

DISSERTATION

submitted to the
Combined Faculties for the Natural Sciences and Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of

Doctor of Natural Sciences

Put forward by

Dipl.-Phys. Paul Seyfert

born in Frankfurt am Main

Oral examination: 14th January 2015

The search for $\tau \rightarrow \mu\mu\mu$ at LHCb

Referees:
Prof. Dr. Ulrich Uwer
Prof. Dr. André Schöning

Abstract

The charged lepton flavour violating decay $\tau \rightarrow \mu\mu\mu$ is searched for, using the LHCb experiment. Violation of lepton flavour in the charged lepton sector is unobserved to date. Within the Standard Model of particle physics including neutrino oscillation, the branching fraction is expected to be unmeasurable small and an observation would be an unambiguous sign for physics beyond the Standard Model.

Over 10^{11} τ leptons have been produced in proton-proton collisions at LHCb during the first run of the LHC. Most of them in decays of D_s mesons. Compared to previous experiments at electron-positron colliders, the signature of $\tau \rightarrow \mu\mu\mu$ is harder to identify in hadronic collisions and background processes are more abundant.

A multivariate event classification has been developed to distinguish a possible signal from background events. The number of τ leptons produced in the LHCb acceptance is estimated by measuring the yield of $D_s \rightarrow \phi(\mu\mu)\pi$ decays. The sensitivity reached by analysing LHCb data corresponding to 3 fb^{-1} is sufficient to constrain the branching fraction of $\tau \rightarrow \mu\mu\mu$ to be smaller than 7.1×10^{-8} at 90 % confidence level.

Kurzfassung

Am LHCb Experiment wird nach dem Leptonfamilienzahl verletzenden Zerfall $\tau \rightarrow \mu\mu\mu$ gesucht. Die Verletzung der Leptonfamilienzahl ist bei geladenen Leptonen bisher nicht beobachtet worden. Auch unter Berücksichtigung der Neutrinooszillation wird ein unbeobachtbar kleines Verzweungsverhältnis im Standardmodell der Teilchenphysik vorher gesagt. Jegliche Beobachtung des Zerfalls wäre ein eindeutiges Zeichen für Physik jenseits des Standardmodells.

Über 10^{11} τ Leptonen wurden in Proton-Proton Kollisionen bei LHCb während der ersten Betriebsperiode des LHC produziert, die meisten davon in Zerfällen von D_s Mesonen. Verglichen mit früheren Experimenten, bei denen Elektron-Positron Kollisionen untersucht wurden, ist die Signatur des $\tau \rightarrow \mu\mu\mu$ Zerfalls in hadronischen Kollisionen schwerer zu identifizieren und Untergrundprozesse sind zahlreicher.

Um ein mögliches Signal von Untergrundereignissen zu unterscheiden wurde eine multivariate Klassifizierung von Ereignissen entwickelt. Die Anzahl an τ Leptonen, die in der LHCb Detektorakzeptanz produziert wurden, wird anhand der beobachteten $D_s \rightarrow \phi(\mu\mu)\pi$ Ereignisse bestimmt. Die Analyse von LHCb Daten, die einer integrierten Luminosität von 3 fb^{-1} entsprechen, ermöglicht es, Verzweungsverhältnisse von $\tau \rightarrow \mu\mu\mu$, die größer als 7.1×10^{-8} sind, bei einem 90 % Konfidenzintervall auszuschließen.

Contents

Introduction	v
1 Theoretical background	1
2 The LHCb experiment	15
3 Branching fraction limits in absence of a signal	33
4 τ production at LHCb	43
5 Analysis strategy	55
6 Event selection	63
7 Online event selection	71
8 Background processes to $\tau \rightarrow \mu\mu\mu$	75
9 $\mathcal{M}_{3\text{body}}$ development	79
10 Likelihood calibrations	95
11 Normalisation	109
12 Limits for the $\tau \rightarrow \mu\mu\mu$ decay	119
13 Model dependence	133
14 Conclusion	139
A Toy studies on classifier binning	141
B Statistics glossary	153
C Tag-and-probe and TISTOS	159
D b hadronisation at LHC	161
E τ production fractions	163

Introduction

The principles of symmetries and conservation laws have been essential for the understanding of nature. The two are connected by the Noether's theorem [1]. Symmetries and gauge invariance are principles of the Standard Model of particle physics, which describes today's knowledge of fundamental particles and their interactions. Despite its success, the Standard Model is believed to be only an effective theory which is valid within the range which is probed by particle physics experiments today. It can however not explain the matter/anti-matter asymmetry in the universe, does not describe gravity, or contains an explanation for dark matter.

To investigate the fundamental laws of particle physics further, the Large Hadron Collider (LHC), situated at CERN, has been built. It is designed to collide proton beams at centre-of-mass energies of up to $\sqrt{s} = 14$ TeV at four interaction points. One of its experiment is the LHCb experiment, an experiment dedicated to investigate beauty and charm particle decays. The during the first run of the LHC, from late 2009 to early 2013, protons have been collided at up to $\sqrt{s} = 8$ TeV and the LHCb experiment collected data corresponding to an integrated luminosity of 3 fb^{-1} .

Lepton flavour is an almost conserved quantity without a fundamental symmetry protecting its conservation within the Standard Model. In extensions of the Standard Model this lack of symmetry protection can lead to observable branching fractions for lepton flavour violating decays, like the decay $\tau \rightarrow \mu\mu\mu$.

Violation of lepton flavour in τ decays has been searched for at e^+e^- colliders and the most stringent limits on the branching fraction $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ have been set by the B factories BaBar and Belle at the order of $10^{-8} - 10^{-7}$ [2]. For the decay $\tau \rightarrow \mu\mu\mu$ the Belle collaboration reports $\mathcal{B} < 2.1 \times 10^{-8}$ [3].

In this work, the decay $\tau \rightarrow \mu\mu\mu$ is searched for using data collected by the LHCb experiment during the first run of the Large Hadron Collider. An essential prerequisite for performing the search is a large τ production. At LHCb about 10^{11} τ leptons have been produced during the first run of the LHC. It was still argued in the past whether a search for $\tau \rightarrow \mu\mu\mu$ is possible at a hadron collider due to the large amount of backgrounds in hadron collisions.

This work is organised in three parts. Firstly, an introductory part recapitulates the theoretical background relevant for $\tau \rightarrow \mu\mu\mu$ (Chap. 1), gives a rough description of the experiment (Chap. 2), and explains the statistics framework (Chap. 3) and the mechanisms of τ production at LHC collisions (Chap. 4).

The second part explains the steps of the search for $\tau \rightarrow \mu\mu\mu$ at LHCb. These are the search strategy (Chap. 5), the event selection (Chaps. 6 and 7), the background processes (Chap. 8), how these are discriminated from a possible signal with

multivariate classification techniques (Chap. 9) and the calibration of that discrimination with $D_s \rightarrow \phi(\mu\mu)\pi$ and $b \rightarrow J/\psi(\mu\mu)X$ decays (Chap. 10). Eventually, the analysis's normalisation factor is determined in Chap. 11 and the amount of residual background events estimated in Sect. 12.1.

The results are presented in the third part. It contains the actual results in Chap. 12 and a review of whether the result is applicable to any new physics model in Chap. 13.

This work is part of the analyses presented by the LHCb collaboration in Refs. [4] and [5]. The author of this thesis is one of the contact authors of both publications. Both analyses were carried out in collaboration with members of LHCb and are internally documented in Refs. [6, 7, 8]. Two multivariate classifiers are used, $\mathcal{M}_{3\text{body}}$ and \mathcal{M}_{PID} , the former is developed for the present analysis and most of Chap. 9 is dedicated to its development. Reference [5] and the present work differ in the multivariate classifier $\mathcal{M}_{3\text{body}}$. A classifier developed by the author is used in the present work. The classifier in [5] is an upgrade of the classifier used here, it combines several TMVA classifiers in an ensemble selection [9] with a MatrixNet boosted decision tree [10].

The choice which classifier to describe in this work was made prior to the finalisation of the development of the classifier in [5]. Furthermore, the classifier choice must be done in an unbiassing way, as outlined in Sect. 5.4. Since all $D_s \rightarrow \phi(\mu\mu)\pi$ decays recorded by LHCb are used for the final calibration of $\mathcal{M}_{3\text{body}}$ in [5], the classifier choice cannot be done by means of the calibrated classifier performance: the classifier choice must be independent of the final calibration and there is no independent calibration data left.

Prior to the calibration, the performance difference of $\mathcal{M}_{3\text{body}}$ in this work and $\mathcal{M}_{3\text{body}}$ in Ref. [5] is estimated to be 6% ([8, 11]). The author regarded 6% as small enough to document his own classifier in the present document, not to change to the classifier from [5] once this number was known.

Comparing the final result of [5] to that in Chap. 12, after unblinding of the signal region, it appears that the analysis presented here observes a weaker exclusion limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ than [5]. A breakdown of the contributions to this difference is given at the end of Chap. 12. The largest part of the difference in the observed exclusion limit is a Poissonian fluctuation of the observed event count.

The work presented in Chaps. 7, 9, 10, and 13 have been developed by the author in the context of this thesis and the results of Chaps. 7, 10, and 13 entered in the LHCb analysis [5]. Furthermore, the work in Chap. 4 is repeated with minor modifications with respect to previous work by Jonathan Harrison. The author adapted previous work by Basem Khanji in Sect. 12.1, and the methods in Chap. 12 have been introduced in LHCb by the $B_s^0 \rightarrow \mu\mu$ analysis group and have been adapted by the author for the search for $\tau \rightarrow \mu\mu\mu$.

Theoretical background

1.1 The Standard Model of particle physics

Today's understanding of the fundamental particles and their interactions is described by the Standard Model of particle physics [13, 14, 15]. The fundamental fermions are either quarks or leptons. Both occur in three generations with two particles in each generation. For each fermion, there exists an antifermion, too. The fermions are listed in Tab. 1.1.

The quark generations are the up (u) and down (d) quark, the charm (c) and strange (s) quark, and the top (t) and bottom (b) quark. One of each generation is a “up type” quark and has the electric charge $2/3 e$, the other member of each generation is a “down type” quark and has the electric charge $-1/3 e$. Each quark occurs as left handed or as right handed particle. Furthermore, each quark occurs with three different colour charges of the strong interaction.

In the lepton sector, there is a charged lepton with charge $-e$ and an uncharged neutrino in each lepton generation. The charged leptons are the electron (e), the muon (μ), and the tau (τ). The corresponding neutrinos are the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ). The charged leptons occur as left handed and right handed particles. The neutrinos are traditionally massless in the Standard Model and occur only left handed. Leptons don't carry colour charge.

The Standard Model is a gauge theory with the gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$. Local gauge invariance is granted by the gauge fields. The excitations of these gauge fields are gauge bosons. These are 8 gluons (g) for $SU(3)_c$ which couple to colour charge. The gauge bosons of $SU(2)_L$ are the $W_{(1,2,3)}$ which are coupling to the weak

Table 1.1: Fermion content of the Standard Model and their electro-weak charges, in natural units ($c = \hbar = e = 1$).

particles	J_3	Y	Q
left handed neutrinos (ν_e, ν_μ, ν_τ) $_L$	$1/2$	-1	0
left handed charged leptons (e, μ, τ) $_L$	$-1/2$	-1	-1
right handed charged leptons (e, μ, τ) $_R$	0	-2	-1
left handed up type quarks (u, c, t) $_L$	$1/2$	$1/3$	$2/3$
left handed down type quarks (d, s, b) $_L$	$-1/2$	$1/3$	$-1/3$
right handed up type quarks (u, c, t) $_R$	0	$4/3$	$2/3$
right handed down type quarks (d, s, b) $_R$	0	$-2/3$	$-1/3$

1. THEORETICAL BACKGROUND

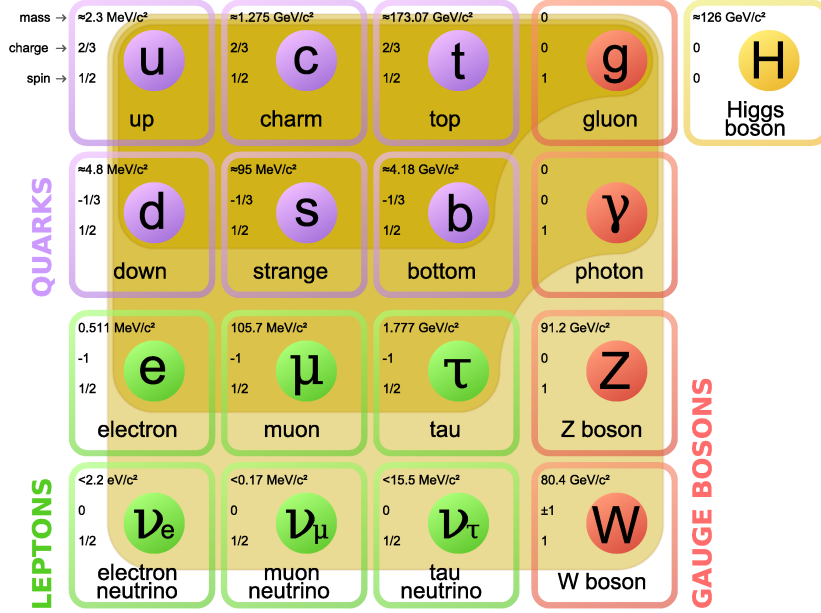


Figure 1.1: Illustration of the particle content of the Standard Model of particle physics. Figure modified from [12]. The current version indicates the fermion–boson couplings with the yellow backgrounds. The Yukawa couplings and the couplings of the bosons among each other are not indicated.

isospin J of left handed fermions, and B , the gauge boson of $U(1)_Y$ which is coupling to the fermions' hypercharge Y .

The latter quantum numbers are related to the electric charge Q by the Gell-Mann-Nishijima relationship

$$Q = J_3 + \frac{Y}{2}$$

where J_3 is the projection of J .

The observable force mediator bosons are linear combinations of the gauge bosons W_i and B , as determined by the weak mixing angle ϑ_W

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \vartheta_W & \sin \vartheta_W \\ -\sin \vartheta_W & \cos \vartheta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \pm W_\mu^2).$$

Since the W bosons couple to J , only left handed particles undergo weak processes with W^\pm bosons. Particles are converted into the other family member by emission or absorption of a W^\pm boson, e.g. $u_L \rightarrow W^+ d_L$ or $W^+ \mu^- \rightarrow \nu_\mu$. Here, u_L and d_L are eigenstates of the weak interaction. The weak eigenstates (flavour eigenstates) of quarks do not coincide with the mass eigenstates as described by the Cabbibo-

Table 1.2: Lepton flavour quantum numbers of Standard Model particles.

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

Kobayashi-Maskawa (CKM) matrix

$$\begin{aligned}
 & (\text{u} \quad \text{c} \quad \text{t})_{\text{weak}} \cdot \begin{pmatrix} \text{d} \\ \text{s} \\ \text{b} \end{pmatrix}_{\text{weak}} \\
 &= (\text{u} \quad \text{c} \quad \text{t})_{\text{mass}} \cdot \underbrace{\begin{pmatrix} U_{ud} & U_{us} & U_{ub} \\ U_{cd} & U_{cs} & U_{cb} \\ U_{td} & U_{ts} & U_{tb} \end{pmatrix}}_{\text{CKM matrix}} \cdot \begin{pmatrix} \text{d} \\ \text{s} \\ \text{b} \end{pmatrix}_{\text{mass}} .
 \end{aligned}$$

The Z couples to left and right handed particles at different strength. Fermion-Z vertices, in contrast to W vertices, do not alter a particle's flavour. The latter holds for photon vertices as well, while photons couple to left and right handed particles equally strong because they have the same electric charge.

The lepton flavour numbers in Tab. 1.2 are assigned to the Standard Model particles. Couplings to Z and γ do not alter the particle type of a fermion and W couplings only convert leptons within a lepton generation. Therefore, lepton flavour is conserved in the electro-weak interaction. The strong interaction, see below, does not involve leptons at all, and therefore cannot change lepton flavour. Lepton flavour eigenstates can be defined to coincide with the mass eigenstates of charged leptons¹. Consequently, charged lepton propagators conserve lepton flavour. If neutrinos are considered massless then the neutrino flavour eigenstates are mass eigenstates, too. Therefore, in that case any neutrino propagator would conserve lepton flavour in the Standard Model.

The additional so-called strong interaction is quantum chromodynamics, QCD [16, 17, 18]. The massless gauge bosons of the $SU(3)_c$ group are eight gluons which couple to the quarks' colour charge. While quarks carry a colour charge and antiquarks carry an anticolour charge, the gluons carry one colour and one anticolour charge.

Two key features of QCD are *confinement* and *asymptotic freedom*. Confinement is the property that quarks do not occur as free particles but are always confined to hadron, i. e. mesons or baryons. The former are bound states of a quark and an antiquark, the latter are bound states of three quarks. In both cases, the colours of the quarks and antiquarks result in a colour neutral state. This is a result of

¹The same is done for up type quarks in the quark sector.

the running coupling behaviour of α_s , the coupling constant of QCD. At very low momentum transfer, α_s becomes large. At large momentum transfer, α_s becomes small and quarks become free particles.

Massive gauge bosons are incompatible with gauge invariance. The W and Z bosons, and the massive fermions gain their masses through the Higgs mechanism [19, 20, 21]. The crucial aspect of the Higgs field is its non vanishing vacuum expectation value. An excitation of the Higgs field was predicted to be observable as a Higgs particle. The Higgs boson has been discovered at the LHC [22, 23]. The mass of fermions is a consequence of the coupling strength of the Higgs to the fermion fields (called Yukawa coupling).

Shortcomings of the Standard Model

The Standard Model is extremely successful at describing particle physics experimental data. It is still only an effective theory since it lacks a description of gravity and does not contain dark matter or dark energy, which make up $\sim 95\%$ of the universe. Additionally, the observed excess of matter over antimatter in the universe is not explained [24, 25].

Another property² of the Standard Model is *hierarchy* [26, 27, 28]. Radiative corrections to the Higgs mass diverge quadratically and should therefore result in a very large Higgs mass. The observed Higgs mass, however, is much smaller than these corrections, which suggests that the Standard Model parameters must be *fine tuned* to very high accuracy.

Neutrino oscillation

The results of all attempts to directly measure the mass of neutrinos so far resulted only in upper limits consistent with vanishing neutrino masses. The discovery of neutrino oscillation [29, 30, 31], however, implies that the neutrino mass differences $(m_{\nu_i} - m_{\nu_j})^2$ are different from zero. It is thus indirectly established that at least two neutrinos are massive. Furthermore, lepton mass and lepton flavour eigenstates can be defined to coincide with each other either in the charged lepton sector or in the neutrino sector, but not simultaneously. Similar to the CKM matrix in the quark sector, the base change between neutrino mass eigenstates (ν_1, ν_2, ν_3) and flavour eigenstates $(\nu_e, \nu_\mu, \nu_\tau)$ is given by a unitary matrix.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \underbrace{\begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix}}_{\text{PMNS matrix}} \cdot \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

The unitary matrix is the unitary PMNS matrix [32]. It has not been resolved yet, whether neutrinos are their own antiparticles, $\nu \stackrel{?}{=} \bar{\nu}$. Fermions which are their own antiparticles are called Majorana particles, as opposed to Dirac particles. If neutrinos are Dirac particles, the PMNS matrix has the same number of degrees of freedom as the CKM matrix. Otherwise there are two additional Majorana phases.

²Since it is often regarded a problematic property, the term *hierarchy problem* is used.

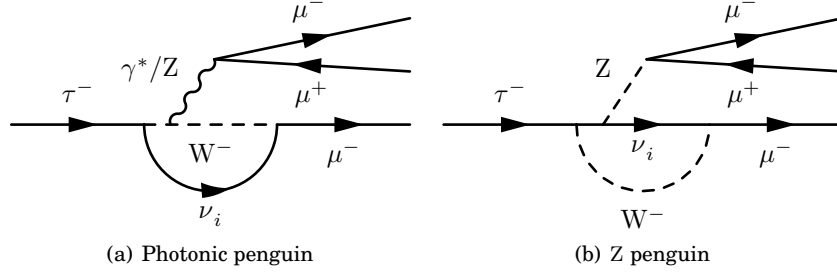


Figure 1.2: Feynman diagrams for $\tau \rightarrow \mu\mu\mu$ in the Standard Model with massive neutrino. The neutrino mass $\nu_{(1,2,3)}$ eigenstates do not oscillate. In the interference of diagrams with different mass eigenstates, the oscillation $\nu_\tau \rightarrow \nu_\mu$ appears.

$\tau \rightarrow \mu\mu\mu$ in the Standard Model

Given the shortcomings of the Standard Model, theories for physics beyond the Standard Model (BSM) have been suggested. One of the observables for which BSM models predict strong deviations from the Standard Model prediction in experimental reach, is the branching fraction $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$.

In the Standard Model without neutrino oscillation, every propagator has a defined lepton flavour and every vertex conserves lepton flavour. Therefore any perturbative Standard Model calculation, i. e. a finite combination of these vertices and propagators, will therefore lead to $\mathcal{B}(\tau \rightarrow \mu\mu\mu) = 0$. Once considering neutrino oscillation, the neutrino propagators, as $\nu_{(1,2,3)}$, do not have a defined lepton flavour anymore and flavour change can occur as in the penguin diagrams in Fig. 1.2. Each of the penguin diagrams is GIM suppressed. The amplitude of each penguin is proportional to the corresponding PMNS matrix elements $U_{\tau,i}U_{\mu,i}^*$ and has a kinematic dependence

$$f\left(\frac{m_{\nu_i}}{m_W}\right) = 1 + a \cdot \frac{m_{\nu_i}^2}{m_W^2}$$

with an expansion coefficient a . Summing over all neutrino generations gives

$$\begin{aligned} \mathcal{A} &\propto \sum U_{\tau,i}U_{\mu,i}^* \left(1 + a \cdot \frac{m_{\nu_i}^2}{m_W^2}\right) \\ &= \underbrace{\sum (U_{\tau,i}U_{\mu,i}^*)}_{=0 \text{ unitarity of the PMNS matrix}} + \sum \left(a \cdot U_{\tau,i}U_{\mu,i}^* \frac{m_{\nu_i}^2}{m_W^2}\right). \end{aligned}$$

The first summand is zero due to the unitarity of the PMNS matrix. The second summand is strongly suppressed; in a two flavour system, it would reduce to $a \cdot b \cdot \Delta m_{\nu_i}^2 / m_W^2$, where b depends on the mixing angle³.

³In the 2 flavour case, the PMNS matrix reduces to a rotation matrix containing a single mixing angle.

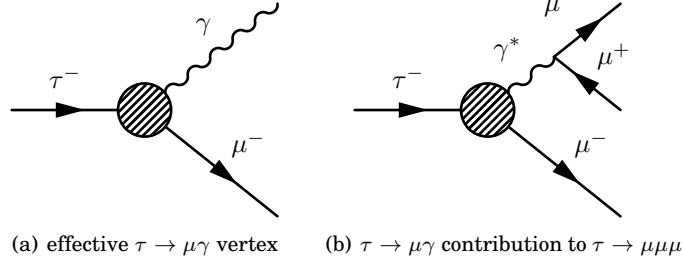


Figure 1.3: Feynman diagram for $\tau \rightarrow \mu\gamma$ in an effective theory and the contribution of that diagram to $\tau \rightarrow \mu\mu\mu$.

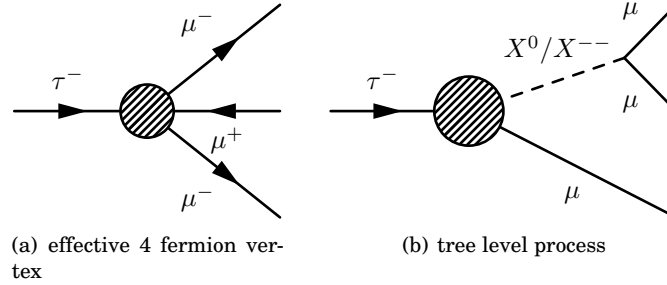


Figure 1.4: Feynman diagrams for effective theories contributing to $\tau \rightarrow \mu\mu\mu$ which do not contribute to $\tau \rightarrow \mu\gamma$.

1.2 Models for physics beyond the Standard Model with charged lepton flavour violation

Overviews over BSM models, which predict a significant deviation from the Standard Model prediction for $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$, have been made from the experimental side by several authors, e. g. [33, 34, 35, 36, 37, 38].

The implications of BSM models for $\tau \rightarrow \mu\mu\mu$ are often studied along with the lepton flavour violating decay $\tau \rightarrow \mu\gamma$. If the latter decay is enhanced in a BSM scenario, then $\tau \rightarrow \mu\mu\mu$ receives a contribution, too, when the photon is off-shell, as shown in Fig. 1.3. The branching fraction is in the absence of additional $\tau \rightarrow \mu\mu\mu$ diagrams expected to be larger for $\tau \rightarrow \mu\gamma$ since the decay with three leptons in the final state requires an additional $\gamma^*\mu\mu$ vertex.

The search for $\tau \rightarrow \mu\mu\mu$ is of particular interest to find evidence or rule out BSM models in which $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is enhanced with respect to $\tau \rightarrow \mu\gamma$ through diagrams from which the latter decay does not arise, i. e. diagrams like Fig. 1.4⁴.

It is pointed out that most of the theory papers referred to in this section have been published prior to most LHC results. The discovery of the Higgs boson, the measurement of $\mathcal{B}(B_s^0 \rightarrow \mu\mu)$, the measurement of the neutrino mixing angle $\sin\theta_{13}$, and the exclusions from searches for BSM effects are therefore not taken into account. Where easily possible, it is commented whether the models in question have been ruled out by recent findings.

⁴This argument has been given already by e. g. [39] or [40].

1.2. Models for physics beyond the Standard Model with charged lepton flavour violation

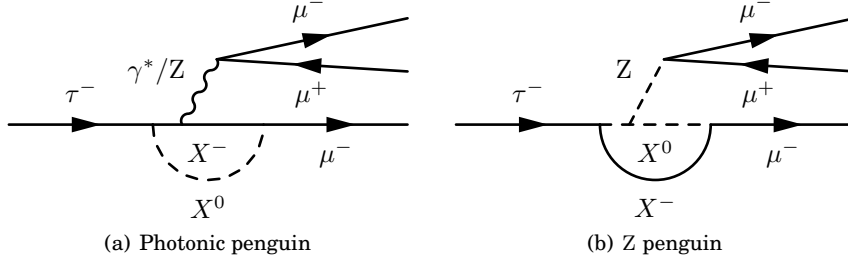


Figure 1.5: Penguin diagrams for $\tau \rightarrow \mu\mu\mu$ in models with extended particle content with respect to the Standard Model. The photonic penguin can contribute to $\tau \rightarrow \mu\gamma$, too.

Table 1.3: Overview over BSM models which predict an enhanced branching fraction for $\tau \rightarrow \mu\mu\mu$. A value of 10^{-8} is given where the current exclusion limit is used as model constrain already.

model	maximal $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$	$\frac{\mathcal{B}(\tau \rightarrow \mu\mu\mu)}{\mathcal{B}(\tau \rightarrow \mu\gamma)}$	reference
Standard Model + neutrino oscillation	10^{-54}	0.002	e. g. (44)
MSSM (photon dominant)	10^{-8}	0.002	(45)
MSSM (Higgs dominant)	4×10^{-10}	0.06...0.1	(46, 47)
<i>R</i> -parity violating SUSY	10^{-8}	> 1	(48)
LHT	2×10^{-8}	0.04...0.4	(49)
L-R symmetry	10^{-7}	$\gg 1$	(50)
4th generation	7.1×10^{-8}	0.06...2.2	(51)
Non-universal Z' / technicolor TC2	4×10^{-8}	3...100	(52)
unparticles	10^{-3}	possibly 10^3	(53)

As mentioned in [41, 42, 43] and [40, Note therein], the Z penguin, shown in Fig. 1.5, is usually neglected although it can lead to a large modifications of the branching fraction $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ and modify the ratio $\mathcal{B}(\tau \rightarrow \mu\mu\mu)/\mathcal{B}(\tau \rightarrow \mu\gamma)$.

The models presented in this section are summarised in Tab. 1.3. For some models the prediction for $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is out of experimental reach, in other cases the experimental constraint for $\mathcal{B}(\tau \rightarrow \mu\gamma)$ implies that $\tau \rightarrow \mu\mu\mu$ will not be seen within this model. An evidence for $\tau \rightarrow \mu\mu\mu$ is still possible due to effects of any of the other models. Conversely, those remaining models can be constrained or ruled out by improving the exclusion limit for $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$.

Minimal Supersymmetric Standard Model (MSSM) without Higgs mediation

Supersymmetry has been introduced as an attempt to solve the hierarchy problem by introducing bosonic partners to all Standard Model fermions and fermionic partners to the bosons.

The decay amplitude for $\tau \rightarrow \mu\mu\mu$ is dominated by photonic diagrams, as shown in Fig. 1.3 (b). Computations in [45] show that branching fractions up to the current exclusion limits are possible within MSSM. In MSSM photonic diagrams are found to dominate the decay amplitude [54] and the relation

$$\begin{aligned} \mathcal{B}(\ell_i \rightarrow 3\ell_j) &\simeq \frac{\alpha}{3\pi} \left(\log \left(\frac{m_{\ell_i}^2}{m_{\ell_j}^2} \right) - \frac{11}{4} \right) \mathcal{B}(\ell_i \rightarrow \ell_j \gamma) \\ &\simeq 0.002 \cdot \mathcal{B}(\ell_i \rightarrow \ell_j \gamma) \end{aligned}$$

is found.

MSSM with Higgs mediated LFV

As pointed out in [55], lepton flavour violation in MSSM may not only occur through a photon penguin diagram, but may be due to a Higgs propagator, as in Fig. 1.4 (b). However, already in [46], the constraints from $\mathcal{B}(B_s^0 \rightarrow \mu\mu)$ and $\mathcal{B}(\ell \rightarrow \ell' \gamma)$ were found to require $\mathcal{B}(\tau \rightarrow \mu\mu\mu) \lesssim 4 \times 10^{-10}$.

R-parity violating SUSY

The Z penguin diagrams (Fig. 1.5 (b)), which cancel in the MSSM, do not cancel once R-parity violation is possible, as derived in [48]. The existing limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ already constrain the parameter space for trilinear R-parity violating couplings in [48]. R-parity violation is reviewed in [56].

Little Higgs with T parity

The little Higgs model with T parity is an alternative to the MSSM. It requires less parameters than MSSM and makes different predictions for charged lepton flavour violation.

As explained in [57], “little Higgs models [58, 59, 60] were proposed as a solution to the little hierarchy problem. In these models, the Higgs doublet field appears as pseudo Nambu-Goldstone (NG) bosons of new strong dynamics at the cutoff scale. [...] In the littlest Higgs model with T parity (LHT) [61, 62], the model is extended to have a Z_2 parity so that the heavy gauge bosons assigned to be T-odd particles do not directly couple with a pair of the SM fermions, and the phenomenological constraints are somewhat relaxed.”

In [49] it is found that $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ can be as large as 2×10^{-8} , depending on the scale of symmetry breaking. The ratio $\mathcal{B}(\tau \rightarrow \mu\mu\mu)/\mathcal{B}(\tau \rightarrow \mu\gamma)$ is between 0.04 and 0.4.

It must however be pointed out that the findings in [57] suggest that searches for $\mu \rightarrow eee$ are better suited for probing LHT. Furthermore, as explained in [49], a large value for $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ requires a small value for the scale of symmetry breaking f . In [63, Fig. 7], small values for f require a large mass of the Higgs boson $m_h \gg 500 \text{ GeV}/c^2$.

Left-Right symmetric model

The Left-Right (LR) symmetric models are extensions of the Standard Model based on the gauge group $SU(3)_c \times SU(2)_R \times SU(2)_L \times U(1)_{B-L}$ [64]. LR models introduce

parity violation as a symmetry breaking and include right handed neutrinos for type I or type II seesaw mechanisms.

The lepton flavour violating decay $\tau \rightarrow \mu\mu\mu$ is established in the LR model at tree level level, $\tau^\pm \rightarrow H^{\pm\pm}(\mu^\pm\mu^\pm)\mu^\mp$, as in Fig. 1.4 (b) with a doubly charged mediator. This leads to an enhancement of $\mathcal{B}(\tau \rightarrow \mu\mu\mu) \gg \mathcal{B}(\tau \rightarrow \mu\gamma)$ and $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ can be as large as 10^{-7} while respecting the exclusion limits for $\tau \rightarrow \mu\gamma$, $\mu \rightarrow e\gamma$, and $\mu \rightarrow eee$, [50]. Lepton flavour violating decays with mesons in the final state are not expected in LR models.

Four generation models (SM4)

Extensions of the Standard Model introducing another generation of quarks and leptons are considered interesting as they can release tension between electro-weak precision tests and the “heavy” Higgs mass [65].

It is found in [51] that SM4 can lead to a detectable branching fraction $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$, as well as a detectable signature in other lepton flavour violating τ and μ decays; with and without mesons in the final state. The predictions for lepton flavour violation in meson decays are within the recent exclusion limits reported in [66], the Higgs mass measurement however rules out this model [67].

Non-universal Z' bosons in technicolor TC2 models

Extensions of the Standard Model gauge group introduce additional gauge bosons. These must not necessarily be flavour universal. They are expected to contribute at tree level to flavour changing neutral currents. In Ref. [52], the implications for a non-universal Z' in topcolour-assisted technicolor models (TC2) have been investigated. Recently, Z' models have been identified as possible explanations for the observed tension with the Standard Model in analyses of flavour changing neutral currents in the quark sector [68, 69].

The branching fraction $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is generally larger than $\mathcal{B}(\tau \rightarrow \mu\gamma)$ while the actual ratio depends strongly on a free parameter of the model. Both branching fractions are proportional to $m_{Z'}^{-4}$ and the values given in Tab. 1.3 are valid for $m_{Z'} = 1$ TeV.

Unparticles

A hidden sector of unparticles is motivated by [70, 71]. The so-called Banks-Zaks fields are supposed to be scale invariant and thus massless [53]. Assuming a coupling between Standard Model particles and Banks-Zaks fields which leads to a behaviour which is similar to the presence of d massless invisible particles, where d is non-integral. The unparticles can either lead to tree level processes, as in Fig. 1.4 (b), or in loop processes, as in Fig. 1.5 (a). The analysis in [53] finds that the current experimental bound on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ constrains the unparticle parameter space, already.

1.3 Model independent description of the decay $\tau \rightarrow \mu\mu\mu$

The search for $\tau \rightarrow \mu\mu\mu$ in this work cannot be conducted without assumptions on the decay kinematics of the decay. Instead of conducting a search for individual models presented above, a model independent description is utilised to obtain a general description and derive exclusion limits which constrain any of the above models. The

1. THEORETICAL BACKGROUND

description by [72] is used in this work. To allow literal quotation of the results in [72], the usual convention $c = \hbar = 1$ applies.

It is argued that for unpolarised τ leptons, the possible contributions from scalar, pseudoscalar, vector, and axial couplings can be Fierz rearranged to vector couplings. This leaves three vector coupling operators with different helicity structures. They all correspond to the Feynman diagram in Fig. 1.4 (a). The three operators have different chirality combinations, indicated by L for left handed leptons and R for right handed leptons.

$$\begin{aligned} H_{\text{eff}}^{(LL)(LL)} &= g_V^{(LL)(LL)} \frac{(\bar{\mu}_L \gamma_\mu \tau_L)(\bar{\mu}_L \gamma^\mu \mu_L)}{\Lambda^2} \\ H_{\text{eff}}^{(RR)(RR)} &= g_V^{(RR)(RR)} \frac{(\bar{\mu}_R \gamma_\mu \tau_R)(\bar{\mu}_R \gamma^\mu \mu_R)}{\Lambda^2} \\ H_{\text{eff}}^{(LL)(RR)} &= g_V^{(LL)(RR)} \frac{(\bar{\mu}_L \gamma_\mu \tau_L)(\bar{\mu}_R \gamma^\mu \mu_R)}{\Lambda^2} + \underbrace{g_V^{(RR)(LL)} \frac{(\bar{\mu}_R \gamma_\mu \tau_R)(\bar{\mu}_L \gamma^\mu \mu_L)}{\Lambda^2}}_{=: H_{\text{eff},2}^{(LL)(RR)}} \end{aligned}$$

Here, g are coupling strengths, Λ is the new Physics scale. The operators $H_{\text{eff}}^{(RR)(RR)}$ and $H_{\text{eff},2}^{(LL)(RR)}$ turn out to be suppressed by the smallness of the muon mass. The relevant operators are thus

$$\begin{aligned} H_{\text{eff}}^{(LL)(LL)} &= g_V^{(LL)(LL)} \frac{(\bar{\mu}_L \gamma_\mu \tau_L)(\bar{\mu}_L \gamma^\mu \mu_L)}{\Lambda^2} \\ H_{\text{eff}}^{(LL)(RR)} &= g_V^{(LL)(RR)} \frac{(\bar{\mu}_L \gamma_\mu \tau_L)(\bar{\mu}_R \gamma^\mu \mu_R)}{\Lambda^2} \end{aligned}$$

Due to the different helicity structures of the operators, the interference between these operators is suppressed by m_μ^2/m_τ^2 . The resulting phase space distributions are given by

$$\begin{aligned} \frac{d^2 \Gamma_V^{(LL)(LL)}}{dm_{+-}^2 dm_{-+}^2} &= \frac{|g_V^{(LL)(LL)}|^2}{\Lambda^4} \frac{(m_\tau^2 - m_\mu^2)^2 - (2m_{--}^2 - m_\tau^2 - 3m_\mu^2)^2}{256 \pi^3 m_\tau^3} \\ \frac{d^2 \Gamma_V^{(LL)(RR)}}{dm_{+-}^2 dm_{-+}^2} &= \frac{|g_V^{(LL)(RR)}|^2}{\Lambda^4} \left[\frac{(m_\tau^2 - m_\mu^2)^2 - 4m_\mu^2(m_\tau^2 + m_\mu^2 - m_{--}^2)}{512 \pi^3 m_\tau^3} \right. \\ &\quad \left. - \frac{(2m_{+-}^2 - m_\tau^2 - 3m_\mu^2)^2 + (2m_{+-}^2 - m_\tau^2 - 3m_\mu^2)^2}{1024 \pi^3 m_\tau^3} \right]. \end{aligned}$$

Here, m_τ and m_μ are the rest masses of the corresponding particles, m_{--} is the invariant mass of the same charge dimuon system, m_{+-} is the invariant mass of one of the opposite sign dimuon systems, and $m_{+-'}$ is the invariant mass of the other opposite sign dimuon system. The latter is not a free parameter and given by $m_{+-'}^2 = m_\tau^2 + 3m_\mu^2 - m_{--}^2 - m_{+-}^2$. The distributions are shown in Fig. 1.6.

