lack of knowledge of the hadronic interaction length of the extra pion from the D_s decay (2.0%). A ratio of correction factors of $0.997 \pm 0.009 \pm 0.026$ and $0.996 \pm 0.009 \pm 0.026$ is obtained for 2011 and 2012 data, respectively.

4.8.1.2 Correction to isMuon efficiency

Because of the precision with which the simulation describes the isMuon requirement [116], the correction factors $c^{\mu\text{ID}}$ are calculated from the tables determined by the LHCb collaboration for the needs of the search for $B_s^0 \to \mu^+\mu^-$ decay [117].

The two-dimensional map of efficiency vs acceptance in bins of momentum (p) and transverse momentum (p_T) obtained both in data and in simulated events is folded into the p, p_T spectra of the muons originating from the reconstructed and selected $\tau^- \to \mu^-\mu^+\mu^-$ and $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ Monte Carlo samples and the average values $< \varepsilon_\mu \times acc_\mu >_{\text{data}}$ and $< \varepsilon_\mu \times acc_\mu >_{\text{MC}}$ are obtained. The ratio

$$c^{\mu\text{ID}} = < \varepsilon_\mu \times \text{acc}_\mu >_{\text{data}} / < \varepsilon_\mu \times \text{acc}_\mu >_{\text{MC}} \tag{4.12}$$

is then used to correct the isMuon efficiency evaluated on the respective Monte Carlo sample for a given decay channel. The differences between data and MC largely cancel in the ratio, and the overall uncertainty comes from the statistical error in the isMuon map and takes into account the correlations.

While it is preferred to use the ratio of data and Monte Carlo simulations as it cancels possible biases originating from the method, to determine the systematic error on this correction, the correction itself is also calculated, using the two-dimensional data map and the measured isMuon efficiency for each of the $\tau^- \to \mu^-\mu^+\mu^-$ and $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ Monte Carlo samples instead of the MC map. This tests the reliability of the method when applied to the MC samples used in this analysis. The difference in the ratio of these correction factors is then assigned as a systematic uncertainty. The resulting ratios of correction factors are

$0.9731 \pm 0.0031 \pm 0.0264$ and $1.0071 \pm 0.0022 \pm 0.0204$ for 2011 and 2012 samples respectively, and are listed in Table 4.45.

4.8.1.3 Correction to $\phi(1020)$ mass cut efficiency

In the MC simulation the $\phi(1020)$ meson is generated for masses above K^+K^- threshold (988 MeV/c^2) and with an upper requirement on the mass at 1085 MeV/c^2. Since no such restrictions exist in data and the processes involving $\phi \to \mu\mu$ decay are also included in this analysis, the measured rates need to be corrected for this effect.

To measure the intensities in the tails of the $\phi(1020)$ Breit-Wigner (BW) distribution, we generated (using Monte Carlo method) a simple non-relativistic BW lineshape, modified by the available phase space in the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ decay. It is appropriate to mention that the decay $\phi \to K^+K^-$ proceeds via an S-wave and there is no need to include mass-dependent factors in the BW. Assuming that no ϕ would be produced below the KK threshold, we find that 1.0% of the generated sample would be truncated by the 1085 MeV/c^2 generator-level cut on the $\mu\mu$ mass. If we allow the BW to go down to $\mu\mu$ threshold, then we lose 2.8% of the generated sample by the cuts at 986 MeV/c^2 and 1085 MeV/c^2.

We fit in a range ± 20 MeV/c^2 around the ϕ peak. For the full BW distribution from $2M_\mu$ up to $M(D_s) - M(\pi)$, the region outside of the fit range corresponds to 6.6% of the total mass range. In the generated truncated Monte Carlo sample, the region outside the fit range corresponds to 3.8% of the generated area. Thus in the fits we do not count 3.8% of the respective MC events, while we do not count 6.6% of the data. As a result, our $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ selection efficiency, as determined on MC, would need to be corrected down by 1.066/1.038, i.e. by a factor of 2.7%.

If we assume that in reality the Breit-Wigner distribution does not go below K^+K^- threshold in the data (this is unlikely, but cannot be excluded), then the above 6.6% loss becomes 4.8% and the correction factor for the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ selection efficiency would amount to 1.048/1.038, i.e. to 1.0%.

As a result, the correction to $\phi(1020)$ mass efficiency was estimated to be in the range from 1% to 2.7%. Conservatively, the value of (2 ± 1)% was attributed as the correction factor, c^ϕ, to allow the BW lineshape to be truncated in the Monte Carlo simulation.

4.8.1.4 Correction to τ mass cut efficiency

To account for differences in the efficiency of τ mass cuts between MC and data samples, the correction factor, c^τ, is applied. It assumes the form:

$$c^\tau = \frac{\text{Integral of MC-resolution Gaussians at } \pm 30\text{MeV}/c^2}{\text{Integral of data-resolution Gaussians at } \pm 20 \text{ MeV}/c^2}.$$

This correction takes into account both the effect of reducing the mass window from ± 30 MeV/c^2 in the selection to ± 20 MeV/c^2 in data and the measured difference in

mass resolutions between data and MC samples. The widths of the double-Gaussian function describing the mass resolutions in data are the raw signal MC widths multiplied by the ratio of data/MC widths, as determined in the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ channel. The resulting values of the widths. which are presented in Table 4.42, are then used as input to the integration. The central value of the Gaussians was fixed at the PDG value of mass of the τ lepton (1776.8 MeV/c^2). However, the corrected central values of the τ mass shown in Table 4.42 were used as the central values of the mass *p.d.f.* in the evaluation of the final result. The efficiency of the $\pm$ 20MeV/c^2 mass cut on the mixed signal MC samples after offline selection and trigger requirements is found to be 95.9% and 95.7% for 2011 and 2012 data, respectively (assuming an uncorrected MC mass resolution). The uncertainty on c^τ factor represents the statistical error propagated through the calculation procedure. Correction factors of 1.032 ± 0.006 and 1.026 ± 0.006 are obtained for 2011 and 2012 data respectively, and are listed in Table 4.45.

Table 4.42: Parameters of Gaussian functions (after applying corrections described in the text) which were used to parametrize the $\tau^- \to \mu^-\mu^+\mu^-$ signal in data.

CMS Energy	7 TeV	8 TeV
Mean (MeV/c^2)	1779.1 ± 0.1	1779.0 ± 0.1
σ_1 (MeV/c^2)	7.7 ± 0.1	7.6 ± 0.1
σ_2 (MeV/c^2)	12.0 ± 0.8	11.5 ± 0.5

4.8.1.5 Correction due to exclusion of low likelihood bins

As described in Sect. 4.7, the so-called trash bins in the distributions of $\mathcal{M}_{3body}$ and $\mathcal{M}_{PID}$ variables were excluded from the evaluation of the $\tau^- \to \mu^-\mu^+\mu^-$ branching fraction. The excluded range corresponds to the requirements $\mathcal{M}_{3body} < 0.26(0.28)$ for 2011 (2012) data, respectively, as well as $\mathcal{M}_{PID} < 0.4$, which may have different efficiencies for data and Monte Carlo samples. To account for this possible discrepancy, the following factor c^{trash} has been introduced.

Errors on the calibrated $\mathcal{M}_{3body}$ distribution are derived as follows: The correction factors to get a $\tau^- \to \mu^-\mu^+\mu^-$ response from the observed $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ distribution have statistical errors which is due to the overall size of the relevant Monte Carlo sample. The $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ distribution in data is extracted using the *sPlot* method. The weights in the *sPlot* approach are designed so that using the sum of weights provides correct error propagation. Thus, propagating the default errors from the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ response distribution in data with the statistical errors on the correction factors gives the statistical uncertainty on each bin of the $\mathcal{M}_{3body}$ variable. This error is then added in quadrature to the systematic uncertainty (1.3%) of the calibration strategy. The cut efficiency for the trash bin removal is $(81.3 \pm 1.8)\%$ for 2011 and $(82.6 \pm 2.0)\%$ for 2012 sample.

The `MultiTrack` tool provides the statistical error on the cut efficiency at $\mathcal{M}_{PID} > 0.4$. Added in quadrature to the systematic error, the combined $\mathcal{M}_{PID}$

cut efficiency is determined to be $(63.6 \pm 2.4)\%$ for 2011 and $(62.2 \pm 2.2)\%$ for 2012 (the correction method from Sect. 4.5.2.2 is used for the efficiency and its error is part of the systematic error).

Adding in quadrature the $\mathcal{M}_{PID}$ and $\mathcal{M}_{3body}$ contributions for the removal of the trash bins gives a cut efficiency of $(52.9 \pm 3.1)\%$ for 2011 and $(51.0 \pm 2.8)\%$ for 2012. This results in values of $c^{trash} = 1.89 \pm 0.12$ for 2011 and 1.96 ± 0.12 for 2012 sample.

4.8.2 Trigger efficiencies

The trigger efficiencies for signal and normalization channels are determined using the mixed MC samples. For 2012 samples, variations of the requirements applied in the trigger lines of interest which took place for different TCKs could potentially cause significant changes to the trigger efficiency for different subsamples of the dataset. For the trigger lines applied to the signal channel, these changes are summarized in Table 4.43. Note that for the Hlt2DiMuonDetached line used in the $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ channel all cuts remained unchanged throughout the whole period of data taking in 2012.

The default TCK (0x009f0045) applied to 2012 MC samples described fully the behaviour of TCKs 0x00990042 onwards. Therefore the minimum set of additional TCKs, relevant for the description of changes in trigger conditions throughout 2012 data taking is 0x008c0040 and either 0x0094003d or 0x0097003d. A further set of $\tau^- \to \mu^-\mu^+\mu^-$ and $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$ MC samples was prepared with the TCK 0x0097003d and with cuts identical to those applied to the default MC samples. These samples were used solely for trigger efficiency calculations. As the TCK 0x008c0040 accounted for less than 5% of the total integrated luminosity, it was not studied separately and was instead assumed to have the same efficiency as 0x0097003d, with a conservative additional systematic 5% included to account for possible differences.

Table 4.43: Trigger requirements for the major TCKs in the 2012 dataset. Cut values which remained constant across all TCKs are excluded for brevity.

		Trigger Line					
TCK	$\mathcal{L}^{-1}$ (pb^{-1})	L0Muon	Hlt1TrackMuon		Hlt2TriMuonTau		
Cut		p_T(MeV/c)	p(GeV/c)	Trk χ^2/ndf	μ IPχ^2	Tr χ^2	$c\tau$ [µm]
0x008c0040	70	1480	8	3	4	6	45
0x0094003d	280	1760	8	2.5	9	4	75
0x0097003d	280	1760	8	2.5	9	4	75
0x00990042	620	1760	3	2.5	9	4	75
0x00990044	140	1760	3	2.5	9	4	75
0x009f0045	20	1760	3	2.5	9	4	75
0x009f0045+	675	1760	3	2.5	9	4	75

The trigger efficiencies for the signal and normalization channels, ${\varepsilon_{\text{sig}}}^{\text{TRIG}}$ and ${\varepsilon_{\text{cal}}}^{\text{TRIG}}$, were then calculated as the weighted averages of the individual TCK efficiencies shown in Table 4.44, according to their integrated luminosity. For each of the luminosities corresponding to each of the given TCKs we attributed an uncertainty of 3.5%, according to the current best knowledge of the absolute luminosity determination in experimental conditions of the LHCb experiment.

Table 4.44: Trigger efficiencies for the signal and normalization decay channels for the different TCKs corresponding to Monte Carlo 2012 samples.

MC TCK	$\mathcal{L}^{-1}$ (pb^{-1})	${\varepsilon_{\text{sig}}}^{\text{TRIG}}$(%)	${\varepsilon_{\text{cal}}}^{\text{TRIG}}$(%)
0x0097003d	630	38.79 ± 0.52	20.06 ± 0.26
0x009f0045	1,455	39.48 ± 0.12	20.86 ± 0.12

For 2011 samples the full trigger efficiency was derived from the mixed MC sample with the default TCK, as described in Sect. 4.3.1.

For the 2012 samples the systematic uncertainty on ${\varepsilon_{\text{sig}}}^{\text{TRIG}}$ and ${\varepsilon_{\text{cal}}}^{\text{TRIG}}$ was determined using $\text{B}^- \to \text{J}/\psi\text{K}^-$ events by comparing the difference between the MC trigger efficiency and that determined from data using the TISTOS method. The behaviour of the trigger in the momentum range typical for the muons originating from the $\tau^- \to \mu^-\mu^+\mu^-$ signal was checked by requiring the muons to have the momentum less than 30 GeV/c and transverse momentum less than 2 GeV/c. Approximately 50% of the signal MC events passed this requirement, with the remaining ones located in the higher momentum range which was known to be well described in other analyses. A full TISTOS study was performed on every existing trigger line, using only a single kinematic bin due to the limited statistics in the $\text{B}^- \to \text{J}/\psi\text{K}^-$ data sample. In the TISTOS method the trigger efficiency was calculated via the formula:

$$\varepsilon^{\text{TRIG}} = \varepsilon^{\text{TIS}} \times \frac{N^{\text{TRIG}}}{N^{\text{TIS}}}, \tag{4.13}$$

where N^{TRIG} is the number of triggered events in the data sample, N^{TIS} is the number of triggered events that would trigger without the decay of interest being present and ε^{TIS} can be explicitly written as:

$$\varepsilon^{\text{TIS}} = \frac{N^{\text{TIS\&TOS}}}{N^{\text{TOS}}}, \tag{4.14}$$

where $N^{\text{TIS\&TOS}}$ is the number of events that would be triggered both with and without the decay of interest, and N^{TOS} is the number of events triggered due to the decay of interest.

A difference of 5.2% between MC and data is found and is used as a conservative error estimate for the ratio $\frac{{\varepsilon_{\text{cal}}}^{\text{TRIG|SEL}}}{{\varepsilon_{\text{sig}}}^{\text{TRIG|SEL}}}$, as shown in Table 4.46.

For the 2011 samples the systematic uncertainty on ${\varepsilon_{\rm sig}}^{\rm TRIG}$ and ${\varepsilon_{\rm cal}}^{\rm TRIG}$ was determined using the same method and further cross-checked for $\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+$ decay channel.

Table 4.45: Factors used in the normalization of the signal yield. The uncertainties are statistical and systematic respectively - if only a single uncertainty is given, then no corresponding systematic error was estimated.

Factor \ CMS Energy	7 TeV	8 TeV
${\varepsilon_{\rm sig}}^{\rm GEN}(\%)$	8.989 ± 0.40	9.21 ± 0.35
${\varepsilon_{\rm cal}}^{\rm GEN}(\%)$	11.19 ± 0.34	11.53 ± 0.32
${\varepsilon_{\rm sig}}^{\rm REC,isMuon,SEL}(\%)$	9.927 ± 0.028	9.261 ± 0.023
${\varepsilon_{\rm cal}}^{\rm REC,isMuon,SEL}(\%)$	7.187 ± 0.022	6.690 ± 0.022
$\frac{{\rm c_{cal}}^{\rm track}}{{\rm c_{sig}}^{\rm track}}$	$0.997 \pm 0.009 \pm 0.026$	$0.996 \pm 0.009 \pm 0.026$
$\frac{{\rm c_{cal}}^{\mu\rm ID}}{{\rm c_{sig}}^{\mu\rm ID}}$	$0.9731 \pm 0.0031 \pm 0.0264$	$1.0071 \pm 0.0022 \pm 0.0204$
c^{ϕ}	0.98 ± 0.01	
c^{τ}	1.032 ± 0.006	1.026 ± 0.006
$\mathrm{c}^{\rm trash}$	1.89 ± 0.12	1.96 ± 0.12
${\varepsilon_{\rm sig}}^{\rm TRIG}(\%)$	$35.52 \pm 0.14 \pm 0.14$	$39.3 \pm 1.7 \pm 2.0$
${\varepsilon_{\rm cal}}^{\rm TRIG}(\%)$	$23.42 \pm 0.14 \pm 0.09$	$20.62 \pm 0.76 \pm 1.07$

4.8.3 Normalization summary

The final normalization factor is given in Table 4.46, including the value of N_{cal} from Sect. 4.5.3. Here the $\varepsilon^{\rm REC\&SEL}$ parameter represents the combined generation, reconstruction, isMuon and selection efficiencies, including the correction factors which were described previously. For technical reasons, in the computation of the final result we use α without c^{trash}, as the effect of the removal of the low likelihood bins is incorporated in the signal $p.d.f$. Therefore two values of the normalization factor are given in Table 4.46: α (α^{trash}), i.e. the one without (with) the trash correction c^{trahs} included, respectively.

Table 4.46: Summary of the terms entering in the overall normalization factor. The symbol α^{trash} indicates the normalization factor with the trash correction included.

Factor \ CMS Energy	7 TeV	8 TeV
$\mathcal{B}(\mathrm{D}_s^+ \to \phi(\mu^-\mu^+)\pi^+)$	$(1.317 \pm 0.099) \times 10^{-5}$	
$f^{\tau}_{D_s}$	0.78 ± 0.04	0.80 ± 0.03
$\mathcal{B}(\mathrm{D_s}^- \to \tau^-\overline{\nu}_\tau)$	0.0561 ± 0.0024	
${\varepsilon_{\mathrm{cal}}}^{\mathrm{REC\&SEL}}/{\varepsilon_{\mathrm{sig}}}^{\mathrm{REC\&SEL}}$	0.898 ± 0.060	0.912 ± 0.054
${\varepsilon_{\mathrm{cal}}}^{\mathrm{TRIG}}/{\varepsilon_{\mathrm{sig}}}^{\mathrm{TRIG}}$	0.6593 ± 0.0058	0.525 ± 0.040
N_{cal}	$28,207 \pm 440$	$52,131 \pm 695$
α	$(3.81 \pm 0.46) \times 10^{-9}$	$(1.72 \pm 0.23) \times 10^{-9}$
α^{trash}	$(7.20 \pm 0.98) \times 10^{-9}$	$(3.37 \pm 0.50) \times 10^{-9}$

4.9 Results on the $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-)$

In this section, we calculate the signal branching fraction using the statistical method called $\mathrm{CL_s}$ [118], described in Sect. 5.2. The inputs for the interpretation of this result are the normalization factor (Sect. 4.8) and the parameters describing the $\mathcal{M}_{3body}$, $\mathcal{M}_{PID}$ and mass $p.d.f.s$ for signal (Sect. 4.5) and background (Sect. 4.7.4).

Each systematic is assigned the so-called nuisance parameter, defined as a parameter that is of no interest to the analysis but needs to be considered in the calculation of the final result. The nuisance parameters considered in the branching fraction evaluation are:

- the production fractions of τ leptons at LHC (5 degrees of freedom), mentioned in Sect. 4.3.3,
- the parameters of the mass resolution: central value and both Gaussian widths, mentioned in Sect. 4.5.3, (3 degrees of freedom),
- the parameters of particle identification calibration as provided by `PIDCalib` package, including the statistical uncertainty from `PIDCalib` and the systematics, mentioned in Sect. 4.5.2.1, (5 degrees of freedom, one for each bin boundary),
- the parameter of PID correction, explained in Sect. 4.5.2.2, (1 degree of freedom),
- the $\mathcal{M}_{3body}$ calibration, from Sect. 4.5.1.4, (1 degree of freedom per bin boundary),

- the normalization factor α, from Sect. 4.8, (1 degree of freedom).

The contribution to α from the τ production is excluded from this last point since it has been already covered in the first point; similarly, the contribution to c^{trash} is also included.

After given permission to unblind by the LHCb collaboration, we did not observe an significant access of events, as can be seen in Tables 4.36, 4.37 and Fig. 4.11 to 4.14. Moreover, the tri-muon distributions corresponding to highest sensitivity to the $\tau^- \to \mu^-\mu^+\mu^-$ decay in bins of $(\mathcal{M}_{3body}, \mathcal{M}_{PID})$ are shown in Fig. 4.15.

In the absence of any evidence for a signal, we set an upper limit on branching fraction of the decay in question.

4.9.1 Calculation of expected upper limit on the $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-)$

The distribution of the expected (dashed lines) and observed (solid lines) CL_s values is shown in Fig. 4.16 as a function of the assumed branching fraction, under the hypothesis to observe background events only. The two bands cover the region of 68% containment (yellow) and 95% containment (green) of compatible observations, respectively.

At 90% confidence level (CL), the expected upper limit (median of all toy experiments in CL_s method) for the branching fraction of $\tau^- \to \mu^-\mu^+\mu^-$ is found to be:

$$\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 5.0 \times 10^{-8}. \tag{4.15}$$

At 95% confidence level, the expected upper limit for the branching fraction of $\tau^- \to \mu^-\mu^+\mu^-$ is determined as:

$$\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) \quad < \quad 6.1 \times 10^{-8}\,. \tag{4.16}$$

It is worthwhile to mention that the expected upper limit obtained with the best standard classifier, instead of the blending technique turned out to be 6% worse.

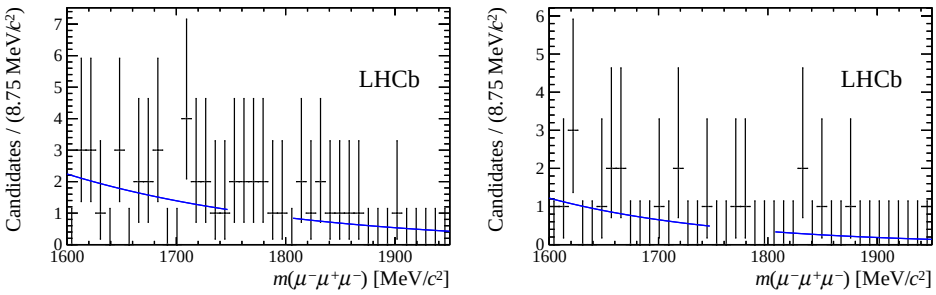
4.9.2 Calculation of observed upper limit on the $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-)$

After opening the signal box, the distributions of observed CL_s values are calculated, including the systematic uncertainties. The observed numbers of events in each bin can be seen in Table 4.36 and 4.37. Since we did not find a significant enhancement of signal candidates the upper limit was calculated. The observed upper limit for the branching fraction of $\tau^- \to \mu^-\mu^+\mu^-$ is found to be

$$\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) \quad < \quad 4.6 \times 10^{-8} \text{ at } 90\% \text{ CL}, \tag{4.17}$$

$$\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) \quad < \quad 5.6 \times 10^{-8} \text{ at } 95\% \text{ CL}. \tag{4.18}$$

These limits are set using $\tau^- \to \mu^-\mu^+\mu^-$ Monte Carlo simulations with phase-space distributions. The results are comparable with those obtained by B factories:



(a) 2011 data (b) 2012 data

Figure 4.15: Distributions of tri-muon invariant mass (data points) in the highest sensitivity bins of $(\mathcal{M}_{3body}, \mathcal{M}_{PID})$ variables for the $\tau^- \to \mu^-\mu^+\mu^-$ data in 2011 and 2012 after $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$veto. The blue-solid line shows the results of the fit to an exponential *p.d.f.* performed in the sidebands of the tri-muon invariant mass.