The contribution from Fig. 1.3 (b) must not be ignored. The operator is

$$H_{\text{eff}}^{\text{rad}} = \frac{e}{4\pi} \frac{v}{\Lambda^2} \sum_{h,s} g_{\text{rad}}^{(s,h)} (\bar{\mu}(-i\sigma_{\mu\nu})\tau_s) F^{\mu\nu}.$$

The sum $\sum_{h,s}$ ranges over the two possible chirality combinations, $(h, s) = (L, R)$ and $(h, s) = (R, L)$. The corresponding phase space distribution is shown in Fig. 1.7 and is given by

1.3. Model independent description of the decay $\tau \rightarrow \mu\mu\mu$

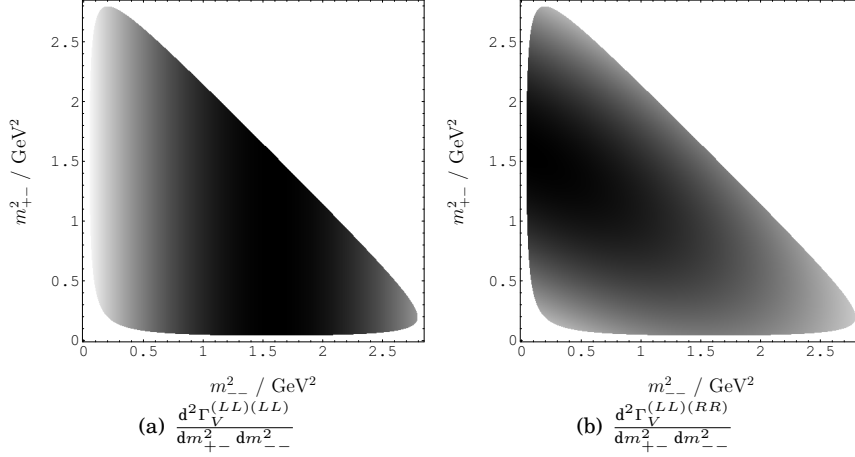


Figure 1.6: Phase space distributions for $\tau \rightarrow \mu\mu\mu$ with effective vector couplings. The $(LL)(LL)$ chirality combination has a broad maximum around $m_{--}^2 \approx 1.5 \text{ GeV}^2$ and is independent of the opposite sign masses. The $(LL)(RR)$ has a maximum around $m_{+-}^2 \approx 1.5 \text{ GeV}^2$ which results in favouring small m_{--}^2 due to the kinematic constraints. Figure taken from [72].

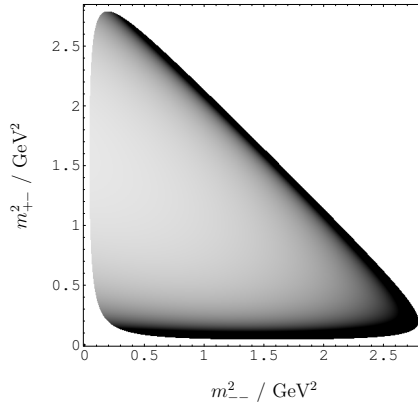


Figure 1.7: Phase space distribution for a photon penguin dominated $\tau \rightarrow \mu\mu\mu$ decay, $\frac{d^2\Gamma_{\text{rad}}^{(LR)}}{dm_{+-}^2 dm_{--}^2}$. The opposite charge muon pair has a small invariant mass – a consequence of the photon pole. Figure taken from [72].

1. THEORETICAL BACKGROUND

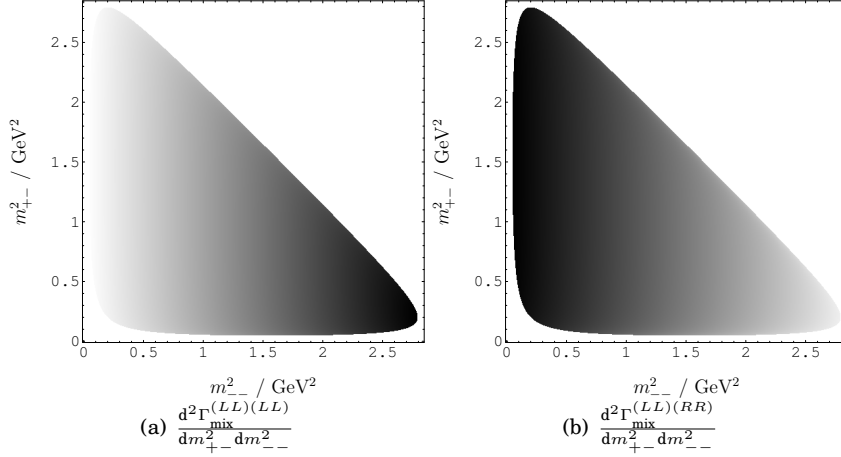


Figure 1.8: Phase space distribution of the interference terms of the photonic contribution to $\tau \rightarrow \mu\mu\mu$ with the vector couplings. Both distributions depend mainly on m_{--} , the m_{+-} dependence is not visible on this scale. The interference with the $(LL)(LL)$ chirality configuration affects mainly the high m_{--} range while the interference with the $(LL)(RR)$ chirality configuration affects the low m_{--} range. Figure taken from [72].

$$\begin{aligned}
 \frac{d^2\Gamma_{\text{rad}}^{(LR)}}{dm_{+-}^2 dm_{--}^2} &= \alpha_{\text{em}}^2 \frac{|g_{\text{rad}}^{(LR)}|^2 v^2}{\Lambda^4} \left[\frac{m_\mu^2 (m_\tau^2 - m_\mu^2)^2}{128 \pi^3 m_\tau^3} \left(\frac{1}{m_{+-}^4} + \frac{1}{m_{+-}^4} \right) \right. \\
 &+ \frac{m_\mu^2 (m_\tau^4 - 3m_\tau^2 m_\mu^2 + 2m_\mu^4)}{128 \pi^3 m_{+-}^2 m_{+-}^2 m_\tau^3} \\
 &+ \frac{(m_{+-}^2 + m_{+-}^2)(m_{--}^4 + m_{+-}^4 + m_{+-}^4 - 6m_\mu^2(m_\mu^2 + m_\tau^2))}{256 \pi^3 m_{+-}^2 m_{+-}^2 m_\tau^3} \\
 &\left. + \frac{2m_{--}^2 - 3m_\mu^2}{128 \pi^3 m_\tau^3} \right]
 \end{aligned}$$

The radiative operator can interfere with the vector couplings $H_{\text{eff}}^{(LL)(LL)}$ and $H_{\text{eff}}^{(LL)(RR)}$. The interference results in the phase space distributions shown in Fig. 1.8 and given

by

$$\begin{aligned} \frac{d^2\Gamma_{\text{mix}}^{(LL)(LL)}}{dm_{+-}^2 dm_{--}^2} &= \alpha_{\text{em}} \frac{2v \text{Re}[g_V^{(LL)(LL)} g_{\text{rad}}^{*(LR)}]}{\Lambda^4} \\ &\times \left[\frac{m_{--}^2 - 3m_\mu^2}{64\pi^3 m_\tau^2} + \frac{m_\mu^2(m_\tau^2 - m_\mu^2)(m_{+-}^2 + m_{+-}^2)}{128\pi^3 m_\tau^2 m_{+-}^2 m_{+-}^2} \right] \\ \frac{d^2\Gamma_{\text{mix}}^{(LL)(RR)}}{dm_{+-}^2 dm_{--}^2} &= \alpha_{\text{em}} \frac{2v \text{Re}[g_V^{(LL)(RR)} g_{\text{rad}}^{*(LR)}]}{\Lambda^4} \\ &\times \left[\frac{m_\tau^2 - m_{--}^2 - 3m_\mu^2}{256\pi^3 m_\tau^2} + \frac{m_\mu^2(m_\tau^2 - m_\mu^2)(m_{+-}^2 + m_{+-}^2)}{256\pi^3 m_\tau^2 m_{+-}^2 m_{+-}^2} \right]. \end{aligned}$$

For the present work, it is assumed that $\tau \rightarrow \mu\mu\mu$ has a constant phase space distribution; i. e. a search for $\tau \rightarrow \mu\mu\mu$ according to $\frac{d^2\Gamma}{dm_{+-}^2 dm_{--}^2} = \text{const}$ is performed. This phase space distribution is simulated, and the results are valid for this distribution because the determination of the event selection efficiency depends on the simulated phase space distribution. Exclusion limits for $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ in the effective description presented here, are given in Chap. 13.

1.4 Placement of $\tau \rightarrow \mu\mu\mu$ within other lepton flavour violating decays

Charged lepton flavour violation can be thought of in many decays as outlined in Sect. 1.5. The decay investigated in this work is one out many lepton flavour violating lepton decays. The relevance of searching for $\tau \rightarrow \mu\mu\mu$ shall briefly be compared to other decay channels. The decays $\tau \rightarrow \mu\gamma$ and $\tau \rightarrow \mu\mu\mu$ have been compared throughout Sect. 1.2.

The decay $\tau \rightarrow \mu\mu\mu$ is only one out of six $\tau \rightarrow \ell_j \ell_k \bar{\ell}_l$ decays. It is stated in [48] that final states with different lepton flavours are typically suppressed with respect to equal final state flavours by combinatorial factors. Surprisingly, the opposite is reported in [45] for MSSM, where $\tau \rightarrow \mu ee$ is enhanced by a factor 4 over $\tau \rightarrow \mu\mu\mu$ due to phase space factors. From the experimental side, however, background processes are not expected for $j = k \neq l$, which might make the final states $\mu^\pm \mu^\pm e^\mp$ and $e^\pm e^\pm \mu^\mp$ more desirable to search for. At the B factories background processes are almost absent and due to comparable reconstruction efficiencies for muons and electrons the sensitivities for all final states are about equal. At LHCb, the reconstruction and identification efficiency as well as the momentum resolution for electrons does not reach that of muons, the three muon final state is thus expected to provide the best sensitivity.

The branching fraction $\mathcal{B}(\mu \rightarrow eee)$ is experimentally much better constrained than $\tau \rightarrow \mu\mu\mu$. One might therefore assume that its constraints to BSM theories are much better. This however strongly depends on the model: as stated above, the LHT model is found to be better constrainable by $\mu \rightarrow eee$. In other models, the strong exclusion of $\mu \rightarrow eee$ is still compatible with a detectable $\tau \rightarrow \mu\mu\mu$ signal, e. g. in the LR model or R -parity violating SUSY. For Higgs mediated LFV in MSSM ([47]), the

expectations are

$$\begin{aligned}\mathcal{B}(\tau \rightarrow \mu\mu\mu) &\approx (1 \times 10^{-7}) \cdot \left(\frac{\tan\beta}{60}\right)^6 \cdot \left(\frac{100 \text{ GeV}/c^2}{m_A}\right)^4 \\ \mathcal{B}(\mu \rightarrow eee) &\approx (5 \times 10^{-14}) \cdot \left(\frac{\tan\beta}{60}\right)^6 \cdot \left(\frac{100 \text{ GeV}/c^2}{m_A}\right)^4 \\ \frac{\mathcal{B}(\tau \rightarrow \mu\mu\mu)}{\mathcal{B}(\mu \rightarrow eee)} &\approx 2 \cdot 10^6.\end{aligned}$$

The current experimental limits are in the ratio $\text{Limit}(\mathcal{B}(\tau \rightarrow \mu\mu\mu))/\text{Limit}(\mathcal{B}(\mu \rightarrow eee)) \approx 2 \cdot 10^4$, which implies that the discovery potential would be better for τ decays than for μ decays, if the large $\tan\beta$ range was not already strongly constrained by flavour changing neutral current measurements of $B_s^0 \rightarrow \mu\mu$, [44].

Generally, lepton flavour violation in τ and μ decays is related to different model parameters. Therefore theory papers concentrate on the ratios $\mathcal{B}(\ell \rightarrow \ell'\ell'\ell')/\mathcal{B}(\ell \rightarrow \ell')$ instead of ratios of different lepton flavour transitions like $\mathcal{B}(\mu \rightarrow eee)/\mathcal{B}(\tau \rightarrow \mu\mu\mu)$.

As explained in [73], searches for $\tau \rightarrow \mu h$ and $\tau \rightarrow \mu hh'$ are complementary to the pure leptonic final state. The operators contributing to semileptonic final states are not constrained by the exclusion limits on $\mathcal{B}(\tau \rightarrow \ell_j \ell_k \bar{\ell}_l)$ and vice versa.

1.5 Current status on lepton flavour violation with $\tau \rightarrow \mu$ transitions

In the neutrino sector, the first direct ν_μ to ν_τ transition has been observed by the OPERA collaboration in 2010. By now only three further events have been reported [74]. The appearance of ν_τ from solar neutrinos has been seen by Super-Kamiokande, too [75].

In the charged lepton sector, the direct transitions $\tau \rightarrow \mu\gamma$ and $\tau \rightarrow \mu\mu\mu$ have been searched for, but no evidence has been reported so far. The best experimental exclusion limits are reported by the B factories BaBar and Belle:

$$\begin{aligned}\mathcal{B}(\tau \rightarrow \mu\gamma) &< 4.4 \times 10^{-8} \quad \text{by BaBar, [76]} \\ \mathcal{B}(\tau \rightarrow \mu\gamma) &< 4.5 \times 10^{-8} \quad \text{by Belle, [77]; soon to be updated [78].}\end{aligned}$$

In addition to the violation of τ lepton flavour in lepton decays, lepton flavour can also be violated in hadron or gauge boson decays. The latest result has been published by the BaBar collaboration searching for $B \rightarrow h\tau^\pm\ell^\mp$, [79]. Exclusion limits on $\Upsilon \rightarrow \tau\mu$, $J/\psi \rightarrow \tau\mu$, as well as $Z \rightarrow \tau\mu$ can be found in [80].

The investigation of Higgs boson decays is rapidly evolving. Recently the CMS collaboration announced a 2.5σ excess of the $H \rightarrow \tau\mu$ branching fraction [81].

The LHCb experiment

The LHCb experiment is one of the experiments at the Large Hadron Collider at CERN. A comprehensive detector description can be found in [82], the physics objectives are described in [83]. A review of the measurements by LHCb and their implications for particle physics are summarised in [84]. Here, short description of the LHCb detector is given. While [82] has been published prior LHC operations, the first run of LHC has ended in the meanwhile and a performance review is being prepared [85] and partially available for subsystems [86, 87, 88, 89, 90, 91, 92, 93, 94]. These reviews of LHCb operations are considered here, too.

2.1 Implications of physics objectives on the detector design

The LHCb detector is designed to serve the needs of flavour physics analyses. At the LHC, heavy flavour hadrons are produced with a large boost in the forward direction, i. e. at small polar angles with respect to the beam axis. The LHCb detector is thus

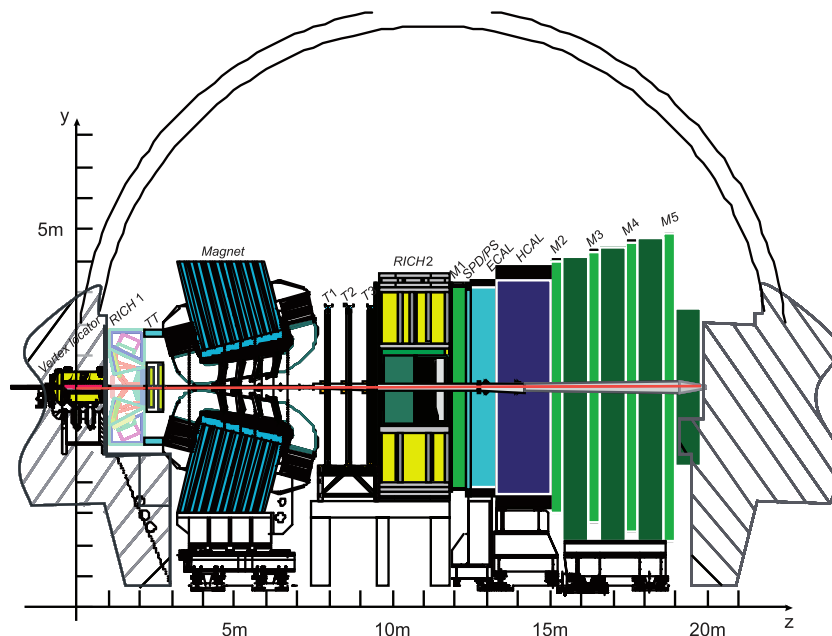


Figure 2.1: Layout of the LHCb detector. Figure taken from [95].

designed as a forward spectrometer along the beam axis, shown in Fig. 2.1. The subsystems are explained in the following sections. This involves a good acceptance for boosted heavy flavour hadron decays and an accurate decay time measurement. A large boost of the hadron leads to a large flight distance of the heavy flavour hadron from the primary vertex to the decay vertex and thus a small relative uncertainty on the flight distance.

The coordinate system of LHCb is defined with respect to the nominal LHC interaction point. The beam axis defines the z -axis. The y -axis is chosen to point vertically upwards, in the plane perpendicular to the z -axis. The x -axis is then defined to result in a right handed Cartesian coordinate system; i. e. when standing at the interaction point and looking along the beam axis into the spectrometer, the x -axis points to the left. The azimuth, φ , and polar angle, ϑ , are defined according to the standard definition of spherical coordinates.

The direction from the interaction point along the beam axis through the spectrometer is called downstream. The opposite direction, towards the interaction point is called upstream. Accordingly, the magnet is the border of the so-called upstream region and the downstream region. The upstream region hosts the vertex locator (VELO), a first RICH detector (RICH1) and a first tracking station (TT). The downstream region hosts three more tracking stations, a second RICH detector (RICH2), the calorimeter system and the muon stations.

2.2 LHC operations

Proton beams are injected at 450 GeV from CERN's accelerator chain into both directions of the LHC ring. Throughout data taking in 2011 and 2012, protons were injected in bunches separated by 50 ns, twice as much as designed. Once the desired number of bunches is circulating in the LHC, the beam energy was increased up to 7 TeV in 2011 and 8 TeV in 2012. During the acceleration, the beams are kept on non-intersecting orbits. After reaching the target energy, the beam optics magnetic fields are adjusted to let the beams intersect at the interaction points of the LHC experiments.

The sequence of injecting protons into the LHC, accelerating them, and letting them collide is called a *fill*. A fill is ended by deflecting the beam out of the beam pipe onto the LHC beam dump which is either done because the safety of the LHC requires it or because the LHC beam current becomes so low that it is advantageous (for the integrated luminosity) to dump and re-inject the beam.

The beam current reduces during a fill as protons which collide in the experiments "leave" the beam, but also interactions with collimators or interaction with residual gas in the beam pipe reduce the number of protons. This would normally lead to a continuous reduction of the interaction rate. At LHCb the beams are displaced vertically with respect to each other, to lower the instantaneous luminosity. The displacement is reduced during a fill to cancel the reduction of the beam current and maintain a constant instantaneous luminosity (*luminosity levelling*). This is only possible until the beams are not vertically displaced anymore, from this point on, the interaction rate decreases. This happened in roughly one fifth of all fills during 2011 and 2012.

The intersection of the beams is chosen such that the beams are separated 7.5 m from the interaction region¹, not to let protons collide in other points than the nominal

¹Bunches are separated by 15 m at 50 ns bunch spacing. Since the beams "move" in opposite directions, a bunch "meets" the next bunch from the opposite beam after 7.5 m.

interaction points.

The data set analysed in this work corresponds to $(1.01 \pm 0.01) \text{ fb}^{-1}$ taken in 2011 at $\sqrt{s} = 7 \text{ TeV}$ and $(1.99 \pm 0.02) \text{ fb}^{-1}$ taken in 2012 at $\sqrt{s} = 8 \text{ TeV}$.

2.3 LHCb subsystems

Magnet

The LHCb spectrometer magnet is a dipole magnet with a magnetic field perpendicular to the beam axis. Charged particles are deflected in the magnetic field. The deflection depends on the particles' momenta and their charges. The electric current in the magnet coil is regularly reversed. Trajectories of positively charged particles with one magnet polarity are almost identical to negatively charged particles' trajectories with the opposite magnet polarity. The reversal reduces the effect of detection asymmetries on studies of CP violation [96]. For the present analysis, possible detector asymmetries are assumed to be irrelevant and the data taken with both polarities is combined. For consistency of simulated events with the actually data taking, however, half of all simulated events have been simulated with either polarity.

The spectrometer magnet deflects the LHC beam, too. Due to the large momentum of the protons of the LHC beam, their deflection is small and the beam remains within the beam pipe. The LHCb magnet's deflection is compensated by additional magnets outside the LHCb cavern along the beam pipe.

Vertex Locator

The LHCb vertex locator (VELO) is a silicon strip detector built in r - φ geometry; i. e. φ -sensors with approximately radial strips (i. e. approximately constant φ coordinate along the strip), and r -sensors with strips which are shaped as circle segments, each of them covering $1/8$ of a circle. Its performance is reviewed in [88]. To achieve the best possible vertex resolution, it is built as close as possible to the beam line. In contrast to other LHC experiments, it is not built around the LHC beam pipe, instead the VELO replaces the beam pipe and only a thin corrugated foil, the RF foil, separates the sensors from the beam. The RF foil provides shielding against RF pickup from the LHC beam and prevents possible outgasings from the detector modules to reach the LHC beam.

The VELO sensors are retracted, away from the beam axis, when the particles are injected into the LHC and during acceleration to provide a larger aperture for the beam.

Two of the VELO stations, the pile-up unit, were designed with only r sensors and the capability to be read out at 40 MHz for consideration in the hardware trigger. A simplified vertex finding algorithm, implemented in hardware, would have identified events with more than one primary interaction, to veto these. Eventually, it has been decided to consider these *pile-up* events as good events for physics analyses – the average number of visible interactions is about 2 in the 2011 and 2012 data.

TT

Downstream of RICH1, a tracking station, the so-called TT is situated, a large area silicon strip detector. It detects charged particles, especially decay products of strange

2. THE LHCb EXPERIMENT

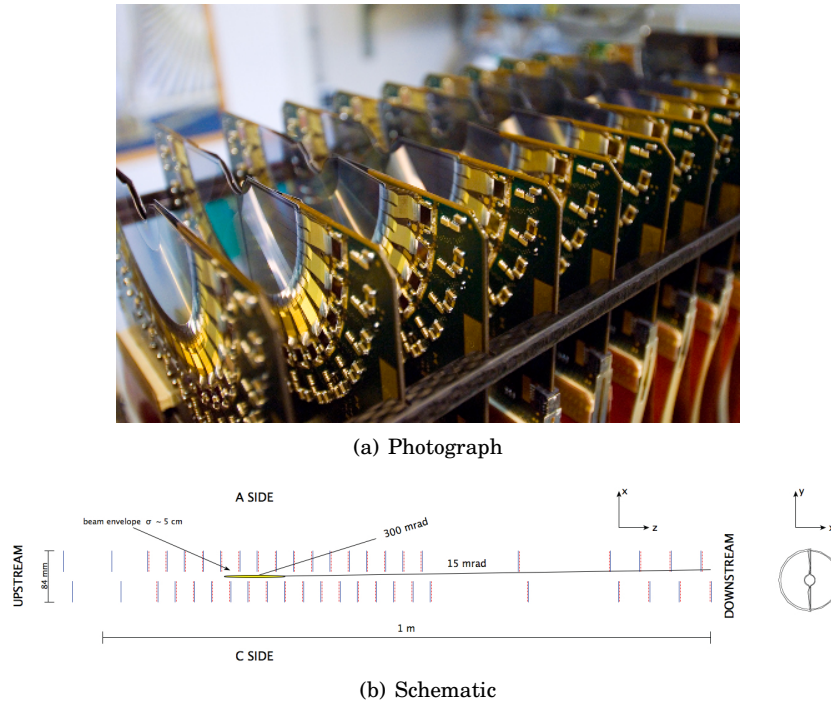


Figure 2.2: The VELO as shown in [88].

particles which decay outside of the VELO and thus cannot be detected in the VELO. Given the fringe field of the nearby dipole magnet, charged particle tracks which are detected in the VELO and in the TT have a measurable curvature.

The layout of the TT is shown in Fig. 2.3. The strips are all approximately oriented in y -direction and thus provide a measurement of the x coordinate of a track at the TT. Those detector layers which are exactly parallel to the y -direction are therefore called x -layers. Due to the orientation of the magnetic field, the slope in the y - z plane is approximately constant and thus the measurement in the VELO sufficient; i. e. the y - z slope resolution cannot be improved by further measurements. The x -coordinate is of interest because it is not constrained by the measurement in the VELO due to the deflection in the magnetic field. Therefore, the x -measurement leads to a momentum estimate. The measurement with the *stereo* layers – those layers which are tilted by 5° with respect to the y -direction – provides a confirmation, whether hits in the TT are compatible with a track extrapolation from the VELO in the y - z plane.

Inner Tracker

The tracking detectors downstream of the detector are arranged in three stations (T stations). The part close to the beam pipe, where the particle flux is high, is covered by a silicon strip detector, the Inner tracker (IT). There are four active layers in each T station, the first and the last with strips in y -direction (x -layers), and the inner two tilted by $\pm 5^\circ$, as in the TT.

Due to the position of the IT, its readout electronics and cooling infrastructure are

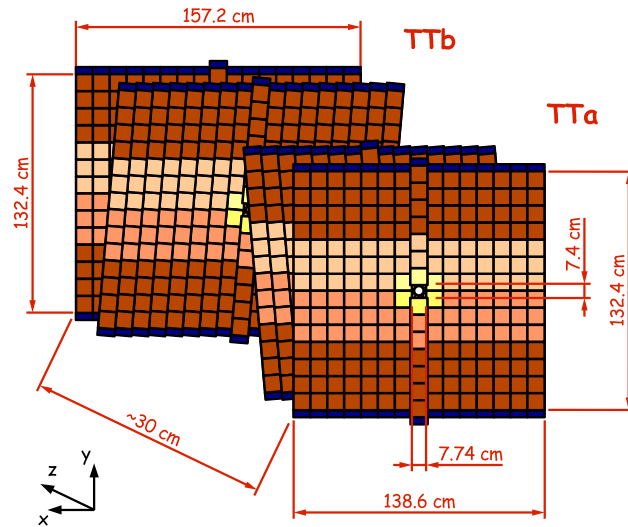


Figure 2.3: Schematic of the layout of the TT. The TT comprises two tracking stations, each with one layer of silicon strips along the y -direction, and one *stereo* layer, where the strips are tilted by $\pm 5^\circ$ in the x - y plane with respect to the y -direction. The figure appears in various theses describing LHCb, the original source could not be identified. The figure does not appear in the LHCb technical design reports or [82].

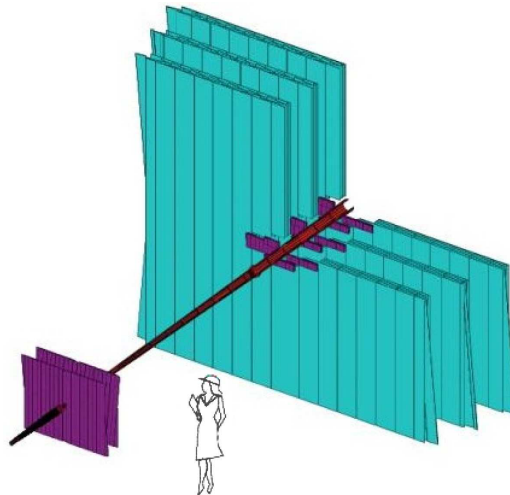


Figure 2.4: Arrangement of the TT and the T station trackers. The OT is shown in turquoise, the IT and TT in purple. Figure taken from [82].

2. THE LHCb EXPERIMENT

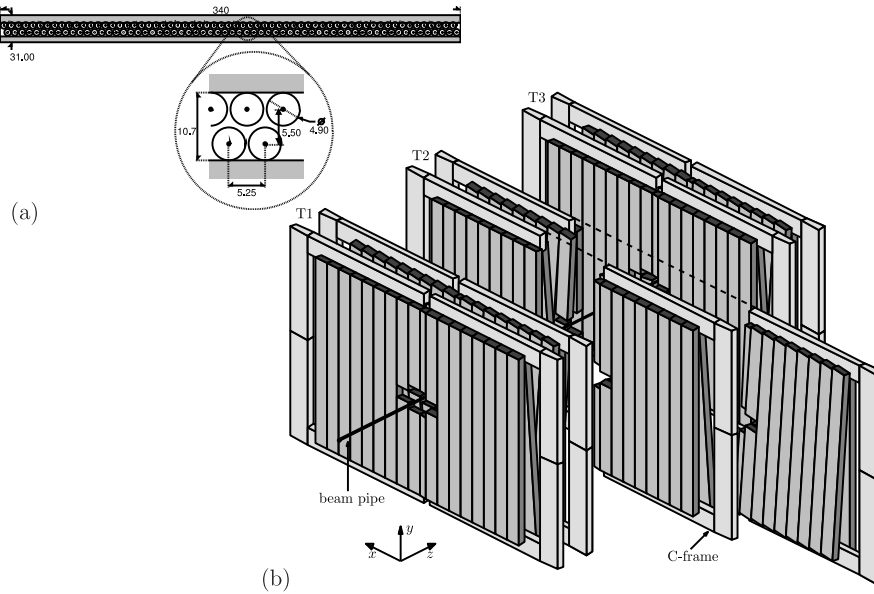


Figure 2.5: The Outer tracker. (a) Module cross section. (b) Arrangement of OT straw-tube modules in layers and stations. The observant reader will notice that one of the two innermost modules in each layer is more narrow than the others, the narrow module is always situated in the negative- x range half of the detector due to spatial constraints by the LHCb cavern. This is different to the deprecated, but still often quoted, design in [82]. Figure taken from [94].

placed in the detector acceptance and contribute significantly to the material budget. The IT is the first detector which is not orthogonal to the z -axis – the z -axis is not horizontal² and the detectors downstream of the magnet are installed along the gravitational field for mechanical reasons. The occupancy of the IT reaches 2.5% close to the beam pipe and the IT has a hit resolution of 50 μm in the direction orthogonal to the strips.

All silicon trackers at LHCb (VELO, TT, and IT) record the pulse height of the hits.

Outer Tracker

The largest part of the T stations is covered by the Outer tracker (OT), a straw tube drift detector [94], shown in Fig. 2.5. Similar to the IT, there are two x layers and two stereo layers in each T station. Differently to the IT, each of these detector layers is made of two layers of straw tubes. The measurement of drift times³ in the straw tubes determines the distance of a charged particle's track to the anode wire. The resulting hit resolution is commonly reported to be 205 μm , an improved detector alignment

²Due to geological constraints, the LEP/LHC tunnel is inclined.

³The OT has no pulse height readout. It measures the time of hits with respect to the bunch crossing signal instead.

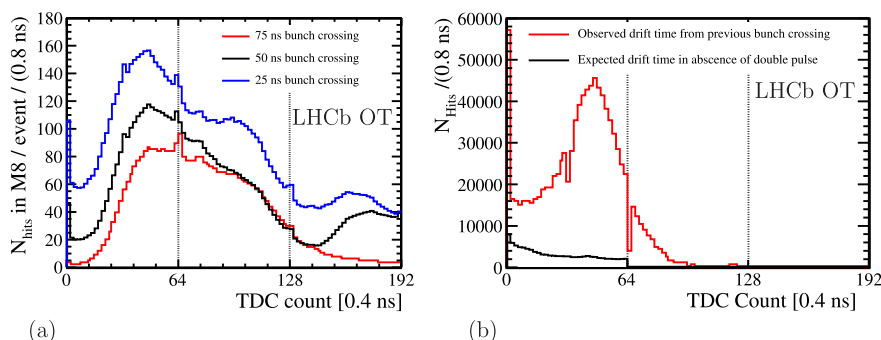


Figure 2.6: Drift time spectrum in the OT. The time between the LHC clock signal and the first signal in an Outer Tracker straw is histogrammed for different bunch spacings (left) and in the absence of a collision (right). For the drift time spectrum at 75 ns bunch spacing, the drift time spectrum without spillover or after-pulse is visible. The shape is the result of the distance-to-anode-wire to drift-time relation, resolution effects, and clock noise, which is visible at the vertical lines. The spectra at 50 ns contain additional contributions from spill-over – best visible at high TDC counts – and after-pulses, which result in a larger number of measurements at small drift times. Figure taken from [94].

with fine granular alignment of the OT monolayers may result in a hit resolution of 179 μm in the future [97].

The OT electronics measures one drift time per bunch crossing if a hit is present. Since the hit occupancy reaches 20% in the innermost modules, a substantial amount of straws with two hits has to be expected. For a given particle track the measured drift time may thus be incorrect, but instead “belong” to a different particle. The track fit at LHCb can therefore ignore drift time measurement if it is found incompatible with the track. Instead of ignoring the hit entirely, the straw is then assumed as a binary object. The introduction of this feature lead to a homogeneous distribution of the track fit χ^2/ndf , independent of the pile-up or whether a particle is tracked in the IT or in the OT.

Drift times in the OT are up to 35 ns, which is sufficient not to observe hits from the previous or following bunch crossing in an event. Due to the drift time resolution and the presence of after-pulses the picture changes. Figure 2.6 shows the drift time spectrum in the OT. At 75 ns bunch spacing, after-pulses are not present, the excess at 50 ns bunch spacing is due to these after-pulses. When no bunches intersect but there was collision present 50 ns ago, the red drift time spectrum in Fig. 2.6 (b) can be observed, this shows the after-pulse spectrum isolated. At 25 ns bunch spacing, so-called *spill-over* hits are observed. This may become the standard in future LHCb data.

2.4 Track reconstruction

Tracks are the fundamental reconstruction objects of LHCb analyses. Therefore their reconstruction is explained given emphasis in the following paragraphs.

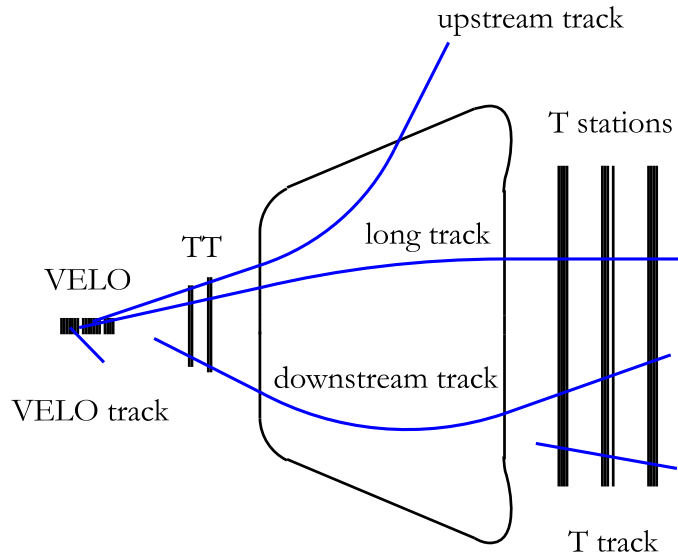


Figure 2.7: Schematic top-view of the tracking detectors and the track types at LHCb. Figure taken from [98].

The track reconstruction at LHCb has been described in e.g. [98, Chap. 3], [87, Sect. 3], [85, Sect. 2.2], and the references therein. Figure 2.7 illustrates the tracking detectors, the curvature of charged particles' tracks in the magnetic field of the dipole magnet, and the definitions of track types at LHCb. For the analysis presented in this work, only long tracks are of interest; their reconstruction is, however, not independent of the reconstruction of other track types, due to the selection of “best” tracks (see below). This section focuses on the reconstruction steps relevant for the present analysis. Unless stated differently, extrapolations mentioned throughout this section always assume the laws of electrodynamics and respect the curvature of charged particles' trajectories in magnetic fields.

The search for charged particles' tracks begins in the VELO, where tracks are straight lines. Stable particles of interest are mostly produced close to the z -axis, either in the primary interaction point or as decay products of long lived particles which are produced in the forward direction. Straight lines intersecting the z -axis are straight in the r - z projection of cylinder coordinates. Measurements in the r -sensors of the VELO are therefore used in a first pattern recognition step to find VELO tracks. Thereby tracks are found which originate from the primary interaction point or from decays of long lived particles, if they decay close to the z -axis⁴. Subsequently, hits in the φ sensors are associated to the r - z tracks. The small angle of the φ -strips with respect to exact lines of constant φ resolves the ambiguity in case two r - z tracks are reconstructed in the same octant.

The laws of electrodynamics allow to compute a particle's trajectory once a single point on the trajectory (\vec{x}), the momentum vector (\vec{p}) at that point, and the charge (q) are known. To be more precise, \vec{p} and q can be reduced to $\vec{p}/|p|$ and $q/|p|$. A VELO track measures \vec{x} and $\vec{p}/|p|$, but not $q/|p|$. A single measurement in the x - z plane⁵

⁴This leads to a decrease in reconstruction efficiency with increased decay time, as explained in [99].

⁵To first order, the y - z projection of a track is a straight line, independent of the particle's momentum.

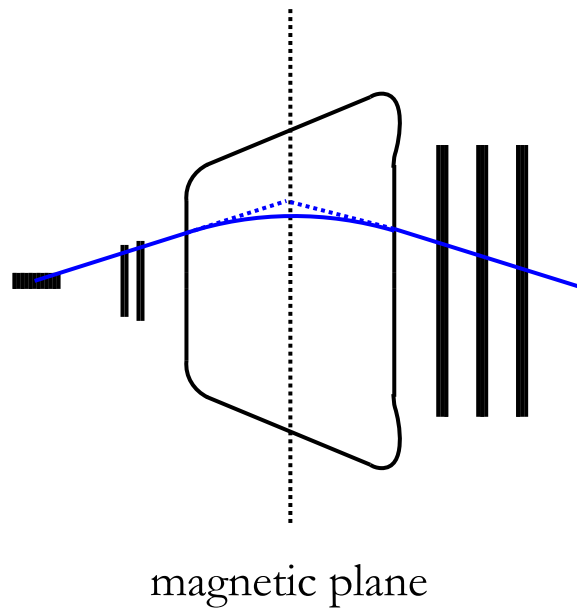


Figure 2.8: Illustration of the optical model, that the magnet is a single plane and charged particles' trajectories are straight lines elsewhere: The solid line shows the actual trajectory of a particle in LHCb, the dashed line shows the straight line extrapolations of the T station segment and VELO segment into the magnet. Figure modified from [98].

downstream of the magnetic field is sufficient to establish a momentum hypothesis. Based on the momentum estimate from every hit in the T stations, a Hough transform [100, 101] is applied to the hits in the T stations. Hits following the track model with the assigned momentum hypothesis are hereby transformed to the same point. A long track through the T stations is thus found as a cluster of hits in the Hough plane. This track finding strategy is used in the software trigger, too, and called *forward tracking* in [85].