$2.1(3.2) \times 10^{-8}$ for Belle (Babar) respectively. We would like to point out that in the case of B factories the Feldman Cousin approach was used to evaluate the limit. In case of a downward fluctuation, this method gives stronger upper limits. We will recalculate the previous B factories limits in Sect. 5.2. As will be shown, the Babar result using the CL_s method will yield:

$$\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) \quad < \quad 4.0 \times 10^{-8} \text{ at } 90\% \text{ CL}, \tag{4.19}$$

making our result very comparable with the BaBar's one.

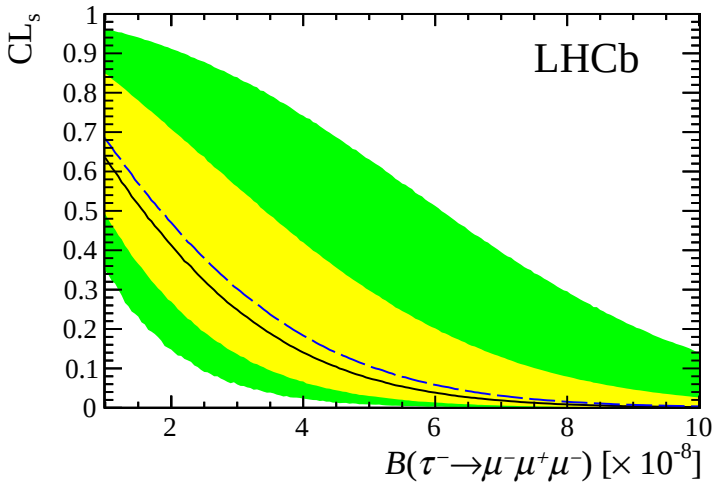


Figure 4.16: The CLs curve with 68% (yellow) and 90% (green) containment bands, under the hypothesis to observe background events only. The black line represents the observed value of CL_s, while the blue dashed one – the expected CL_s value.

The author has estimated that after Run 2 of the LHC the expected upper limit of the LHCb experiment on the decay in question will be at the level of the current Belle estimate.

4.10 Dependence of the $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-)$ on the predictions of the effective field theory

The upper limits on the $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-)$ can also be evaluated in terms of an effective field theory, including BSM operators with different lepton chirality structures, as described in Sect. 2.3.4. The contributions from these operators can include the kinematic properties of the decay in question. The Monte Carlo samples containing the decay $\tau^- \to \mu^-\mu^+\mu^-$ were generated assuming a flat phase space distribution. As described in Sect. 2.3.4, there are three operators and two interferences that can introduce LFV in the effective field theory model. As the models introduced to TAUOLA by the author were not implemented in the LHCb simulations at the time when the described analysis was finalized, we estimated their influence on the $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-)$ in the following way. For Eq. 2.47 - 2.51 we introduced the corresponding Dalitz densities: $\rho_{\mathrm{V}}^{(LL)(LL)}$, $\rho_{\mathrm{V}}^{(LL)(RR)}$, $\rho_{\mathrm{rad}}^{(LR)}$, $\rho_{\mathrm{mix}}^{(LL)(LL)}$ and $\rho_{\mathrm{mix}}^{(LL)(RR)}$ respectively. The difference is that the differential decay rates are not normalized to unity when integrating over full phase space, where Dalitz densities are, so there is no dependence on coupling constants, mass scale Λ etc.

To implement the given MFLV model, each event simulated with a flat phase space acquired an appropriate weight, applied to the density value ρ for the corresponding phase space point $(m_{01}^2, m_{02}^2, m_{12}^2)$. The weights are normalized in such a way that the integral over the whole Dalitz phase space is equal to 1. With such normalization the weights allow us to determine the respective reconstruction efficiencies in the framework of the MLFV model, using the values evaluated from the flat phase space simulated samples, as given in the formula:

$$\varepsilon_{\mathrm{gen\&rec}}^{\mathrm{model}} = \frac{\varepsilon_{\mathrm{gen\&rec}}^{\text{LHCb MC}}}{cN} \sum_{N} \rho_{\mathrm{gen}}^{\mathrm{model}}(m_{12}, m_{23}), \tag{4.20}$$

where $\varepsilon_{\mathrm{gen\&rec}}^{\text{LHCb MC}}$ is the reconstruction and generator cut efficiency of the flat phase space model, N is the number of events, c is the constant providing the normalization for the flat phase space model ($\int c \mathrm{d}m_{01}^2 \mathrm{d}m_{02}^2 = 1$) and hence making ρ/c dimensionless and independent of the phase space parametrisation.

Given that the efficiency ratio $\varepsilon_{\mathrm{gen\&rec}}^{\mathrm{model}}/\varepsilon_{\mathrm{gen\&rec}}^{\text{LHCb MC}}$ is a sample average of ρ/c, the error is the respective uncertainty:

$$\sigma_\varepsilon = \sqrt{\frac{1}{N}\left(\left(\frac{1}{N}\sum\left(\frac{\rho}{c}\right)^2\right) - \left(\frac{1}{N}\sum\frac{\rho}{c}\right)^2\right)}.$$

Table 4.47: Ratio of efficiencies $\varepsilon^{\rho}/\varepsilon^{\text{LHCb MC}}$ for 2011 Monte Carlo sample obtained for five different chiralities of a model independent MLFV. The consecutive rows correspond to the application of subsequent cuts.

Sample	$\rho_{\text{V}}^{(LL)(LL)}$	$\rho_{\text{V}}^{(LL)(RR)}$	$\rho_{\text{rad}}^{(LR)}$	$\rho_{\text{mix}}^{(LL)(LL)}$	$\rho_{\text{mix}}^{(LL)(RR)}$
Gen & rec	1.05531 ± 0.00044	1.02015 ± 0.00028	1.01848 ± 0.0013	1.03352 ± 0.00070	0.98537 ± 0.00031
Stripping	1.05988 ± 0.00094	1.03074 ± 0.00056	1.0210 ± 0.0027	1.0269 ± 0.0014	0.98873 ± 0.00063
Offline	1.0170 ± 0.0011	1.08502 ± 0.00059	0.62249 ± 0.00098	0.9240 ± 0.0016	1.00959 ± 0.00074
Trigger	1.0049 ± 0.0019	1.0772 ± 0.0010	0.6318 ± 0.0017	0.9193 ± 0.0027	1.0124 ± 0.0013

Table 4.48: Ratio of efficiencies $\varepsilon^{\rho}/\varepsilon^{\text{LHCb MC}}$ for 2012 Monte Carlo sample obtained for five different chiralities of a model independent MLFV. The consecutive rows correspond to the application of subsequent cuts.

Sample	$\rho_{\text{V}}^{(LL)(LL)}$	$\rho_{\text{V}}^{(LL)(RR)}$	$\rho_{\text{rad}}^{(LR)}$	$\rho_{\text{mix}}^{(LL)(LL)}$	$\rho_{\text{mix}}^{(LL)(RR)}$
Gen & rec	1.05475 ± 0.00048	1.01913 ± 0.00031	1.0204 ± 0.0014	1.03354 ± 0.00077	0.98550 ± 0.00034
Stripping	1.05832 ± 0.00081	1.02850 ± 0.00050	1.0244 ± 0.0024	1.0263 ± 0.0012	0.98923 ± 0.00055
Offline	1.01522 ± 0.00097	1.08315 ± 0.00051	0.62431 ± 0.00085	0.9236 ± 0.0014	1.00992 ± 0.00064
Trigger	1.0060 ± 0.0016	1.06798 ± 0.00082	0.6422 ± 0.0014	0.9280 ± 0.0023	1.0089 ± 0.0010

The corresponding efficiency ratios for the 2011 and 2012 mixed Monte Carlo samples are listed in Table 4.47 and 4.48 for the combined generator level cuts and reconstruction, stripping, offline cuts (ϕ veto, clone veto, η veto, fiducial cuts) and the trigger requirements.

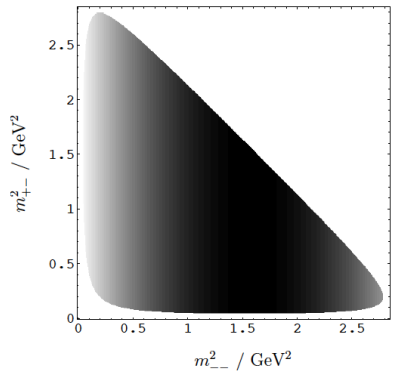
The effect of the $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$ veto is particularly large for the $\rho_{\rm rad}^{(LR)}$ scenario, which is due to the preference for low values of opposite-sign dimuon masses in the Dalitz plot. Correlations between the Dalitz variables and the multivariate classifiers are considered in the following way. The correction factors in the calibration of $\mathcal{M}_{3body}$ are determined on MC samples as in a standard analysis. However, in the $\tau^- \to \mu^-\mu^+\mu^-$ MC signal sample events are weighted with each of the Dalitz distributions. The resulting signal $p.d.f.$ is used in the evaluation of the final result together with the corrected normalization factors from Table 4.47 and 4.48.

These factors show how much the expected and observed limits are modified in each individual chirality structure of the effective field theory approach. Their values, shown in Table 4.49, are in general close to the ratios listed in Table 4.47 and 4.48. (NB: the limits are inversely proportional to efficiencies. One thus has to compare them to the respective reciprocal values).

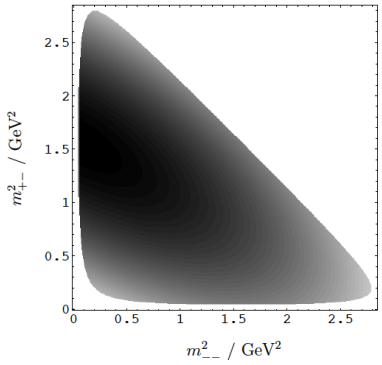
In conclusion, we investigated how the limit on $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-)$ changes in the presence of contributions due to five combinations of new physics operators, as predicted in the MLFV model. We found that four operators do not change our limit significantly as the limit is within: $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < (4.1 - 4.6) \times 10^{-8}$. The radiative operator due to $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$veto was found to yield the value $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 6.8 \times 10^{-8}$.

Table 4.49: Observed and expected limits $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) \times 10^{-8}$ in the five combinations of chiralities in MLFV model.

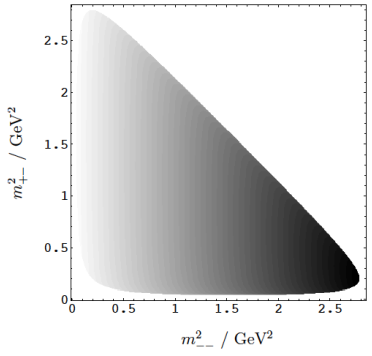
Dalitz distribution	$\rho_{\rm V}^{(LL)(LL)}$	$\rho_{\rm V}^{(LL)(RR)}$	$\rho_{\rm rad}^{(LR)}$	$\rho_{\rm mix}^{(LL)(LL)}$	$\rho_{\rm mix}^{(LL)(RR)}$
expected	4.7	4.6	7.6	5.1	5.0
observed	4.2	4.1	6.8	4.4	4.6



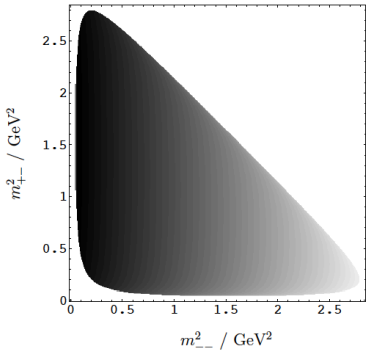
(a) Dalitz density for $\rho_{\mathrm{V}}^{(LL)(LL)}$



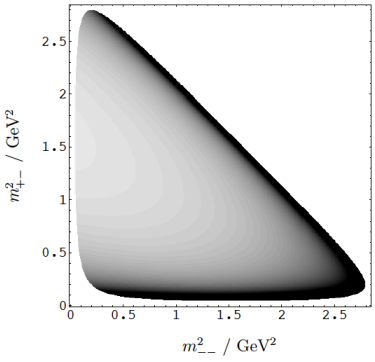
(b) Dalitz density for $\rho_{\mathrm{V}}^{(LL)(RR)}$



(c) Dalitz density for $\rho_{\mathrm{mix}}^{(LL)(LL)}$



(d) Dalitz density for $\rho_{\mathrm{mix}}^{(LL)(RR)}$



(e) Dalitz density for $\rho_{\mathrm{rad}}^{(LR)}$

Figure 4.17: Dalitz distributions in the effective field approach for five different BSM operators corresponding to different lepton chirality structures Fig. 2.5.

Chapter 5

Experimental limits on branching fractions for τ lepton flavour violating τ decays

5.1 Review of current experimental limits on lepton flavour violating τ decays

This chapter summarizes briefly the results of searches for lepton flavour violation in τ lepton decays as well as provides the combined limits for the processes in question. In 2013 the author of this thesis was invited to participate in Heavy Flavour Averaging Group (HFAG) [119] and took part in the preparation of the combined results and other averages, including those presented in this chapter.

Over the past fifteen years the searches for lepton flavour violating decays of τ lepton have been carried out by the BaBar, Belle and LHCb collaborations. The collection of all relevant results, as given by the HFAQ group, is provided in Table 5.1 and Fig. 5.1. It is worth underlining that the result of studies described in this thesis is also included in this „Official" HFAG report [120].

Table 5.1: Collection of upper limits on branching fractions for the LFV τ decay modes. For convenience, the decay modes are grouped in categories labelled according to their particle content (S-scalar, P-pseudovector, V-vector particle). The mark "L" in the category column means that the decay mode implies lepton number violation as well as lepton flavour violation, while "BNV" indicates that the channel is Baryon Number Violating [120]. The result of the study described in this thesis is marked with an asterisk (*).

Decay mode	**Category**	**90% CL Limit**	**Exp.**	**Ref.**
$\Gamma_{156} = e^-\gamma$	$l\gamma$	$< 12.0 \cdot 10^{-8}$	Belle	[121]
		$< 3.3 \cdot 10^{-8}$	BaBar	[122]
$\Gamma_{157} = \mu^-\gamma$		$< 4.5 \cdot 10^{-8}$	Belle	[121]
		$< 4.4 \cdot 10^{-8}$	BaBar	[122]
$\Gamma_{158} = e^-\pi^0$	lP^0	$< 2.2 \cdot 10^{-8}$	Belle	[123]
		$< 13.0 \cdot 10^{-8}$	BaBar	[124]
$\Gamma_{159} = \mu^-\pi^0$		$< 2.7 \cdot 10^{-8}$	Belle	[123]

Table 5.1 – continued from previous page

Decay mode	Category	90% CL Limit	Exp.	Ref.
		$< 11.0 \cdot 10^{-8}$	BaBar	[124]
$\Gamma_{162} = e^- \eta$		$< 4.4 \cdot 10^{-8}$	Belle	[123]
		$< 16.0 \cdot 10^{-8}$	BaBar	[124]
$\Gamma_{163} = \mu^- \eta$		$< 2.3 \cdot 10^{-8}$	Belle	[123]
		$< 15.0 \cdot 10^{-8}$	BaBar	[124]
$\Gamma_{172} = e^- \eta'(958)$		$< 3.6 \cdot 10^{-8}$	Belle	[123]
		$< 24.0 \cdot 10^{-8}$	BaBar	[124]
$\Gamma_{173} = \mu^- \eta'(958)$		$< 3.8 \cdot 10^{-8}$	Belle	[123]
		$< 14.0 \cdot 10^{-8}$	BaBar	[124]
$\Gamma_{160} = e^- K_S^0$		$< 2.6 \cdot 10^{-8}$	Belle	[125]
		$< 3.3 \cdot 10^{-8}$	BaBar	[126]
$\Gamma_{161} = \mu^- K_S^0$		$< 2.3 \cdot 10^{-8}$	Belle	[125]
		$< 4.0 \cdot 10^{-8}$	BaBar	[126]
$\Gamma_{174} = e^- f_0(980)$	lS^0	$< 3.2 \cdot 10^{-8}$	Belle	[127]
$\Gamma_{175} = \mu^- f_0(980)$		$< 3.4 \cdot 10^{-8}$	Belle	[127]
$\Gamma_{164} = e^- \rho^0$	lV^0	$< 1.8 \cdot 10^{-8}$	Belle	[128]
		$< 4.6 \cdot 10^{-8}$	BaBar	[129]
$\Gamma_{165} = \mu^- \rho^0$		$< 1.2 \cdot 10^{-8}$	Belle	[128]
		$< 2.6 \cdot 10^{-8}$	BaBar	[129]
$\Gamma_{168} = e^- K^*(892)^0$		$< 3.2 \cdot 10^{-8}$	Belle	[128]
		$< 5.9 \cdot 10^{-8}$	BaBar	[129]
$\Gamma_{169} = \mu^- K^*(892)^0$		$< 7.2 \cdot 10^{-8}$	Belle	[128]
		$< 17.0 \cdot 10^{-8}$	BaBar	[129]
$\Gamma_{170} = e^- \bar{K}^*(892)^0$		$< 3.4 \cdot 10^{-8}$	Belle	[128]
		$< 4.6 \cdot 10^{-8}$	BaBar	[129]
$\Gamma_{171} = \mu^- \bar{K}^*(892)^0$		$< 7.0 \cdot 10^{-8}$	Belle	[128]
		$< 7.3 \cdot 10^{-8}$	BaBar	[129]
$\Gamma_{176} = e^- \phi$		$< 3.1 \cdot 10^{-8}$	Belle	[128]
		$< 3.1 \cdot 10^{-8}$	BaBar	[129]
$\Gamma_{177} = \mu^- \phi$		$< 8.4 \cdot 10^{-8}$	Belle	[128]
		$< 19.0 \cdot 10^{-8}$	BaBar	[129]
$\Gamma_{166} = e^- \omega$		$< 4.8 \cdot 10^{-8}$	Belle	[128]
		$< 11.0 \cdot 10^{-8}$	BaBar	[130]
$\Gamma_{167} = \mu^- \omega$		$< 4.7 \cdot 10^{-8}$	Belle	[128]
		$< 10.0 \cdot 10^{-8}$	BaBar	[130]
$\Gamma_{178} = e^- e^+ e^-$	lll	$< 2.7 \cdot 10^{-8}$	Belle	[83]
		$< 2.9 \cdot 10^{-8}$	BaBar	[84]
$\Gamma_{181} = \mu^- e^+ e^-$		$< 1.8 \cdot 10^{-8}$	Belle	[83]
		$< 2.2 \cdot 10^{-8}$	BaBar	[84]
$\Gamma_{179} = e^- \mu^+ \mu^-$		$< 2.7 \cdot 10^{-8}$	Belle	[83]

Table 5.1 – continued from previous page

Decay mode	Category	90% CL Limit	Exp.	Ref.
		$< 3.2 \cdot 10^{-8}$	BaBar	[84]
$\Gamma_{183} = \mu^-\mu^+\mu^-$		$< 2.1 \cdot 10^{-8}$	Belle	[83]
		$< 3.3 \cdot 10^{-8}$	BaBar	[84]
		$< 4.6 \cdot 10^{-8*}$	LHCb	[76]
$\Gamma_{182} = e^-\mu^+e^-$		$< 1.5 \cdot 10^{-8}$	Belle	[83]
		$< 1.8 \cdot 10^{-8}$	BaBar	[84]
$\Gamma_{180} = \mu^-e^+\mu^-$		$< 1.7 \cdot 10^{-8}$	Belle	[83]
		$< 2.6 \cdot 10^{-8}$	BaBar	[84]
$\Gamma_{184} = e^-\pi^+\pi^-$	lhh	$< 2.3 \cdot 10^{-8}$	Belle	[131]
		$< 12.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{186} = \mu^-\pi^+\pi^-$		$< 2.1 \cdot 10^{-8}$	Belle	[131]
		$< 29.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{188} = e^-\pi^+K^-$		$< 3.7 \cdot 10^{-8}$	Belle	[131]
		$< 32.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{194} = \mu^-\pi^+K^-$		$< 8.6 \cdot 10^{-8}$	Belle	[131]
		$< 26.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{189} = e^-K^+\pi^-$		$< 3.1 \cdot 10^{-8}$	Belle	[131]
		$< 17.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{195} = \mu^-K^+\pi^-$		$< 4.5 \cdot 10^{-8}$	Belle	[131]
		$< 32.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{192} = e^-K^+K^-$		$< 3.4 \cdot 10^{-8}$	Belle	[131]
		$< 14.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{198} = \mu^-K^+K^-$		$< 4.4 \cdot 10^{-8}$	Belle	[131]
		$< 25.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{191} = e^-K^0_SK^0_S$		$< 7.1 \cdot 10^{-8}$	Belle	[125]
$\Gamma_{197} = \mu^-K^0_SK^0_S$		$< 8.0 \cdot 10^{-8}$	Belle	[125]
$\Gamma_{185} = e^+\pi^-\pi^-$	L	$< 2.0 \cdot 10^{-8}$	Belle	[131]
	L	$< 27.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{187} = \mu^+\pi^-\pi^-$	L	$< 3.9 \cdot 10^{-8}$	Belle	[131]
	L	$< 7.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{190} = e^+\pi^-K^-$	L	$< 3.2 \cdot 10^{-8}$	Belle	[131]
	L	$< 18.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{196} = \mu^+\pi^-K^-$	L	$< 4.8 \cdot 10^{-8}$	Belle	[131]
	L	$< 22.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{193} = e^+K^-K^-$	L	$< 3.3 \cdot 10^{-8}$	Belle	[131]
	L	$< 15.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{199} = \mu^+K^-K^-$	L	$< 4.7 \cdot 10^{-8}$	Belle	[131]
	L	$< 48.0 \cdot 10^{-8}$	BaBar	[132]
$\Gamma_{211} = \pi^-\Lambda$	BNV	$< 3.0 \cdot 10^{-8}$	Belle	[133]
		$< 5.8 \cdot 10^{-8}$	BaBar	[134]