T tracks are searched for by fitting a parabola through all three-station combinations of hits in x layers of the T stations. The parabola's parameters are constrained to the parameter range observed in simulated events for particles originating in the primary interaction point. Unless enough additional hits are found in the T stations around the parabola, the track candidate is rejected. Otherwise, hits in the stereo layers which are compatible with the x - z track candidate are added to contribute information about the position of the track in y . An origin of the track near the nominal interaction point in the y - z plane is assumed in this stage again.

A second long track finding algorithm applies an optical model. In this optical model, the magnetic field is a single plane. Straight line extrapolations of VELO tracks and T tracks into the dipole intersect in the bending plane if they reconstruct the same particle, as illustrated in Fig. 2.8. To increase the track finding efficiency beyond the efficiency of the forward tracking, VELO tracks and T tracks which are compatible with this optical method are combined to long tracks.

VELO tracks can be extrapolated as a straight line into the TT. For not too low particle momenta, hits from that exact particle which was found as VELO track can be found close to the track extrapolation. If this is the case, these hits are added to the VELO track to build an upstream track.

Using the optical model again, T tracks can be extrapolated straight to the magnet's bending plane. Hits in the TT can be found close to a straight line from this bending-plane point to the nominal interaction point if the particle originated from a decay vertex in the VELO or RICH1. This upgrades the T track to a downstream track.

If hits in the TT are found around the interpolation of a long track into the TT, these hits are added to the track, to improve the track parameter estimation.

A final selection of “best” tracks is made from the tracks found by all these methods. If two different tracks are built from (partially) the same detector hits, the “longest” is kept, e. g. long tracks are preferred over downstream tracks, which are preferred over T tracks, long tracks with TT hits are preferred over long track without TT hits. The criterion is slightly loosened for track types which are generically difficult to compare. In this case a momentum difference criterion is used. E. g. long tracks and downstream tracks are supposed to reconstructed the same particle if they “share” half of their hits in the T stations and at least one of the following two conditions is met: at least one common TT hit or their $\Delta q/p$ is within $5 \cdot 10^{-6}$. To determine the particle momentum and trajectory, the tracks are fitted with a Kalman filter fit [102]. Only tracks which can be fitted with a good fit quality, $\chi^2/\text{ndf} < 3$, qualify as “best”; i. e. if a correctly reconstructed downstream track is built from the same hits in the T stations as a fake long track⁶, the bad fit quality of the long track will identify the downstream track as “better” track.

For the remaining tracks are compared by computing their relative entropy [103] given by their Shannon information content. For two tracks i , it is computed from their track parameters $\vec{x} = (x, y, p_x/p, p_y/p, q/p)$ computed by the track fit and from the covariance matrix C estimated by the track fit:

$$D(1, 2) = \text{Tr} \left[(C_1 - C_2)(C_2^{-1} - C_1^{-1}) \right] + (x_1 - x_2)^T (C_1^{-1} + C_2^{-1})(x_1 - x_2).$$

This relative entropy is small for tracks with similar track parameters. From each two tracks within a relative entropy of 5 000 or less, the track with more hits (or the better fit quality in the case of equally many hits) is kept which reduces the track finding efficiency by two per mille at a residual rate of doubly reconstructed tracks of three per mille [104].

The rejection of reconstructed tracks which do not correspond to any charged particle, *fake tracks*, may not be sufficient when using only the aforementioned χ^2/ndf . A neural network has been trained to identify fake tracks. Input variables to this network are mainly the fit χ^2 contributions in the subdetectors, the number of outlier hits which were not considered in the track fit, the χ^2 contribution of combining the segments in the upstream region with the downstream region, and the number of active detector elements through which the track passes but in which no hit is observed. Also the number of ignored drift time measurements enters the neural network. The response is calibrated on simulated events to be the fake track acceptance rate. E. g. 30% of all fake tracks pass a selection of $\mathcal{M}_{\text{fake track}} < 0.3$. The inefficiency – not accepting good tracks – is at the percent level for this requirement.

⁶i. e. where the hits in VELO are from a different particle than those in the T stations

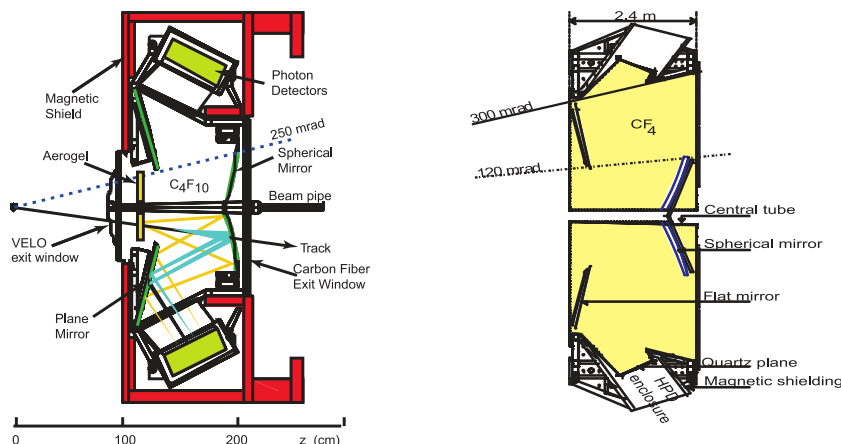


Figure 2.9: Schematic of the RICH detectors: side view of RICH1 (left) and top view of RICH2 (right). Figure rearranged from [82] as inspired by [38].

Ring Imaging Cherenkov Detector

Downstream of the VELO, a first ring imaging Cherenkov detector (RICH1) is situated. Charged particles emit Cherenkov light in two radiators if they are faster than the speed of light in the radiator material. The focusing optics of a mirror system focuses the Cherenkov photons of a particle to a ring in the focus plane where they are detected by hybrid photon detectors [105]. A second single-radiator RICH is placed between the T stations and the calorimeter system (RICH2). The three radiators provide particle identification in different ranges of particle momenta, together covering the entire range of interest for LHCb.

Based on the observed photon pattern in the RICH detectors, an event likelihood of particle hypotheses is computed and maximised. A particle hypothesis is assigned to each reconstructed track, for it being a pion, kaon, proton, or muon. The likelihood for the observed hit pattern is computed for the expected Cherenkov angles for the measured track momenta. The event likelihood is maximised by changing the particle hypotheses for the tracks until a maximum is found. Starting from the maximal likelihood mass-hypotheses configuration, the likelihood loss for each single hypothesis change is computed to quantify how much more likely a track is a particle of its assigned hypothesis with respect to an alternative hypothesis [90].

This procedure is not equivalent to reconstructing Cherenkov rings and the measurement of their radii, the likelihood computation considers only possible Cherenkov angles for the four particle hypotheses (π , K , p , μ). The ring of a hypothetical 300 MeV/ c^2 particle would remain unobserved. Consequently there is no unambiguous ordering of particle hypothesis likelihoods, as shown in [90, Fig. 16].

Cherenkov angles are measured for detector performance studies. For isolated tracks, which are defined to be so far away from other tracks that their Cherenkov rings cannot intersect, a measurement of the Cherenkov radius has been performed in [90] to illustrate the Cherenkov angle resolution of RICH2. The result in Fig. 2.10 shows a clear separation of the four particle species. It is remarkable that the RICH

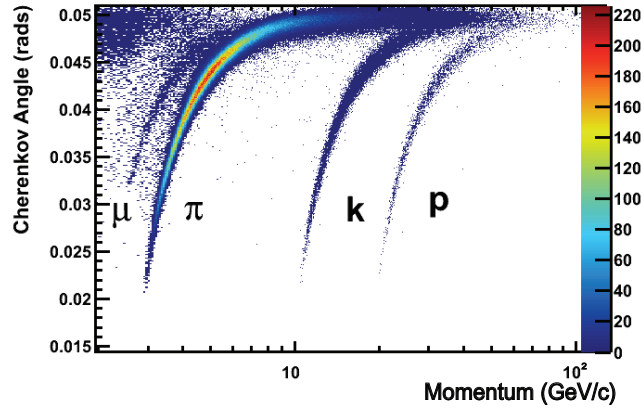


Figure 2.10: Reconstructed Cherenkov angle as a function of track momentum in the gas radiator of RICH1. Figure taken from [90].

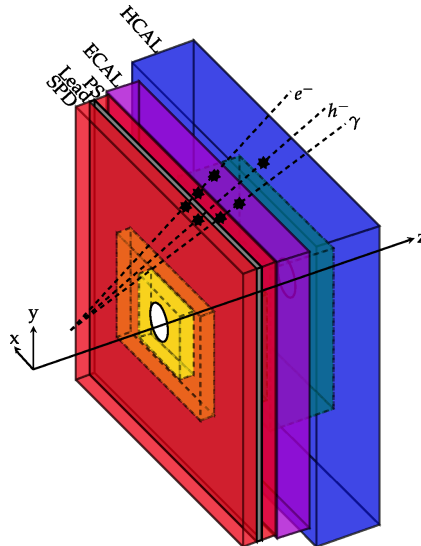


Figure 2.11: Schematic of the calorimeter system at LHCb. The z -scale of the SPD and PS are exaggerated. Figure taken from [38].

distinguishes muons from pions to some extent, despite the wide-spread belief that RICH detectors could not contribute to pion-muon separation.

Calorimeters

The calorimeter system [89], shown in Fig. 2.11, absorbs particles, except muons and neutrinos, and measures their energy deposit. For the present analysis, the main purpose of the calorimeter system is therefore to filter charged particles. Hadrons and electrons are filtered out, muons pass the filtering and reach the muon system.

The calorimeter system consists of four subdetectors, a scintillator pad detector (SPD), a preshower detector (PS), an electromagnetic calorimeter (ECAL), and a hadronic calorimeter (HCAL). The layout is projective with respect to the nominal interaction point, i. e. the transverse size of the subsystems increases linearly with their distance to the nominal interaction point. The sizes of individual detector cells scale with increasing distance to the interaction point accordingly. The SPD, PS, and ECAL are divided into three regions of different granularity. Close to the z axis, where the particle multiplicity is highest, the detector cells are smallest, in the outer region the cells are largest. The HCAL is divided into two regions of different granularity.

The subdivision into SPD, PS, and ECAL provides the possibility to analyse the longitudinal shower profile of electromagnetic showers. E. g. showers of neutral particles begin in the lead plate between SPD and PS, the absence/presence of a hit in the SPD separates neutral particle showers from charged particle showers.

The ECAL and HCAL are shashlik calorimeters, i. e. successive scintillator and absorber (lead or iron) plates. Light signals in the scintillators are collected by wavelength-shifting fibres and guided to photomultiplier tubes or multi-anode photomultiplier tubes. The gains of the calorimeter cells are adjusted to achieve that the calorimeter's response is a measure of $E_T := E \cdot \sin \vartheta$ because the hardware trigger is selecting signatures according to their transverse energy. The angle ϑ is the polar angle of the position of the calorimeter cell⁷.

The energy resolution of the calorimeters is given in [89] as:

$$\begin{aligned} \left(\frac{\sigma(E)}{E} \right)_{\text{ECAL}} &= \frac{(9.0 \pm 0.5) \%}{\sqrt{E/\text{GeV}}} \oplus (0.8 \pm 0.2) \% \oplus \frac{0.003}{E/\text{GeV} \cdot \sin \vartheta} \\ \left(\frac{\sigma(E)}{E} \right)_{\text{HCAL}} &= \frac{(69 \pm 5) \%}{\sqrt{E/\text{GeV}}} \oplus (9 \pm 2) \%. \end{aligned}$$

The first term is due to the stochastic nature of showers, the second due to residual mis-calibrations, non-linearities, and other effects, and the third term is due to noise of the electronics.

Muon chambers

The muon system [91] comprises five stations, named M1 to M5, as shown in Fig. 2.12. The first, M1, plays a special role. It is placed in front of the calorimeter system and thus detects any charged particle, not only muons. Therefore, M1 cannot serve particle identification purposes. It is used in the hardware trigger to improve the momentum estimate. M1 is not used in the event reconstruction⁸.

The muon stations are divided into four regions of different granularity, to account for the higher/lower particle density close-to/far-from the beam axis. The regions are designed projectively with respect to the nominal interaction point. Iron absorbers of 80 cm thickness are placed between muon stations. The stations themselves are multi-wire proportional chambers for all stations and regions except for the innermost region of M1, where triple-GEM detectors [106] are used.

For every track, hits in the muon chambers are searched for in a field of interest around the track extrapolation into the muon system. The field of interests are parametrised as a function of the track momentum and tracks are identified as muons

⁷For charged particles this implies the assumption of negligible deflection in the magnetic field.

⁸Up to corner cases of detector performance studies, e. g. [87].

2. THE LHCb EXPERIMENT

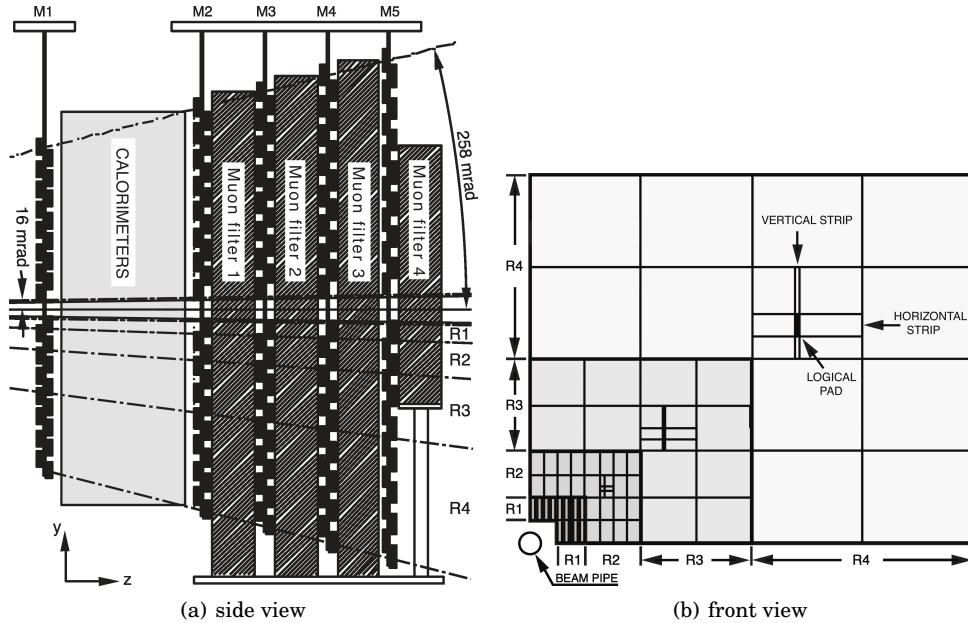


Figure 2.12: Layout of the LHCb muon system. Figure taken from [82].

if hits are found in M2 and M3 for $3 \text{ GeV}/c < p < 6 \text{ GeV}/c$, an additional hit in M4 or M5 is required up to $10 \text{ GeV}/c$, and hits in all four stations are required above $10 \text{ GeV}/c$.

This muon identification serves as a muon pre-selection [93]. For further quantification of the track's likelihood to be a muon, the mean squared distance of the muon hits to the track extrapolation is computed. The distance of these hits to other tracks in the event is a measure for the misidentification likelihood.

2.5 Particle identification

Eventually, the information from the particle identification systems is combined. In this work, the ProbNN variables ([85]) are used. They are neural networks, one for the identification of each of electrons, muons, pions, kaons, protons, and fake tracks. The muon network is named \mathcal{M}_{PID} for unified typesetting in this work. The neural network is only evaluated for tracks passing the muon pre-selection, described in the previous section. The following input quantities are used by \mathcal{M}_{PID} :

- The track's likelihood to be a muon, by means of the distance of hits in the muon chambers from the track extrapolation.
- The likelihood of these muon chamber hits to originate from a different track, by means of the distance of these muon chamber hits from other charged particle track extrapolations.
- The number of charged particle tracks compatible with these muon chamber hits.

- The information which of the three RICH radiators have been used in the RICH identification of the track.
- The information whether the track momentum is above the threshold for Cherenkov radiation for different mass hypotheses.
- The different mass hypotheses likelihood differences in the RICH identification (for the muon, pion, kaon, proton, electron, and the below-threshold hypotheses).
- The information whether the track is in the acceptance of the calorimeter subsystems.
- The energy deposit in the ECAL and HCAL at the track extrapolation, to evaluate the compatibility with a minimum ionising particle.
- dE/dx from the VELO ADC measurements.

Given that the behaviour of all these variables depends not only on the true origin of track, but depend on the track momentum and the accuracy to which the track position in the PID detectors is estimated, also variables of the track reconstruction are used. Additionally, a track does not necessarily correspond to a charged particle, but might be a fake track. The tracking variables used in \mathcal{M}_{PID} are:

- The track momentum
- The track transverse momentum
- The track fit χ^2/ndf
- The number of degrees of freedom in the track figure
- The likelihood that the track is a fake track $\mathcal{M}_{\text{fake track}}$

This neural network particle identification is expected to provide better muon from non-muon separation. Furthermore, the present analysis's aim is not to separate muons from a single other particle type, but all other particle types (including fake tracks). In $\Delta \log \mathcal{L}$, several discriminations would have been needed to be applied ($\Delta \log \mathcal{L}(\mu - \pi)$, $\Delta \log \mathcal{L}(\mu - K)$, ...), while \mathcal{M}_{PID} allows the usage of a single discrimination criterion.

The true performance of \mathcal{M}_{PID} becomes visible after its calibration – the particle identification is subject of significant differences between simulated and real events, as shown in Fig. 10.8. The background of the present analysis is shown in Fig. 10.9 for comparison. The background is extremely concentrated at small \mathcal{M}_{PID} response values.

2.6 Trigger system

The trigger system consists of two stages [86, 92]. A hardware trigger system is designed to evaluate information from the calorimeters and the muon chambers at 40 MHz input rate. If the hardware trigger accepts a proton-proton collision, then the detector is read out, at a maximum rate of 1 MHz. A fast event reconstruction is then done in a computing farm (“event filter farm”, EFF), to select events for later

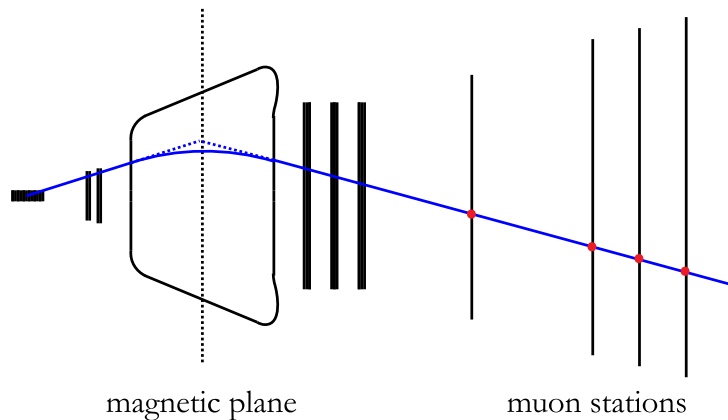


Figure 2.13: Schematic of the momentum determination in the hardware muon trigger. Figure taken from [98].

analyses. The software trigger output rate⁹ has been 3 kHz in 2011. To not leave the computing farm unused in the time between LHC fills, events were buffered in 2012 [107]; i. e. the software trigger spent more computing time for the analysis of events than available at an input rate of 1 MHz. Consequently, evaluation of events continued after the end of a fill until the event buffer was drained. The resulting output rate of the software trigger, normalised to the fill duration, reached 5 kHz.

Hardware stage

Most detector components of LHCb can be read out at a maximum rate of 1 MHz while proton bunches cross each other at the LHCb interaction point at a 20 MHz frequency¹⁰. The actual bunch crossing rate is lower than this frequency, as explained in the footnote on this page. The muon system and the calorimeters can be read out at up to 40 MHz. A custom built electronic system decides by means of the information from these subdetectors, whether the entire detector is read out and its data is sent to the EFT, or not.

A positive trigger decision requires a hadronic shower with $E_T > 3.5$ GeV, an electromagnetic shower with $E_T > 2.5$ GeV, or one or two muon tracks. For muon tracks, the track slope in the x - z plane and the assumption that the muon originates in the primary interaction region constrain the curvature in the magnetic field and thus determine the muon momentum, as illustrated in Fig. 2.13 and explained in more detail in [95]. The muon hardware trigger requires one muon with $p_T > 1.48$ GeV/ c or two muons with $\sqrt{p_{T,1} \cdot p_{T,2}} > 1.296$ GeV/ c . There is no requirement that the two muons must have opposite charge.

These thresholds refer to the 2011 data taking period. In the 2012 data taking period, the thresholds were raised to adapt to the higher instantaneous luminosity and

⁹Different definitions of the output rate exist in the literature because not only events relevant for physics analyses are stored – an additional part of the output bandwidth is spent for calibration, monitoring, and luminosity measurement purposes.

¹⁰The RF frequency of the LHC is actually 40 MHz. At a bunch spacing of 50 ns, however, every second well of the RF wave is not occupied by a proton bunch. Additionally, gaps between bunch trains decrease the bunch crossing rate even below the bunch crossing frequency.

higher collision energy in 2012. The thresholds were raised to 3 GeV for the hadronic trigger, 3 GeV for the electromagnetic trigger, 1.76 GeV/ c for the single muon trigger, and 1.6 GeV/ c for the di-muon trigger.

Software stage

The software trigger is subdivided into two main components, HLT1 and HLT2. In each stage various so-called trigger lines are run. Each line in HLT1 runs its own reconstruction and applies a simple selection. In general a single muon or hadron track with a non-vanishing impact parameter with respect to the primary vertex is required. If an event is accepted by at least one HLT1 line, the event is passed on to HLT2. For more details about the HLT1 lines, see [86]. A full event reconstruction is done in HLT2, which is simplified with respect to the offline reconstruction due to the constraints of the available computing resources in the EFF. The HLT2 applies about 300 different event selections.

While the bandwidth of the trigger stages becomes smaller with each stage, the selection requirements are not restricted throughout the trigger system; e.g. the transverse momentum requirement for the 2012 single muon hardware trigger is 1.76 GeV/ c , but only 1 GeV/ c in the subsequent HLT1. This inconsistency has advantages as explained in the following example. Events which contain $B \rightarrow \mu X$ decays, where the muon may have a transverse momentum of, for example, “only” 1.2 GeV/ c can be accepted if a higher momentum prompt muon is present in the event, i. e. one muon “triggers” the hardware trigger and another muon “triggers” the software trigger. Lowering the p_T threshold in the hardware trigger to consistency with HLT1 would not be possible within the available bandwidth.

The same is true for the selections applied by analyses to the recorded data. The efficiency of the trigger system is commonly defined as the probability to accept an event in the trigger system, if the event fulfils the selection requirements of an analysis. If selection requirements were more restrictive than the trigger requirement, these efficiencies would have to be $\equiv 100\%$. The advantage of applying selections which are less restrictive than the trigger is to access signal candidates below the trigger thresholds if there is an independent signature in the event leading to a positive trigger decision.

Assume an event e which is accepted by trigger line L and contains a three track combination $c = (t_1, t_2, t_3)$ which fulfils the event selection in Chap. 6. There are three ways¹¹ in which the positive trigger decision of L is related to the $\tau \rightarrow \mu\mu\mu$ candidate c .

TOS: The three tracks $t_{1,2,3}$ are sufficient to fulfil the requirements of L . This includes cases where a proper subset of $t_{1,2,3}$ is sufficient to trigger L . (**T**riggered **O**n **S**ignal)

TIS: Another set of tracks T which is independent of c (i. e. $t_i \notin T$) is sufficient to fulfil the requirements of L , e.g. for single track triggers. (**T**riggered **I**ndependent **S**ignal)

TOB: Neither c nor any independent set of tracks is sufficient to fulfil the requirements of L . But the combination of t_i with more tracks from the event is suffi-

¹¹See [108] for a detailed explanation of the abbreviations and the application of the former two.

cient. This includes cases where a proper subset of $t_{1,2,3}$ and other tracks trigger L . (**Triggered On Both**)

Additionally, if L finds multiple candidates in e , combinations of the above are possible. The trigger lines have been used in various configurations throughout data taking, to adapt to increasing computing resources or to incorporate improvements of the reconstruction software. Due to the luminosity levelling (see above) a change of the trigger configuration is not necessary during a fill. The trigger lines relevant for the present analysis are explained in Chap. 7.

2.7 Simulation

In the simulation, pp collisions are generated using Pythia [109, 110] with a specific LHCb configuration [111]. Decays of hadronic particles are described by EvtGen [112], in which final-state radiation is generated using Photos [113]. The interaction of the generated particles with the detector and its response are implemented using the Geant toolkit [114, 115] as described in [116]. The decay $\tau \rightarrow \mu\mu\mu$ is simulated as a phase space decay of a spin-less particles into spin-less particles, i. e. with a constant Dalitz distribution.

Branching fraction limits in absence of a signal

This chapter gives a general explanation how an upper limit on a branching fraction is obtained in case no signal events are observed. The statistics framework is the CLs method¹ [117] in the `mclimit` implementation [118] using a simple likelihood ratio.

3.1 The fraction of decays and its branching fraction

A branching fraction \mathcal{B} is the fraction of a given set of particles of size N_{tot} which is *expected* to undergo a certain decay N_{exp} . In the present case:

$$N_{\text{exp}}(\tau \rightarrow \mu\mu\mu) = \mathcal{B}(\tau \rightarrow \mu\mu\mu) \times N_{\text{tot}}(\tau). \quad (3.1)$$

In practice, a real $\tau \rightarrow \mu\mu\mu$ decay can only be observed² at a probability $\varepsilon \lesssim 1$, called *efficiency*.

$$N_{\text{exp.det.}}(\tau \rightarrow \mu\mu\mu) = \varepsilon \cdot N_{\text{exp}}(\tau \rightarrow \mu\mu\mu) \quad (3.2)$$

It is matter of statistics to define an estimate for the expected number of observed $\tau \rightarrow \mu\mu\mu$ decays. The most trivial approach is to use the observed mean as an estimate for the true mean; i. e. the number of observed decays $N_{\text{obs}}(\tau \rightarrow \mu\mu\mu)$.

$$\mathcal{B}_{\text{est}}(\tau \rightarrow \mu\mu\mu) = \frac{N_{\text{obs}}(\tau \rightarrow \mu\mu\mu)}{\varepsilon N_{\text{tot}}(\tau)} \quad (3.3)$$

In the absence of observed signal events ($N_{\text{obs}} = 0$), or in the absence of a significant number of signal events, upper limits are set for the branching fraction. In Poissonian statistics, the upper limit UL on the expectation value of can be taken from tables. For $N_{\text{obs}} = 0$ and a 90 % confidence level, UL is 2.3025. Equation 3.3 reads for upper limits:

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < \frac{UL(\tau \rightarrow \mu\mu\mu)}{\varepsilon N_{\text{tot}}(\tau)}, \quad (3.4)$$

$$=: \alpha \times UL(\tau \rightarrow \mu\mu\mu). \quad (3.5)$$

The proportionality factor α is also referred to as *normalisation factor*.

¹The letters ‘‘CLs’’ are used as the name of the method, the CLs method, and as a name for a variable which is used in the method, the CL_s variable.

²abbreviated with ‘‘det.’’ for *detection* to avoid confusion with the actual observed event count in an experiment

The normalisation factor measures an experiment’s capability to observe a decay, it combines the number of particles to begin with and the efficiency to observe an decay given that it occurred. Obviously, a small normalisation factor is desirable and the sensitivity is a linear function of α . The normalisation factor allows to compute the expected number of $\tau \rightarrow \mu\mu\mu$ decays visible to the experiment by rewriting Eqs. 3.1, 3.2, and 3.5 to

$$\begin{aligned} N_{\text{exp.det.}}(\tau) &= \varepsilon \cdot \mathcal{B}(\tau \rightarrow \mu\mu\mu) \times N_{\text{tot}}(\tau) \\ N_{\text{exp.det.}}(\tau) &= \mathcal{B}(\tau \rightarrow \mu\mu\mu) \times \alpha^{-1}. \end{aligned} \quad (3.6)$$

In other words, the normalisation factor is the branching fraction of exactly one expected visible decay. It is emphasised, that in this simple case, the sensitivity depends linearly on the efficiency (as contained in α).

The computation of UL can be found in Appendix B.3. This description is not applicable to the present analysis because of the presence of backgrounds. The implications of the presence of backgrounds for the statistical treatment of UL are illustrated in the next section.

3.2 Presence of backgrounds

Several possible treatments of background events in the determination of an upper limit on a branching fraction have been suggested in the literature. The CLs method suggested by [117] is one of the methods considered state-of-the-art. How the CLs method is used in the present analysis³ is described in Sect. 3.5.

Naïvely, when b background events are expected the number of observed background events will be within $b \pm \sqrt{b}$ events⁴. A conservative interpretation of actually observing what is expected – i. e. b events – is therefore that $b - \sqrt{b}$ background events and \sqrt{b} true $\tau \rightarrow \mu\mu\mu$ decays occurred. From two experiments with the same normalisation factor, the one with less expected background events will thus have the better sensitivity. Furthermore, a large normalisation factor can be overcompensated if only very few background events are expected.

3.3 Background discrimination

Given that a small number of background events can overcompensate a large normalisation factor, an experiment can benefit from artificially decreasing its efficiency if this is compensated by a sufficient reduction of background events. In addition to the term efficiency, the term *background efficiency* is used throughout this document for the probability that a background event passes the selection requirement. The letter ξ is used as short form.

“The use of multivariate classifiers has become commonplace in particle physics.”⁵ A multivariate classifier is a real-valued function of several candidate properties; the values of the function are called *response*, in this document \mathcal{M} is used for both, the multivariate classifier and its response. Typically, \mathcal{M} is designed to be withing $[0, 1]$.

³It is pointed out in [119] that the term “CLs method” only describes a part of the statistics framework and further information must be given to unambiguously specify what is done.

⁴Modelling the event count as a Poissonian process, the variance σ^2 of the event count is equal to the expectation value. For large expectation values, the $\pm 1\sigma$ interval has the desirable coverage of 68 %.

⁵Quoting the abstract of [120]. The usage and outlook of multivariate methods in particle physics can be found in [121, 122, 123].

For the purpose of background discrimination, the responses for signal events are desired to be large, while background event responses are desired to be small:

$$\varepsilon(\mathcal{M} > x) > \xi(\mathcal{M} > x) \quad \forall x.$$

The simplest application of a multivariate classifier is a selection criterion x_1 and evaluating the event count only in events with $\mathcal{M} > x_1$. This reduces the efficiency of the analysis $\varepsilon_{\max} \searrow \varepsilon_1$ with respect to an analysis without any selection beyond having reconstructed a candidate. A formula to determine the optimal value for x_1 is given in [124] by FOM_{Punzi} . It is given in Appendix A.3.1.

3.4 Event classification

An analysis strategy can be optimised beyond the application of a multivariate event selection. Assuming the data with $\mathcal{M} > x_1$ have been analysed, an experiment can also investigate the complementary data with $\mathcal{M} < x_1$. I. e. two experiments are performed:

$$\begin{aligned} E_1 &\text{ analyses data with } \mathcal{M} > x_1 \\ E_2 &\text{ analyses data with } \mathcal{M} < x_1. \end{aligned}$$

The efficiencies of the two experiments are ε_1 and $\varepsilon_{\max} - \varepsilon_1$. The first experiment, E_1 , can be assumed to be optimal: a multivariate event selection is applied with an optimal selection value x_1 , and therefore E_1 cannot be improved further. The other experiment, E_2 , will have less sensitivity than E_1 , but not vanishing sensitivity, and E_2 has not been optimised. When combining the results of E_1 and E_2 an improvement over the single experiment E_1 can be achieved. The combination of experiments is explained in Sect. 3.6.

The argument of applying a multivariate event selection in the first place holds for E_2 . In other words, E_2 can be improved by introducing a selection requirement $\mathcal{M} > x_2$. The value can be optimised with FOM_{Punzi} , again. Since FOM_{Punzi} is protected against divergence to vanishing efficiency, the value of x_2 must turn out different from x_1 (The data of E_2 is pre-selected with $\mathcal{M} < x_1$ and $\varepsilon(x_2 < \mathcal{M} < x_1) = 0$ for $x_1 = x_2$). One arrives at:

$$\begin{aligned} E_1 &\text{ analyses data with } \mathcal{M} > x_1 && \text{with an efficiency } \varepsilon_1 \\ E_2 &\text{ analyses data with } x_2 < \mathcal{M} < x_1 && \text{with an efficiency } \varepsilon_2. \end{aligned}$$

The efficiencies are bound by the original experiments efficiency $\varepsilon_1 + \varepsilon_2 \lesssim \varepsilon_{\max}$. And there is a data set with $\mathcal{M} < x_2$ “left over” which can be analysed. Iterating further (and setting $x_0 := 1$) one arrives at N experiments

$$E_i \text{ analyses data with } x_i < \mathcal{M} < x_{i-1} \quad \text{with an efficiency } \varepsilon_i.$$

Adding a last non-optimal experiment with $x_{N+1} := 0$ one achieves “completeness” $\sum_{i=1}^{N+1} \varepsilon_i = \varepsilon_{\max}$. The contribution of the last experiment is tiny once N is sufficiently large. Furthermore, practical aspects like constrained computing resources are strong arguments against E_{N+1} and only the best N data ranges are analysed: $\sum_{i=1}^N \varepsilon_i \lesssim \varepsilon_{\max}$.

The individual data sets with $x_i < \mathcal{M} < x_{i-1}$ can be called *classes*, which motivates the word “multivariate *classifier*”. The division of a large data set into classes

by means of a classifier is consequently an event *classification*. An alternative naming is the word *bin* for the individual ranges of the \mathcal{M} response. The values x_i are consequently called *bin boundaries*.

The above assumption, that E_1 is optimal, is only partially correct. The experiment E_1 can not be improved as an individual experiment. The derivation of FOM_{Punzi} , however, assumes that the data which is not selected by $\mathcal{M} > x_1$ will not be analysed at all. But once E_1 is combined with E_2 , it turns out that a larger value for x_1 would have reached a better *combined* sensitivity, as shown in numeric examples in Appendix A. Consequently, a fully optimal *combined* analysis will have larger values x_i for all bin boundaries than in the motivational example here.

3.5 The CLs method

Abstractly speaking, the purpose of recording and investigating data for the search for $\tau \rightarrow \mu\mu\mu$ is to distinguish whether the data is an evidence for either the hypothesis “ $\tau \rightarrow \mu\mu\mu$ occurs at a pre-defined branching fraction \mathcal{B} ” or the hypothesis “ $\tau \rightarrow \mu\mu\mu$ does not occur, there is only background”⁶. I. e. one branching fraction hypothesis is tested at a time (*test hypothesis*). The restriction of testing only one hypothesis is dropped at the very end of this section.

It is shown in [126] that the best way of doing so is the computation of the likelihood ratio

$$\lambda = \frac{P_{H_1}(\vec{x})}{P_{H_0}(\vec{x})},$$

where H_i are the respective hypotheses, \vec{x} is the observed data, and $P_{H_i}(\vec{x})$ is the probability to observe the data \vec{x} under the hypothesis H_i . For easier distinction, instead of P_{H_i} the notation L_b and L_{s+b} is used for the hypothesis of vanishing $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ (b for “background”) and the hypothesis of the presence of a signal ($s + b$ for “signal and background”), respectively.

The data observed which is relevant for the search for $\tau \rightarrow \mu\mu\mu$ is the number of observed events; $\vec{x} = n$. The numbers of expected background events and signal events are part of the probability functions. For each branching fraction, one likelihood ratio λ needs to be defined to reject or not reject that branching fraction hypothesis. For an analysis using event classification, as motivated in Sect. 3.4, \vec{x} is a vector of observed events in each bin: $\vec{x} = (n_1, n_2, \dots)$.

Motivated by Gaussian approximation of probabilities by the central limit theorem, instead of λ itself one uses

$$Q_{\text{obs}} = -2 \ln \lambda = -2 \ln \frac{L_{s+b}(\vec{x})}{L_b(\vec{x})}.$$

Q is called *test statistics* and takes small values⁷ for data which is more likely occur under the $s + b$ hypothesis than under the b hypothesis, the index obs indicates that it is the observed value of Q and not the variable of the probability density function under a hypothesis.

Different test statistics are possible as well. The most simple one would be the number of observed events. Q has the advantage that it is additive. Combining analyses from two experiments, the combined test statistics is the sum of the individual

⁶This is not to be confused with profile likelihood ratio methods, [125].

⁷In the sense that “ x is more negative than y ” counts as “smaller”, too.

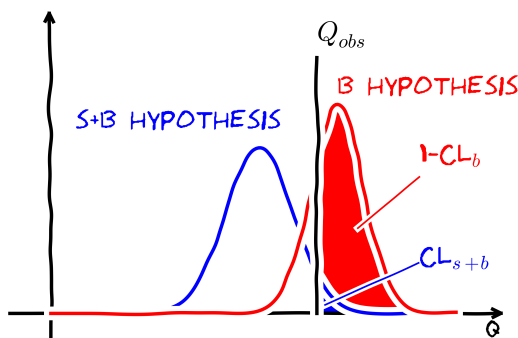


Figure 3.1: Example of the distribution of Q values under the $s + b$ hypothesis and the b hypothesis for a fixed branching fraction. The integrals CL_{s+b} and $1 - CL_b$ are indicated as well.

test statistics. As previously described by [127, Sect. 3.1-3.2], Q of an analysis which is performed in bins, is the sum of the Q values of the bins. Moreover, Q is invariant under increasing the number of bins by randomly splitting a bin into two bins (see Eq. B.1). Studying Q values allows to compare analysis strategies with different numbers of bins.

The $s + b$ hypothesis is usually rejected if Q_{obs} is larger than a critical value Q_{crit} which is chosen such that the probability of observing a Q value larger or equal to Q_{crit} under the $s + b$ hypothesis is $1 - c$, where c is the pre-defined, desired confidence level. The probability of observing a Q value larger or equal to Q_{obs} value under the $s + b$ hypothesis is called CL_{s+b} :

$$CL_{s+b} := \mathcal{P}(Q \geq Q_{\text{obs}} | s + b)$$

An illustration is given in Fig. 3.1. It is visible, that CL_{s+b} takes small values if Q_{obs} is in the range of typical values as expected under the background hypothesis.