Table 5.1 – continued from previous page

Decay mode	Category	90% CL Limit	Exp.	Ref.
$\Gamma_{212} = \pi^- \bar{\Lambda}$		$< 2.8 \cdot 10^{-8}$	Belle	[133]
		$< 5.9 \cdot 10^{-8}$	BaBar	[134]
$\Gamma_{213} = K^- \Lambda$		$< 4.2 \cdot 10^{-8}$	Belle	[133]
		$< 15. \cdot 10^{-8}$	BaBar	[134]
$\Gamma_{214} = K^- \bar{\Lambda}$		$< 3.1 \cdot 10^{-8}$	Belle	[133]
		$< 7.2 \cdot 10^{-8}$	BaBar	[134]
$\Gamma_{215} = \mathrm{p}\mu^-\mu^-$		$< 44.0 \cdot 10^{-8}$	LHCb	[135]
$\Gamma_{216} = \overline{\mathrm{p}}\mu^+\mu^-$		$< 33.0 \cdot 10^{-8}$	LHCb	[135]

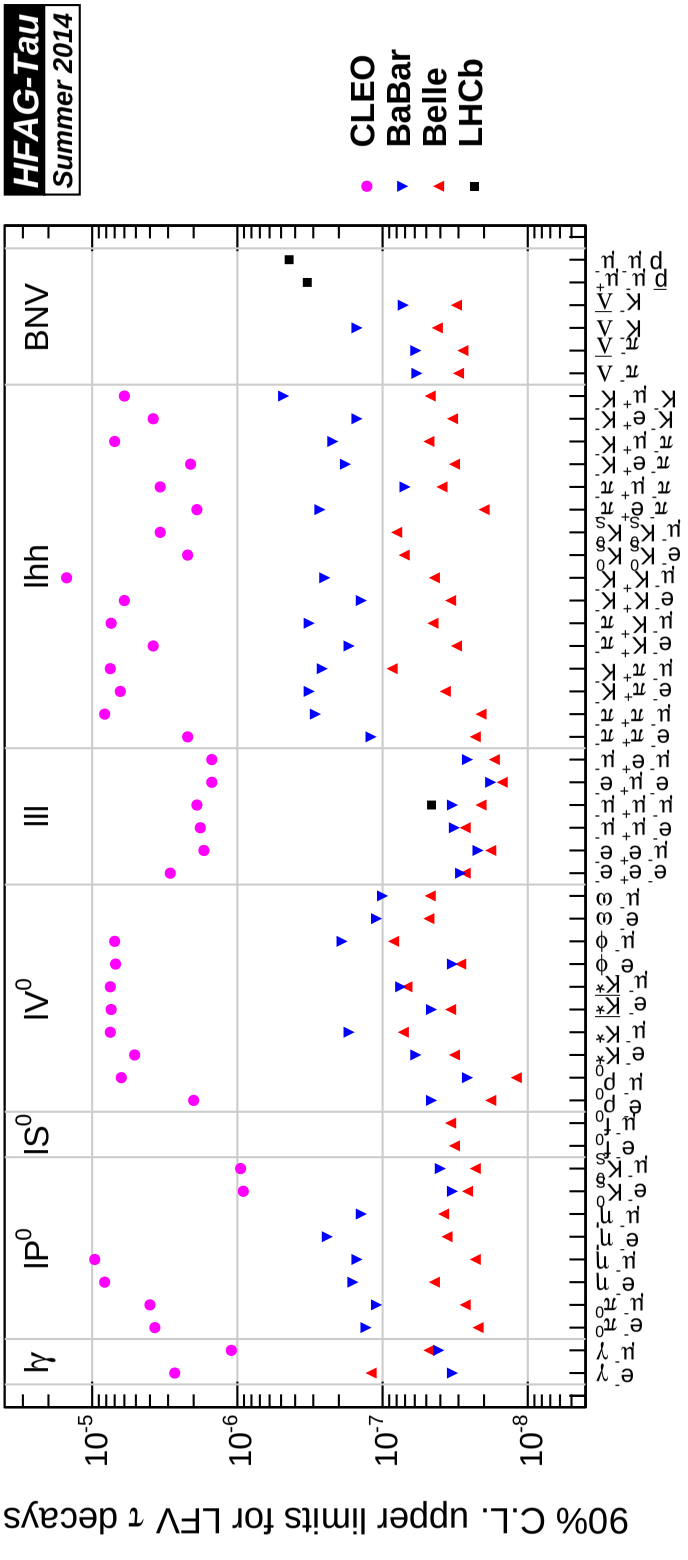


Figure 5.1: Summary plot of upper limits on branching fraction for lepton flavour violating decays of the τ lepton, as collected by the HFAG [120].

5.2 Combination of limits on lepton flavour violating τ decays

In the collection of upper limits presented in Table 5.1 several statistical methods have been used. In order to provide the combined limits in a consistent way, the HFAQ group decided to re-compute the result of each individual study with a single common approach and then perform the combination for each decay channel in question. The following procedure has been adopted:

- for each published limit a new limit was computed with the CL_s method [136] (described below), using the published information on the observed candidates, the expected background, the signal efficiency, and the number of recorded τ lepton decays,
- combination of new limits was performed with the same method to provide the HFAG combined limits.

The limits evaluated with the CL_s method are easily combined (see below) and are resilient and conservative in the presence of a downward fluctuation of the number of detected signal candidates. This has been observed for most of the published studies, corresponding to searches for τ lepton LFV decays, with the consequence that the published limits are more restrictive than the experimental sensitivity when computed with the methods such as e.g. Feldman-Cousins approach [137] (which was used in the substantial part of the discussed studies). For this reason we decided to use the CL_s approach for the limit combination.

The CL_s method is based on study of two hypotheses: the so called "signal plus background" (denoted as $s + b$) and "background only" (marked as b). The former describes the cases in which the observed spectrum can be explained in the presence of signal and background components. The latter corresponds to the existence of the background contribution with the lack of the signal. The observed confidence levels for these two hypotheses read:

$$\mathrm{CL_{s+b}} = \mathrm{P_{s+b}}(\mathrm{Q} \leq \mathrm{Q_{obs}}) = \int_{-\infty}^{\mathrm{Q_{obs}}} \frac{\mathrm{dP_{s+b}}}{\mathrm{dQ}} \mathrm{dQ}, \tag{5.1}$$

$$\mathrm{CL_{b}} = \mathrm{P_{b}}(\mathrm{Q} \leq \mathrm{Q_{obs}}) = \int_{-\infty}^{\mathrm{Q_{obs}}} \frac{\mathrm{dP_{b}}}{\mathrm{dQ}} \mathrm{dQ}, \tag{5.2}$$

where $\frac{dP_{s+b}}{dQ}$ and $\frac{dP_b}{dQ}$ are the probability distribution functions ($p.d.f.s$) for the two corresponding hypotheses and Q is called the test statistics. The CL_s value is defined as the ratio between the confidence level for the signal plus background hypothesis to the confidence level for the background hypothesis:

$$\mathrm{CL_s} = \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}}. \tag{5.3}$$

When multiple results are combined, the *p.d.f.s* in Equations 5.1 and 5.2 are the product of the individual *p.d.f.s*,

$$\mathrm{CL_s} = \frac{\prod_{\mathrm{i=1}}^{\mathrm{N_{chan}}} \sum_{\mathrm{n=0}}^{\mathrm{n_i}} \frac{\mathrm{e}^{-(\mathrm{s_i}+\mathrm{b_i})}(\mathrm{s_i}+\mathrm{b_i})^{\mathrm{n}}}{\mathrm{n!}}}{\prod_{\mathrm{i=1}}^{\mathrm{n_{chan}}} \sum_{\mathrm{n=0}}^{\mathrm{n_i}} \frac{\mathrm{e}^{-\mathrm{b_i}}\mathrm{b_i^n}}{\mathrm{n!}}} \cdot \frac{\prod_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{s_i S_i(x_{ij})} + \mathrm{b_i B_i(x_{ij})}}{\prod_{\mathrm{j=1}}^{\mathrm{n_i}} \mathrm{B_i(x_{ij})}} \ , \tag{5.4}$$

where N_{chan} is the number of results (or decay channels) and for each channel i, n_i is the number of the observed candidates, s_i and b_i are the numbers of signal and background events and S_i, B_i are the probability distribution functions of the discriminating variables. The x_{ij} are the values of the discriminating variables (with index j).

The extraction of the upper limits is performed using the code implemented by Tom Junk [138]. For each experiment we estimated the number of expected signal events using the formula:

$$\begin{aligned} s_i &= 2\mathcal{L}_i \sigma_{\tau\tau} \mathcal{B}(\tau \to \mathrm{LFV}) \\ &= \frac{\mathcal{B}(\tau \to \mathrm{LFV})}{\alpha}, \end{aligned} \tag{5.5}$$

where $\mathcal{L}_i$ is the integrated luminosity of a given experiment, $\sigma_{\tau\tau}$ is the cross-section of the process $e^+e^- \to \tau^+\tau^-$, $\mathcal{B}(\tau \to \mathrm{LFV})$ is the branching fraction of the searched process and α is the so-called normalization factor[1]. The systematics uncertainties are evaluated using Monte Carlo method by running several simulations with different values of nuisance parameters (s_i, b_i). The values are varied according to Gaussian distribution with the width equal to the corresponding systematic uncertainty.

Since most of the limits provided by B factories used Feldman-Cousin method, for the purpose of comparison in Table 5.2 we reported also individual limits calculated with $\mathrm{CL_s}$ method as well as our combined limit. These results are also summarized in Fig. 5.2. In these considerations older results obtained by the CLEO collaboration [139] were not taken into account (although they are shown in Fig. 5.1) as they gave a negligible impact. Additionally, in the cases when the upper limits of BaBar and Belle exhibited a difference at the level of an order of magnitude, we did not perform a combination.

In summary, the combined limits, provided by HFAG, illustrate a wide range of processes investigated in recent years as well as impressive progress in these searches. As can be seen in Fig. 5.1 and 5.2, the limits went down typically by two orders of magnitude over the past two decades.

For $\tau^- \to \mu^-\mu^+\mu^-$ decay, which is the main topic of this thesis, we show the full $\mathrm{CL_s}$ scan that we performed in Fig. 5.3.