It is argued in [117] that this leads to false rejections at a rate of $1 - c$, even if there was no sensitivity to distinguish between b and $s + b$. This behaviour is undesired, although it is perfectly correct and unavoidable in frequentist statistics. The CLs method ([117, 128]) is an ad-hoc correction. In addition to the CL_{s+b} value, the corresponding value for the b hypothesis is defined; $1 - CL_b$ is the probability of observing a Q value larger or equal to Q_{obs} value under the b hypothesis.

$$1 - CL_b := \mathcal{P}(Q \geq Q_{\text{obs}} | b)$$

The variable which the decision of rejecting $s + b$ is then based on is

$$CL_s := \frac{CL_{s+b}}{1 - CL_b}, \quad (3.7)$$

which is treated like CL_{s+b} above: $s + b$ is rejected if CL_s is smaller than $1 - c$. In the extreme case, where there is no sensitivity to distinguish the hypotheses, $1 - CL_b$ and CL_{s+b} are equal, $CL_s = 1$, which will never lead to a rejection in the case of no sensitivity.

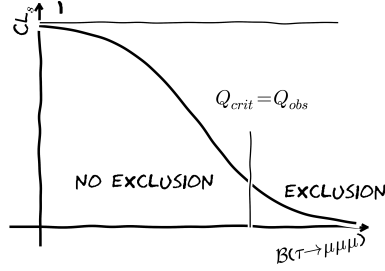


Figure 3.2: Example for CL_s as a function of $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$. There is no sensitivity for very small branching fractions, where CL_s is 1, and large branching fractions can certainly be excluded, where CL_s takes very small values. In between the upper limit on the branching fraction depends on the desired confidence level.

It is pointed out, that the CLs method only rejects (or does not reject) the $s + b$ hypothesis for a fixed branching fraction. When searching for an upper limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$, the smallest branching fraction which can be rejected is interpreted as upper limit. This means, that for several branching fraction hypotheses CL_s is computed, as illustrated in Fig. 3.2.

3.6 Nuisance parameters

The CLs method still leaves freedom about the actual computation of L_{s+b} and L_b , as well as the computation of CL_{s+b} and CL_b . Natural choices for the former are Poisson probabilities. The combination of several bins is done by multiplying the individual probabilities, as motivated by [128]

$$L_{s+b} = \prod_i \frac{(s_i + b_i)^{n_i}}{n_i!} e^{-(s_i + b_i)}$$

$$L_b = \prod_i \frac{(b_i)^{n_i}}{n_i!} e^{-b_i}$$

where i ranges over the analysis bins, b_i is the number of expected background events in bin i , s_i is the expectation for the number of observable $\tau \rightarrow \mu\mu\mu$ decays in bin i assuming a certain branching fraction, and n_i is the number of observed events in bin i .

In practice, there are no certain background expectations b_i . The true mean of background events is estimated⁸ by an estimate \hat{b}_i with some error σ_{b_i} . Furthermore, even setting $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ to a fixed value for the $s + b$ hypothesis does not fix s_i due to the uncertainty on $N_{\text{tot}}(\tau)$ and the efficiency, i. e. the normalisation factor. There is an efficiency for each analysis bin, ε_i .

The number of expected background events and the true efficiencies are so-called nuisance parameters. Their deviation from their estimated values is denoted by $\vec{\theta}$.

⁸Consequently, from here on the cumbersome formulation of “estimated number of expected background events” is used.

The full test statistics is therefore

$$Q_{\text{obs}}(\vec{\vartheta}) = -2 \ln \frac{L_{s+b}(\vec{x}, \vec{\vartheta})}{L_b(\vec{x}, \vec{\vartheta})}.$$

The original description of the statistical method used in this analysis is given in [118]. The possibility of profiling and marginalising the nuisance parameters, described in [118], are not used because that requires significantly more computing resources and no sensitivity improvement is expected.

For this analysis, $\vec{\vartheta}$ is fixed to the expected value for the definition of Q , i. e. the uncertainties on $\vec{\vartheta}$ is not considered in the computation of L_{s+b} and L_b .

$$Q_{\text{obs}} = -2 \ln \frac{L_{s+b}(\vec{x}, \vec{\vartheta} = 0)}{L_b(\vec{x}, \vec{\vartheta} = 0)}.$$

No systematic error is introduced by fixing the nuisance parameters in the computation of Q_{obs} as long as the integration of the Q distribution considers the nuisance parameters, i. e. the computation of CL_{s+b} and CL_b . The computation of CL_{s+b} and CL_b is done by Monte Carlo integration, i. e. a large number of pseudo-experiments is generated.

CL_b : For each pseudo-experiment, the number of observed events x_i is generated as a Poissonian random number with mean b_i . The expected number of background events b_i is itself a random number in each pseudo-experiment; it is Gaussian distributed around the estimated number of expected background events \hat{b}_i and with a width as the estimated uncertainty of the estimated expected background level $\hat{\sigma}_b$. Then, CL_b is the fraction of pseudo-experiments in which Q is larger than Q_{obs} .

CL_{s+b} : For CL_{s+b} , the Poissonian random number x_i is generated for a mean $b_i + s_i$; where b_i is defined as before and s_i is the number of expected signal events; it is a Gaussian random number with mean $\mathcal{B} \times \alpha_i^{-1}$ and width $\mathcal{B} \times \sigma_{\alpha_i^{-1}}$, i. e. the fluctuated mean of $\tau \rightarrow \mu\mu\mu$ decays in that bin as computed by Eq. 3.6, using the estimated normalisation factor (uncertainties stem from the efficiency determination, but also from the estimation of $N_{\text{tot}}(\tau)$).

To keep the expected numbers of signal and background events positive, the aforementioned Gaussian distributions are truncated at zero because negative numbers of expected events are unphysical. The truncation is implemented by rejecting the Gaussian random numbers b_i or s_i if they are negative and generating a new random number. The truncated Gaussian distribution thereby becomes

$$g(x) = \mathcal{N} \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & \forall x > 0 \\ 0 & \forall x \leq 0 \end{cases}$$

where \mathcal{N} is a normalisation factor to ensure $1 = \int_0^{\infty} dx g(x)$.

3.7 Expected sensitivity

In addition to the actual result – the observed exclusion limit – also the expected exclusion limits are reported when publishing the outcome of a rare decay search. The

3. BRANCHING FRACTION LIMITS IN ABSENCE OF A SIGNAL

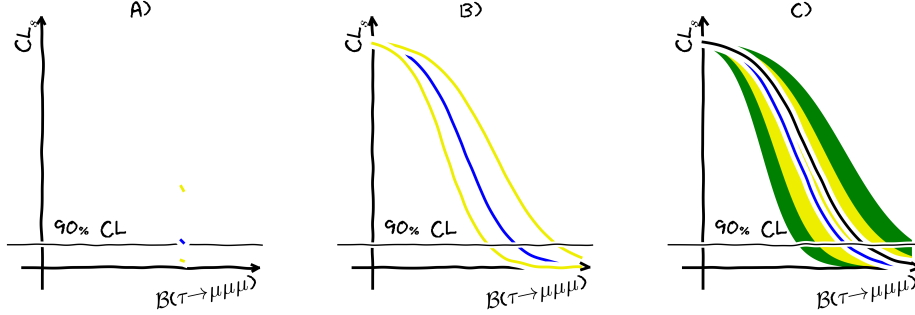


Figure 3.3: a) For each branching fraction, percentiles of the expected CL_s distribution indicate the boundaries of the range of a given coverage. The blue dot indicates the median of the CL_s distribution for the indicated branching fraction test hypothesis. The yellow dots indicate the 16 % and 84 % percentiles of the CL_s distribution for that test hypothesis. The CL_s distribution is generated for the absence of a signal. b) The same, evaluated for several branching fraction test hypotheses, the yellow lines can be read horizontally for the 1σ range in which the exclusion level is expected. c) In addition to the 1σ coverage in yellow, the 2σ coverage is shown in green, additionally. The black line is the observed CL_s value as a function of the test hypothesis's branching fraction.

expected exclusion limit is based on a hypothesis which is assumed (expected) to be true. Under this hypothesis, pseudo-experiments are generated as for the treatment of the nuisance parameters. In the case of $\tau \rightarrow \mu\mu\mu$, these are pseudo-experiments containing only background events.

The observed CL_s distribution from the pseudo-experiments is then interpreted as an expectation for the outcome of the experiment, by determining percentiles of the distribution. A centred 68 % coverage interval is the range between the percentiles at 16 % and at 84 %. Therefore, the 1σ range in which CL_s is expected, is the range between these percentiles. This is illustrated in Fig. 3.3 a). In the illustrated example, a 90 % confidence level is investigated, half of the pseudo-experiments had a CL_s value larger than 0.1, the expected CL_s value (in blue) is thus 0.1. At the 1σ level, CL_s is expected to be within the two yellow markers.

Applying these definitions to a whole range of branching fraction hypotheses, the upper limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is at the 1σ level expected to be larger than the branching fraction, for which CL_s is expected to be larger than $1 - c$ at the 1σ level. In Fig. 3.3 b) this is where the lower yellow line intersects the $1 - c$ line. The upper limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is furthermore expected to be smaller than the smallest branching fraction for which CL_s is expected to be smaller than $1 - c$ at the 1σ level. This is where the upper yellow line intersects the $1 - c$ line. This means that the upper limit is expected to be within the intersections of the yellow lines with the $1 - c$ line.

The full information about the sensitivity and the outcome of the experiment is given in figures like Fig. 3.3 c). Not only the expected 1σ range of CL_s values as a function of $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is shown as the yellow band, but also the observed values in black, the 0σ expected range (the median of the expected CL_s distribution) in blue, and the 2σ expected range in green. In the example the exclusion limit turned out

slightly worse than the expected exclusion limit, but still within the expected range.

3.8 Model dependence

A statement of the kind $\mathcal{B}(\tau \rightarrow \mu\mu\mu) < X$ does not contain any model dependence. Moreover, any of the models introduced in Sect. 1.2 predicts a value for $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$. One would therefore expect that an upper limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ can immediately be used to rule out models which predict a higher value. This assumption is not correct! The introduction of efficiencies in Eq. 3.2 is not model independent because the efficiency typically depends on the dimuon invariant masses and appears integrated over the Dalitz plot in Eq. 3.2. This integration can only be done assuming a Dalitz distribution.

Correspondingly, the upper limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ determined in the present document is only valid for New Physics models which predict a flat phase space distribution of the decay products. This model will be referred to as the *nominal analysis*. Exclusion limits for the Dalitz distributions introduced in Sect. 1.3 are given in Chap. 13.

τ production at LHCb

Proton collisions are governed by strong interaction processes. Most particles produced in proton-proton collisions are therefore hadrons. The production of τ leptons is dominated by secondary processes. This chapter describes the processes relevant for τ lepton production at LHCb.

Terminology Whenever a production mode is written as $X \rightarrow \tau$, it is assumed that X is produced in the primary pp interaction vertex. E. g. the τ source D_s is not meant to include D_s originating in b hadron decays. Particles produced in the primary interaction vertex are called *prompt* particles. When both components, prompt and non-prompt production, are addressed, this will be called the *inclusive* production.

Throughout this section charge conjugation and charge conservation are implied. E. g. $D_s \rightarrow \tau\nu_\tau$ is written instead of $D_s^+ \rightarrow \tau^+\nu_\tau$ and $D_s^- \rightarrow \tau^-\bar{\nu}_\tau$. The D^\pm mesons will still be written as D^+ , with a charge indicated, to avoid possible confusion with D^0 .

4.1 Heavy flavour τ production

$D_s \rightarrow \tau\nu_\tau$ production of τ leptons

The largest part of the τ leptons produced at LHCb comes from leptonic D_s decays. Their production cross section has been measured by the LHCb collaboration to be $\sigma(D_s) = (197 \pm 31) \mu\text{b}$ [129]. The only $D_s \rightarrow \tau X$ decay is the leptonic decay shown in Fig. 4.1. The $D_s \rightarrow \tau\nu_\tau$ branching fraction is $(5.43 \pm 0.31)\%$, averaged by [80] from the measurements by the BaBar and CLEO collaborations.

The expected τ production cross section for $D_s \rightarrow \tau$ transitions within the LHCb acceptance is thus

$$\sigma(\text{pp} \rightarrow \tau + X)_{D_s} = (10.7 \pm 1.8) \mu\text{b}.$$

$D^+ \rightarrow \tau\nu_\tau$ production of τ leptons

Similarly to the previous decay channel, τ leptons can be produced in D^+ decays. The decay $D^+ \rightarrow \tau\nu_\tau$ is unobserved, so far. The branching fraction is therefore computed.

The decay width for any charged pseudo-scalar meson to lepton decay is

$$\Gamma(P \rightarrow \ell\nu) = \frac{G_F^2}{8\pi} f_P^2 m_\ell^2 M_P \left(1 - \frac{m_\ell^2}{M_P^2}\right)^2 |V_{q_1 q_2}|^2, \quad (4.1)$$

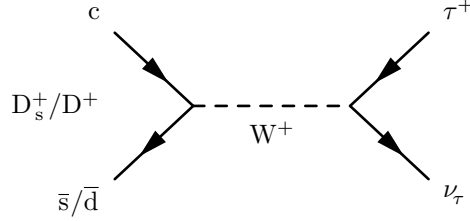


Figure 4.1: Feynman diagram of the leptonic $D_s^+ \rightarrow \tau \nu_\tau$ and $D^+ \rightarrow \tau \nu_\tau$ decay.

where G_F is the Fermi coupling constant, P denotes the pseudo-scalar meson, M_P its mass, f_P the meson's decay constant, m_ℓ the final state lepton mass, $V_{q_1 q_2}$ the CKM matrix element for the meson's quark flavours. Only the mass dependence, $\Gamma(P \rightarrow \ell \nu) \propto m_\ell^2 (1 - m_\ell^2/M_P^2)^2$, is relevant, here.

The $D^+ \rightarrow \mu \nu_\mu$ branching fraction is $(3.82 \pm 0.33) \times 10^{-4}$ [130]. Using the above expression to calculate the ratio of decay widths for decays to μ leptons to τ leptons, it is found that

$$\mathcal{B}(D^+ \rightarrow \tau \nu_\tau) = \frac{m_\tau^2 (M_{D^+}^2 - m_\tau^2)^2}{m_\mu^2 (M_{D^+}^2 - m_\mu^2)^2} \times \mathcal{B}(D^+ \rightarrow \mu \nu_\mu) \quad (4.2)$$

$$= (1.02 \pm 0.09) \times 10^{-3}. \quad (4.3)$$

This is in agreement with the exclusion limit $\mathcal{B}(D^+ \rightarrow \tau \nu_\tau) < 1.2 \times 10^{-3}$ reported in [130].

The D^+ production cross section has been measured by the LHCb collaboration in [129]. It is $(645 \pm 74) \mu\text{b}$. The expected τ production cross section, for τ from D^+ decays is therefore

$$\sigma(\text{pp} \rightarrow \tau + X)_{D^+} = (0.66 \pm 0.10) \mu\text{b}.$$

Charm cross section extrapolation

The charm production cross section has been measured in the phase space $p_T < 8 \text{ GeV}/c$. From the p_T dependence of the production cross section in [129] it is assumed that the cross section of charm mesons with $p_T > 8 \text{ GeV}/c$ can be neglected.

Charm production cross section measurements have not been performed at LHCb for $\sqrt{s} = 8 \text{ TeV}$. Simulations of pp collisions using Pythia show that the charm cross section scales with the centre-of-mass energy, i. e. $\sigma(c\bar{c}, 8 \text{ TeV})/\sigma(c\bar{c}, 7 \text{ TeV}) \approx 8/7$.

Using the uncertainties on the proton structure functions in the simulation to assess the uncertainty of the scaling factor, it is found that the uncertainty on the scaling factor is negligible with respect to the charm cross section measurement at 7 TeV.

$b \rightarrow \tau X$ production of τ leptons

Branching fractions for b hadron decays into a τ lepton are not available for all b-flavoured particle. There are, however, inclusive measurements of $\mathcal{B}(b \rightarrow \tau \bar{\nu}_\tau X)$ measured at LEP and averaged in [80] to $(24.1 \pm 2.3) \times 10^{-3}$. This branching fraction

is valid for the present analysis, if the hadronisation fractions at LHC and LEP are similar.

It has been found in [131] that the b hadronisation for light spectator mesons is in agreement with the value reported in [80, page 1105]. The production of B_c^+ is compatible with the exclusion limits on B_c^+ production at LEP. The measurements of the b hadronisation to baryons, see [132] for the latest result, is different in LHC collisions than in e^+e^- collisions at the Z resonance¹. The effect is neglected here, assuming $\mathcal{B}(\Lambda_b \rightarrow \tau X) \approx \mathcal{B}(B \rightarrow \tau X)$. See Appendix D for an approximate calculation of the effect of the Λ_b hadronisation and the B_c^+ production.

The b production cross section in the LHCb acceptance has been measured by the LHCb collaboration in [133]. It amounts to²

$$\sigma(b\bar{b}) = (49 \pm 8) \mu\text{b}.$$

The resulting expected τ cross section is

$$\sigma(\text{pp} \rightarrow \tau + X)_b = 2 \cdot (1.2 \pm 0.2) \mu\text{b}$$

where the factor 2 accounts for the pair production of b quarks.

At $\sqrt{s} = 8$ TeV the b production measurement from [134] is used³. The τ production cross section from b decays is

$$\sigma(\text{pp} \rightarrow \tau + X)_b = (2.7 \pm 0.4) \mu\text{b}.$$

The cross section measurements are accessing the phase space where the b hadron decays into a J/ψ with transverse momentum range below $14 \text{ GeV}/c$. Given the p_T distributions in [133] and [134], it is concluded that the fraction of b hadrons not within that transverse momentum range is insignificant.

$b \rightarrow D_{(s)}^\pm \rightarrow \tau^\pm$ production

The b hadrons created in LHC collisions can not only directly decay into τ leptons, but also decay into lighter hadrons which then decay into τ leptons. The only non- b flavoured hadrons which are heavier than τ leptons are c flavoured hadrons. Assuming baryon number conservation, the Λ_c^+ cannot decay into a τ due to the small mass difference $m_{\Lambda_c^+} - m_p < m_\tau$. The relevant $b \rightarrow D$ contributions to the τ lepton production are

$$\mathcal{B}(b \rightarrow D^+ X) = (23.3 \pm 1.7) \%$$

$$\mathcal{B}(b \rightarrow D^- X) \quad \text{unobserved}$$

$$\mathcal{B}(b \rightarrow D_s^+ X) = (14.7 \pm 2.1) \%$$

$$\mathcal{B}(b \rightarrow D_s^- X) = (10.1 \pm 3.1) \%.$$

¹c. f. [2, Sect. 3.1.3]

²Using Eq. (12) and the factor $\alpha_{4\pi}$ from [133].

³Dividing the 4π cross section in [134] by the factor $\alpha_{4\pi}$ from that reference.

4. τ PRODUCTION AT LHCb

With the above $D_{(s)}^+ \rightarrow \tau \nu_\tau$ branching fraction the $b \rightarrow D_{(s)}^\pm \rightarrow \tau^\pm$ branching fractions are evaluated to

$$\begin{aligned}\mathcal{B}(b \rightarrow \tau^+ X)_{D^+} &= (2.38 \pm 0.27) \times 10^{-4} \\ \mathcal{B}(b \rightarrow \tau^- X)_{D^-} &= \text{n/a} \\ \mathcal{B}(b \rightarrow \tau^+ X)_{D_s^+} &= (8.0 \pm 1.2) \times 10^{-3} \\ \mathcal{B}(b \rightarrow \tau^- X)_{D_s^-} &= (5.5 \pm 1.7) \times 10^{-3}.\end{aligned}$$

With the above b cross section the expected τ production through $b \rightarrow c$ transitions is

$$\begin{aligned}\sigma(\text{pp} \rightarrow \tau + X)_{b \rightarrow D^\pm, 7 \text{ TeV}} &= (23 \pm 4) \text{ nb} \\ \sigma(\text{pp} \rightarrow \tau + X)_{b \rightarrow D_s^\pm, 7 \text{ TeV}} &= (1.32 \pm 0.36) \text{ } \mu\text{b} \\ \sigma(\text{pp} \rightarrow \tau + X)_{b \rightarrow D^\pm, 8 \text{ TeV}} &= (26 \pm 4) \text{ nb} \\ \sigma(\text{pp} \rightarrow \tau + X)_{b \rightarrow D_s^\pm, 8 \text{ TeV}} &= (1.48 \pm 0.37) \text{ } \mu\text{b}.\end{aligned}$$

τ production from charmonium decays

Additionally to the decays of hadrons with a single c or b quark, τ leptons can also occur in the decay of quarkonium resonances. The $J/\psi \rightarrow \tau^+ \tau^-$ decay is not possible due to energy conservation. The lightest quarkonium which can decay into two τ leptons is the $\psi(2S)$ with $\mathcal{B}(\psi(2S) \rightarrow \tau^+ \tau^-) = (3.1 \pm 0.4) \times 10^{-3}$, [135].

The production cross section at $\sqrt{s} = 7 \text{ TeV}$ is $(1.44 \pm 0.32) \text{ } \mu\text{b}$ for $\psi(2S)$ produced in the primary interaction, and $(0.25 \pm 0.02) \text{ } \mu\text{b}$ for $b \rightarrow \psi(2S)$ production, as reported by LHCb in [136]. The resulting expected τ cross section is, with a factor 2 to take the production of 2 τ per $\psi(2S)$ into account,

$$\begin{aligned}\sigma(\text{pp} \rightarrow \tau X)_{\psi(2S)} &= 2 \cdot (4.5 \pm 1.1) \text{ nb} \\ \sigma(\text{pp} \rightarrow \tau X)_{b \rightarrow \psi(2S)} &= 2 \cdot (0.78 \pm 0.12) \text{ nb}.\end{aligned}$$

No cross section measurement at $\sqrt{s} = 8 \text{ TeV}$ is known to the author, given the insignificant contribution of τ production in charmonium decays with respect to the D production, no attempts is made to obtain an estimate for 8 TeV collision energy.

τ production from bottomonium decays

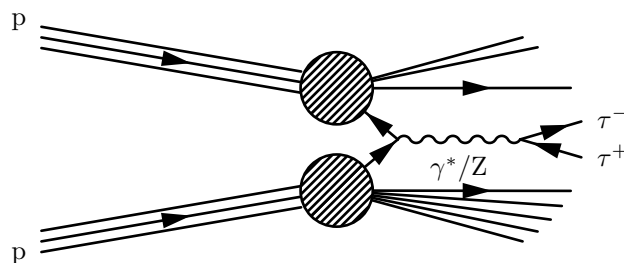
The $\Upsilon(nS) \rightarrow \tau\tau$ branching fractions are listed in [80], but the ratio of branching fractions $\mathcal{B}(\Upsilon \rightarrow \tau\tau)/\mathcal{B}(\Upsilon \rightarrow \mu\mu)$ is measured at higher accuracy and is more suitable for the present study, given that the Υ production cross section at LHCb is measured in $\Upsilon \rightarrow \mu\mu$ decay.

The product of the branching fraction $\Upsilon \rightarrow \mu\mu$ times the production cross section has been measured at LHCb at 7 TeV [137] and 8 TeV [134].

The cross section measurements and branching fractions are summarised in Tab. 4.1. Again, the LHCb cross section measurements are limited to $p_T < 15 \text{ GeV}/c$. The resulting expected τ cross sections are listed in Tabs. 4.2 and 4.3, where a factor 2 is taken into account for the number of τ in the final state.

Table 4.1: $\sigma(pp \rightarrow \tau\tau X)_\Upsilon$ contributions.

resonance	$\frac{\mathcal{B}(\Upsilon \rightarrow \tau\tau)}{\mathcal{B}(\Upsilon \rightarrow \mu\mu)}$	$\sigma(pp \rightarrow \Upsilon X) \times \mathcal{B}(\Upsilon \rightarrow \mu\mu)$
$\sqrt{s} = 7 \text{ TeV}$		
$\Upsilon(1S)$	1.008 ± 0.023	$(2.29 \pm 0.30) \text{ nb}$
$\Upsilon(2S)$	1.04 ± 0.06	$(0.56 \pm 0.07) \text{ nb}$
$\Upsilon(3S)$	1.05 ± 0.09	$(0.28 \pm 0.04) \text{ nb}$
$\sqrt{s} = 8 \text{ TeV}$		
$\Upsilon(1S)$	1.008 ± 0.023	$(3.24 \pm 0.23) \text{ nb}$
$\Upsilon(2S)$	1.04 ± 0.06	$(0.76 \pm 0.05) \text{ nb}$
$\Upsilon(3S)$	1.05 ± 0.09	$(0.369 \pm 0.027) \text{ nb}$

Figure 4.2: Drell-Yan production of τ leptons.

4.2 Electroweak τ production

Additionally to the τ production in hadron decays, there is a contribution from virtual photons, weak gauge bosons, or Higgs decays. The latter is ignored in this work because of its rather small contribution, the others are addressed in this section. These, so called, electro-weak processes are usually measured at large lepton momenta, for the present analysis, the low momentum regime is important as well, therefore corrections to the measured values are made to assess the τ production cross section.

Drell-Yan processes

Virtual photons and Z bosons can be produced in proton-proton collisions through $q\bar{q}$ annihilation. The boson can then decay into a lepton pair (Drell-Yan process), shown in Fig. 4.2.

The Drell-Yan cross section of μ pairs has been measured by LHCb [138]. The cross section in the LHCb acceptance is $(0.64 \pm 0.12) \text{ nb}$ for dimuon invariant masses between $5 \text{ GeV}/c^2$ and $40 \text{ GeV}/c^2$ with transverse muon momenta $> 1 \text{ GeV}/c$. For dimuon invariant masses between $40 \text{ GeV}/c^2$ and $120 \text{ GeV}/c^2$ it amounts to $(24 \pm 6) \text{ pb}$ with transverse muon momenta $> 3 \text{ GeV}/c$.

The different rest masses of μ leptons and τ leptons needs to be taken into account and the following effects are estimated to extrapolate to the phase space accessible for the present analysis. Leptons with a transverse momentum below $1 \text{ GeV}/c$ are relevant, Drell-Yan events with only one lepton in the acceptance need to be considered,

the invariant mass ranges above $120 \text{ GeV}/c^2$ and below $5 \text{ GeV}/c^2$ contain additional leptons.

Simulated Drell-Yan processes are used which have been generated as $pp \rightarrow \gamma^*/Z + X \rightarrow \mu\mu + X$ events with $m_{\mu\mu} > 2 \text{ GeV}/c^2$; at least one muon in the approximate LHCb acceptance, $2 < \eta_\mu < 5$; and at least one of the muons with a transverse momentum larger than $1 \text{ GeV}/c$. For the first part of the following corrections, it is additionally required that the second muon is within the approximate LHCb acceptance.

Close to the τ threshold a significant correction to the $m_{\bar{\ell}\ell}$ distribution of the simulated processes is necessary. Under the assumption that the annihilating quarks are massless and that QCD corrections to the $q\bar{q}\gamma^*$ vertex are identical for the μ and τ final state, the mass dependence of the Drell-Yan process is described by the QED cross section $\sigma(e^+e^- \rightarrow \bar{\ell}\ell)$. In the limit of small e and μ masses, the τ mass dependence of the tree level QED cross section is given by a phase space factor, a spinor-structure term, and the Sommerfeld-Sakharov factor, F_c . With the Mandelstam variable $s = m_{\bar{\ell}\ell}^2$, the correction is⁴

$$\begin{aligned} R &:= \frac{\sigma(e^+e^- \rightarrow \tau^+\tau^-)}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} \\ R &= R_0 \cdot F_c + \text{higher order corrections} \\ R_0 &= \sqrt{1 - 4m_\tau^2/s} \cdot \left(1 - \frac{2m_\tau^2}{s}\right), \end{aligned}$$

where the first factor of R_0 is the available phase factor, the second is due to the spinor structure of the QED matrix element; m_τ is the τ rest mass and F_c is given by

$$F_c = \frac{\pi\alpha/\sqrt{1 - 4m_\tau^2/s}}{1 - \exp\left(-\pi\alpha/\sqrt{1 - 4m_\tau^2/s}\right)}.$$

It parametrises the Coulomb attraction between the τ leptons. Higher order corrections are not considered here. R is used to weight the simulated events. To correctly assess the statistical power of the weighted sample, the weights are normalised with

$$\begin{aligned} w_i &= w_{\text{eff}} \cdot R_i \\ w_{\text{eff}} &:= \frac{\sum_i R_i}{\sum_i R_i^2}, \end{aligned}$$

where the sum ranges over all simulated events. The resulting invariant ditau mass distribution is shown in Fig. 4.3. Averaging R over the simulated events with $5 \text{ GeV}/c^2 < m_{\bar{\ell}\ell} < 120 \text{ GeV}/c^2$ a suppression of 0.776 ± 0.001 is expected.

The transverse momentum distribution of simulated Drell-Yan processes is shown in Fig. 4.4. In Fig. 4.4 (b) is visible that the requirement that one of the leptons must have a transverse momentum $> 1 \text{ GeV}/c$ in the simulation does not introduce a discontinuity of the p_T distribution. It is concluded that the phase space which is not simulated is negligible for τ leptons in the final state.

⁴The cross section has been computed by several authors, the formula here is taken from [139].

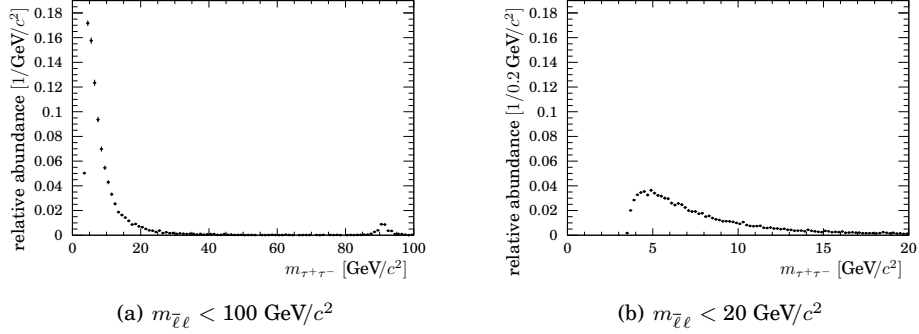


Figure 4.3: Simulated invariant ditau mass distribution in the wide mass range including the Z resonance and a zoom into the threshold region.

The measured Drell-Yan cross section is valid for processes where both leptons have a transverse momentum larger than $1 \text{ GeV}/c$. Averaged over the full spectrum, the probability that both τ are in the measured phase space, $p_T > 1 \text{ GeV}/c$ for $m_{\bar{\ell}\ell} < 40 \text{ GeV}/c^2$ or $p_T > 5 \text{ GeV}/c$ for $m_{\bar{\ell}\ell} > 40 \text{ GeV}/c^2$, is $(96.05 \pm 0.11) \%$.

Using the simulation, it is found that $(22.25 \pm 0.24) \%$ of all $\gamma^*/Z \rightarrow \tau^+\tau^-$ processes are not within the measured invariant mass range, from $5 \text{ GeV}/c^2$ to $120 \text{ GeV}/c^2$.

The reported Drell-Yan cross section is thus changed to

$$\sigma(\text{pp} \rightarrow \gamma^*/Z + X \rightarrow \tau^+\tau^- + X) = \frac{0.776 \cdot \sigma_{\text{measured}}}{0.9605 \cdot 0.7775} = (0.74 \pm 0.12) \text{ nb}$$

Since there are two τ leptons produced in each Drell-Yan process, the cross section has to be multiplied by two:

$$\sigma(\tau \text{ from Drell-Yan}) = 2 \times \sigma(\text{Drell-Yan}).$$

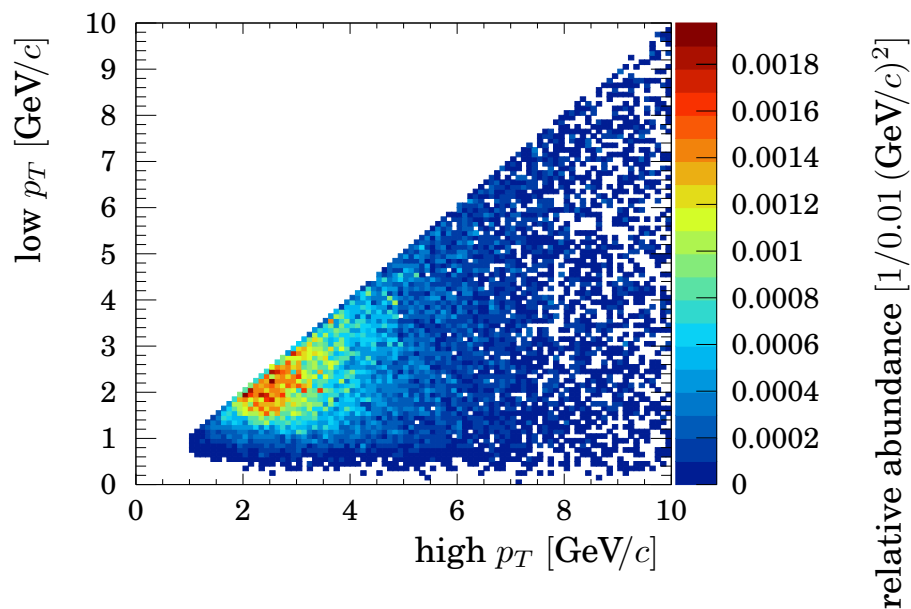
This, however, is still not fully valid because the Drell-Yan cross section only considers Drell-Yan processes with both leptons in the detector acceptance.

The effect from Drell-Yan events with only one lepton in the η acceptance is estimated using the simulated Drell-Yan events, removing the requirement of the second muon to be within the acceptance. There are about as many events where one lepton is in the LHCb acceptance as events where both leptons are in the LHCb acceptance.

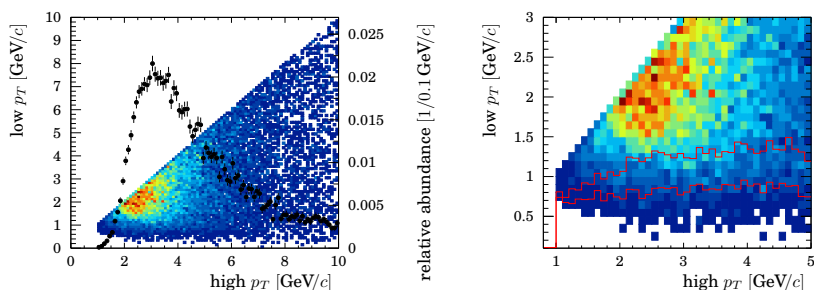
$$\sigma(\text{one Drell-Yan } \tau \text{ in acceptance}) \approx \sigma(\text{two Drell-Yan } \tau \text{ in acceptance})$$

The single τ Drell-Yan cross section is therefore assumed to be not larger than three times the above value because of

$$\begin{aligned} \sigma(\text{one or two Drell-Yan } \tau \text{ in acceptance}) &= \underbrace{\sigma(\text{two Drell-Yan } \tau \text{ in acceptance})}_{\text{contributes two leptons}} \\ &+ \underbrace{\sigma(\text{one Drell-Yan } \tau \text{ in acceptance})}_{\text{contributes one leptons}}. \end{aligned}$$



(a) Two dimensional distribution.



(b) With the one-dimensional distribution overlaid.

(c) With percentiles overlaid.

Figure 4.4: Lepton transverse momentum distribution of $\gamma^*/Z \rightarrow \tau^+\tau^-$ processes. The larger transverse momentum on the abscissa, the smaller on the ordinate. In (b) the one-dimensional distribution of the larger transverse lepton momentum is overlaid, in (c) the 10% and 2% percentiles are overlaid. I.e. for each high- p_T value, the line indicates below which low- p_T value 10% (2%) of the events are events.

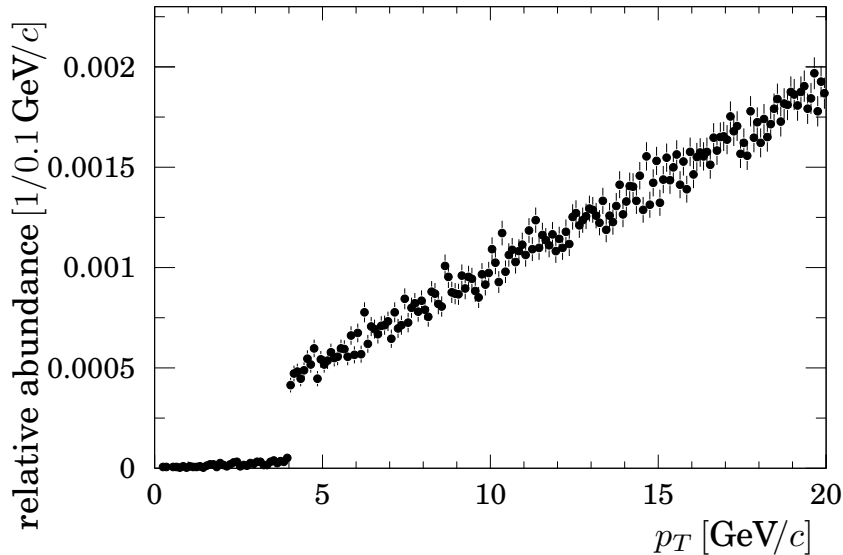


Figure 4.5: Transverse τ momentum distribution for simulated $W \rightarrow \tau\nu_\tau$ decays close to edge of the simulated phase space at $p_T > 4$ GeV/ c . The small amount of entries below the threshold is due to the presence of multiple $W \rightarrow \tau\nu_\tau$ decays in an event. One is sufficient to fulfil the simulation requirement.

The expected τ cross section from Drell-Yan processes is then

$$\sigma(\text{pp} \rightarrow \tau + X)_{\text{Drell-Yan}} = (2.2 \pm 0.4) \text{ nb.}$$

$W \rightarrow \tau\nu_\tau$

The W cross section in the LHCb acceptance is reported in [140, 141]. The combined W^+ and W^- cross section is (1.51 ± 0.05) nb (Ref. [141] has not been available at the time of writing, the values are taken from the corresponding internal documentation in [142]) for $W \rightarrow \ell\nu$ decays with a transverse lepton momentum larger than 20 GeV/ c .

To extrapolate the result into the low transverse momentum region, a simulated $W \rightarrow \tau\nu_\tau$ sample is used, which serves for background studies in [142]. The sample is generated for transverse τ momenta larger than 4 GeV/ c . The fraction of τ with a transverse momentum larger than 20 GeV/ c in the simulated sample is 81.2%, the $W \rightarrow \tau\nu_\tau$ cross section thus increases to (1.86 ± 0.06) nb for transverse lepton momenta larger than 4 GeV/ c .