[1] In the case of the LHCb results we take the normalization directly from Sect. 4.8.

Table 5.2: Collection of upper limits for the lepton flavour violating τ decay modes as combined by the HFAG group. Individual experiments limits are recalculated using CL_s method and the final combination is reported. For convenience, the decay modes are grouped in categories labelled according to their particle content (S-scalar, P-pseudovector, V-vector particle). The label "BNV" indicates that the channel is Baryon Number Violating. The result of the study described in this thesis is marked with an asterisk (*).

Decay mode	**Category**	**90% CL Limit**	**Exp.**
$\Gamma_{156} = e^-\gamma$	$l\gamma$	$< 5.4 \cdot 10^{-8}$	HFAG
		$< 22.0 \cdot 10^{-8}$	Belle
		$< 6.1 \cdot 10^{-8}$	BaBar
$\Gamma_{157} = \mu^-\gamma$		$< 5.0 \cdot 10^{-8}$	HFAG
		$< 17.0 \cdot 10^{-8}$	Belle
		$< 5.9 \cdot 10^{-8}$	BaBar
$\Gamma_{160} = e^- K_S^0$	lP^0	$< 1.4 \cdot 10^{-8}$	HFAG
		$< 1.8 \cdot 10^{-8}$	Belle
		$< 4.7 \cdot 10^{-8}$	BaBar
$\Gamma_{161} = \mu^- K_S^0$		$< 1.5 \cdot 10^{-8}$	HFAG
		$< 1.7 \cdot 10^{-8}$	Belle
		$< 6.9 \cdot 10^{-8}$	BaBar
$\Gamma_{164} = e^- \rho^0$	lV^0	$< 1.5 \cdot 10^{-8}$	HFAG
		$< 1.9 \cdot 10^{-8}$	Belle
		$< 5.2 \cdot 10^{-8}$	BaBar
$\Gamma_{165} = \mu^- \rho^0$		$< 1.5 \cdot 10^{-8}$	HFAG
		$< 2.1 \cdot 10^{-8}$	Belle
		$< 6.2 \cdot 10^{-8}$	BaBar
$\Gamma_{168} = e^- K^*(892)^0$		$< 2.3 \cdot 10^{-8}$	HFAG
		$< 3.4 \cdot 10^{-8}$	Belle
		$< 6.1 \cdot 10^{-8}$	BaBar
$\Gamma_{169} = \mu^- K^*(892)^0$		$< 6.0 \cdot 10^{-8}$	HFAG
		$< 6.6 \cdot 10^{-8}$	Belle
		$< 17.0 \cdot 10^{-8}$	BaBar
$\Gamma_{170} = e^- \bar{K}^*(892)^0$		$< 2.2 \cdot 10^{-8}$	HFAG
		$< 3.3 \cdot 10^{-8}$	Belle
		$< 5.6 \cdot 10^{-8}$	BaBar
$\Gamma_{171} = \mu^- \bar{K}^*(892)^0$		$< 4.2 \cdot 10^{-8}$	HFAG
		$< 6.3 \cdot 10^{-8}$	Belle
		$< 9.1 \cdot 10^{-8}$	BaBar
$\Gamma_{176} = e^- \phi$		$< 2.0 \cdot 10^{-8}$	HFAG
		$< 3.5 \cdot 10^{-8}$	Belle
		$< 4.3 \cdot 10^{-8}$	BaBar
$\Gamma_{177} = \mu^- \phi$		$< 6.8 \cdot 10^{-8}$	HFAG

Table 5.2 – continued from previous page

Decay mode	Category	90% CL Limit	Exp.
		$< 7.6 \cdot 10^{-8}$	Belle
		$< 18.0 \cdot 10^{-8}$	BaBar
$\Gamma_{166} = e^-\omega$		$< 3.3 \cdot 10^{-8}$	HFAG
		$< 5.2 \cdot 10^{-8}$	Belle
		$< 9.4 \cdot 10^{-8}$	BaBar
$\Gamma_{167} = \mu^-\omega$		$< 4.0 \cdot 10^{-8}$	HFAG
		$< 6.1 \cdot 10^{-8}$	Belle
		$< 10.0 \cdot 10^{-8}$	BaBar
$\Gamma_{178} = e^-e^+e^-$	*lll*	$< 1.4 \cdot 10^{-8}$	HFAG
		$< 2.7 \cdot 10^{-8}$	Belle
		$< 3.1 \cdot 10^{-8}$	BaBar
$\Gamma_{181} = \mu^-e^+e^-$		$< 1.1 \cdot 10^{-8}$	HFAG
		$< 1.7 \cdot 10^{-8}$	Belle
		$< 3.0 \cdot 10^{-8}$	BaBar
$\Gamma_{179} = e^-\mu^+\mu^-$		$< 1.6 \cdot 10^{-8}$	HFAG
		$< 2.6 \cdot 10^{-8}$	Belle
		$< 4.1 \cdot 10^{-8}$	BaBar
$\Gamma_{183} = \mu^-\mu^+\mu^-$		$< 1.2 \cdot 10^{-8}$	HFAG
		$< 2.1 \cdot 10^{-8}$	Belle
		$< 4.0 \cdot 10^{-8}$	BaBar
		$< 4.6 \cdot 10^{-8*}$	LHCb
$\Gamma_{182} = e^-\mu^+e^-$		$< 8.4 \cdot 10^{-9}$	HFAG
		$< 1.4 \cdot 10^{-8}$	Belle
		$< 2.1 \cdot 10^{-8}$	BaBar
$\Gamma_{180} = \mu^-e^+\mu^-$		$< 9.8 \cdot 10^{-9}$	HFAG
		$< 1.6 \cdot 10^{-8}$	Belle
		$< 2.6 \cdot 10^{-8}$	BaBar
$\Gamma_{211} = \pi^-\Lambda$	BNV	$< 1.9 \cdot 10^{-8}$	HFAG
		$< 3.2 \cdot 10^{-8}$	Belle
		$< 6.7 \cdot 10^{-8}$	BaBar
$\Gamma_{212} = \pi^-\bar{\Lambda}$		$< 1.8 \cdot 10^{-9}$	HFAG
		$< 2.9 \cdot 10^{-8}$	Belle
		$< 6.5 \cdot 10^{-8}$	BaBar
$\Gamma_{213} = K^-\Lambda$		$< 3.7 \cdot 10^{-9}$	HFAG
		$< 4.4 \cdot 10^{-8}$	Belle
		$< 9.2 \cdot 10^{-8}$	BaBar
$\Gamma_{214} = K^-\bar{\Lambda}$		$< 2.0 \cdot 10^{-9}$	HFAG
		$< 3.4 \cdot 10^{-8}$	Belle
		$< 5.0 \cdot 10^{-8}$	BaBar

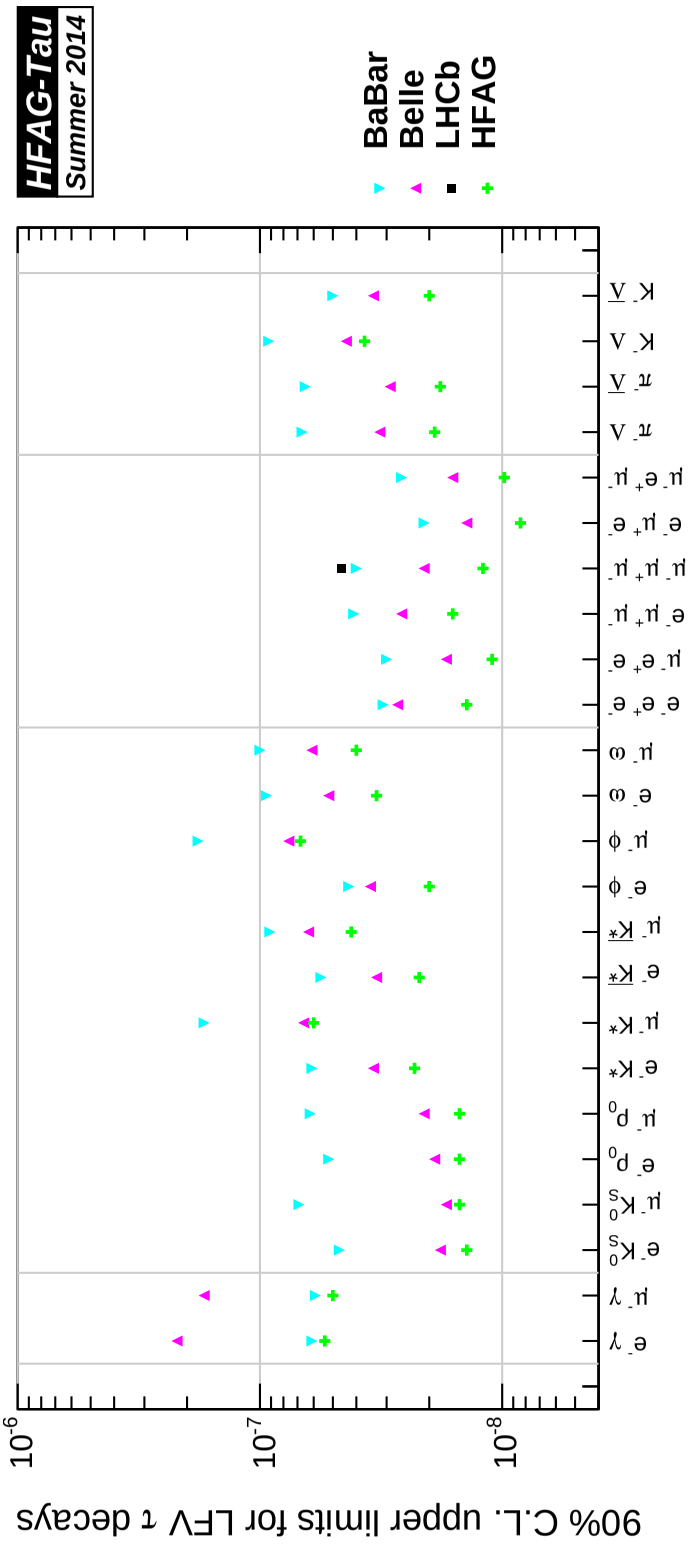


Figure 5.2: Summary plot of combined upper limits on branching fraction for lepton flavour violating decays of the τ lepton, as calculated by the HFAG [120].