The fraction of τ with a transverse momentum smaller than 4 GeV/ c is estimated as follows. The lepton p_T distribution is assumed to be proportional to p_T for $p_T < 20$ GeV/ c , judged by eye from Fig. 4.5. Under this assumption, that the proportionality is a smooth continuation until 0, the amount of τ with $p_T < 4$ GeV/ c is a third of those with $4 \text{ GeV}/c < p_T < 8 \text{ GeV}/c$. This rule-of-thumb calculation is considered to be sufficiently correct for the present work because it turns out that the contribution of τ

from W decays is over a factor 10 000 smaller than the total τ cross section. About 2.5% of all simulated W decays have transverse τ momenta between 4 GeV/ c and 8 GeV/ c . The cross section is therefore increased by another 0.84% to

$$\sigma(\text{pp} \rightarrow \tau + X)_W = (1.88 \pm 0.06) \text{ nb.}$$

4.3 τ polarisation

A possible polarisation of τ leptons can introduce a model dependence of the analysis as explained in Sect. 3.8. The polarisation is suggested in [36] to discriminate between New Physics scenarios once $\tau \rightarrow \mu\mu\mu$ is observed. This has been suggested for an analysis at CMS. In that analysis, the analysed τ leptons originate from gauge boson decays.

In the present analysis, τ leptons originating from D_s decays are the major contribution. Since D_s is a pseudo-scalar particle, the distribution of τ is isotropic in the D_s rest frame and the spin orientation is along (against) the τ^+ (τ^-) flight direction in the D_s rest frame. The τ spin projection on any direction in the D_s rest frame is zero when averaging over the decay distribution. Boosting into the lab frame does not alter the average polarisation. The τ leptons selected at LHCb are thus unpolarised, assuming that the detection efficiency does not depend on the τ flight direction in the D_s rest frame.

4.4 Summary

The production of τ leptons at LHCb is dominated by $D_s \rightarrow \tau\nu_\tau$ decays. The second largest contribution are $b \rightarrow \tau X$ decays. The smallest considerable contribution are from $D^+ \rightarrow \tau\nu_\tau$ decays, which are not observed yet. Its branching fraction is calculated considering phase space suppression and helicity enhancement with respect to $D^+ \rightarrow \mu\nu_\mu$.

Only the τ production in D_s and D^+ decays – both prompt and from b decays – and production $b \rightarrow \tau$ decays are considered in the remainder of this analysis.

The production cross sections depend on the collision’s centre-of-mass energy. For collisions at $\sqrt{s} = 7$ TeV production cross section measurements of heavy flavour hadrons is used to determine the τ production. Simulated collisions are used to extrapolate the charm production cross section to $\sqrt{s} = 8$ TeV.

An overview over the τ production channels at LHCb is given in Tabs. 4.2 and 4.3 for $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV respectively. The majority, from heavy flavour modes, contributes unpolarised τ leptons.

The relative fractions of production channels

$$f_\tau^h := \frac{\sigma(\text{pp} \rightarrow h(\tau X)Y)}{\sigma(\text{pp} \rightarrow \tau X)}$$

will be abbreviated as *production fractions*. If h is a charm meson, only prompt charm mesons are referred to. Production of τ leptons from non-prompt charm mesons will be written as $f_\tau^{b \rightarrow X_c}$. If both, prompt and non-prompt charm mesons are meant, the addition “incl.” is added, e. g. $f_\tau^{\text{incl.}D_s}$.

The study has originally been carried out by Jonathan Harrison [38, 143]. With respect to that reference, the contributions from $\psi(2S)$, Υ , Z , and W decays have been revised and the Drell-Yan process has been added to the consideration.

Table 4.2: Summarised τ production cross sections at the LHC for $2 < y < 4.5$ of the parent hadron (heavy flavour mode) or the lepton (electroweak modes), at $\sqrt{s} = 7$ TeV. The cross section uncertainties are partially correlated. The correlation is considered in the sums of cross sections. The modes of the lower block (given in nb) are neglected in the remainder of this work. The sum of cross sections of the channels considered in this analysis (first sum) and the sum of cross sections of all channels (last row) are indeed identical within the given rounding precision. The production fractions and their uncertainties are computed considering the correlation among the cross section uncertainties. The production fractions f_τ deviate from those in Appendix E, because the calculation there is performed within the simulated detector acceptance, while here $2 < y < 4.5$ is referred to. The computation is executed in the same computer algebra program by setting $\alpha_{4\pi} \cdot \varepsilon_{\text{sim}}$ to unity.

τ source	cross section	f_τ
D_s	$(10.7 \pm 1.8) \mu\text{b}$	0.710 ± 0.037
D^+	$(0.66 \pm 0.10) \mu\text{b}$	0.044 ± 0.006
b	$(2.4 \pm 0.4) \mu\text{b}$	0.157 ± 0.024
$b \rightarrow D_s$	$(1.32 \pm 0.36) \mu\text{b}$	0.088 ± 0.016
$b \rightarrow D^+$	$(0.023 \pm 0.004) \mu\text{b}$	0.00155 ± 0.00026
\sum until here	$(15.1 \pm 2.6) \mu\text{b}$	$\equiv 1$
$\psi(2S)$	$(8.9 \pm 2.3) \text{nb}$	
$b \rightarrow \psi(2S)$	$(1.55 \pm 0.24) \text{nb}$	
$\Upsilon(1S)$	$(4.62 \pm 0.61) \text{nb}$	
$\Upsilon(2S)$	$(1.16 \pm 0.16) \text{nb}$	
$\Upsilon(3S)$	$(0.59 \pm 0.10) \text{nb}$	
W	$(1.88 \pm 0.06) \text{nb}$	
γ^*/Z	$(2.2 \pm 0.4) \text{nb}$	
\sum all channels	$(15.1 \pm 2.6) \mu\text{b}$	

4. τ PRODUCTION AT LHCb

Table 4.3: Same as Tab. 4.2, except for the collision energy of $\sqrt{s} = 8$ TeV. The entry n/a indicates that the cross section has not been evaluated at 8 TeV.

τ source	cross section	f_τ
D_s	$(12.2 \pm 2.1) \mu\text{b}$	0.713 ± 0.035
D^+	$(0.75 \pm 0.11) \mu\text{b}$	0.044 ± 0.006
b	$(2.7 \pm 0.4) \mu\text{b}$	0.155 ± 0.023
$b \rightarrow D_s$	$(1.48 \pm 0.37) \mu\text{b}$	0.087 ± 0.016
$b \rightarrow D^+$	$(0.026 \pm 0.004) \mu\text{b}$	0.00153 ± 0.00026
\sum until here	$(17.2 \pm 2.9) \mu\text{b}$	$\equiv 1$
$\psi(2S)$	n/a	
$b \rightarrow \psi(2S)$	n/a	
$\Upsilon(1S)$	$(6.53 \pm 0.49) \text{nb}$	
$\Upsilon(2S)$	$(1.58 \pm 0.14) \text{nb}$	
$\Upsilon(3S)$	$(0.77 \pm 0.09) \text{nb}$	
W	n/a	
γ^*/Z	n/a	
\sum all evaluated channels	$(17.2 \pm 2.9) \mu\text{b}$	

Analysis strategy

Following the introduction to set a limit on a decay's branching fraction, this chapter summarises how the search for $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is performed at LHCb. While the statistics chapter, Chap. 3, suggests that only a $\tau \rightarrow \mu\mu\mu$ signature must be searched for, a few other aspects need to be considered when designing the analysis. These are the normalisation, the calibration of the event classification, the analysis optimisation, and the treatment of the uncertainties of the τ production. These aspects are explained in this chapter.

5.1 The method of relative normalisation

Any of the formulae to determine a branching fraction, Eq. 3.1 or Eq. 3.4, involve the number of τ leptons which eventually decayed $N_{\text{tot}}(\tau)$. Often, it is undesirable to use the production cross section from Sect. 4 to evaluate $N_{\text{tot}}(\tau)$ since it has a large uncertainty and would involve a precise determination of the absolute reconstruction efficiency in Eq. 3.3. A normalisation relative to an observable decay signature can often be done in a more accurate way.

A commonly used alternative is the measurement of a ratio of branching fractions

$$\frac{\mathcal{B}(\tau \rightarrow \mu\mu\mu)}{\mathcal{B}(\tau \rightarrow \pi\pi\pi\nu_\tau)}.$$

Thereby, the normalisation is given by the number of observed decays into a different final state, here $\tau \rightarrow \pi\pi\pi\nu_\tau$. The decay $D^+ \rightarrow \pi\pi\pi\pi^0$ is however indistinguishable from $\tau \rightarrow \pi\pi\pi\nu_\tau$ and occurs three orders of magnitude more often. It is shown in [144] that the normalisation of a τ branching fraction at LHCb cannot be done in this way. In a similar manner, there is a $D^+ \rightarrow hh'h''\pi^0$ decay for every other $\tau \rightarrow hh'h''\nu_\tau$ decay.

Given that τ leptons are produced in heavy flavoured hadron decays the number of produced τ leptons can also be determined from the number of heavy hadrons produced. For the dominant $D_s \rightarrow \tau$ contribution this reads

$$N_{\text{tot}}(\tau \text{ from } D_s \text{ decays}) = \mathcal{B}(D_s \rightarrow \tau\nu_\tau) \times N_{\text{tot}}(D_s).$$

Measuring the D_s production by means of a D_s decay mode with a known branching fraction, one arrives at

$$\begin{aligned} N_{\text{tot}}(\tau \text{ from } D_s \text{ decays}) \\ = \frac{\mathcal{B}(D_s \rightarrow \tau\nu_\tau)}{\mathcal{B}(D_s \rightarrow \phi\pi) \times \mathcal{B}(\phi \rightarrow \mu\mu)} \times N_{\text{tot}}(D_s \rightarrow \phi(\mu\mu)\pi). \end{aligned}$$

The fraction of τ from the dominant contribution $f_\tau^{\text{incl.D}_s} = N(\tau \text{ from D}_s)/N(\tau)$ is given in Sect. 4.4 leading to

$$N_{\text{tot}}(\tau) = f_\tau^{\text{incl.D}_s} \frac{\mathcal{B}(\text{D}_s \rightarrow \tau \nu_\tau)}{\mathcal{B}(\text{D}_s \rightarrow \phi \pi) \times \mathcal{B}(\phi \rightarrow \mu \mu)} \times N_{\text{tot}}(\text{D}_s \rightarrow \phi(\mu \mu)\pi). \quad (5.1)$$

Using ε_τ for the efficiency to observe a $\tau \rightarrow \mu \mu \mu$ decay and ε_{D_s} for the efficiency to observe a $\text{D}_s \rightarrow \phi(\mu \mu)\pi$ decay, the normalisation factor is given as

$$\alpha^{-1} = \frac{\varepsilon_\tau}{\varepsilon_{\text{D}_s}} \left(f_\tau^{\text{incl.D}_s} \right)^{-1} \frac{\mathcal{B}(\text{D}_s \rightarrow \tau \nu_\tau)}{\mathcal{B}(\text{D}_s \rightarrow \phi \pi) \times \mathcal{B}(\phi \rightarrow \mu \mu)} \times N_{\text{obs}}(\text{D}_s \rightarrow \phi(\mu \mu)\pi). \quad (5.2)$$

For this analysis therefore two decays are reconstructed. The decay sought for $\tau \rightarrow \mu \mu \mu$ and the so-called *normalisation channel* $\text{D}_s \rightarrow \phi(\mu \mu)\pi$.

It also plays a crucial role that the signal channel ($\tau \rightarrow \mu \mu \mu$) and the normalisation channel ($\text{D}_s \rightarrow \phi(\mu \mu)\pi$), c.f. Chap. 10. Namely, another advantage of the relative normalisation is that instead of an efficiency, a ratio of efficiencies $\varepsilon_\tau/\varepsilon_{\text{D}_s}$ is used. A single efficiency would be determined with large systematic uncertainty, these are correlated uncertainties for the numerator and denominator of an efficiency ratio and thus cancel.

5.2 Calibratable event classification

As motivated in Sect. 3.3, the sensitivity of the search for $\tau \rightarrow \mu \mu \mu$ significantly depends on how well $\tau \rightarrow \mu \mu \mu$ decays can be discriminated from background events. These assign a number \mathcal{M} to each observed event. \mathcal{M} is designed to be large for $\tau \rightarrow \mu \mu \mu$ decays and small for background events. Using only events with $\mathcal{M} > x$ instead of using all observed events, reduces the efficiency

$$\varepsilon(\mathcal{M} > x) < \varepsilon(\text{no requirement on } \mathcal{M})$$

but is more effective in removing background events

$$\xi(\mathcal{M} > x) < \varepsilon(\mathcal{M} > x).$$

The efficiency must be known accurately because it enters linearly in the normalisation factor¹. Not to rely on simulated events, the selection efficiencies for multivariate classifiers must be determined using data. If all information from the detector (i. e. information from the particle identification system) were used in a multivariate classifier, the efficiency determination would require a real $X \rightarrow \mu \mu \mu$ decay. None of the decays in Tab. 8.1 can be used because $X \rightarrow \mu \mu \mu$ must not include additional neutrinos in the final state. Two separate classifiers are used instead. These are:

¹NB: The classification is not part of the selection, thus not applied to the normalisation channel and the above argument for efficiency ratios does not hold. Moreover, a priori it is not clear whether an application of the classifier to $\text{D}_s \rightarrow \phi(\mu \mu)\pi$ is sensible. As explained in the remainder of this section, the layout of the classification is constrained by the demand to be meaningfully applicable to $\text{D}_s \rightarrow \phi(\mu \mu)\pi$ decays

Three body decay classification ($\mathcal{M}_{3\text{body}}$): This classifier is designed to identify displaced three body decays. No information from the particle identification system appears. Both, $\tau \rightarrow \mu\mu\mu$ and the normalisation channel $D_s \rightarrow \phi\pi$, are such three body decays. Therefore, at first order the efficiency $\varepsilon_\tau(\mathcal{M}_{3\text{body}} > x)$ can be observed in the $\mathcal{M}_{3\text{body}}$ distribution on data:

$$\varepsilon_\tau(\mathcal{M}_{3\text{body}} > x) \approx \varepsilon_{D_s}(\mathcal{M}_{3\text{body}} > x)$$

Chapter 9 is dedicated to the development of $\mathcal{M}_{3\text{body}}$. A more accurate calibration is given in Chap. 10.

Particle identification (\mathcal{M}_{PID}): One classifier is designed to separate muon tracks from other tracks. Corresponding to the application of a selection criterion of $\mathcal{M}_{\text{PID}} > x$ for all three tracks, the classification of $\tau \rightarrow \mu\mu\mu$ candidates is done by means of the smallest of the three response values:

$$\mathcal{M}_{\text{PID}}(\tau \rightarrow \mu_1\mu_2\mu_3) := \min_i(\mathcal{M}_{\text{PID}}(\mu_i)).$$

In data, the \mathcal{M}_{PID} distribution can be observed for single muon tracks. The calibration, as developed in [145], is described in Sect. 10.3.

\mathcal{M}_{PID} has been developed by the charged particle identification working group in LHCb, called ProbNNmu in [85].

A event classification is done in both of these classifiers, i. e. the response range is split into bins as motivated in Sect. 3.4.

5.3 Expected number of background events

To establish whether a signal is seen or not, not only the number of candidates must be known but also the number of expected candidates, in particular expected background events given that only these are expected, as explained in Chap. 3. A common way of determining the number of expected background events in a data sample is an analysis of the invariant mass spectrum beyond the range in which a signal is sought for.

An example of an invariant mass distribution is shown in Fig. 5.1. Background candidates in this example are exponentially distributed, the signal is peaked. Once a model for the distribution of background events as a function of the invariant mass is known, the background distribution can be fitted in a range where no signal events are expected and the distribution is extrapolated into the region where the signal is sought for, as indicated by the black line. This extrapolation gives an estimate for the number of background events in the signal region.

5.4 Unbiased analysis optimisation

The multivariate classifiers, as well as their usage must be optimised in an unbiased way. This means that any decision or training process must be statistically independent from the actual observation of the experiment, the estimate of the number of expected background events, and the signal efficiency determination.

The following counter example shall illustrate how an optimisation can bias the analysis' outcome.

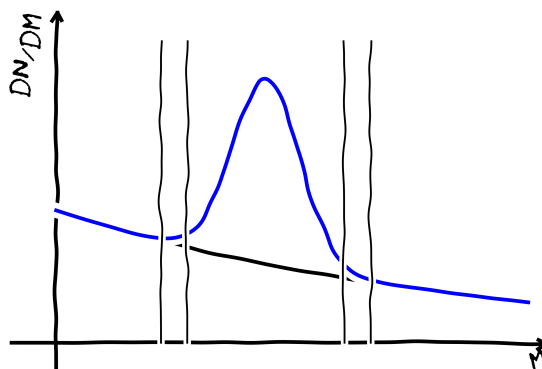


Figure 5.1: Illustration of the invariant mass spectrum. The hypothetical distribution for the presence of a signal is shown in blue, the distribution for the background hypothesis in black. The innermost region is the signal box. The distribution in the outer regions is used to interpolate the background contribution in the signal box. The narrow regions are used for the analysis optimisation.

Assume, there was a choice between using one of two different classifiers, \mathcal{M}_A and \mathcal{M}_B , as $\mathcal{M}_{\text{body}}$. Assume as well that these have the same true efficiency for the same true background efficiency. Mathematically speaking²

$$\xi(\mathcal{M}_A > x) = \xi(\mathcal{M}_B > x).$$

These true background efficiencies are unknown and can only be estimated to finite precision using the event count in the sidebands. If this is done, then this decision must be the consequence of a statistical fluctuation. Obviously, the classifier for which the background fluctuates downward³ would be chosen. As a consequence, the estimated number of expected background events is biased to small values since the fluctuation must be in the downward direction with respect to an unbiased classifier choice – the sensitivity is then estimated to be better than it actually is. Furthermore, the number of observed events would be larger than its expectation, even in the absence of a signal. The optimisation process thus can cause the observation of an excess of events.

Other, more complicated, examples can be constructed in which the sensitivity is biased, but where no difference between the expected and observed numbers of events in the signal region is created.

Blind rare decay searches

A commonly used method of avoiding biases in an analysis is performing a blind analysis as reviewed in e. g. [146].

²Without loss of generality, it is assumed that $\varepsilon(\mathcal{M}_A > x) = \varepsilon(\mathcal{M}_B > x) = 1 - x$, which can be achieved with a probability integral transform.

³or more downward than for the other or less upward than for the other

In rare decay searches, blinding is done by defining the entire analysis prior to evaluating the part of the data in which a signal is sought for, this part is also referred to as *signal box*. In the case of $\tau \rightarrow \mu\mu\mu$, the candidates with an invariant mass between $m_\tau - 20 \text{ MeV}/c^2$ and $m_\tau + 20 \text{ MeV}/c^2$ were removed from the datasets for the development of the analysis strategy and the classifier optimisation⁴. Once the analysis is defined, the signal box is analysed. This means that the number of candidates is evaluated and can be compared to the expectation, i. e. Q_{obs} and CL_s are computed.

The problem arising from the classifier choice in the above example is not solved by performing a blind analysis. As a remedy, a dedicated *optimisation data set* is used as explained in the following section.

Dedicated optimisation data

Experiences from early versions of the search for $\tau \rightarrow \mu\mu\mu$ at LHCb, [147], showed that decisions about the analysis strategy involved knowledge about the background in the data. E. g. the bin boundaries of $\mathcal{M}_{\text{body}}$ and \mathcal{M}_{PID} have to be optimised. Claims were that $\mathcal{O}(1)$ background event in the sensitive analysis bins should be aimed for. Simulated background events corresponded to less than 10 pb^{-1} while on data 1 fb^{-1} was to be studied. This implies that the bin boundaries had to be optimised in response ranges of less than one simulated background event.

The suggested remedy was using background events from data, which were neither part of the signal box, nor used to determine the estimated number of expected background events to the CLs method. The signal box ranges only from $m_\tau - 20 \text{ MeV}/c^2$ to $m_\tau + 20 \text{ MeV}/c^2$, which leaves the range of candidates with a mass difference to the τ mass between $20 \text{ MeV}/c^2$ and $30 \text{ MeV}/c^2$ for other studies. This range is called *inner sideband*.

An example invariant mass distribution is shown in Fig. 5.1. The invariant mass distribution for the presence of a signal is shown in blue; the invariant mass distribution for the absence of a signal is shown in black. The outer region, which is used to estimate the number of background events in the signal box is identical for both hypotheses. The narrow regions are not suited as part of the signal box because the fraction of signal decays would be insignificant while the amount of background is not insignificant. The narrow regions cannot be used for the background description either because in the presence of a signal, the background prediction would be biased to higher values. These events are used for optimisations of the analysis strategy assuming they are all background events.

This strategy leaves one corner case: What if the number of real $\tau \rightarrow \mu\mu\mu$ decays in the inner sidebands is high enough to deteriorate the optimisation? In this case the analysis would be sub-optimal, i. e. the sensitivity of the analysis would lower than an optimal sensitivity could be. If, however, $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ was large enough to contribute a significant number of true $\tau \rightarrow \mu\mu\mu$ decays to the inner sidebands, then $\tau \rightarrow \mu\mu\mu$ will even be discovered with a sub-optimal analysis.

5.5 Signal simulation

At several places, the analysis requires simulated events with $\tau \rightarrow \mu\mu\mu$ decays. Events are simulated for the five most abundant production mechanisms, as listed

⁴Here, m_τ is the mass of the τ lepton from [80]

5. ANALYSIS STRATEGY

in Tabs. 4.2 and 4.3. The numbers of simulated events for each production mode are hereby chosen in the proportions as in the last column of these tables.

For technical reasons, the simulation only generates events where the final τ decay products are within an approximate detector acceptance⁵. Furthermore only τ leptons with a minimum momentum, $p > 2.5$ GeV/ c , and minimum transverse momentum, $p_T > 250$ MeV/ c , are simulated. The probability for a $\tau \rightarrow \mu\mu\mu$ decay to fulfil these criteria has been determined by [38, 11], it depends on the production channel and the collision energy. For technical reasons, these probabilities are calculated with respect to τ produced anywhere in the full phase space and assumes the η distribution of the τ as simulated by the standard LHCb simulation. These probabilities are called *simulation efficiencies*, ε_{SIM} here. The cross sections from Tabs. 4.2 and 4.3 are thus extrapolated to the full phase space before the efficiency factors are applied, see Tab. E.3. Once the cross sections for the simulated phase space are known, the final fractions in which the simulated $\tau \rightarrow \mu\mu\mu$ decays must be composed from the production channels can be computed.

The extrapolation to the full phase space, as well as the determination of the simulation efficiency involve strong assumptions on the heavy flavour production far outside of the LHCb acceptance, e. g. the extrapolation of the meson production to the full phase space requires the knowledge of the production at very large rapidities:

$$\varepsilon(\text{meson in LHCb acceptance}) = \underbrace{\varepsilon(\text{meson in LHCb acceptance} \mid \text{meson in } y < 6)}_{\text{well simulated}} \cdot \underbrace{\varepsilon(\text{meson in } y < 6)}_{\text{not trivial}},$$

and correspondingly for the simulation efficiency for a $\tau \rightarrow \mu\mu\mu$ decay where the τ originates in a meson decay:

$$\begin{aligned} \varepsilon(\tau \rightarrow \mu\mu\mu \text{ decay in LHCb acceptance}) &= \\ \varepsilon(\tau \rightarrow \mu\mu\mu \text{ decay in LHCb acceptance} \mid \text{parent meson in } y < 6) & \\ \cdot \varepsilon(\text{parent meson in } y < 6). & \end{aligned}$$

Consequently, extrapolating to the full phase space and determining the acceptance in simulated events is solely sensitive to the ratio of the former factors, while the non-trivial second factors cancel:

$$\begin{aligned} \sigma(\tau \rightarrow \mu\mu\mu \text{ in acceptance}) & \\ \propto \frac{\varepsilon(\tau \rightarrow \mu\mu\mu \text{ decay in LHCb acceptance} \mid \text{parent meson in } y < 6)}{\varepsilon(\text{meson in LHCb acceptance} \mid \text{meson in } y < 6)} & \\ \times \underbrace{N(D_s)}_{\text{yield in cross section measurement}}. & \end{aligned}$$

Physically relevant is, that for each collision energy a nominal data sample is simulated with the expected τ production mixture. Furthermore, the covariance matrix of the mixing fractions is diagonalised to determine five uncorrelated difference vectors $\delta\vec{f} = (\delta f_{\tau}^{D_s}, \delta f_{\tau}^{D^+}, \delta f_{\tau}^b, \delta f_{\tau}^{b \rightarrow D_s}, \delta f_{\tau}^{b \rightarrow D^+})$. For each of these difference vectors, a τ mixture with mixing fractions $\vec{f} \pm \delta\vec{f}$ corresponding to 1σ deviations are determined, too. These are used to determine the influence of the systematic uncertainty of the analysis due to the knowledge of the τ production.

⁵A generous safety factor ensures that a considerable part of the simulated $\tau \rightarrow \mu\mu\mu$ decays is not fully in the detector acceptance, but no decay which is in the detector acceptance is rejected in the simulation.

5.6 Strategy

Events recorded by LHCb are analysed if they pass a selection, which is described in Chap. 6. The selection has been developed by Matt Jaffe [148] as a CERN Summer Student project in 2011 using $\mathcal{O}(50 \text{ nb}^{-1})$ of data as background and simulated $\tau \rightarrow \mu\mu\mu$ decays as signal description. The simulated events used in that study are not used in the present analysis. The data is considered too small to introduce a bias. The selection mainly serves to reduce the data recorded by LHCb to a manageable amount while being efficient for real $\tau \rightarrow \mu\mu\mu$ decays, i. e. the major purpose of the selection is compliance with the computing resources.

Candidates passing the selection are classified using multivariate classifiers as outlined in Sect. 5.2. $\mathcal{M}_{3\text{body}}$ is trained using simulated $\tau \rightarrow \mu\mu\mu$ decays and simulated background events. The development is explained in detail in Chap. 9.

The expected number of background events in the signal box is estimated fitting the invariant mass spectrum of candidates with an invariant mass $m > m_\tau + 30 \text{ MeV}/c^2$ or $m < m_\tau - 30 \text{ MeV}/c^2$ and interpolating the fit result to the signal box. The fit is described in Sect. 12.1.

The CLs method, described in Chap. 3, is used to convert the agreement of the number of observed events with the expected background into a upper limit on the branching fraction $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$.

To avoid biases which are introduced by the design process of the analysis, three blinding steps are done:

1. Blinding of the calibration data: The $D_s \rightarrow \phi(\mu\mu)\pi$ decays in data are only analysed once the design of $\mathcal{M}_{3\text{body}}$ is finalised.
2. Blinding of the side-bands of $\tau \rightarrow \mu\mu\mu$: To avoid a bias of the estimated background level in the signal region, the side-bands of $\tau \rightarrow \mu\mu\mu$ are only analysed once design of the $\mathcal{M}_{3\text{body}}$ classifier and its binning are finalised.
3. Blinding of the signal region: The $\tau \rightarrow \mu\mu\mu$ signal region is only analysed once the full analysis is finalised and an expected limit has been computed.

The analysis is not strictly blind, in that sense that the data has already been analysed in [147, 4, 5].

Event selection

6.1 Signal properties

For the development of the selection criteria, the qualitative physics properties of a $\tau \rightarrow \mu\mu\mu$ are relevant. These are briefly described here. Variables quantifying these properties are described in Sect. 6.2 if they are used in the signal selection and in Sect. 9.4 if they are considered as input for $\mathcal{M}_{3\text{body}}$, while avoiding double listing.

The decay topology for a $\tau \rightarrow \mu\mu\mu$ decay from the most relevant production mechanism (τ from D_s) is shown in Fig. 6.1.

Displacement: Due to the lifetime of the τ lepton and due to the kinematics of its production, τ leptons fly a short distance from their production point before they decay. The production point itself is the decay vertex of a heavy flavour hadron, which also flies a short distance from its production vertex (the primary interaction point or again a heavy flavour hadron decay). Therefore $\tau \rightarrow \mu\mu\mu$ vertex is measurably displaced from the primary interaction point. The muon tracks, extrapolated from the first measurement towards the interaction region, consequently pass through the τ decay vertex and thus have a measurable impact parameter with respect to the primary interaction point.

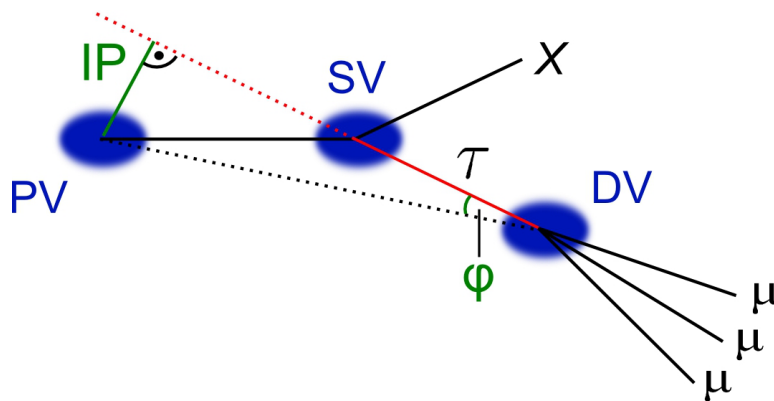


Figure 6.1: Decay topology of a $\tau \rightarrow \mu\mu\mu$, where the τ is produced in a leptonic D_s decay. Figure taken from [149].

Isolation: As a consequence of the displacement, there are only three charged particles coming from the point where a τ decays. For the main production modes, the inclusive $D_s \rightarrow \tau \nu_\tau$ as well as the inclusive $D^+ \rightarrow \tau \nu_\tau$, there is also no charged particle produced in the production point, except for the τ itself. Consequently it is not expected that a charged particle is reconstructed in the vicinity of the $\tau \rightarrow \mu\mu\mu$ decay vertex. This property is called isolation. An isolation of $\tau \rightarrow \mu\mu\mu$ candidates is not required in the event selection, it is part of the event classification in Chap. 9.

Pointing: The difference between the D_s mass and the τ mass is small, $1969 \text{ MeV}/c^2$ and $1778 \text{ MeV}/c^2$ [80]. In the D_s rest frame, after a $D_s \rightarrow \tau \nu_\tau$ decay, the τ will therefore be almost at rest. Consequently, a τ lepton being produced in the decay of a boosted D_s meson, will follow the D_s flight direction in the lab system. The sum of the three muon momenta of a $\tau \rightarrow \mu\mu\mu$ decay will therefore be parallel to the straight connection of the primary interaction with the $\tau \rightarrow \mu\mu\mu$ decay vertex if the τ is produced in the decay of a prompt D_s .

Particle identification: Muons interact differently with matter than other particles. All three tracks of a $\tau \rightarrow \mu\mu\mu$ decay are therefore identified as muons by the particle identification system.

Invariant mass: Due to energy conservation, the invariant mass of the three muons in a $\tau \rightarrow \mu\mu\mu$ decay is equal to the τ mass. This property can not be used for either the event selection or the event classification because the invariant mass spectrum is used to determine the number of background events in the signal box, as mentioned in Sect. 5.3. The invariant mass of the three muons is computed with a vertex fit which assumes a common origin vertex of the tracks.

Reconstruction quality: The tracks from particles originating in a common vertex should fit to a common reconstructed vertex. The tracks of particles from different vertices do not necessarily lead to a good fit quality when fitting a common vertex. Similarly, the track fit of the track of a real particle should have a good quality, while fake tracks can have bad fit qualities.

6.2 Selection variables

The above properties are quantified by means of the following variables. Unless otherwise specified, the requirement is applied to the $\tau \rightarrow \mu\mu\mu$ reconstruction as well as the $D_s \rightarrow \phi(\mu\mu)\pi$ reconstruction. Three distributions are shown for each selection variable

- background candidates from data sidebands which pass the selection (red)
- simulated signal candidates (blue)
- simulated signal candidates which pass the selection (green).

Displacement variables

Track impact parameter χ^2 : The compatibility of a track's origin in the pp interaction point is measured by the difference of the primary interaction vertex fit

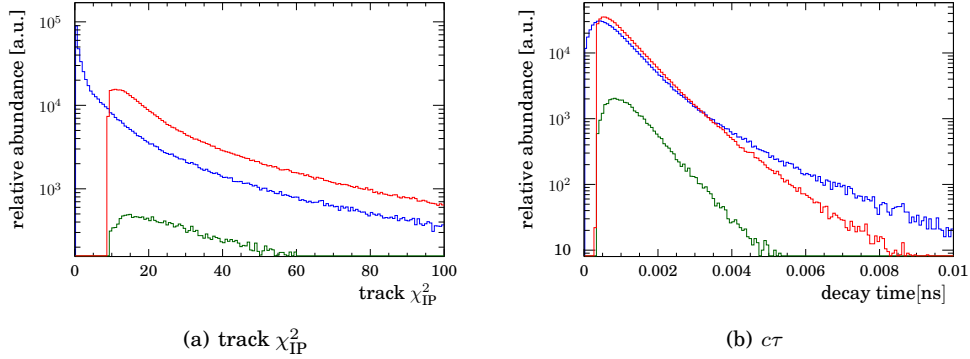


Figure 6.2: Distribution of displacement selection variables. Red shows the background distribution observed on data sidebands, blue shows all reconstructed simulated $\tau \rightarrow \mu\mu\mu$ decays, green shows those simulated decays which fulfil all selection requirements. The latter two are correctly scaled with respect to each other.

The background processes passing the selection are similar to the signal. Prompt background candidates, which are rejected by the selection, are not visible.

with and without using the track in the vertex fit, abbreviated χ_{IP}^2 . The event selection requires

$$\chi_{\text{IP}}^2(\text{tracks}) > 9.$$

Reconstructed decay time Under the assumption that a candidate originates from the primary interaction vertex, the decay time in the particle’s rest frame is computed from the reconstructed momentum and the distance between the primary vertex and the decay vertex.

$$\tau > 0.3336 \text{ ps}$$

Both variables are shown in Fig. 6.2.

Pointing

τ impact parameter χ^2 : The compatibility of the $\tau \rightarrow \mu\mu\mu$ candidate to origin in the primary interaction vertex can be measured as well by including the τ candidate in the fit of the primary vertex. The change of the vertex fit’s χ^2 under inclusion of the τ candidate is again abbreviated χ_{IP}^2 . If the reconstructed τ trajectory is incompatible with originating from the primary interaction point, then the primary vertex fit with the τ flight direction as fit input has a large χ^2 . Candidates with “good pointing” therefore have small χ_{IP}^2 and are selected by

$$\chi_{\text{IP}}^2(\tau) < 225.$$

Direction angle: The angle between the reconstructed τ momentum and a straight line from the reconstructed primary vertex to the reconstructed decay vertex is

6. EVENT SELECTION

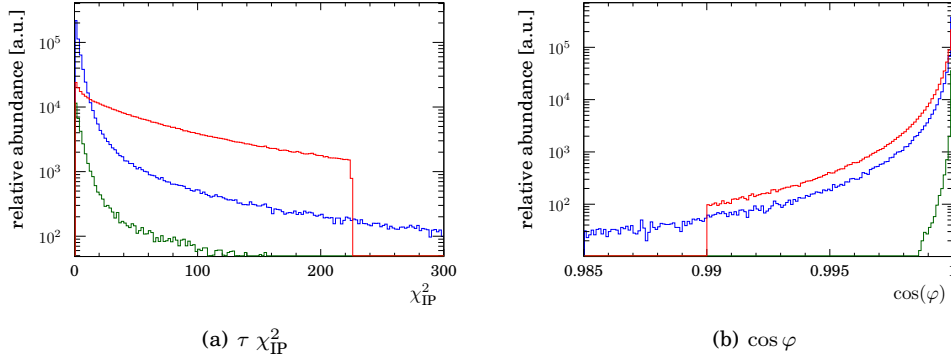


Figure 6.3: Distribution of pointing selection variables. Red shows the background distribution observed on data sidebands, blue shows all reconstructed simulated $\tau \rightarrow \mu\mu\mu$ decays, green shows those simulated decays which fulfil all selection requirements. The latter two are correctly scaled with respect to each other.

In both variables, the signal is much better “pointing”, i. e. the signal is concentrated at small $\tau \chi_{\text{IP}}^2$ and large $\cos \varphi$, while the background has a wide tail where the selection requirement is visible, in contrast to the green signal distribution.

another quantification of the pointing property. Candidates are required to have

$$\cos \varphi > 0.99.$$

Both variables are shown in Fig. 6.3.

Particle Identification

The standard muon pre-selection is applied for all muon tracks as described in Sect. 2.4.

Reconstruction quality

Track fit χ^2/ndf : A track fit is considered good when its reduced fit χ^2 is small,

$$\chi^2/\text{ndf} < 3.$$

Fake track probability: To suppress fake tracks further, only tracks with a low fake track probability are considered,

$$\mathcal{M}_{\text{fake track}} < 0.3.$$

Double reconstruction veto: Whenever a muon is reconstructed as two tracks, the invariant mass of the wrong muon pair is close to twice the muon rest mass. These fake muon pairs are rejected by the requirement

$$m_{\mu^\pm\mu^\pm} > 250 \text{ MeV}/c^2.$$

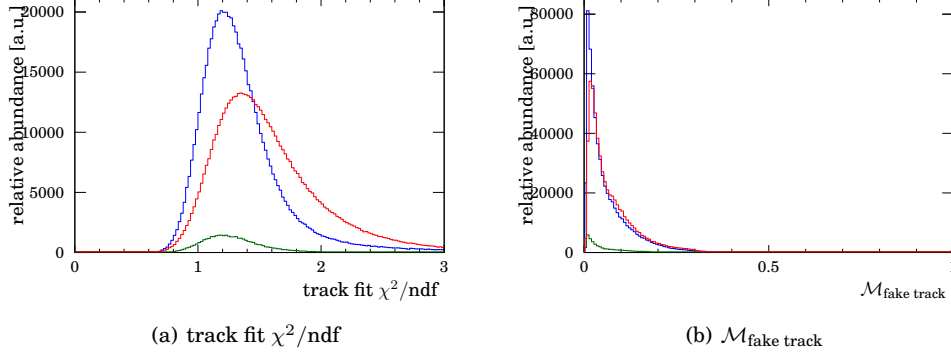


Figure 6.4: Distribution of reconstruction quality selection variables. Red shows the background distribution observed on data sidebands, blue shows all reconstructed simulated $\tau \rightarrow \mu\mu\mu$ decays, green shows those simulated decays which fulfil all selection requirements. The latter two are correctly scaled with respect to each other.

The average reconstruction quality of simulated signal tracks is better than of background candidates in data. This is partially due to the presence of fake tracks in data but also due to resolution effects.

3 μ vertex fit χ^2 : Whether all muon tracks originate from a common vertex is measured by the fit χ^2 of the 3 muon vertex fit:

$$\chi_{3\mu}^2 < 15.$$

The variables are shown in Figs. 6.4, 6.5 (a), and 6.6 (a).

Other variables

In addition to the above selection requirements, the following requirements are imposed to address abundant background processes.

Muon transverse momentum: To reduce the number of tracks from soft QCD production, a minimum transverse momentum requirement is applied.

$$p_T(\text{tracks}) > 300 \text{ MeV}/c$$

The muon p_T distribution is shown in Fig. 6.6 (b).

Veto against $\phi \rightarrow \mu\mu$ decays / selection of $\phi \rightarrow \mu\mu$: Background candidates from $\phi \rightarrow \mu\mu$ decays are rejected in the $\tau \rightarrow \mu\mu\mu$ selection by requiring that both combinations of oppositely charged muons are incompatible with a $\phi \rightarrow \mu\mu$ decay

$$|m_{\mu^+\mu^-} - 1019.46 \text{ MeV}/c^2| > 20 \text{ MeV}/c^2.$$

The opposite charge muon invariant mass distribution is shown in Fig. 6.5 (b).

6. EVENT SELECTION

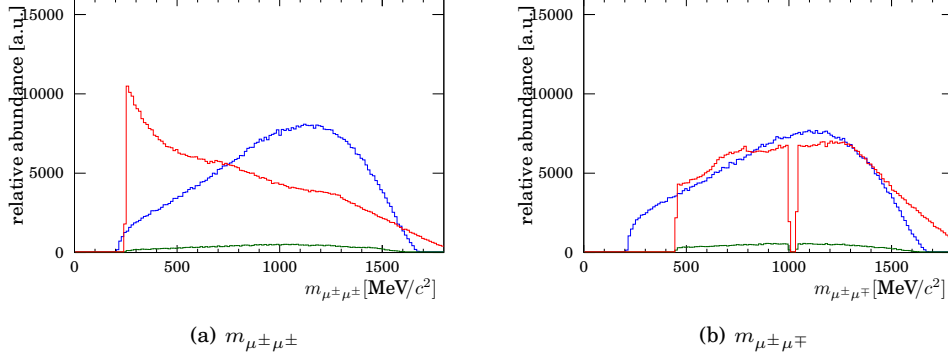


Figure 6.5: Dimuon invariant mass distribution. Red shows the background distribution observed on data sidebands, blue shows all reconstructed simulated $\tau \rightarrow \mu\mu\mu$ decays, green shows those simulated decays which fulfil all selection requirements. The latter two are correctly scaled with respect to each other.

In the same sign spectrum, it can be seen that the efficiency loss due to the double reconstruction veto is tiny for signal candidates.

In the opposite sign spectrum, the ϕ and η vetos are visible. In the background spectrum between 500 MeV/c^2 and 1000 MeV/c^2 , a contribution from ρ and ω decays is visible.

The background spectra extend beyond the range covered by the signal, since the constrain on the invariant three muon mass is different.

This requirement is not applied for $D_s \rightarrow \phi(\mu\mu)\pi$ decays. Here, the purity of the selection is enhanced by imposing the opposite requirement

$$|m_{\mu^+\mu^-} - 1019.46 \text{ MeV}/c^2| < 20 \text{ MeV}/c^2.$$

Veto against $\eta \rightarrow \mu\mu\gamma$ decays: Background candidates from $\eta \rightarrow \mu\mu\gamma$ decays are rejected in the $\tau \rightarrow \mu\mu\mu$ selection by requiring that both combinations of oppositely charged muons are incompatible with a $\eta \rightarrow \mu\mu\gamma$ decay. A $\mu\mu\gamma$ candidate cannot be vetoed since the photon reconstruction efficiency is too low to allow for an effective veto.

$$m_{\mu^+\mu^-} > 450 \text{ MeV}/c^2$$

The selection has been developed by Matt Jaffe [148] except for the η veto (Marcin Chruszcz [11]), the double reconstruction veto (the author, following the suggestion by Marco Meißner [149]), the fake track probability (the author in collaboration with Johann Brehmer and Johannes Albrecht), and the ϕ veto (the author).

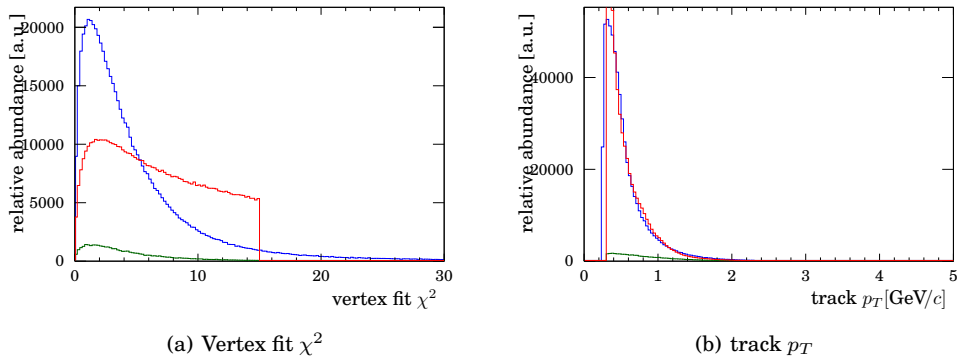


Figure 6.6: Distribution of the vertex fit χ^2 and the muon transverse momentum. Red shows the background distribution observed on data sidebands, blue shows all reconstructed simulated $\tau \rightarrow \mu\mu\mu$ decays, green shows those simulated decays which fulfil all selection requirements. The latter two are correctly scaled with respect to each other.

A large contribution of bad vertex fit qualities to the background spectrum is visible, due to combinations of tracks from different origins.

After the selection, the track transverse momentum does not provide additional background discrimination from the background.

Online event selection

The choice of trigger lines, from which events are used in this analysis, is based on the following evaluation of the trigger efficiency and background yield.

7.1 Ranking of HLT2 trigger lines

The Punzi figure of merit is computed for each HLT2 trigger line l individually by evaluating its efficiency (ε) in simulated $\tau \rightarrow \mu\mu\mu$ events and the expected number of background events (b) in the inner sidebands¹. For both, events are required to fulfil the event selection described in the previous chapter (sel) and to be accepted by at least one of all trigger lines as TOS, TIS, or TOB event (any trig).

$$FOM_{\text{Punzi}}(l) := \frac{\varepsilon(l|\text{sel} \wedge \text{any trig})}{1 + \sqrt{2 \cdot b(\text{acc.by } l \wedge \text{sel} \wedge \text{any trig})}}$$

The best trigger line is than that with the largest FOM_{Punzi} and other trigger lines are sorted by decreasing FOM_{Punzi} .

The procedure is done separately considering only TOS candidates ($FOM_{\text{Punzi}}^{\text{TOS}}$) and for any possible positive trigger decision ($FOM_{\text{Punzi}}^{\text{TOS|TIS|TOB}}$). In all cases, the TOS figure of merit is found better:

$$FOM_{\text{Punzi}}^{\text{TOS}}(l) \stackrel{\text{exp. found}}{>} FOM_{\text{Punzi}}^{\text{TOS|TIS|TOB}}(l) \quad \forall l$$

In the case of HLT2, the dedicated $\tau \rightarrow \mu\mu\mu$ HLT2 line is the best HLT2 line in for the 2012 data taking period. For the 2011 data taking period, the $\tau \rightarrow \mu\mu\mu$ HLT2 line ranks second, slightly worse than the $D \rightarrow h\mu\mu$ HLT2 line.

The outcome differs between the two data taking periods due to a re-optimisation of the software trigger between 2011 and 2012. The $D \rightarrow h\mu\mu$ trigger was not used in data taking in 2012.

7.2 Accumulation of HLT2 trigger lines

From the ranking of HLT2 trigger lines, it is decided whether more than one HLT2 line can be used in the analysis. Let $L_n := \{l_i | 1 \leq i \leq n\}$ be the set of the n best

¹The invariant mass range of the signal box is twice as large as the inner sidebands. Therefore a factor 2 is introduced.

Table 7.1: Requirements of the HLT1 single muon trigger.

variable	requirement in 2011	requirement in 2012
track impact parameter	> 0.1 mm	> 0.1 mm
track χ_{IP}^2	> 16	> 16
p_T	> 1 GeV/c	> 1 GeV/c
p	> 8 GeV/c	> 8/3 GeV/c (reduced after the first 630 pb ⁻¹)
track fit χ^2/ndf	< 2	< 3/2.5 (reduced after the first 70 pb ⁻¹)

ranked HLT2 lines. The decision how many HLT2 lines to consider is then based on the figure of merit

$$FOM_{\text{Punzi}}^{\text{TOS}}(n) := \frac{\varepsilon(\text{TOS for any of the } l_i \mid \text{sel} \wedge \text{any trig})}{1 + \sqrt{2 \cdot b(\text{TOS for any of the } l_i) \wedge \text{sel} \wedge \text{any trig}}}.$$

The condition of the conditional probability ε still does only require the event selection from the previous chapter and is not further restricted to the HLT2 selection from the previous section.

In the case of the 2011 data, the figure of merit is maximised when using the two best HLT2 lines; candidates selected by the $D \rightarrow h\mu\mu$ or the $\tau \rightarrow \mu\mu\mu$ HLT2 line. This is not an exclusive or, the two trigger lines have a large overlap.

In the case of the 2012 data, the figure of merit is maximised when using one HLT2 lines; candidates selected by the $\tau \rightarrow \mu\mu\mu$ HLT2 line.

7.3 Extension to HLT1 and L0

The procedure is repeated for the lower trigger stages. To ensure compatible optimisation results in all stages, the computation of FOM_{Punzi} for HLT1 counts only $\tau \rightarrow \mu\mu\mu$ candidates which fulfil the selection requirements and the HLT2 TOS requirement, as established in the previous section. Consequently, the L0 optimisation is based on candidates being selected offline, in HLT2, and in HLT1.

7.4 Description of the chosen trigger lines

The L0 single muon trigger is recapitulated from Chap. 2 for completeness. A muon with $p_T > 1480$ MeV/c is required in the 2011 data taking period and in the first 70 pb⁻¹ of the 2012 data taking period. In the remainder of the 2012 data taking period the threshold has been raised to 1760 MeV/c.

The HLT1 single muon trigger requires a VELO track which is compatible with hits in the muon stations. This VELO track is then used as seed to online forward tracking algorithm. The candidate is accepted by the single muon trigger if it fulfils the additional requirements in Tab. 7.1 and passes the muon pre-selection from Sect. 2.5. This trigger line is called `TrackMuon` in [86, 150].

The HLT2 $\tau \rightarrow \mu\mu\mu$ trigger requires a three track candidate, each of them identified as a muon, by the offline muon pre-selection, each of them not pointing to a primary interaction vertex. The exact numeric requirements are listed in Tab. 7.2.

Table 7.2: Requirements of the HLT2 $\tau \rightarrow \mu\mu\mu$ trigger.

variable	requirement in 2011	requirement in 2012
track χ_{IP}^2	> 16	$> 4/9$ (increased after the first 70 pb $^{-1}$)
one track with p_T	$> 1 \text{ GeV}/c$	n/a
track fit χ^2/ndf	< 6	$< 6/4$ (reduced after the first 70 pb $^{-1}$)
reconstructed decay time $c\tau$	$> 45 \mu\text{m}$	$> 45/75 \mu\text{m}$ (increased after the first 70 pb $^{-1}$)
vertex fit χ^2	< 25	< 25

Table 7.3: Requirements of the HLT2 $D \rightarrow h\mu\mu$ trigger.

variable	requirement in 2011
track requirements	
track fit χ^2/ndf	< 5
muon requirements	
χ_{IP}^2	> 2
p	$> 5 \text{ GeV}/c$
p_T	$> 500 \text{ MeV}/c$
hadron requirements	
only implicit constrains due to the pattern recognition	
muon pair requirements	
flight distance χ^2	> 20
track distance of closest approach	$< 0.1 \text{ mm}$
D candidate requirements	
$\sum_{\text{tracks}} p_T$	$> 1.5 \text{ GeV}/c$
track distance of closest approach	$< 0.25 \text{ mm}$
vertex fit χ^2/ndf	< 20
pointing angle ($\cos(\varphi)$)	> 0.9998
$\sum_{\text{tracks}} \sqrt{\chi_{\text{IP}}^2}$	> 15
χ_{IP}^2	< 36

The additional $D \rightarrow h\mu\mu$ trigger requires a di-muon candidate, where the two muons have a small distance of closest approach and each of the muons passes a minimum χ_{IP}^2 and a minimum momentum requirement. The exact numeric requirements are listed in Tab. 7.3.

This choice of trigger lines has first been performed by the author as documented in [8].

7.5 $D_s \rightarrow \phi(\mu\mu)\pi$ trigger

To achieve a small uncertainty in the relative normalisation, the same trigger requirement is used for $D_s \rightarrow \phi(\mu\mu)\pi$ as for $\tau \rightarrow \mu\mu\mu$. For the L0 and HLT1, this is accomplished by requiring $D_s \rightarrow \phi(\mu\mu)\pi$ candidates to be selected by the single muon L0 trigger and single muon HLT1 line. The HLT2 $\tau \rightarrow \mu\mu\mu$ line is not expected to behave in the same way for $\tau \rightarrow \mu\mu\mu$ and for $D_s \rightarrow \phi(\mu\mu)\pi$ due to the absence of a third muon in the final state. In HLT2 the $D_s \rightarrow \phi(\mu\mu)\pi$ candidates are selected by a displaced dimuon trigger line. Its selection requirements are listed in Tab. 7.4. The usage of this HLT2 line has been suggested by Jonathan Harrison [38].

Table 7.4: Selection requirements of the displaced di-muon trigger, used for the $D_s \rightarrow \phi(\mu\mu)\pi$ selection event selection.

variable	requirement
muon requirements	
track fit χ^2/ndf	< 4
χ_{IP}^2	> 25
p_T	> 300 MeV/c
muon pair requirements	
vertex fit χ^2/ndf	< 8
decay length significance	> 7
p_T	> 600 MeV/c

Background processes to $\tau \rightarrow \mu\mu\mu$

8.1 Processes with three muons from a common vertex

The most relevant background processes for this analysis are decays of heavy flavour hadrons with three muons and no other charged particles in the final state because these will have very similar properties as the $\tau \rightarrow \mu\mu\mu$ decay.

The decay of a b hadron into three muons without other charged particles in the final state will likely result in a candidate with a large invariant mass, larger than the τ mass and larger than the invariant mass range in which the background spectrum is fitted. Therefore, only c meson decays are investigated further.

The list in Tab. 8.1 is based on the study made in [34] and the updates for [5, 8]. The values are quoted literally. All decays listed are partially reconstructed, i. e. there are particles in the final state which are not reconstructed. Therefore, the invariant mass of the three muons is smaller than the mass of the meson which decays. Given their abundance, it is worth considering to remove them from the data sample. The only feasible way of removing these decays is the two muon invariant mass. Decays involving a $\phi \rightarrow \mu\mu$ decay are excluded from the data by only selecting candidates where neither of the two dimuon combinations has an invariant mass within $20 \text{ MeV}/c^2$ of the ϕ mass. Charm meson decays with a subsequent $\eta \rightarrow \mu\mu\gamma$ decay are excluded by requiring that both dimuon pairs have an invariant mass of at least $450 \text{ MeV}/c^2$.

8.2 Background sources in simulated events

Given the event selection, background candidates are combinations of non-prompt charged particle tracks. These originate from c and b hadron decays. Since a positive muon identification is required in the selection, a generic background is a pp collision with a $c\bar{c}$ or $b\bar{b}$ pair production where muons occur in the decay chains of the heavy flavour hadrons. Two or more real muons originating from c or b hadron decays are required in each simulated event. Due to limited computing resources, additional restrictions are imposed on the simulated muons, to ensure that the simulated events are selected once the events are reconstructed. Both muons to have a transverse momentum larger than $280 \text{ MeV}/c$, a momentum larger than $2900 \text{ MeV}/c$, the dimuon invariant mass must be smaller than $4500 \text{ MeV}/c^2$, and a distance of closest approach smaller than 0.35 mm . There is no requirement imposed that the two muons have the same charge or the opposite charge.

Assuming similar backgrounds for collision energies of 7 TeV and 8 TeV, only background at 8 TeV is simulated.

8. BACKGROUND PROCESSES TO $\tau \rightarrow \mu\mu\mu$

Table 8.1: Branching fractions of tri-muon final states from c meson decays [80]. The expected production cross-sections are calculated by multiplying the 4π production cross-section $\sigma(D^+) = 3.16$ mb and $\sigma(D_s) = 976$ μb 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 . The table is taken literally from [8].

D decay	$\mathcal{B}_1^{(*)}$	Secondary decay	\mathcal{B}_2	$\mathcal{B}_1 \times \mathcal{B}_2$	$\sigma(3\mu X)$
D_s					
$\eta\mu\nu_\mu$	2.67×10^{-2}	$\eta \rightarrow \mu\mu$	5.8×10^{-6}	1.5×10^{-7}	0.03 nb
$\eta\mu\nu_\mu$	2.67×10^{-2}	$\eta \rightarrow \mu\mu\gamma$	3.1×10^{-4}	8.2×10^{-6}	1.5 nb
$\eta\mu\nu_\mu$	2.67×10^{-2}	$\eta \rightarrow \pi^0\mu\mu\gamma$	$< 3 \times 10^{-6}$	$< 8.0 \times 10^{-8}$	< 0.02 nb
$\eta'\mu\nu_\mu$	9.9×10^{-3}	$\eta' \rightarrow \mu\mu\gamma$	1.09×10^{-4}	1.1×10^{-6}	0.20 nb
$\phi\mu\nu_\mu$	2.49×10^{-2}	$\phi \rightarrow \mu\mu$	2.87×10^{-4}	7.1×10^{-6}	1.3 nb
$\phi\mu\nu_\mu$	2.49×10^{-2}	$\phi \rightarrow \mu\mu\gamma$	1.4×10^{-5}	3.5×10^{-7}	0.06 nb
$\phi\mu\nu_\mu$	2.49×10^{-2}	$\phi \rightarrow \mu\mu\pi^0$	$1.12 \times 10^{-5}(\dagger)$	2.8×10^{-7}	0.05 nb
D⁺					
$\eta\mu\nu_\mu$	1.14×10^{-3}	$\eta \rightarrow \mu\mu$	5.8×10^{-6}	6.6×10^{-9}	< 0.01 nb
$\eta\mu\nu_\mu$	1.14×10^{-3}	$\eta \rightarrow \mu\mu\gamma$	3.1×10^{-4}	3.5×10^{-7}	0.20 nb
$\eta\mu\nu_\mu$	1.14×10^{-3}	$\eta \rightarrow \pi^0\mu\mu\gamma$	$< 3 \times 10^{-6}$	$< 3.4 \times 10^{-9}$	< 0.01 nb
$\eta'\mu\nu_\mu$	2.2×10^{-4}	$\eta' \rightarrow \mu\mu\gamma$	1.09×10^{-4}	2.4×10^{-8}	0.01 nb
$\omega\mu\nu_\mu$	1.6×10^{-3}	$\omega \rightarrow \mu\mu$	9.0×10^{-5}	1.4×10^{-7}	0.09 nb
$\omega\mu\nu_\mu$	1.6×10^{-3}	$\omega \rightarrow \mu\mu\pi^0$	1.3×10^{-4}	2.1×10^{-7}	0.13 nb
$\rho^0\mu\nu_\mu$	2.4×10^{-3}	$\rho^0 \rightarrow \mu\mu$	4.55×10^{-5}	1.1×10^{-7}	0.07 nb
$\phi\mu\nu_\mu$	$< 9 \times 10^{-5}$	$\phi \rightarrow \mu\mu$	2.87×10^{-4}	2.6×10^{-8}	0.02 nb
(*) : given branching ratios are from corresponding $e\nu_e$ decays					
(\dagger) : given branching ratio is from $\phi \rightarrow e^+e^-\pi^0$ decays					

The probability that a simulated $pp \rightarrow q\bar{q}X \rightarrow (\mu\mu Y)_{q\bar{q}}X$ event is accepted by the generator selection is evaluated using simulated events. The probabilities are $(5.875 \pm 0.047) \times 10^{-4}$ for charm and $(2.788 \pm 0.029) \times 10^{-3}$ for bottom quarks. The determination has been carried out by Jonathan Harrison [8].

The inclusive charm cross section is taken from [129], scaled up by $8/7$ to account for the different collision energy. It is (6.95 ± 1.07) mb. The inclusive bottom cross section is taken from [134] and amounts to (298 ± 36) μb .

The background simulation of 10 million charm events and 5 million bottom events corresponds to (2.5 ± 0.4) pb^{-1} and (6.2 ± 1.0) pb^{-1} , respectively. The charm events contain 4641 candidates fulfilling the selection from Chap. 6. And 4936 candidates are in the bottom events.

8.3. Background from misidentified particles

Table 8.2: Charm decay modes which contribute to the misidentified background. The number of events produced in 1 fb^{-1} in the LHCb acceptance shown in the fourth column is estimated using the LHCb measured inclusive D^+ , D_s , and D^{*+} cross sections in 4π (summed over charm and beauty contributions), the PDG branching fraction and a 20% acceptance efficiency. The table is taken literally from [8].

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

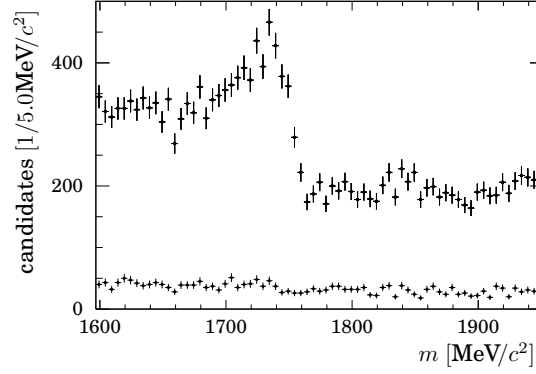
† : These are sufficiently suppressed by the \mathcal{M}_{PID} classifier.

8.3 Background from misidentified particles

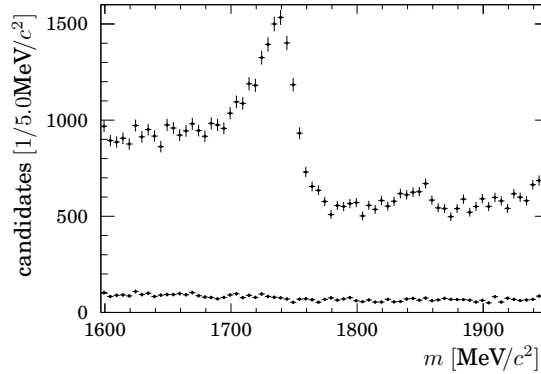
Background processes with misidentified particles are decay processes with three charged particles in the final state, which are not muons. The particle identification should prevent tracks of these particles to be considered in the analysis, a residual misidentification rate can, however, never be ruled out.

The decays which are relevant for the $\tau \rightarrow \mu\mu\mu$ analysis are listed in Tab. 8.2. Most background processes are observable when no requirement on \mathcal{M}_{PID} is imposed. When the lowest \mathcal{M}_{PID} bin is removed from the analysis, only $D^+ \rightarrow K^- \pi^+ \pi^+$ needs to be considered in the background description. The three muon invariant mass spectrum in data is shown in Fig. 8.1 for candidates with $\mathcal{M}_{3\text{body}} > 0.3$ – to enhance the visibility of the backgrounds above the combinatorial background – with and without a requirement of $\mathcal{M}_{\text{PID}} > 0.2$. The values in Tab. 8.2 are literally quoted from [8].

8. BACKGROUND PROCESSES TO $\tau \rightarrow \mu\mu\mu$



(a) 2011



(b) 2012

Figure 8.1: Invariant mass spectrum of $\tau \rightarrow \mu\mu\mu$ candidates in data for both data taking periods. All candidates are required to have $\mathcal{M}_{3\text{body}} > 0.3$. The respective upper points show are for candidates without an additional \mathcal{M}_{PID} criterion, the lower points for candidates with $\mathcal{M}_{\text{PID}} > 0.2$. The large structure at $1740 \text{ MeV}/c^2$ is the $D^+ \rightarrow K^- \pi^+ \pi^+$ background. It is visible that the contribution of this background is drastically reduced by the \mathcal{M}_{PID} criterion and the combinatorial background dominates the data above \mathcal{M}_{PID} values of 0.2.

$\mathcal{M}_{3\text{body}}$ development

Various multivariate classification techniques are implemented in the software toolkits NeuroBayes@[151, 152] and TMVA[153]. These two toolkits are considered for the development of $\mathcal{M}_{3\text{body}}$. The development is based on the simulated signal and background for the data taking conditions in 2012. The classifier is applied to the data taken in 2011 without changes. A different binning is applied nevertheless.

9.1 Training of $\mathcal{M}_{3\text{body}}$

Multivariate classification techniques for optimal signal/background separation use machine learning for the construction of the classifier. A sample of signal and a sample of background events are used as input of the learning algorithm, called *training samples*. The learning algorithm constructs a classifier based on the properties observed in the training samples. Most learning algorithms are iterative procedures, where the classifier is iteratively improved until no further improvement can be reached. A review of the current state of the usage of multivariate techniques in particle physics can be found in [121, 122]. The non-automised part of the training are the following aspects.

variable selection: From a set of variables which are believed to provide information for the signal/background discrimination, several combinations are formed and all classification methods are trained on each variable set. As explained before, it is expected that different variable sets will be optimal for boosted decision trees (BDT) and neural networks (NN), the latter performing better when irrelevant inputs are removed while the former are invariant.

training data: The training samples have to describe the actual signal and background correctly. A background which is missing in the background training might not be discriminated from the signal by the trained classifier and correspondingly non-simulated signal components would not be discriminated from the background. Furthermore, components of the signal and background samples have to be represented in the right proportions in the training samples – a background which is under-represented in the training sample might not be discriminated well by the trained classifier, if this allows for a better discrimination of other backgrounds. It therefore needs to be ensured that the training data is accurate in terms of signal production (as described in Chap. 4) and background components (as described in Sect. 8.2). This background sample contains $b\bar{b}$ and $c\bar{c}$ production with two muons in the final state. Consequently, events without heavy flavour production and events without muons in the final state are not

correctly represented. The selection and the application of \mathcal{M}_{PID} , as explained in Sect. 5.2 excludes these events from the analysis and they are therefore not necessary in the training data.

candidate selection: The training can be improved further by restricting the candidates in the training samples to not using all candidates. E. g. Background candidates which are made from three charged particles from a c hadron decay without further charged particles in the final state are not expected to be distinguishable from $\tau \rightarrow \mu\mu\mu$ decays, as further explained in Sect. 9.5.

Ultimately, the performance of different classifications cannot be distinguished due to limited sensitivity of the performance evaluation, as visible in Fig. 9.4. At that point further improvement can hardly be obtained in a purposeful optimisation and the development is stopped.

9.2 Data sample division

The training data is divided into disjoint subsamples to avoid the introduction of biases due to the classifier development (c. f. Sect. 5.4).

Training sample 60 % of the simulated signal events and 50 % of the simulated background events are used to train multivariate classifiers

Test sample 30 % of the simulated signal events and 50 % of the simulated background events are used as “test” sample. These events are used to determine which classifier is the best, how to configure the training algorithms and develop the classifier binning.

Calibration sample 10 % of the simulated signal events are left out of the development process. The performance of the classifier on this sample is therefore unbiased and serves as input to the calibration procedure outlined in Sect. 10.2.

The numbers are chosen in this way because the machine learning stage should get most “knowledge” of the events by receiving the largest part of the simulated statistics. The development is on the other hand easily limited by the statistical significance at which classifiers can be compared to each other, therefore the test sample must not be too small either. The remaining calibration sample kept at a size which is expected to be large enough to keep the statistical uncertainty of the $\mathcal{M}_{3\text{body}}$ calibration at an acceptable level.

9.3 The boosted decision tree technique

The boosted decision tree used in [147], for example, uses 452 cut values. Understanding every single value of that tree is unfeasible and unnecessary. The analysis must be able to only use the classifier response and phenomenologically understand the physics of the response distribution.

For the search for $\tau \rightarrow \mu\mu\mu$, the author studied boosted decision trees with adaptive boosting as implemented in TMVA, while following instructions by colleagues [154, 155] about which configurations to test when tuning a multilayer perceptron (the neural network, implemented in TMVA) and following instruction manuals for NeuroBayes® neural networks.

Given that eventually a boosted decision tree is chosen in Sect. 9.7, a description of boosted decision trees is given here.

Boosted decision trees

Decision trees [156] with adaptive boosting [157] have first been used in particle physics by the MiniBooNE collaboration [158].

Simple decision trees: Simple decision trees are binary trees, as shown in Fig. 9.1. When classifying events, at each node one variable of the event is compared to a fixed value to decide which branch to follow. Eventually, a leaf is reached which either assigns the signal hypothesis or the background hypothesis to the event.

The difference between a selection and a decision tree is that the former selects one hyper cube in the variable space while the latter selects multiple hypercubes.

In the training process, it is decided which variable is considered in each node and at which value the branches are separated. The variable and value are chosen which provide the best separation of the training at that node. The separation is measured with the Gini-Index, $g = p \cdot (1 - p)$, where p is the decision purity¹, as explained in e. g. [158]. If the training is not constrained to a maximum tree size, the tree can be as large as necessary to achieve a perfect classification of the training sample². In this case that leaf assigns the signal hypothesis if the training events are signal events and, conversely, the background hypothesis if the events are background events. If the tree size is limited and multiple training events end up in a leaf, a majority vote, illustrated in Fig. 9.2, decides the leaf type.

The tree depth can be limited to avoid overtraining. This, however, reduces the maximum achievable separation power significantly. In practise, it is found to be preferable to limit the tree depth and apply a boosting procedure to overcompensate the performance loss. In addition to enhancing the separation power, boosting results in classifiers which are less vulnerable to overtraining than single decision trees.

The training is not necessarily based on event counts, weights can be assigned to the training events which are considered in the computation of the Gini-index and in the majority vote.

Adaptive boosting: Adaptive boosting can be understood as a power series, summing classifier responses to converge to the best possible classifier [159, 160]. It can be applied to any classifier, while “weak” classifiers profit most, e. g. decision trees with limited depth.

An initial decision tree is trained with weights \vec{w}_0 , where the vector indicates individual weights for individual training events while the index stands for the training iteration. The initial weights are 1 for signal events and background events from b quark pair production. For c quark pair production, the weights is chosen to give charm and bottom events, which corresponds to the same integrated luminosity the same total weight. The charm background weights are set to

$$2.48 \left(= \frac{6.2 \text{ pb}^{-1}}{2.5 \text{ pb}^{-1}} \right).$$

¹ $p = N(\text{correct decisions}) / (N(\text{correct decisions}) + N(\text{incorrect decisions}))$

²This is always possible, even if the signal and background samples do not differ in any physics property. Random decisions at each node will create a perfect tree at a tree depth of $\log_2(N)$ for N training events.

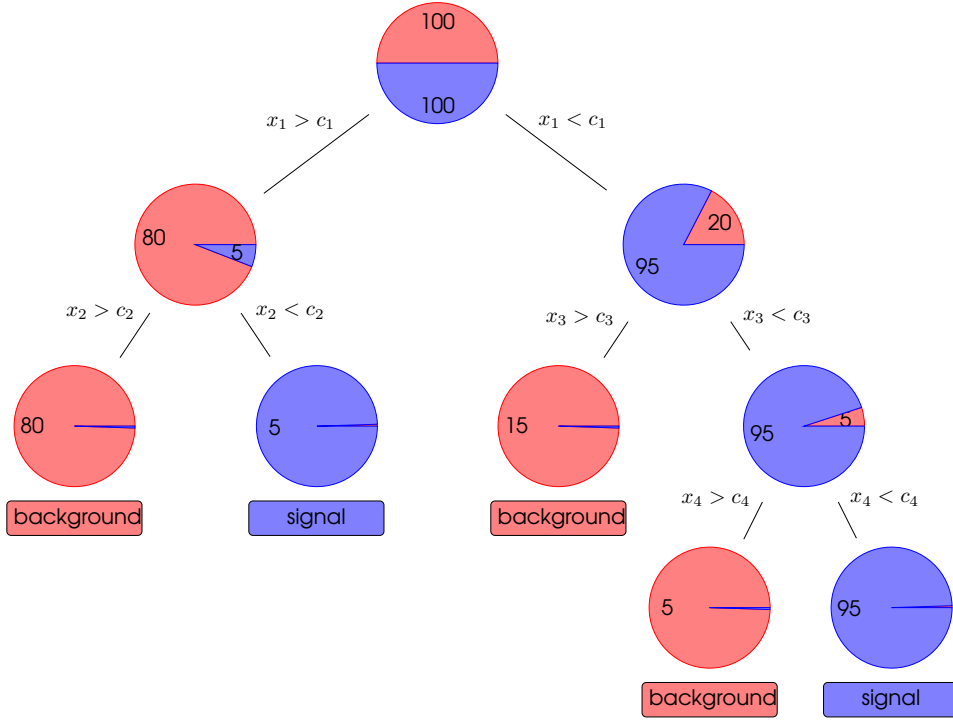


Figure 9.1: Example of a full grown decision tree. The numbers are to be understood as absolute numbers of signal and background events in the training data, x_i are event variables, c_i are constants which determine which branch to follow.

After a training iteration i , the error rate³ e_i of the training is computed and the weights of the training events are updated. The speed at which the weights are updated is steered by the predefined parameter β :

$$(w_{i+1})_j = (w_i)_j \cdot \begin{cases} 1 & \text{event } j \text{ is classified correctly in iteration } i \\ \left(\frac{1-e_i}{e_i}\right)^\beta & \text{event } j \text{ is classified wrongly in iteration } i \end{cases}$$

In the next iteration, a decision tree is trained with the updated weights w_{i+1}^\rightarrow and the procedure is repeated. The number of iterations, N , is predefined and usually at the order of a few hundreds.

Once all training iterations are performed, the boosted decision tree response $y(j)$ of an event j is computed as a weighted average of the individual tree responses $h_i(j)$ using the error rate e_i .

$$h_i(j) = \begin{cases} 0 & \text{event } j \text{ reaches a background leaf in the tree from iteration } i \\ 1 & \text{event } j \text{ reaches a signal leaf in the tree from iteration } i \end{cases} \quad (9.1)$$

$$y(j) = \frac{1}{N} \sum_i \ln \left(\frac{1-e_i}{e_i} \right)^\beta h_i(j) \quad (9.2)$$

³ $e = N(\text{incorrect decisions}) / (N(\text{correct decisions}) + N(\text{incorrect decisions}))$

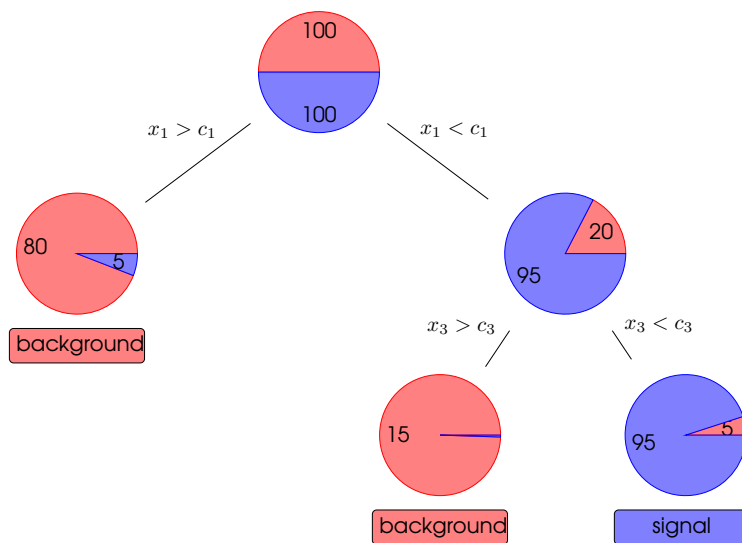


Figure 9.2: Example for a majority vote for one of the nodes in the decision tree from Fig. 9.1

A common misconception is that the last training iteration h_N should provide a better discrimination than the weighted average y . In contrast it must be pointed out that during the reweighting procedure, the weighted training data represents the actual classification problem less and less accurately. The best individual tree is, in fact, the initial iteration h_0 . All subsequent iterations “only” correct residual misclassification. See [159] or [160] for simple examples for boosting which illustrate this.

The description here is based on [161], the same information can be found in [153, Sects. 7.1 and 8.12].

9.4 Input variables for $\mathcal{M}_{3\text{body}}$

The signal properties, listed in Sect. 6.1, can be quantified in several variables. Background candidates do not necessarily have these signal properties and can therefore be differently distributed in the variable space than the signal. The variables which are considered for the $\mathcal{M}_{3\text{body}}$ development are listed in Tab. 9.1. They are partially covered in Chap. 6. For some variables, different options of usage are listed, e. g. the impact parameters of all three muon tracks of the $\tau \rightarrow \mu\mu\mu$ candidate are used as input variables, or the maximal and minimal value, or only the minimal value, or none of them. The option of not using a variable is implied for all variables. The variables which need additional explanation with respect to the list in Chap. 6 are:

Reconstructed flight distance: The distance between the τ decay vertex and the primary interaction vertex, measured in the lab system. It is a measure of the displacement similar to the reconstructed decay time.

Track distances of closest approach: For signal events, the extrapolations of the muon tracks meet in a single point up to resolution effects. Consequently, the

9. $\mathcal{M}_{3\text{BODY}}$ DEVELOPMENT

Table 9.1: List of variables from which the $\mathcal{M}_{3\text{body}}$ input variables are chosen. For many variables there are several options to use them. Those options tested are listed in the individual rows of the second column.

variable	options	property
	all three values	
track impact parameter	all min&max	displacement
	min	
	all three values	
track χ_{IP}^2	min&max	displacement
	min	
	best fit & uncert.	
flight distance wrt. PV	best fit	displacement
	best fit uncert.	
decay time		displacement
3μ vertex fit χ^2		rec. quality
	all three values	
track distance of closest approach	max&min	rec. quality
	max	
	IP&uncert.	
τ impact parameter	IP	pointing
	χ_{IP}^2	
τ transverse momentum		other
	all three values	
track isolation	\sum three values	isolation
	max	
cone isolation		isolation
	all three values	
track fit χ^2/ndf	three "match" contribs.	rec. quality
	max "match" contrib.	

minimal distance of closest approach of each pair of these tracks is zero, up to resolution effects. For background candidates where tracks from different vertices are combined, this is not the case.

τ transverse momentum: The production of τ leptons has a characteristic p_T spectrum, as a consequence of the p_T spectrum of b and c production and the decay kinematics. The spectrum of background candidates can be different.

Track isolation: There are exactly three charged particle tracks originating in a $\tau \rightarrow \mu\mu\mu$ decay vertex. Background processes can have more than three charged particles in the final state.