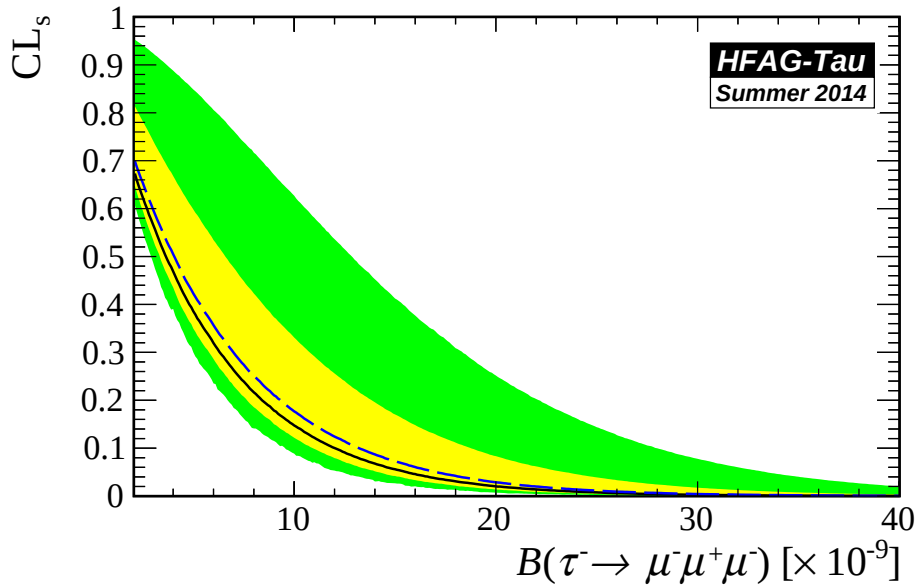


Figure 5.3: The expected CL_s curve with 68% (yellow) and 90% (green) containment bands, under the hypothesis to observe background events only for combination of LHCb, BaBar and Belle. The observed (expected) CL_s values are marked with black (blue dashed) line.

Chapter 6

Summary

The presented thesis has undertaken the search for the lepton flavour violation phenomenon in $\tau^- \to \mu^-\mu^+\mu^-$ decay using data collected in 2011 and 2012 by the LHCb collaboration at LHC. Studies of LFV are of paramount importance to elucidate new theories constituting the extensions of the Standard Model and to shed some light on such phenomena like neutrino oscillations and baryogenesis. No statistically significant signal of the decay $\tau^- \to \mu^-\mu^+\mu^-$ has been found and, as a result, the following upper limit on the branching ratio was set: $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 4.6 \times 10^{-8}$ at 90% CL. The dependence of this limit on the presence of contributions due to BSM operators with different lepton chirality structures, as described in terms of an effective field theory, was also calculated. The limit presented in this study, although it was calculated in a harsh hadronic environment, is competitive with the results obtained at B factories.

The selection criteria were implemented for the signal decay mode $\tau^- \to \mu^-\mu^+\mu^-$ and for the normalization channel, which was $D_s^+ \to \phi(\mu^-\mu^+)\pi^+$. The discrimination between a potential signal of the decay in question and the background was based on a three-dimensional binned distribution in the mass of the τ lepton candidate and two multivariate classifiers. The first classifier exploited the three-body decay topology, while the second one was based on information about muon identification. The author developed and incorporated a number of new experimental techniques, such as blending method. The analysis required a thorough discussion of trigger lines to be applied as well as careful studies of various background sources to be constructed. The latter were performed using both Monte Carlo and data events. In particular the background source originating from $D_s^- \to \eta\,(\mu^-\mu^+\gamma)\,\mu^-\nu_\mu$decays was found to be the most pronounced in this study.

The author was invited to participate in the work of the Heavy Flavour Averaging Group, where he performed a limit combination of all lepton flavour violating τ decays. In particular, the combination of the HFAG group performed on the decay in question turned out to be $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 1.2 \times 10^{-8}$ at 90% CL.

Appendix A

Isolation variables

In this appendix we discuss briefly three variables which parametrize the degree of isolation of muons w.r.t. the rest of the event. These variables have been used as input for multivariate classifiers in the first stage of the blending procedure, as described in Sect. 4.5.1.1.

A.1 Track isolation variable

The track isolation (TI) variable is constructed on the basis of the respective studies performed by the LHCb collaboration for the needs of $B_s^0 \to \mu^+\mu^-$ analysis [140]. The TI is defined as the number of extra tracks (i.e. excluding tracks that are attributed to the $\tau^- \to \mu^-\mu^+\mu^-$ candidate) that can form a vertex with a muon track. The assignment to the above SV is based on the selection criteria imposed on the following variables:

- minimum distance between the **track** and the PV (**pvdist**),
- minimum distance between the **track** and the $\tau^- \to \mu^-\mu^+\mu^-$ vertex (**svdist**),
- the distance of the closest approach between the **muon** and the **track** (DOCA),
- IP χ^2,
- angle between the **muon** and the **track** (β),
- the quantity

$$f_c = \frac{|\overrightarrow{p}_h + \overrightarrow{p}_{trk}|\alpha^{h+trk,PV}}{|\overrightarrow{p}_h + \overrightarrow{p}_{trk}|\alpha^{h+trk,PV} + p_{\mathrm{T},h} + p_{\mathrm{T},trk}}, \tag{A.1}$$

  where $\alpha^{h+trk,PV}$ is the angle between the **muon** and the **track** candidate, $P_{\mathrm{T},h}$ and $P_{\mathrm{T},trk}$ are the transverse momentum with respect to the beam line.

The track is considered as "isolated" if it satisfies the following requirements (imposed on the above mentioned variables):

- pvdist $\in [0.5, 40]$ mm,
- svdist $\in [-0.15, 30]$ mm,
- DOCA < 0.13 mm,

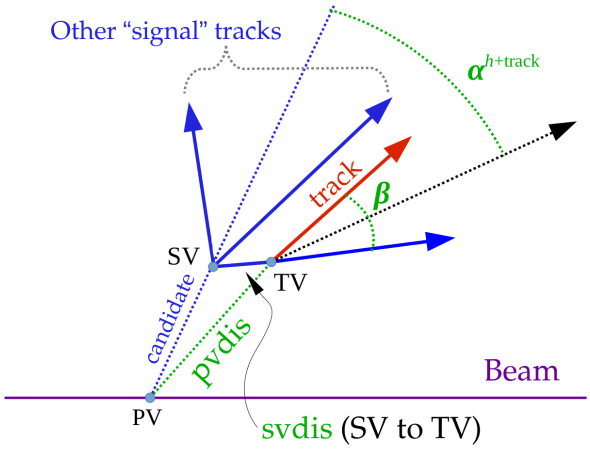


Figure A.1: Diagram showing the variables used in the track isolation variable. The description of the variables marked in the plot is provided in the text.

- Track IP significance > 3,
- $\beta < 0.27$ rad,
- $f_c < 0.6$.

A.2 BDT isolation variable

BDT isolation variable is constructed on the same set of variables as the ones used for the track isolation. However, instead of using the sequential selection criteria on the input variables, they are feed into a BDT algorithm, as described in [141].

A.3 Cone isolation variable

The cone isolation variable (CI) is defined entirely on the basis of the information about the $\tau^- \to \mu^-\mu^+\mu^-$ candidate and reads:

$$\text{Cone Isolation} = \frac{\text{p}_\text{T}(\tau)}{\text{p}_\text{T}(\tau) + \sum_\text{track} \text{p}_\text{T}(\text{track})}, \tag{A.2}$$

where $p_T(\tau)$ is the transverse momentum of the $\tau^- \to \mu^-\mu^+\mu^-$ candidate and $\sum_{track} p_T(track)$ is the algebraic sum of transverse momenta of all tracks (excluding those forming the$\tau^- \to \mu^-\mu^+\mu^-$ candidate), which satisfies the requirement $\sqrt{\delta\eta^2 + \delta\phi^2} < 1$. The $\delta\eta(\delta\phi)$ is the difference in pseudorapidity (azimuthal angle) between the track and τ candidate, respectively. This variable was first introduced by the CDF collaboration [142].

Bibliography

[1] S. Neddermeyer and C. Anderson, Phys. Rev. 51 (1937) 884.

[2] J. Street and E. Stevenson, Phys. Rev. 52 (1937) 1003.

[3] Y. Hideka, Proceedings of the Physico-Mathematical Society of Japan 17 (1935) 139.

[4] G. Feinberg, Phys. Rev. 110 (1958) 1482.

[5] AGS, `www.bnl.gov/rhic/ags.asp`.

[6] G. Danby et al., Phys. Rev. Lett. 9 (1962) 36.

[7] SLAC, `www.slac.stanford.edu/research/accelerator-research.aspx`.

[8] M. Perl et al., Phys. Rev. Lett. 35 (1975) 1489.

[9] ARGUS Collaboration, H. Albrecht et al., Phys. Lett. B292 (1992) 221.

[10] DONUT Collaboration, `www.donut.fnal.gov/`.

[11] DONUT Collaboration, K. Kodama et al., Phys. Lett. B504 (2001) 218, arXiv:hep-ex/0012035.

[12] U. Dore and L. Zanello, (2009), arXiv:0910.1657.

[13] R. Davis, J. Evans and B. Cleveland, (1978), CONF-780453–11.

[14] Kamiokande Collaboration, K. Hirata et al., Phys. Lett. B280 (1992) 146.

[15] A. Balantekin and W. Haxton, Prog. Part. Nucl. Phys. 71 (2013) 150, arXiv:1303.2272.

[16] W. Cottingham and D. Greenwood, An Introduction to Nuclear Physics (Cambridge University Press, 2001).

[17] M. Leon, Particle Physics: an Introduction (Academic Press, 1973).

[18] D. Perkins, Introduction to High Energy Physics (Cambridge University Press, 2000).

[19] K. Wilson, Phys. Rev. D 10 (1974) 2445.

[20] Q. Ho-Kim and X. Pham, Elementary Particles and Their Interactions: Concepts and Phenomena (Springer, 1998).

[21] T. Lee and C. Yang, Phys. Rev. 104 (1956) 254.

[22] C. Wu et al., Phys. Rev. 105 (1957) 1413.

[23] Particle Data Group, J. Beringer et al., Phys. Rev. D 86 (2012) 010001.

[24] A. Salam, Conf.Proc. C680519 (1968) 367.

[25] S. Weinberg, Phys. Rev. Lett. 19 (1967) 1264.

[26] S. Glashow, Nucl. Phys. 22 (1961) 579.

[27] S. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D 2 (1970) 1285.

[28] M. Kobayashi and T. Maskawa, Progress of Theoretical Physics 49 (1973) 652.

[29] Z. Maki, M. Nakagawa and S. Sakata, Progress of Theoretical Physics 28 (1962) 870.

[30] T2K Collaboration, K. Abe et al., Phys. Rev. D88 (2013) 032002, arXiv:1304.0841.

[31] M. Gonzalez-Garcia et al., JHEP 12 (2012) 123, arXiv:1209.3023.

[32] P. Skands, (2013) 341, arXiv:1207.2389.

[33] G. Zweig, (1964) 22, CERN-TH-412, NP-14146, PRINT-64-170.

[34] A. Bazavov et al., Reviews of Modern Physics 82 (2010) 1349, arXiv:0903.3598.

[35] C. Dominguez, Mod. Phys. Lett. A28 (2013) 1360002, arXiv:1305.7047.

[36] F. Hussain and G. Thompson, (1994), arXiv:hep-ph/9502241.

[37] S. Scherer, Adv. Nucl. Phys. 27 (2003) 277, arXiv:hep-ph/0210398.

[38] F. Close, Reports on Progress in Physics 42 (1979) 1285.

[39] I. Aitchison, (2005), arXiv:hep-ph/0505105.

[40] A. Melfo et al., Phys. Rev. D85 (2012) 055018, arXiv:1108.4416.

[41] M. Schmaltz, Nucl. Phys. Proc. Suppl. 117 (2003) 40, arXiv:hep-ph/0210415.

[42] L. Carpenter, A. Rajaraman and D. Whiteson, (2010), arXiv:1010.1011.

[43] S. Dimopoulos and H. Georgi, Nucl. Phys. B193 (1981) 150.

[44] G. Raz, Phys. Rev. D66 (2002) 037701, arXiv:hep-ph/0205310.

[45] P. Paradisi, JHEP 0608 (2006) 047, arXiv:hep-ph/0601100.

[46] F. Feruglio et al., Nucl. Phys. B832 (2010) 251, arXiv:0911.3874.

[47] S. Antusch et al., JHEP 0611 (2006) 090, arXiv:hep-ph/0607263.

[48] S. Martin, Adv.Ser.Direct.High Energy Phys. 21 (2010) 1, arXiv:hep-ph/9709356.

[49] M. Blanke, J. Andrzej and B. Duling, JHEP 0705 (2007) 013, hep-ph/0702136.

[50] T. Goto, Y. Okada and Y. Yamamoto, Phys. Rev. D83 (2011) 053011, arXiv:1012.4385.

[51] A. Buras et al., JHEP 1009 (2010) 104, arXiv:1006.5356.

[52] SuperB Collaboration, M. Baszczyk et al., (2013), arXiv:1306.5655.

[53] S. Turczyk, (2008), arXiv:0812.3830.

[54] B. Dassinger et al., JHEP 0710 (2007) 039, arXiv:0707.0988.

[55] R. Dalitz, Phys. Rev. 94 (1954) 1046.

[56] Z. Was, Nucl. Phys. Proc. Suppl. 98 (2001) 96, arXiv:hep-ph/0011305.

[57] CERN, `home.web.cern.ch/`.

[58] LINAC, `home.web.cern.ch/about/accelerators/linear-accelerator-2`.

[59] Booster, `home.web.cern.ch/about/accelerators/proton-synchrotron-booster`.

[60] PS, `home.web.cern.ch/about/accelerators/proton-synchrotron`.

[61] SPS, `home.web.cern.ch/about/accelerators/super-proton-synchrotron`.

[62] Accelerators at CERN, `www.nature.com/nature/journal/v448/n7151/box/nature06077_BX2.html`.

[63] LEP, `sl-div.web.cern.ch/sl-div/history/lep_doc.html`.

[64] ATLAS Collaboration, `home.web.cern.ch/about/experiments/atlas`.

[65] CMS Collaboration, `home.web.cern.ch/about/experiments/cms`.

[66] LHCb Collaboration, `home.web.cern.ch/about/experiments/lhcb`.

[67] ALICE Collaboration, `home.web.cern.ch/about/experiments/alice`.

[68] TOTEM Collaboration, `home.web.cern.ch/about/experiments/totem`.

[69] LHCf Collaboration, `home.web.cern.ch/about/experiments/lhcf`.

[70] MOEDAL Collaboration, `home.web.cern.ch/about/experiments/moedal`.

[71] R. Aaij et al., JINST 8 (2013) P04022, arXiv:1211.3055.

[72] LHCb Collaboration, A.A. Alves et al., JINST 3 (2008) S08005.

[73] Hybrid Photon Detectors, `tilde-gys.web.cern.ch/~gys/LHCb/PixelHPDs.htm`.

[74] J. Alves et al., JINST 8 (2013) 2022P, arXiv:1211.1346.

[75] Rare Decays Working Group, `twiki.cern.ch/twiki/bin/viewauth/LHCbPhysics/RareDecays?redirectedfrom=LHCb.RareDecays`.

[76] LHCb Collaboration, R. Aaij et al., (2014), arXiv:1409.8548.

[77] LHCb Collaboration, R. Aaij et al., Phys. Lett. B724 (2013) 36, arXiv:1304.4518.

[78] CKM 2012, `indico.cern.ch/event/208832`.

[79] Rencontres de Blois, `indico.cern.ch/event/296546/`.

[80] Heavy Quark Lepton, `hql2014.uni-mainz.de`.

[81] TAU 2012, `tau2012.hepl.phys.nagoya-u.ac.jp`.

[82] TAU 2014, `tau2014.physik.rwth-aachen.de`.

[83] Belle Collaboration, K. Hayasaka et al., Phys. Lett. B687 (2010) 139, arXiv:1001.3221.

[84] BaBar Collaboration, J. Lees et al., Phys. Rev. D81 (2010) 111101, arXiv:1002.4550.

[85] BRUNEL, `lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel`.

[86] DAVINCI, `lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci`.

[87] GAUSS, `lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/`.

[88] T. Sjostrand, S. Mrenna and P. Skands, JHEP 0605 (2006) 026, arXiv:hep-ph/0603175.

[89] T. Sjostrand et al., (2014), arXiv:1410.3012.

[90] Cyfronet, `cyfronet.krakow.pl/en/4421,main.html`.

[91] Cracow Cloud One, `cc1.ifj.edu.pl/en/`.

[92] C. Fitzpatrick, CERN preprint LHCb-INT-2012-031. CERN-LHCb-INT-2012-031 (2012).

[93] LHCb Collaboration, R. Aaij et al., Eur.Phys.J. C71 (2011) 1645, arXiv:1103.0423.

[94] LHCb Collaboration, R. Aaij et al., Nucl. Phys. B (2013), arXiv:1302.2864.

[95] LHCb collaboration, R. Aaij et al., JHEP 1306 (2013) 064, arXiv:1304.6977.

[96] J. Pumplin et al., JHEP 0207 (2002) 012, arXiv:hep-ph/0201195.

[97] G. Punzi, Statistical Problems in Particle Physics, Astrophysics, and Cosmology, edited by L. Lyons, R. Mount and R. Reitmeyer, p. 79, 2003, arXiv:physics/0308063.

[98] PIDCalib, `twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage`.

[99] R. Caruana et al., Proceedings of the Twenty-first International Conference on Machine Learning, ICML '04, p. 18, New York, NY, USA, 2004, ACM.

[100] T. Anderson, An Introduction to Multivariate Statistical Analysis, second ed. (Wiley, New York, NY, 1984).

[101] A. Hoecker et al., PoS ACAT (2007) 040, arXiv:physics/0703039.

[102] A. Gulin, I. Kuralenok and D. Pavlov, JMLR: Workshop and Conference Proceedings 14 (2011) 63.

[103] MatrixNet, `company.yandex.com/technologies/matrixnet.xml`.

[104] C. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics) (Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006).

[105] M. Pivk and F. Le Diberder, Nucl. Instrum. Meth. A555 (2005) 356, arXiv:physics/0402083.

[106] R. Arnaldi et al., Phys. Lett. B 677 (2009) 260.

[107] NA60 Collaboration, A. Uras, J. Phys. Conf. Ser. 270 (2011) 012038, arXiv:1108.0968.

[108] J. Rosner and S. Stone, (2012), arXiv:hep-ex/1201.2401.

[109] Belle Collaboration, A. Zupanc et al., JHEP 1309 (2013) 139, arXiv:1307.6240.

[110] CLEO Collaboration, P. Onyisi et al., Phys. Rev. D88 (2013) 032009, arXiv:1306.5363.

[111] BaBar Collaboration, P. del Amo Sanchez et al., Phys. Rev. D82 (2010) 091103, arXiv:1008.4080.

[112] Belle Collaboration, A. Zupanc, (2012), arXiv:1212.3942.

[113] BaBar Collaboration, P. del Amo Sanchez et al., Phys. Rev. D83 (2011) 052001, arXiv:1011.4190.

[114] Tracking Efficiency, . `twiki.cern.ch/twiki/bin/view/LHCb/LHCbTrackingEfficiencies`.

[115] A. Jaeger et al., CERN preprint LHCb-PUB-2011-025. CERN-LHCb-PUB-2011-025 (2012).

[116] G. Lanfranchi et al., CERN preprint LHCb-PUB-2009-013. CERN-LHCb-PUB-2009-013 (2009).

[117] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 108 (2012) 231801.

[118] A. Read, Journal of Physics G: Nuclear and Particle Physics 28 (2002) 2693.

[119] HFAG Collaboration, `www.slac.stanford.edu/xorg/hfag/`.

[120] HFAG web report, `www.slac.stanford.edu/xorg/hfag/tau/summer-2014/index.html`.

[121] Belle Collaboration, K. Hayasaka et al., Phys. Lett. B666 (2008) 16, arXiv:0705.0650.

[122] BaBar Collaboration, B. Aubert et al., Phys. Rev. Lett. 104 (2010) 021802, arXiv:0908.2381.

[123] Belle Collaboration, K. Hayasaka, J. Phys. Conf. Ser. 335 (2011) 012029.

[124] BaBar Collaboration, B. Aubert et al., Phys. Rev. Lett. 98 (2007) 061803, arXiv:hep-ex/0610067.

[125] Belle Collaboration, Y. Miyazaki et al., Phys. Lett. B692 (2010) 4, arXiv:1003.1183.

[126] BaBar Collaboration, B. Aubert et al., Phys. Rev. D79 (2009) 012004, arXiv:0812.3804.

[127] Belle Collaboration, Y. Miyazaki et al., Phys. Lett. B672 (2009) 317, arXiv:0810.3519.

[128] Belle Collaboration, Y. Miyazaki, Phys.Lett. B699 (2011) 251, arXiv:1101.0755.

[129] BaBar Collaboration, B. Aubert et al., Phys. Rev. Lett. 103 (2009) 021801, arXiv:0904.0339.

[130] BaBar Collaboration, B. Aubert et al., Phys. Rev. Lett. 100 (2008) 071802, arXiv:0711.0980.

[131] Belle Collaboration, Y. Miyazaki et al., Phys. Lett. B719 (2013) 346, arXiv:1206.5595.

[132] BaBar Collaboration, B. Aubert et al., Phys. Rev. Lett. 95 (2005) 191801, arXiv:hep-ex/0506066.

[133] K. Hayasaka, Nuclear Physics B - Proceedings Supplements 225 (2012) 169 , Proceedings of the International Workshop on e^+e^- collisions from Phi to Psi.

[134] BaBar, G. Lafferty, Nucl. Phys. Proc. Suppl. 169 (2007) 186.

[135] LHCb, R. Aaij et al., Phys. Lett. B724 (2013) 36, arXiv:1304.4518.

[136] B. Mistlberger and F. Dulat, (2012), arXiv:1204.3851.

[137] G. Feldman and R. Cousins, Phys. Rev. D 57 (1998) 3873.

[138] CDF, T. Junk, (2007), CDF/DOC/STATISTICS/PUBLIC/8128.

[139] CLEO Collaboration, `www.phys.ufl.edu/hee/cleo/`.

[140] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 110 (2013) 021801, arXiv:1211.2674.

[141] L. Gavardi, Search for lepton flavour violation in τ decays at the LHCb experiment, Master's thesis, Milan Bicocca U., 2013, Presented 28 Nov 2013, `cds.cern.ch/record/1645251`.

[142] CDF Collaboration, A. Abulencia et al., Phys. Rev. Lett. 95 (2005) 221805, arXiv:hep-ex/0508036.