For each of the three tracks of a $\tau \rightarrow \mu\mu\mu$ candidate i and for each other track in the event j a multivariate classifier \mathcal{M}_{iso} is evaluated which measures the likelihood that i and j originate from a common heavy flavour decay vertex. It is trained on simulated background events which pass the event selection. The training uses pairs of tracks which do (not) originate from the same heavy flavour decay chain as signal (background) events. In background events where the $\tau \rightarrow \mu\mu\mu$ candidate tracks originate from a heavy-flavour decay with more than 3 charged particles in the final state, the classifier \mathcal{M}_{iso} thus identifies the track j which is most likely a decay product from a background process.

\mathcal{M}_{iso} has been developed by Laura Gavardi in [35].

Cone isolation: Another commonly used variable to measure the abundance of tracks around a decay candidate is the cone isolation. For all particles within a circle in the ϕ - η -plane around the candidate, the transverse momenta are summed. The fraction $p_T(\text{cand.}) / \sum_{\text{cone}} p_T$ is the cone isolation. The size of the cone is a tuning parameters. The training of the final $\mathcal{M}_{3\text{body}}$ classifier has been repeated for different cone radii from 0.1 to 2.0. Optimal $\mathcal{M}_{3\text{body}}$ performance is found for a cone radius of 1.0 by Laura Gavardi [35]. The cone isolation has been used first in [162].

Track match χ^2/ndf : A significant fraction of fake tracks are wrong combinations of correctly reconstructed VELO tracks with correctly reconstructed T-tracks. The quality χ^2/ndf of the track fit of these fake tracks is usually good because the two segments, in the VELO and in the T stations, can be fit the track model quite accurately. The contribution of how well the VELO segment and the T station segment are compatible is measured as χ_{match}^2 .

The variables v_i in Tab. 9.1 are potentially useful input variables for $\mathcal{M}_{3\text{body}}$. For $\{v_i\}$, the different options are considered as individual variables:

$$\begin{aligned} v_1 &:= (\text{IP}_1, \text{IP}_2, \text{IP}_3) \\ v_2 &:= (\min(\text{IP}_i), \max(\text{IP}_i)) \\ v_3 &:= \min(\text{IP}_i) \\ v_4 &:= (\chi_{\text{IP},1}^2, \chi_{\text{IP},3}^2, \chi_{\text{IP},3}^2) \\ &\vdots \end{aligned}$$

For each element V of the powerset

$$V \in \mathcal{P}(\{v_i\}) := \{V | V \subseteq \{v_i\}\}$$

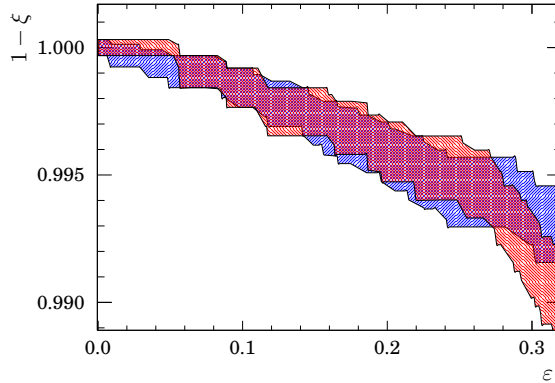


Figure 9.3: Classification performance of $\mathcal{M}_{3\text{body}}$ when irreducible background candidates are included (blue) and excluded (red) from the training sample. The bands indicated the statistical uncertainties due to the amount of events in the inner sidebands. The same test events are used for both classifiers, the shown uncertainty does therefore not illustrate if the curves differ significantly.

a multivariate classifier is trained to determine the optimal set of input variables. Combinations of different options of a variable⁴ in V are excluded. Those variable sets V where the decay length, the decay time, and the τ transverse momentum are in V are excluded, too, because this combination could enable the multivariate framework to determine the candidate's invariant mass. For all remaining V , a TMVA boosted decision tree, a TMVA neural network, and a NeuroBayes® neural network are trained.

For each multivariate technique, the ten best variable sets V are determined by measuring the signal efficiency and background rejection on the test sample. The decision is made by means of the background efficiency at a signal efficiency of 40%. For each technique, the best ten variable sets are identified and the resulting thirty classifiers are studied further.

9.5 Candidate selection

Charm hadron decays with three charged particles in the final state, which are in the background training sample, cannot be expected to be discriminatable from the signal by $\mathcal{M}_{3\text{body}}$. This is a consequence of the signal properties for which $\mathcal{M}_{3\text{body}}$ is developed because these charm backgrounds share these properties. To avoid this, background candidates which are correct reconstructions of decays with three charged particles in the final state are removed from the training sample.

In [8, 5] it has been found that many classifiers profit from this candidate selection. Before the configuration optimisation of the classifier chosen in this work, only an insignificant improvement is observed, as shown in Fig. 9.3. The curves⁵ are obtained

⁴as indicated by the horizontal lines in Tab. 9.1

⁵These curves are called Receiver Operating Characteristic curves (ROC curves).

by scanning $0 \leq x \leq 1$ and drawing $1 - \xi(\mathcal{M} > x)$ as a function of $\varepsilon(\mathcal{M} > x)$. Using

$$\begin{aligned} N_b &:= N(\text{bkg. evts.}) \\ N_s &:= N(\text{sig. evts.}) \\ k_b &:= N(\text{bkg. evts. with } \mathcal{M} > x) \\ k_s &:= N(\text{sig. evts. with } \mathcal{M} > x) \end{aligned}$$

the signal efficiency ε and the background rejection $1 - \xi$ are given by the following Bayesian estimators (for prior parameters $\alpha = \beta = 0.5$):

$$\varepsilon(\mathcal{M}, x) = \frac{k_i}{N_i} \pm \underbrace{\sqrt{\frac{(k_i + \alpha)(N_i - k_i + \beta)}{(N_i + \alpha + \beta)^2(N_i + \alpha + \beta + 1)}}}_{=: \sigma_\varepsilon(\mathcal{M}, x)} \quad \text{for } i = s. \quad (9.3)$$

And ξ correspondingly for $i = b$.

The formula is given in e. g. [163]. The signal test sample is three times larger than the background test sample, therefore the simplification that all uncertainties are dominated by the background sample is made.

The comparison of two classifiers is done by comparing the background efficiency at a response of a fixed signal efficiency, here $\varepsilon = 0.25$; i. e. for two classifiers \mathcal{M}_A and \mathcal{M}_B the response values x_A and x_B are defined by

$$\begin{aligned} \varepsilon(\mathcal{M}_A > x_A) &= 0.25 \\ \varepsilon(\mathcal{M}_B > x_B) &= 0.25. \end{aligned}$$

The classifier decision function \mathcal{D} for a classifier \mathcal{M}_A and for an event j is defined as

$$\mathcal{D}_A(j) := \begin{cases} 0 & \text{if } \mathcal{M}_A(j) < x_A \\ 1 & \text{if } \mathcal{M}_A(j) \geq x_A \end{cases}$$

and correspondingly for \mathcal{M}_B .

The correlation between classifier decisions is the Pearson correlation coefficient⁶ of the classifier decision functions:

$$\text{corr}(A, B) := \frac{E[(\mathcal{D}_A - E(\mathcal{D}_A))(\mathcal{D}_B - E(\mathcal{D}_B))]}{\sigma[\mathcal{D}_A]\sigma[\mathcal{D}_B]}$$

Since the classifier comparison here is done by means of observed background events in the simulated background, the observed correlation coefficient is evaluated, and the expression can be simplified with the background efficiency:

$$\text{corr}(A, B) = \frac{E[(\mathcal{D}_A - \xi(\mathcal{M}_A, x_A))(\mathcal{D}_B - \xi(\mathcal{M}_B, x_B))]}{\sigma_\xi(\mathcal{M}_A, x_A)\sigma_\xi(\mathcal{M}_B, x_B)}.$$

This correlation is necessary to compute the significance of the difference of two classifier background efficiencies⁷, as given by error propagation:

$$\begin{aligned} D &= \xi(\mathcal{M}_A, x_A) - \xi(\mathcal{M}_B, x_B) \\ \sigma_D^2 &= \sigma_\xi^2(\mathcal{M}_A, x_A) + \sigma_\xi^2(\mathcal{M}_B, x_B) - 2 \text{corr}(A, B) \sigma_\xi(\mathcal{M}_A, x_A) \sigma_\xi(\mathcal{M}_B, x_B). \end{aligned}$$

At the benchmark point $\varepsilon = 0.25$, the background efficiency changes by 0.6σ when excluding three body charm decays from the background training sample.

⁶ E is the expectation value, σ the standard deviation from Eq. 9.3.

⁷Rigorously correct would be the correlation of the efficiency estimates. It has been verified in a toy study that the observed decision correlation is a faithful estimate of the efficiency estimator correlation.

9.6 Training parameters of $\mathcal{M}_{3\text{body}}$

To achieve an optimal performance, the following four BDT training parameters are optimised by repeating the training various times while varying these parameters. The figure of merit is the signal efficiency at 1% background efficiency, measured on the test sample. This is one of the figures of merit implemented in TMVA for the optimisation, it is regarded to be the one with the largest relevance for rare decay searches. For each of the trainings, the performance is measured on the test sample. The optimised parameters are the number of decision trees (N), the learning speed (β), the minimum number of events for a node creation, and the maximum tree depth.

It is expected that the optimal set of variables depends on the maximum tree depth and the number of trees. Therefore, for each of the ten best variable sets identified before, a parameter optimisation is performed to achieve a performance gain. Whether a performance maximum is found is verified for the best variable set after the parameter optimisation as follows. For the optimal training parameters of the most performant variable set, the training is repeated removing one variable from the variable set at a time, or adding one variable to the variable set at a time. This test identifies if the optimal variable set was only locally optimal, i. e. optimal for the default training parameters of TMVA. No performance improvement from updating the variable selecting after the parameter optimisation is observed.

The final boosted decision tree which uses 762 trees with a maximal tree depth of 3. The boost parameter is set to $\beta = 0.5$. Its variables are listed in the following section.

9.7 Final $\mathcal{M}_{3\text{body}}$

After the classifier optimisation, the best classifier is identified. For x in steps of 0.005, the signal efficiency and the background efficiency are determined for $\mathcal{M} > x$. Both efficiencies are measured as the observed efficiency on the simulated test sample and the inner sidebands. The uncertainties are statistical uncertainties evaluated as before.

The background efficiency is measured on the inner sidebands instead of the simulated background to consider possible background processes in the selections which are missing in the simulation. It is expected that there's a significant contribution of $D^+ \rightarrow K^-\pi^+\pi^+$ decays present in the data. Since these are not discriminatable by the classifier their contribution to the background sample needs to be reduced. A suppression of hadronic backgrounds is achieved by considering only candidates with $\mathcal{M}_{\text{PID}} > 0.2$. The effectivity of this selection cut is visible in Fig. 8.1: after the application of the cut, the remaining data is dominated by combinatorial backgrounds.

The variables used by the best classifier of each kind (TMVA boosted decision tree, TMVA neural network, and NeuroBoyes® neural network) are listed in Tab. 9.2. The ROC curves for the best classifier of each kind are shown in Fig. 9.4. The classifiers show similar performance. The boosted decision tree is chosen due to the authors familiarity with this technique.

The uncorrelated uncertainties of the classifiers' background efficiencies is esti-

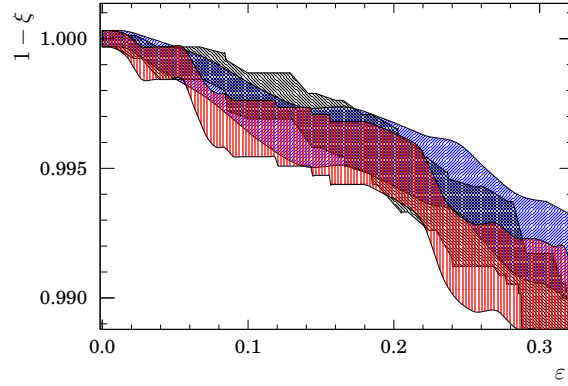


Figure 9.4: Discrimination performance of the best TMVA neural network (blue), NeuroBayes[®] neural network (red), and TMVA boosted decision tree (black). The background efficiency, $1 - \xi$, is measured on the inner sidebands, the signal efficiency, ε , is measured on the test sample. The uncertainties, indicated by the widths of the bands ignore correlations between the classifiers.

Table 9.2: Input variables of the best performing classifier of each classification technique. The variables are ordered according to their importance in the boosted decision tree, the numeric importance is also given for the boosted decision tree. The horizontal line indicates the median of the importances.

variable	boosted decision tree	TMVA neural network	NeuroBayes [®] neural network
track isolation	0.225	✓	✓
pointing angle	0.140		
track distances of closest approach (3 variables)	0.050 +0.046 +0.043		
τ flight distance	0.071	✓	✓
decay time	0.069	✓	✓
τ transverse momentum	0.064		✓
minimal track impact parameter χ^2	0.063	✓	✓
vertex χ^2	0.062	✓	✓
cone isolation	0.061	✓	✓
maximal track match χ^2	0.056	✓	✓
τ flight distance uncertainty	0.050	✓	✓
τ impact parameter χ^2		✓	✓
track fit χ^2 / ndf (3 variables)			✓

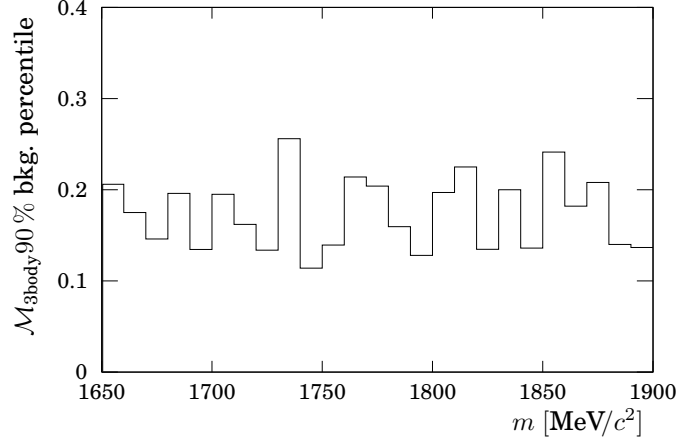


Figure 9.5: The background 90% percentiles of $\mathcal{M}_{3\text{body}}$ for different invariant mass ranges. I. e. the $\mathcal{M}_{3\text{body}}$ value, below which 90% of the simulated background events are ranked, in bins of the invariant 3 muon mass. No peak at the τ rest mass is visible.

mated as outlined in Sect. 9.5 for $\varepsilon = 0.25$:

$$\begin{aligned} 1 - \xi_{\text{TMVA NN}} &= (99.46 \pm 0.11) \% \\ 1 - \xi_{\text{TMVA BDT}} &= (99.28 \pm 0.13) \% \\ 1 - \xi_{\text{NeuroBayes}^{\text{®}} \text{ NN}} &= (99.21 \pm 0.13) \%. \end{aligned}$$

The importance of an input variables of a boosted decision tree is defined as the sum of the separation gain squared in each node multiplied by the number of training events in the node and the tree's boosting weight from Eq. 9.2, summed over all trees and nodes using the variable. The variables in Tab. 9.2 are ordered by the importance of the variables in $\mathcal{M}_{3\text{body}}$.

The true performance of $\mathcal{M}_{3\text{body}}$ is known after its calibration in the following chapter: Figures 10.5 and 10.6 show the response distributions for the $\tau \rightarrow \mu\mu\mu$ signal and the background in data respectively.

9.8 $\mathcal{M}_{3\text{body}}$ characterisation

Mass correlation

In Sect. 12.1, the number of background events in the signal region is estimated with a fit to the invariant mass spectrum in the sidebands. If the background efficiency of $\mathcal{M}_{3\text{body}}$ was different for candidates with an invariant mass in the signal region than for candidates in the sidebands, then the fitted function would not predict the number of expected background events correctly. $\mathcal{M}_{3\text{body}}$ percentiles on the simulated background are determined as a function of the invariant mass. The 90% percentiles in Fig. 9.5 distributions do not show signs for a problematic correlation.

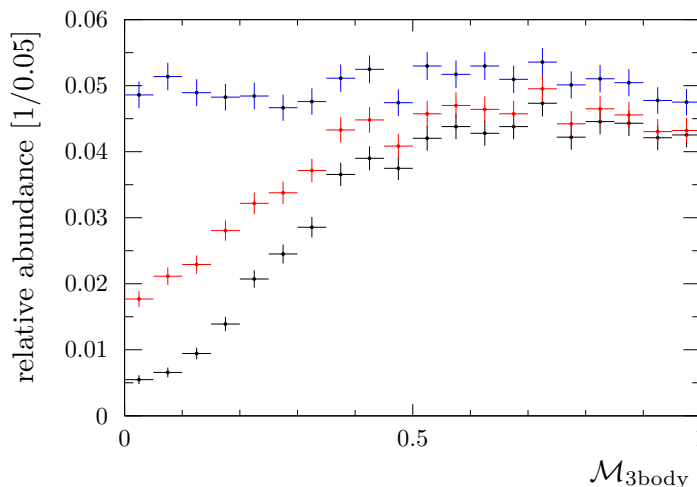


Figure 9.6: Distribution of $\tau \rightarrow \mu\mu\mu$ candidates on the simulated 2011 signal test sample. The full distribution is given in blue, the contribution of τ from inclusive D_s decays in red, the contribution of τ from prompt D_s decays in black. The error bars are the statistical uncertainty due to the size of the simulated sample.

Production preference

When assessing the systematic uncertainty due to the τ production fractions in Sect. 11.2 the calibration of $\mathcal{M}_{3\text{body}}$ is repeated for varied τ production fractions. This is necessary because $\tau \rightarrow \mu\mu\mu$ decays where the τ originated in a $D_s \rightarrow \tau\nu_\tau$ decay are preferred by $\mathcal{M}_{3\text{body}}$ over other production modes. This is a consequence of the signal properties, listed in Sect. 6.1. In other words, τ from D_s decays can be distinguished from background candidates much more efficiently than other τ . The τ from B decays are “sacrificed” by $\mathcal{M}_{3\text{body}}$ for an optimal performance on the integrated sample of all τ . The preference is visualised in Fig. 9.6.

9.9 Binning

The response ranges of \mathcal{M}_{PID} and $\mathcal{M}_{3\text{body}}$ are split into bins. The method to choose an optimal binning is to compute the difference of the median Q values under the b and the $s+b$ hypotheses for possible binnings, called ΔQ . The method has been introduced in LHCb in the search for $B_s^0 \rightarrow \mu\mu$ as documented in [127]. Numeric examples for the optimisation method can be found in Appendix A.

Before the optimisation, a probability integral transform is applied to $\mathcal{M}_{3\text{body}}$. A grid with a 0.05 grid point spacing in the two dimensional $(\mathcal{M}_{3\text{body}}, \mathcal{M}_{\text{PID}})$ space is used to define possible bin boundary positions. The expected number of background events under the b hypothesis is twice⁸ the number of events in the inner sidebands in each bin. The number of signal events in each bin is estimated by the number

⁸The inner sidebands’ invariant mass range covers 20 MeV/c^2 while the signal region covers 40 MeV/c^2 , therefore the event count in the inner sidebands needs to be multiplied by two.

of candidates in that bin in the simulated test sample. It is shown in Appendix A.3 that the number of expected signal events should be chosen at the order of the final sensitivity. The sensitivity of [5] is used as sensitivity prediction for the present work:

$$\frac{\text{limit}(\mathcal{B}(\tau \rightarrow \mu\mu\mu))}{\alpha} \approx \mathcal{O}(30 \text{ evts.}) \quad (\text{numbers from [5]})$$

using the normalisation factor for the 2012 data due to the assumption that the sensitivity is driven by the 2012 data and rounding up as an ad-hoc correction. The simulated signal events for the 2012 data are therefore weighted to $\sum w_i = 30$. Roughly half as many signal events are expected in the 2011 data because it corresponds to only half the integrated luminosity.

The optimisation procedure determines the optimal bin boundaries for a given number of bins in each $\mathcal{M}_{3\text{body}}$ and \mathcal{M}_{PID} . The maximal figure of merit for each given number of \mathcal{M}_{PID} bins and number of $\mathcal{M}_{3\text{body}}$ bins combination is shown in Fig. 9.7.

One bin in either dimension means that the entire \mathcal{M} range from 0 to 1 is a single bin. This is equivalent to not using the classifier at all. The figure of merit has small values if one or both classifiers are not used.

Two bins in either dimension mean that the range $[0, 1]$ is split into two bins $b_l = [0, x]$ and $b_h = [x, 1]$. As stated in Sect. 3.4, the data in the lower bin is removed from the analysis. This is equivalent to applying a selection requirement $\mathcal{M} > x$. It can be seen that a selection with both classifiers improves the figure of merit significantly over ignoring one of the classifiers (change from blue to grass-green in both data taking periods).

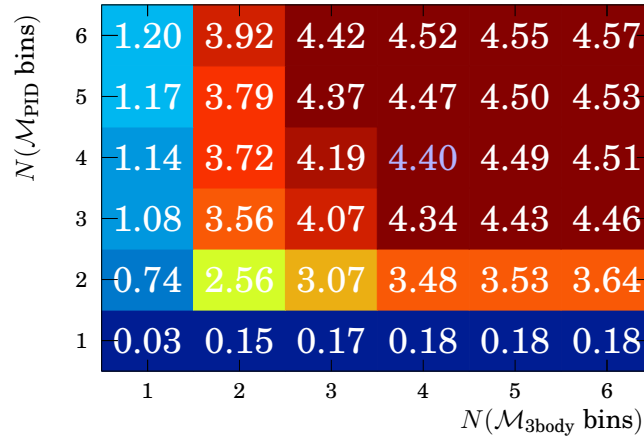
N bins with $N \geq 3$ are a division of the \mathcal{M} range into N bins. As before the lowest bin is not used in the analysis, all other bins are statistically independent counting experiments. The figure of merit saturates in the red region between 8 and 9 for the 2012 data and between 4 and 5 for the 2011 data. It is expected that the 2011 data reaches a smaller figure of merit, since the sensitivity for the data set which corresponds to the smaller integrated luminosity should be smaller.

The figure of merit must, by construction, increase when increasing the number of bins, therefore no stagnancy or downward fluctuations can be seen in the plateau region of Fig. 9.7.

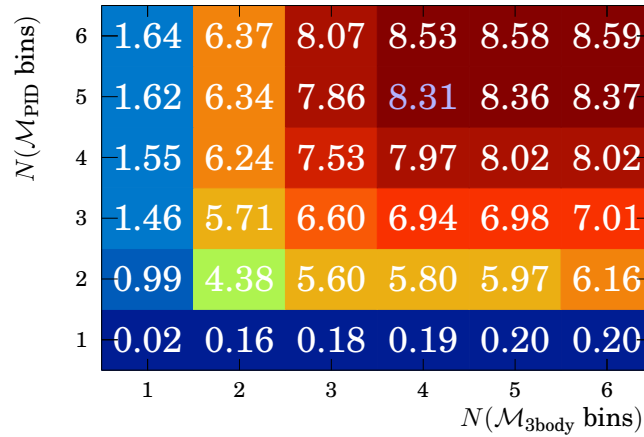
2011 data: The figure of merit saturates at 4 bins in $\mathcal{M}_{3\text{body}}$ and 4 bins in \mathcal{M}_{PID} . The figure of merit is 4.40.

2012 data: The figure of merit saturates at 4 bins in $\mathcal{M}_{3\text{body}}$ and 5 bins in \mathcal{M}_{PID} . The figure of merit is 8.31. Figure 9.7 does not show the behaviour for larger numbers of \mathcal{M}_{PID} bins because the number of possible binnings increases drastically and therefore increasing the plot range would require more than a day of computing time. As a single point the figure of merit for 4 $\mathcal{M}_{3\text{body}}$ bins and 7 \mathcal{M}_{PID} bins was computed and found to be 8.62.

The optimal bin boundaries are listed in Tab. 9.3.



(a) 2011 data



(b) 2012 data

Figure 9.7: ΔQ as a function of the number of bins. If no classification is applied in one of the classifiers (i. e. the number of bins is one), ΔQ takes small values. ΔQ increases once a classification is applied in both classifiers and increases further when the number of bins is increased. Saturation is reached at four bins in $\mathcal{M}_{3\text{body}}$, and 4 bins in \mathcal{M}_{PID} for 2011 data, and 5 bins for 2012 data.

Table 9.3: Optimal bin boundaries for the two classifiers.

$\mathcal{M}_{3\text{body}}$	\mathcal{M}_{PID}
2012 data	
0.15	0.35
0.35	0.65
0.70	0.70
n/a	0.85
2011 data	
0.30	0.15
0.70	0.40
0.85	0.80

The binning optimisation uses the same data for the signal description and the background description which is already used for the choice of the multivariate classifier. This can lead to “overtrained bin boundaries”. I. e. If two classifiers have identical performance, the choice is made based on a statistical fluctuation in the simulated test sample or the inner sidebands. The background response distribution is thus biased to small values in this sample, the signal response distribution to large values, which affects the positions of the bin boundaries. The bin boundaries can therefore be at sub-optimal positions, the analysis is not biased because the signal response is calibrated using an unbiased sample.

Invariant mass: The invariant mass binning is an ad-hoc choice of equidistant bins of $5 \text{ MeV}/c^2$ width covering the $\pm 20 \text{ MeV}/c^2$ signal region.

Likelihood calibrations

10.1 Invariant mass

The distribution of $\tau \rightarrow \mu\mu\mu$ decays in the invariant mass bins is dominated by detector resolution effects. The largest contribution to the detector resolution is the track momentum resolution¹ as described in [85]. The track momentum resolution depends on the track momentum. Therefore the mass resolution for strongly boosted τ is wider than for τ with a small boost. The sum of two Gaussian functions with a common mean describes the $\tau \rightarrow \mu\mu\mu$ mass resolution:

$$f \cdot \mathcal{G}(m; \mu, \sigma_1) + (1 - f) \cdot \mathcal{G}(m; \mu, \sigma_2) \quad \text{with } \sigma_1 < \sigma_2. \quad (10.1)$$

This model is found to describe the $D_s \rightarrow \phi(\mu\mu)\pi$ invariant mass distribution in simulated events and in data, too. The normalisation ratio f can be fixed to 0.7 without degrading the shape description. The mass model fitted to the $D_s \rightarrow \phi(\mu\mu)\pi$ candidates is shown in Fig. 10.1 where an additional exponential function is added to describe the background.

The phase space covered by $D_s \rightarrow \phi(\mu\mu)\pi$ decay products is similar to that of $\tau \rightarrow \mu\mu\mu$ decay products. Furthermore, the invariant mass resolution of $\tau \rightarrow \mu\mu\mu$ and $D_s \rightarrow \phi(\mu\mu)\pi$ in simulated events are compatible, as shown in Fig. 10.2. A first-order correction is obtained from $D_s \rightarrow \phi(\mu\mu)\pi$ decays, correcting the simulated invariant mass resolution to the observed mass resolution:

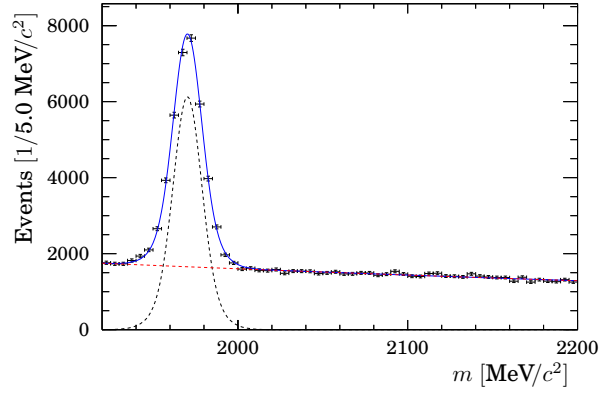
$$\begin{aligned} c_\mu &:= \frac{\mu_{D_s, \text{data}}}{\mu_{D_s, \text{sim}}} \\ c_{\sigma_1} &:= \frac{\sigma_{1, D_s, \text{data}}}{\sigma_{1, D_s, \text{sim}}} \\ c_{\sigma_2} &:= \frac{\sigma_{2, D_s, \text{data}}}{\sigma_{2, D_s, \text{sim}}} \end{aligned}$$

The factors c_μ , c_{σ_1} , and c_{σ_2} are then applied as correction to the simulated $\tau \rightarrow \mu\mu\mu$ mass resolution.

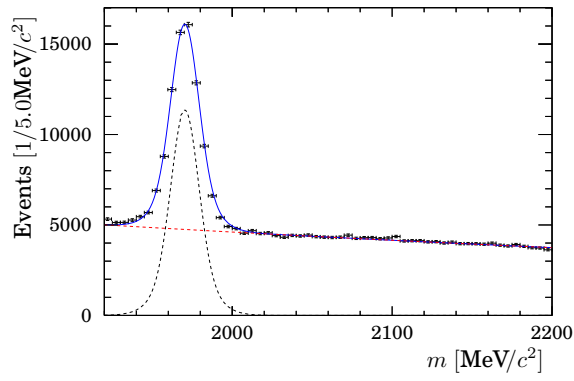
$$\begin{aligned} \mu_{\tau, \text{data}} &= c_\mu \cdot \mu_{\tau, \text{sim}} \\ \sigma_{1, \tau, \text{data}} &= c_{\sigma_1} \cdot \sigma_{1, \tau, \text{sim}} \\ \sigma_{2, \tau, \text{data}} &= c_{\sigma_2} \cdot \sigma_{2, \tau, \text{sim}} \end{aligned}$$

The assumed invariant mass resolution parameters are listed in Tab. 10.1 along with the mass resolution in simulated events.

¹The resolution of the track slopes is not dominant in the kinematic regime of $\tau \rightarrow \mu\mu\mu$.



(a) 2011



(b) 2012

Figure 10.1: Fit to the $D_s \rightarrow \phi(\mu\mu)\pi$ invariant mass distribution, used for the calibration of the invariant mass resolution and the sPlot of the $\mathcal{M}_{3\text{body}}$ response.

Table 10.1: Invariant mass resolution parameters for $\tau \rightarrow \mu\mu\mu$ in simulated events and after the application of the calibration.

year	μ [MeV/ c^2]	σ_1 [MeV/ c^2]	σ_2 [MeV/ c^2]
2011 data	1779.1 ± 0.1	7.7 ± 0.1	12.0 ± 0.8
2011 sim	1779.1 ± 0.05	6.74 ± 0.031	12.12 ± 0.07
2012 data	1779.0 ± 0.1	7.6 ± 0.1	11.5 ± 0.5
2012 sim	1779.0 ± 0.05	7.11 ± 0.03	12.5 ± 0.1

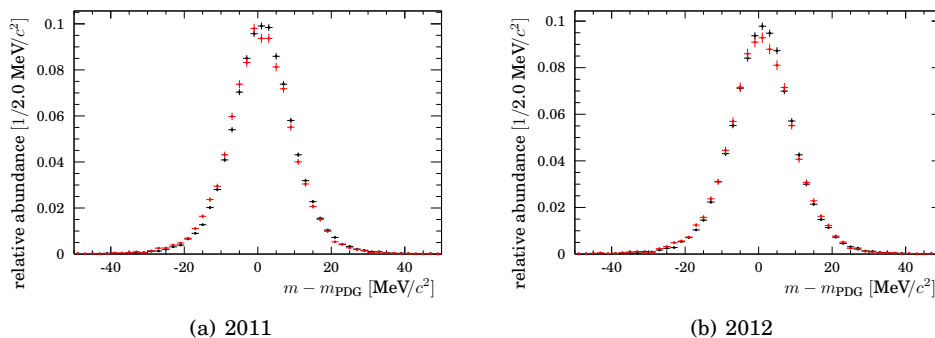


Figure 10.2: Mass resolution for simulated events for $\tau \rightarrow \mu\mu\mu$ decays (black) compared to $D_s \rightarrow \phi(\mu\mu)\pi$ decays (red). The reconstructed invariant mass is m , and the τ lepton or D_s meson mass from [80] is m_{PDG} .

The integral of the invariant mass distribution in the signal region for 2011 is 0.966, i. e. 3.4% of possible $\tau \rightarrow \mu\mu\mu$ decays would not appear in the signal window (0.961 and 3.9% for 2012). The uncertainty on the invariant mass parameters is a correlated uncertainty between this inefficiency and the signal distribution in the invariant mass binning.

For the limit calculation three nuisance parameters arise from the invariant mass shape calibration, one for each of μ , σ_1 , and σ_2 . They are assumed to be uncorrelated among each other.

This assumption corresponds to the following seven invariant mass distributions of the signal for each data taking period: A nominal one with the central values from Tab. 10.1. Furthermore, one for an upward change of each parameter and one for a downward change of each parameter. For the six latter ones, the normalisation factor is corrected according to the corresponding change of the inefficiency of the $\pm 20 \text{ MeV}/c^2$ mass window.

Same parameters for the two data taking periods are assumed to be maximally correlated. Different parameters are assumed to be uncorrelated. The mass shape parameters are assumed to be uncorrelated to all other systematic uncertainties.

10.2 $\mathcal{M}_{3\text{body}}$ calibration

For the usage of $\mathcal{M}_{3\text{body}}$ in the analysis, the fraction of expected $\tau \rightarrow \mu\mu\mu$ decays in each $\mathcal{M}_{3\text{body}}$ bin needs to be known. The fraction of $\tau \rightarrow \mu\mu\mu$ decays in a $\mathcal{M}_{3\text{body}}$ bin with bin boundaries $[a, b]$ is:

$$\varepsilon(\mathcal{M}_{3\text{body}} > a | \text{rec\&sel\&trig}) - \varepsilon(\mathcal{M}_{3\text{body}} > b | \text{rec\&sel\&trig}).$$

These efficiencies cannot be measured for $\tau \rightarrow \mu\mu\mu$ decays. Instead the measured efficiency for $D_s \rightarrow \phi(\mu\mu)\pi$ is used.

Differences between the $D_s \rightarrow \phi(\mu\mu)\pi$ and $\tau \rightarrow \mu\mu\mu$ response distribution arise from the following aspects:

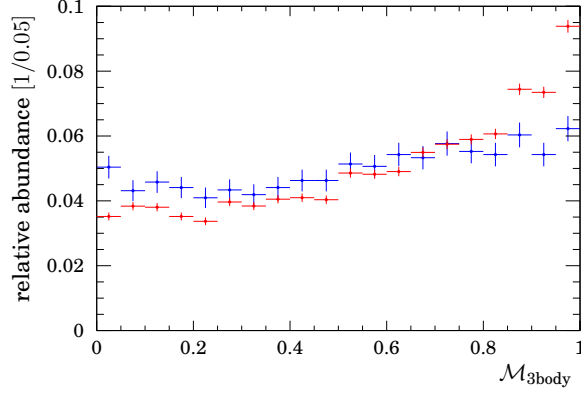


Figure 10.3: Comparison of the $\mathcal{M}_{3\text{body}}$ response in simulated events for $\tau \rightarrow \mu\mu\mu$ decays (blue) and $D_s \rightarrow \phi(\mu\mu)\pi$ decays (red). The responses of $D_s \rightarrow \phi(\mu\mu)\pi$ decays are slightly higher than $\tau \rightarrow \mu\mu\mu$ responses, as expected because D_s are pointing better to the primary interaction vertex.

- The ratio of prompt D_s and non-prompt D_s is the same as the ratio of τ leptons produced in the decays of prompt D_s and non-prompt D_s , but there is no production mechanism corresponding to $b \rightarrow \tau$ transitions.
- Prompt $D_s \rightarrow \phi(\mu\mu)\pi$ decays are perfectly pointing. The undetected neutrino in $D_s \rightarrow \tau\nu_\tau$ leads to a small direction angle. The measured direction angle distribution of $D_s \rightarrow \phi(\mu\mu)\pi$ decays will therefore be better concentrated at 0.
- Considering that a D_s meson can undergo a $\phi(\mu\mu)\pi$ decay or a $\tau(\mu\mu\mu)\nu_\tau$ decay, it is clear that the reconstructed flight distance² of a $\tau \rightarrow \mu\mu\mu$ decay is longer than that of a $D_s \rightarrow \phi(\mu\mu)\pi$ decay.

These differences are assumed to be correctly described in simulated events because they are consequences of the τ life time, the τ and D_s mass, and the $B \rightarrow \tau X$ decay kinematics. The response distributions for both decays in simulated events for the 2011 simulation are shown in Fig. 10.3. The efficiency ratio

$$\frac{\varepsilon_\tau(\mathcal{M}_{3\text{body}} > x)}{\varepsilon_{D_s}(\mathcal{M}_{3\text{body}} > x)}$$

is assumed to be correctly simulated. A calibration of $\mathcal{M}_{3\text{body}}$ is therefore obtained as a first-order correction to the observed $\mathcal{M}_{3\text{body}}$ response distribution of $D_s \rightarrow \phi(\mu\mu)\pi$ decays.

$$\varepsilon_\tau^{\text{data}}(\mathcal{M}_{3\text{body}} > x) = \frac{\varepsilon_\tau^{\text{sim}}(\mathcal{M}_{3\text{body}} > x)}{\varepsilon_{D_s}^{\text{sim}}(\mathcal{M}_{3\text{body}} > x)} \cdot \varepsilon_{D_s}^{\text{data}}(\mathcal{M}_{3\text{body}} > x) \quad (10.2)$$

²As in Chap. 9, the *reconstructed flight distance* is always the measured distance between the three-track vertex and the primary interaction vertex.

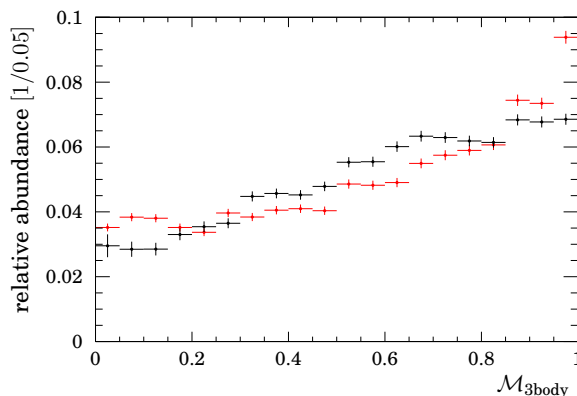


Figure 10.4: $\mathcal{M}_{3\text{body}}$ response distribution for $D_s \rightarrow \phi(\mu\mu)\pi$ decays in simulated events (red) compared to the sPlot'ed distribution in data (black). The former is identical to that shown in Fig. 10.3. The responses for simulated events slightly better than for real decays. This observed difference is used as correction to the simulated $\tau \rightarrow \mu\mu\mu$ response distribution.

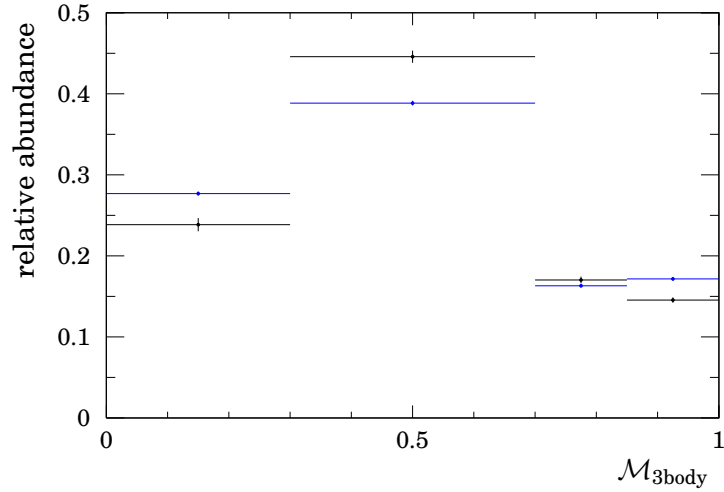
An alternate interpretation of the calibration formula is that if the response of D_s decays is correctly simulated, then also τ decays are correctly simulated. Consequently, if there is only a small difference between the $\mathcal{M}_{3\text{body}}$ response of D_s decays in data and the simulation, then a first-order correction to the simulated response distribution of τ decays is sufficient

$$\varepsilon_{\tau}^{\text{data}}(\mathcal{M}_{3\text{body}} > x) = \frac{\varepsilon_{D_s}^{\text{data}}(\mathcal{M}_{3\text{body}} > x)}{\varepsilon_{D_s}^{\text{sim}}(\mathcal{M}_{3\text{body}} > x)} \cdot \varepsilon_{\tau}^{\text{sim}}(\mathcal{M}_{3\text{body}} > x). \quad (10.2 \text{ rearranged})$$

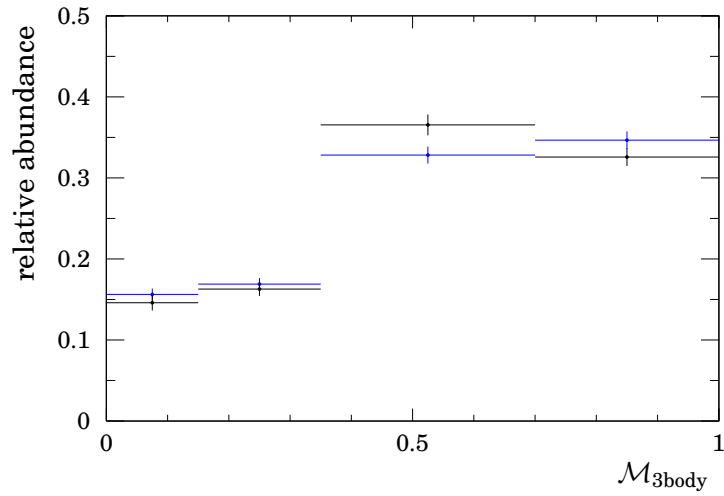
The $\mathcal{M}_{3\text{body}}$ distribution for $D_s \rightarrow \phi(\mu\mu)\pi$ in data is obtained with the sPlot technique (see Appendix B.4 or Ref. [164]). The invariant mass distribution of $D_s \rightarrow \phi(\mu\mu)\pi$ decays is fitted as in the calibration of the invariant mass shape. The sPlot'ed $\mathcal{M}_{3\text{body}}$ response distribution for $D_s \rightarrow \phi(\mu\mu)\pi$ in data is shown in Fig. 10.4 along with the response distribution for simulated $D_s \rightarrow \phi(\mu\mu)\pi$ decays.

The final calibrated $\mathcal{M}_{3\text{body}}$ response distribution, as expected for $\tau \rightarrow \mu\mu\mu$ decays in data, is shown in Fig. 10.5. The figure shows the response distribution in simulated events, too. The error bars indicate the statistical uncertainty.

Three contributions of statistical uncertainty are considered, the fit to the D_s data, the finite size of the simulated $D_s \rightarrow \phi(\mu\mu)\pi$ sample, and the finite size of the simulated $\tau \rightarrow \mu\mu\mu$ calibration sample. The size of the calibration sample is the dominant contribution to the calibration for the 2012 data. For the 2011 data, no classifier development is performed. There are therefore more simulated events for the simulated calibration sample available. For the 2011 calibration, the fit to the $D_s \rightarrow \phi(\mu\mu)\pi$ limits the statistic uncertainty.

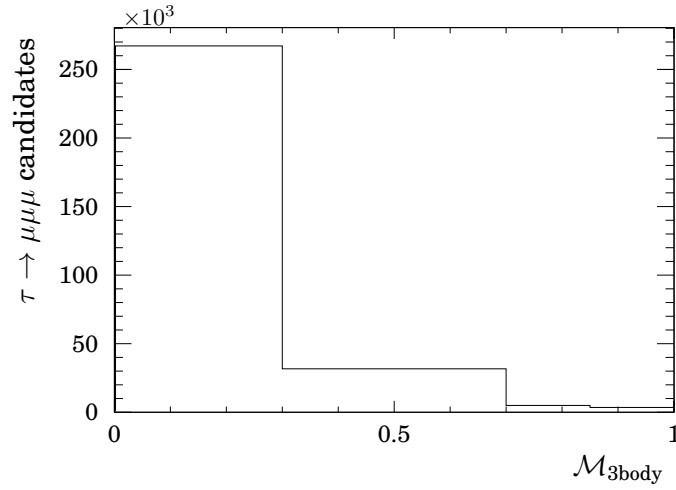


(a) 2011

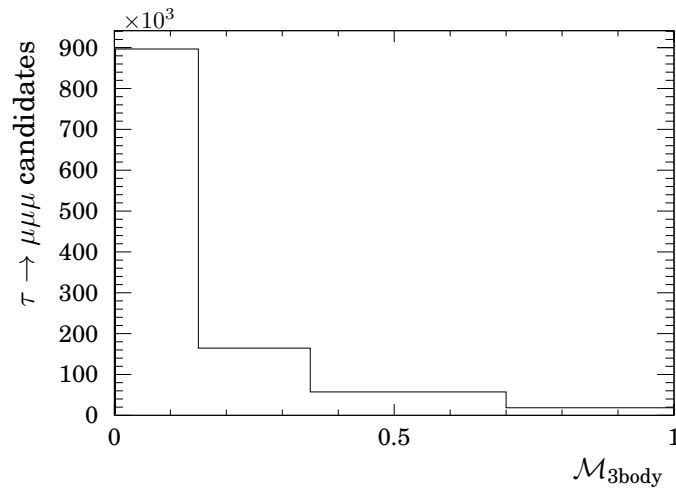


(b) 2012

Figure 10.5: Calibrated $\mathcal{M}_{3\text{body}}$ response in the binning used in the analysis (black) compared to the simulated response (blue). The error bars indicate the statistical uncertainty, an additional systematic uncertainty of 0.01 is assigned to each bin. The calibrated response distribution and the simulated response distribution do not agree, which is why the calibration is applied in the first place.



(a) 2011



(b) 2012

Figure 10.6: $\mathcal{M}_{3\text{body}}$ response distribution in data, i. e. the background distribution. Background candidates occur much more at small $\mathcal{M}_{3\text{body}}$ values than the signal, which is shown in Fig. 10.5.

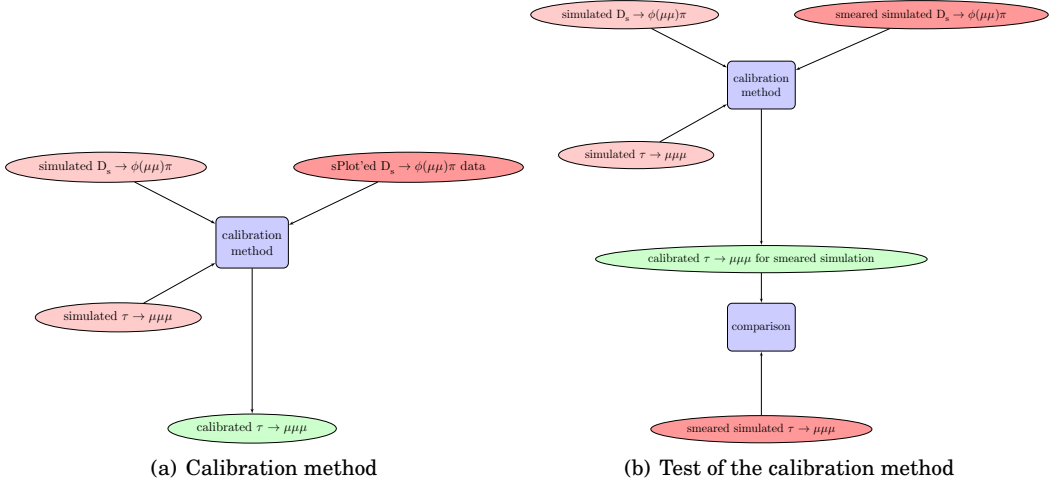


Figure 10.7: Illustration of the test of the calibration method. The smeared simulated $D_s \rightarrow \phi(\mu\mu)\pi$ is treated like the data in the actual calibration. The resulting calibrated $\tau \rightarrow \mu\mu\mu$ response distribution can be compared to the simulated $\tau \rightarrow \mu\mu\mu$ response distribution in the smeared simulation.

Systematic uncertainty

The calibration method is supposed to determine the $\tau \rightarrow \mu\mu\mu$ response distribution in an experiment which can measure the $D_s \rightarrow \phi(\mu\mu)\pi$ response, while the “true” $\tau \rightarrow \mu\mu\mu$ response behaviour is only known in a reference experiment (= the simulated experiment), in which the $D_s \rightarrow \phi(\mu\mu)\pi$ response can be measured, too. Whether the calibration method predicts $\tau \rightarrow \mu\mu\mu$ responses correctly is tested with a second simulated experiment, where the VELO resolution is altered. The VELO resolution is chosen as test parameter because the VELO resolution affects most of the input variables of $\mathcal{M}_{3\text{body}}$. The test is illustrated in Fig. 10.7.

The VELO resolution is worsened in simulated events by Gaussian smearing of the impact parameters of all reconstructed tracks, [154, Sect. 6.4]. This results in a degraded resolution of the track impact parameters, but also of the τ impact parameter and direction angle, the τ decay time, its flight distance, and the track distances of closest approach.

Using $D_s \rightarrow \phi(\mu\mu)\pi$ decays in smeared simulated events instead of real decays in the calibration method leads to a prediction for the $\tau \rightarrow \mu\mu\mu$ response distribution in smeared simulated events. The calibrated response distribution is compared to the simulated $\mathcal{M}_{3\text{body}}$ distribution of the simulated $\tau \rightarrow \mu\mu\mu$ decays. The efficiencies

$$\varepsilon_{\text{smear, sim}}(\mathcal{M}_{3\text{body}} > x) \quad \text{and} \quad \varepsilon_{\text{smear, calib}}(\mathcal{M}_{3\text{body}} > x)$$

are compatible within 1%, which is assigned as estimate for the systematic uncertainty of the calibration method.

Nuisance parameters for the limit calculation

For each of the data taking periods, the $\mathcal{M}_{3\text{body}}$ calibration determines the efficiency of three bin boundaries (for the 2011 data sample 0.30, 0.70, and 0.85; for the 2012 data sample 0.15, 0.35, and 0.70). The resulting nuisance parameters are changes of these efficiencies. They are considered uncorrelated among each other and uncorrelated among data taking periods. The systematic and statistical uncertainty for each efficiency are added in quadrature to result in one uncertainty on each efficiency.

10.3 \mathcal{M}_{PID} calibration

Calibration of multiple particle \mathcal{M}_{PID} efficiencies from single particle \mathcal{M}_{PID} efficiencies

The response distribution of \mathcal{M}_{PID} for a single muon depends on the particle's momentum, p , its pseudorapidity η , and the occupancy of the particle identification system. The latter is measured by the number of reconstructed charged particle tracks in the event, N_{tracks} .

$$\varepsilon_{\text{calib}}(\mathcal{M}_{\text{PID}}(\mu_i) > x) = \varepsilon_{\text{calib}}(p_i, \eta_i, N_{\text{tracks}}) \quad (10.3)$$

This dependency introduces a correlation between the responses for several muons from a decay. For *a single* given $\tau \rightarrow \mu\mu\mu$ decay with an event occupancy N_{tracks} , and the first muon's momentum p_1 and pseudorapidity η_1 , and correspondingly for the other muons, the probability of all three muons to have a response larger than x is therefore

$$\begin{aligned} \varepsilon_{\text{calib}}(\mathcal{M}_{\text{PID}}(\tau \rightarrow \mu_1\mu_2\mu_3) > x) = \\ \varepsilon_{\text{calib}}(p_1, \eta_1, N_{\text{tracks}}) \cdot \varepsilon_{\text{calib}}(p_2, \eta_2, N_{\text{tracks}}) \cdot \varepsilon_{\text{calib}}(p_3, \eta_3, N_{\text{tracks}}). \end{aligned} \quad (10.4)$$

Equation 10.4 can be applied for *a single* $\tau \rightarrow \mu\mu\mu$ decay with fixed $p_i, \eta_i, N_{\text{tracks}}$. The fraction ε of $\tau \rightarrow \mu\mu\mu$ decays *among the full sample* of $\tau \rightarrow \mu\mu\mu$ decays with $\mathcal{M}_{\text{PID}} > x$ is determined by summation over the simulated $\tau \rightarrow \mu\mu\mu$ decays.

$$\begin{aligned} \varepsilon_{\text{calib}}(\mathcal{M}_{\text{PID}}(\tau) > x) = \\ \frac{1}{N} \cdot \sum_{\text{simulation}} \left(\varepsilon_{\text{calib}}(p_1, \eta_1, N_{\text{tracks}}) \cdot \varepsilon_{\text{calib}}(p_2, \eta_2, N_{\text{tracks}}) \cdot \varepsilon_{\text{calib}}(p_3, \eta_3, N_{\text{tracks}}) \right), \end{aligned} \quad (10.5)$$

The factors $\varepsilon_{\text{calib}}(p_i, \eta_i, N_{\text{tracks}})$ remain to be determined in data, which is done in the following steps.

Calibration of single particle \mathcal{M}_{PID} efficiencies

A calibration, $\varepsilon_{\text{calib}}(p, \eta, N_{\text{tracks}})$, is provided by the charged particle identification group of LHCb, [85], in ranges of p, η , and N_{tracks} . A sample of $b \rightarrow J/\psi(\mu\mu)$ decays is selected and the \mathcal{M}_{PID} response distribution $\varrho(\mathcal{M}_{\text{PID}})$ is determined with the sPlot method (see Appendix B.4). When the J/ψ sample is restricted to a narrow range of muon momenta p , muon pseudorapidities η , and a narrow N_{tracks} range, then the single muon $\mathcal{M}_{\text{PID}} > x$ efficiency is an integral of the observed sPlot

$$\varepsilon_{\text{calib}}(p, \eta, N_{\text{tracks}}) = \int_x^1 \varrho(\mathcal{M}_{\text{PID}}) d\mathcal{M}_{\text{PID}}.$$

Table 10.2: Comparison of the calibrated \mathcal{M}_{PID} efficiency (using $b \rightarrow J/\psi(\mu\mu)$ decays) to the validation (using $D_s \rightarrow \phi(\mu\mu)\pi$ decays). The uncertainties are described in the text.

\mathcal{M}_{PID}	ε_{fit}	$\varepsilon_{\text{calib}}$	$\varepsilon_{\text{fit}}/\varepsilon_{\text{calib}} \equiv c$
2011			
0.15	0.98 ± 0.01	0.987 ± 0.001	0.99 ± 0.01
0.40	0.90 ± 0.01	0.909 ± 0.002	0.99 ± 0.01
0.80	0.47 ± 0.01	0.528 ± 0.002	0.89 ± 0.02
2012			
0.35	0.85 ± 0.01	0.927 ± 0.002	0.92 ± 0.01
0.65	0.66 ± 0.01	0.726 ± 0.002	0.91 ± 0.01
0.70	0.59 ± 0.01	0.672 ± 0.002	0.88 ± 0.02
0.85	0.37 ± 0.01	0.422 ± 0.002	0.88 ± 0.02

The classifiers \mathcal{M}_{PID} and $\mathcal{M}_{3\text{body}}$ have a small correlation because the τ transverse momentum is an input variable to $\mathcal{M}_{3\text{body}}$ and the \mathcal{M}_{PID} classification depends on the muon momentum. Therefore, \mathcal{M}_{PID} is calibrated in each $\mathcal{M}_{3\text{body}}$ bin individually. The sum in Eq. 10.5 then ranges only over simulated $\tau \rightarrow \mu\mu\mu$ candidates with $x_1 < \mathcal{M}_{3\text{body}} < x_2$. The calibrations of both classifiers are combined as:

$$\varepsilon_{\text{calib}}(x_1 < \mathcal{M}_{3\text{body}} < x_2 \wedge y_1 < \mathcal{M}_{\text{PID}} < y_2) = \quad (10.6)$$

$$\varepsilon_{\text{calib}}(x_1 < \mathcal{M}_{3\text{body}} < x_2) \cdot \varepsilon_{\text{calib}}\left(y_1 < \mathcal{M}_{\text{PID}} < y_2 \middle| x_1 < \mathcal{M}_{3\text{body, simulated}} < x_2\right)$$

Validation of the \mathcal{M}_{PID} calibration

To validate whether the \mathcal{M}_{PID} calibration using displaced $J/\psi \rightarrow \mu\mu$ decays is accurate for the expected $\tau \rightarrow \mu\mu\mu$ phase space coverage, the calibration is tested using $D_s \rightarrow \phi(\mu\mu)\pi$ decays as follows. For each $D_s \rightarrow \phi(\mu\mu)\pi$ decay, a Bernoulli³ distributed random number r is generated, and if $r = 0$ the positively charged muon is required to have $\mathcal{M}_{\text{PID}} > x$, otherwise the requirement is imposed on the negatively charged muon. The ‘‘average single-muon efficiency’’ is then the ratio

$$\varepsilon_{\text{fit}} := \frac{N(D_s \rightarrow \phi(\mu_1\mu_2)\pi | \mathcal{M}_{\text{PID}}(\mu_r) > x)}{N(D_s \rightarrow \phi(\mu_1\mu_2)\pi)}$$

The denominator is determined with the aforementioned fit of the invariant mass spectrum. The invariant mass spectrum of the D_s sample fulfilling the $\mathcal{M}_{\text{PID}}(\mu_r)$ selection is fitted with the same fit model as before to determine the numerator.

This efficiency can also be determined with the calibration method, $\varepsilon_{\text{calib}}(x)$, the resulting numbers are listed in Tab. 10.2.

The uncertainties in Tab. 10.2 are purely statistical for $\varepsilon_{\text{calib}}$. For ε_{fit} a conservative estimate for the total uncertainty is given: the statistical uncertainty is always found below 0.005, the systematic uncertainty due to the fit model is estimated by varying the fit range. The uncertainty due to the fit model cancels at first order in the ratio of the yields with and without the \mathcal{M}_{PID} requirement.

³with $p = 1/2$

The results in Tab. 10.2 imply that the standard calibration is not fully suited for the present analysis.

It has been investigated whether the difference between the two methods is momentum dependent. No significant trend, that the difference depends on p , p_T , or η is observed, limited by the accuracy of the fitted $D_s \rightarrow \phi(\mu\mu)\pi$ yield. Since no *trend* is observed, it is sufficient to apply a zeroth order correction (as opposed to e.g. a momentum dependent correction). In Eq. 10.4 the occurrence of $\varepsilon_{\text{calib}}(p, \eta, N_{\text{tracks}})$ is *replaced* by

$$\varepsilon_{\text{calib}}(p, \eta, N_{\text{tracks}}) \overset{\text{replace}}{\rightsquigarrow} \underbrace{\frac{\varepsilon_{\text{fit}}(x)}{\varepsilon_{\text{calib}}(x)}}_{=: c_x} \varepsilon(p, \eta, N_{\text{tracks}}). \quad (10.7)$$

The usage of a phase space averaged correction is motivated by the observation that the correlation of the two \mathcal{M}_{PID} responses in $D_s \rightarrow \phi(\mu\mu)\pi$ is small. This is observed as follows: the above fit is repeated after requiring *both* muons to fulfil the requirement $\mathcal{M}_{\text{PID}} > x$. The fraction of $D_s \rightarrow \phi(\mu\mu)\pi$ passing the selection is called $\varepsilon_{\text{fit}, 2}(x)$, i. e. it is the efficiency for both muons passing the \mathcal{M}_{PID} selection. Comparing $\varepsilon_{\text{fit}, 2}(x)$ to the expectation in absence of a correlation it is experimentally found that $\varepsilon_{\text{fit}, 2}(x)$ and $\varepsilon_{\text{fit}}^2(x)$ agree up to $\mathcal{O}(1\%)$.

There are three muons in the final state of $\tau \rightarrow \mu\mu\mu$, the final \mathcal{M}_{PID} calibration, Eq. 10.5, is therefore corrected by the third power of c_x

$$\varepsilon(\mathcal{M}_{\text{PID}}(\tau) > x) = \frac{c_x^3}{N} \cdot \sum_{\text{simulation}} \left(\varepsilon(p_1, \eta_1, N_{\text{tracks}}) \cdot \varepsilon(p_2, \eta_2, N_{\text{tracks}}) \cdot \varepsilon(p_3, \eta_3, N_{\text{tracks}}) \right). \quad (10.8)$$

The resulting \mathcal{M}_{PID} distributions are shown in Fig. 10.8, it shows how a $\tau \rightarrow \mu\mu\mu$ signal would be distributed over the \mathcal{M}_{PID} bins in data (if the branching fraction would be high enough to observe it); i. e. Fig. 10.8 is the \mathcal{M}_{PID} projection of the signal pdf.

Nuisance parameters for the limit calculation

The uncertainty on the \mathcal{M}_{PID} efficiencies are dominated by the uncertainty on c_x , i. e.:

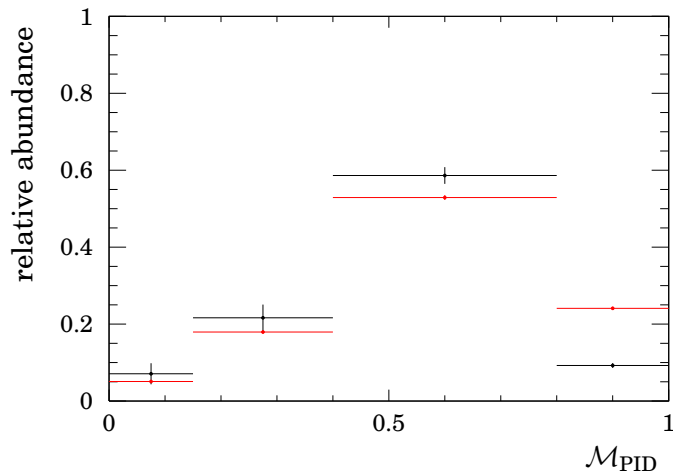
$$\sigma(\varepsilon(\mathcal{M}_{\text{PID}}(\tau) > x)) = 3 \cdot \sigma_{c_x}.$$

The efficiencies for different bin boundaries and different data taking periods are assumed to be uncorrelated. For each pseudo-experiment (PE) in the CLs method, and for each data taking period y , and each \mathcal{M}_{PID} cut value x , a Gaussian random number $\vartheta_{\mathcal{M}_{\text{PID}}, y, x}$ is generated; i. e. seven in total:

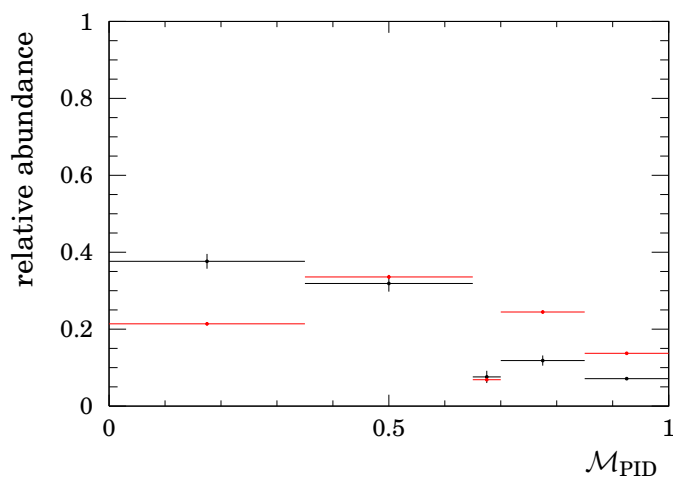
$$\begin{aligned} &\vartheta_{\mathcal{M}_{\text{PID}}, 2011, 0.15}, \vartheta_{\mathcal{M}_{\text{PID}}, 2011, 0.4}, \vartheta_{\mathcal{M}_{\text{PID}}, 2011, 0.8} \\ &\vartheta_{\mathcal{M}_{\text{PID}}, 2012, 0.35}, \vartheta_{\mathcal{M}_{\text{PID}}, 2012, 0.65}, \vartheta_{\mathcal{M}_{\text{PID}}, 2012, 0.7}, \vartheta_{\mathcal{M}_{\text{PID}}, 2012, 0.85} \end{aligned}$$

In the pseudo-experiment, the $\mathcal{M}_{\text{PID}} > x$ efficiencies are then changed by $\vartheta_{\mathcal{M}_{\text{PID}}, y, x}$ standard deviations from nominal value, i. e.:

$$\underbrace{\varepsilon(\mathcal{M}_{\text{PID}} > x)_{\text{PE}}}_{\text{used in PE generation}} = \underbrace{\varepsilon(\mathcal{M}_{\text{PID}} > x)}_{\text{nominal value}} + \underbrace{\vartheta_{\mathcal{M}_{\text{PID}}, y, x}}_{\text{random number}} \cdot \underbrace{\sigma(\varepsilon(\mathcal{M}_{\text{PID}} > x))}_{\text{fixed number from Tab. 10.2}}.$$

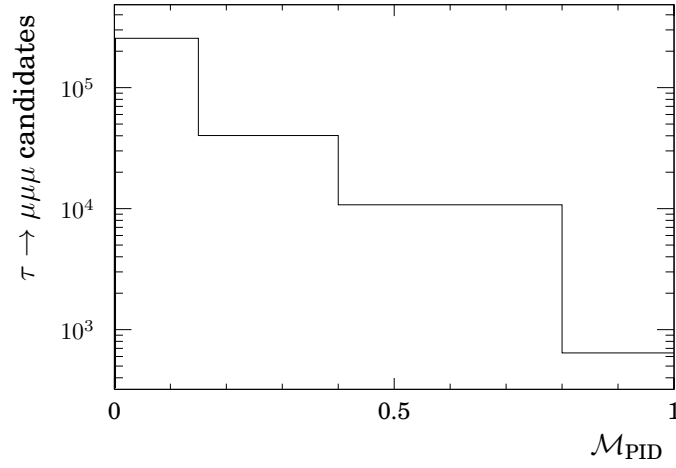


(a) 2011

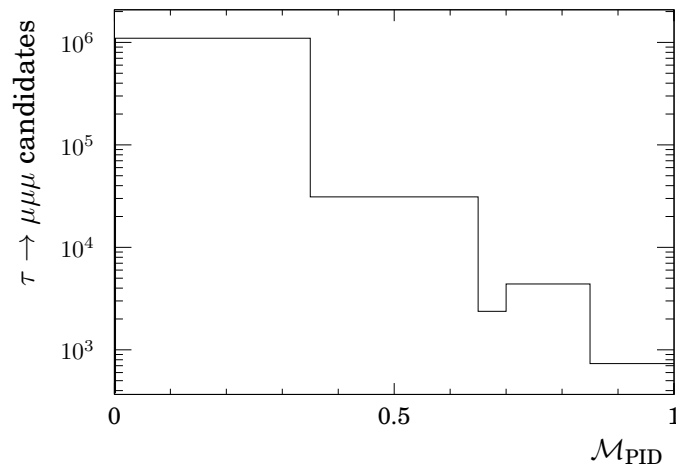


(b) 2012

Figure 10.8: Calibrated \mathcal{M}_{PID} response in the binning used in the analysis (black). It shows how the responses of real $\tau \rightarrow \mu\mu\mu$ decays would be distributed over the \mathcal{M}_{PID} bins in data, if there was a signal. The large visible difference between the two data taking periods is a feature of using different binnings in both years. For comparison the \mathcal{M}_{PID} response in simulated events without any correction is shown in red. The large discrepancy between the calibrated response and the simulated response is the reason for the necessity of a data-driven calibration. The distribution of background candidates is shown in Fig. 10.9 for comparison.



(a) 2011



(b) 2012

Figure 10.9: $\mathcal{M}_{\text{body}}$ response distribution in data, i.e. the background distribution. Background candidates occur much more at small \mathcal{M}_{PID} values than the signal, which is shown in Fig. 10.8 for comparison. A logarithmic scale is chosen because the high \mathcal{M}_{PID} bins would not be visible otherwise.

10.4 Role of the nuisance parameters in the limit computation

The uncertainties of the likelihood calibration are considered in the generation of pseudo-experiments for the CL_s computation, described in Sect. 3.6. The number of expected signal events in an analysis bin s_i is thereby the product of the total number of signal events under a given signal hypothesis $\mathcal{B}(\tau \rightarrow \mu\mu\mu) \cdot \alpha^{-1}$ times the fraction of signal events in that bin. The latter is related to the uncertainties determined in this chapter. For each pseudo experiment and each uncertain parameter, a Gaussian random number is generated to decide by how many standard deviations the parameter is fluctuated.

Example: The mean of the invariant mass distribution for data taken in 2011 is

$$\mu = (1779.1 \pm 0.1) \text{ MeV}/c^2.$$

In a pseudo-experiment the Gaussian random number ϑ_μ will be generated and μ is then assumed to be

$$\mu = (1779.1 + \vartheta_\mu \cdot 0.1) \text{ MeV}/c^2.$$

Similarly, the widths σ_1 and σ_2 are set to $\sigma_1 = (7.7 + \vartheta_{\sigma_1} \cdot 0.1) \text{ MeV}/c^2$ and $\sigma_2 = (12.0 + \vartheta_{\sigma_2} \cdot 0.8) \text{ MeV}/c^2$ correspondingly. The fraction of $\tau \rightarrow \mu\mu\mu$ decays in each mass bin is then computed for this distribution.

The nuisance parameters of the likelihood calibration are assumed to be uncorrelated among the data taking periods. Separate random numbers $\vec{\vartheta}_{2011}$ and $\vec{\vartheta}_{2012}$ are therefore generated.

Normalisation

The determination of the normalisation factor α is subject of this chapter. From the analysis strategy it is given by

$$\alpha^{-1} = \frac{N_{\text{obs}}(\text{D}_s \rightarrow \phi(\mu\mu)\pi)}{\mathcal{B}(\text{D}_s \rightarrow \phi(\mu\mu)\pi)} \frac{\mathcal{B}(\text{D}_s \rightarrow \tau\nu_\tau)}{f_\tau^{\text{incl.D}_s}} \frac{\varepsilon_\tau}{\varepsilon_{\text{D}_s}}. \quad (5.2 \text{ revisited})$$

The factors are determined one after the other in the following sections. Afterwards, the full normalisation factor is determined and its uncertainties are discussed.

11.1 $\text{D}_s \rightarrow \phi(\mu\mu)\pi$ yield and branching fraction

The yield of $\text{D}_s \rightarrow \phi(\mu\mu)\pi$ decays is determined with the fit to the invariant mass spectrum in Sect. 10.1. The integral of the signal component is¹

$$\begin{aligned} N_{\text{obs}, 2011}(\text{D}_s \rightarrow \phi(\mu\mu)\pi) &= (2.321 \pm 0.044) \times 10^4 \\ N_{\text{obs}, 2012}(\text{D}_s \rightarrow \phi(\mu\mu)\pi) &= (5.213 \pm 0.070) \times 10^4. \end{aligned}$$

The branching fraction $\mathcal{B}(\text{D}_s \rightarrow \phi(\mu\mu)\pi)$ is obtained by multiplying $\mathcal{B}(\text{D}_s \rightarrow \phi\pi)$ and $\mathcal{B}(\phi \rightarrow \mu\mu)$ from [80]. The former uses an amplitude analysis of $\text{D}_s^+ \rightarrow \text{K}^+\text{K}^-\pi^+$ to account for the interference of $\text{D}_s^+ \rightarrow \phi(\text{K}^+\text{K}^-\pi^+)$ with other processes, the most dominant of which is $\text{D}_s^+ \rightarrow \overline{\text{K}}^*(892)(\text{K}^-\pi^+)\text{K}^+$.

A non- ϕ -resonant contribution to $\mathcal{B}(\text{D}_s^+ \rightarrow \mu^+\mu^-\pi^+)$ is assumed to be negligible due to the exclusion limits on this contribution [165]. This leads to

$$\mathcal{B}(\text{D}_s \rightarrow \phi(\mu\mu)\pi) = (1.29 \pm 0.14) \times 10^{-5}.$$

11.2 τ production fractions and $\mathcal{B}(\text{D}_s \rightarrow \tau\nu_\tau)$

To determine the uncertainty on the τ production fractions, the correlated sources of uncertainties to the individual production modes need to be taken into account. The uncertain input quantities to the cross section computation are the branching fractions

- $\mathcal{B}(\text{D}_s \rightarrow \tau\nu_\tau) = 0.0543 \pm 0.0031$
- $\mathcal{B}(\text{D}^+ \rightarrow \tau\nu_\tau) = 0.00102 \pm 0.00009$

¹These have been determined by Jonathan Harrison in [38].

- $\mathcal{B}(b \rightarrow \tau X) = 0.0241 \pm 0.0023$
- $\mathcal{B}(b \rightarrow D_s X) = 0.248 \pm 0.037$
- $\mathcal{B}(b \rightarrow D^+ X) = 0.233 \pm 0.017$

and the cross sections

- $\sigma(\text{pp} \rightarrow D_s X) = (197 \pm 31) \mu\text{b} (7 \text{ TeV})$
- $\sigma(\text{pp} \rightarrow D_s X) = (225 \pm 35) \mu\text{b} (8 \text{ TeV})$
- $\sigma(\text{pp} \rightarrow D^+ X) = (645 \pm 74) \mu\text{b} (7 \text{ TeV})$
- $\sigma(\text{pp} \rightarrow D^+ X) = (737 \pm 85) \mu\text{b} (8 \text{ TeV})$
- $\sigma(\text{pp} \rightarrow bX) = (49 \pm 8) \mu\text{b} (7 \text{ TeV})$
- $\sigma(\text{pp} \rightarrow bX) = (55 \pm 7) \mu\text{b} (8 \text{ TeV})$.

The former are assumed uncorrelated. The latter have correlated sources of uncertainties.

Writing the fractions $\sigma(\text{pp} \rightarrow \tau X)_h$ as functions of these quantities, the cross sections can be rearranged to only contain the ratios $\sigma(\text{pp} \rightarrow D^+ X)/\sigma(\text{pp} \rightarrow D_s X)$ and $\sigma(\text{pp} \rightarrow bX)/\sigma(\text{pp} \rightarrow D_s X)$ and no other cross sections.

The ratio $\frac{\sigma(\text{pp} \rightarrow D_s X)}{\sigma(\text{pp} \rightarrow D^+ X)}$ is given in [129] under consideration of the correlation of the systematic uncertainties of the two individual measurements, it is (0.305 ± 0.042) .

The uncertainty on the ratio $\sigma(\text{pp} \rightarrow bX)/\sigma(\text{pp} \rightarrow D_s X)$ is estimated from the uncertainty budgets in the respective papers² as follows. The statistical uncertainties are uncorrelated. Since the luminosity measurement at LHCb is assumed to be dominated by systematic uncertainties [166], the uncertainties due to the luminosity measurement are assumed to be maximally correlated despite the fact that the measurements have been performed on disjoint data samples.

The systematic uncertainty due to the track finding efficiency is correlated as well. Considering the updated measurement of the track finding efficiency, [87], the systematic uncertainty of the ratio $\sigma(\text{pp} \rightarrow bX)/\sigma(\text{pp} \rightarrow D_s X)$ is reduced as follows. The uncertainty on the track reconstruction cancels twice in the ratio – the bottom cross section is measured in a two track final state, the charm cross sections are measured in a three track final state. – The uncertainty on the ratio of track reconstruction efficiencies is therefore *once* the uncertainty of the single track reconstruction efficiency.

The difference of the numbers of final state hadron tracks between the two measurements is three. Therefore, the uncertainty on the track reconstruction due to hadronic interactions with the detector material is considered three times in the ratio of cross section. These two effects add up to 2.9% uncertainty on the ratio due to the track reconstruction.

All other systematic uncertainties of the cross section measurements are assumed to be uncorrelated.

The covariance matrix for the production fractions is computed by first order error propagation as follows. The vector of production fractions is expressed as function of

²[129, 137, 134]

the uncertain input quantities:

$$\begin{pmatrix} f_\tau^{D_s} \\ f_\tau^{b \rightarrow D_s} \\ f_\tau^{D^+} \\ f_\tau^{b \rightarrow D_s} \\ f_\tau^b \end{pmatrix} =: \vec{f}_\tau \left(\mathcal{B}(D_s \rightarrow \tau), \mathcal{B}(D^+ \rightarrow \tau), \dots, \frac{\sigma(D^+)}{\sigma(D_s)}, \frac{\sigma(b)}{\sigma(D_s)} \right) =: \vec{f}_\tau(\vec{x})$$

The Jacobi matrix $A := \partial \vec{f} / \partial \vec{x}$ is determined with computer algebra software. The covariance matrix of \vec{x} is diagonal with the uncertainties' squares of the input quantities on the diagonal, up to one off-diagonal entry for the correlation of the cross section ratios³(see Appendix E).

$$\text{Cov}(\vec{x}) = \begin{pmatrix} \sigma_{\mathcal{B}(D_s \rightarrow \tau)}^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma_{\mathcal{B}(D^+ \rightarrow \tau)}^2 & & \vdots & \vdots \\ \vdots & & \ddots & 0 & 0 \\ 0 & \dots & 0 & \sigma^2 \left(\frac{\sigma(D^+)}{\sigma(D_s)} \right) & 0.0347 \\ 0 & \dots & 0 & 0.0347 & \sigma^2 \left(\frac{\sigma(b)}{\sigma(D_s)} \right) \end{pmatrix}$$

The covariance matrix $\text{Cov}(\vec{f}_\tau)$ is given by

$$\text{Cov}(\vec{f}_\tau) = A \text{Cov}(\vec{x}) A^T.$$

The correlation matrix the production fractions f_τ^h is computed using computer algebra software and given in Appendix E. The uncertainty on $\mathcal{B}(D_s \rightarrow \tau \nu_\tau)$ is hereby excluded since it appears a second time in α . It is addressed separately below. The tables in Appendix E refer to the τ production in the acceptance of the LHCb simulation, because the ‘‘mixture’’ at that stage is needed for the simulation. Converting to 4π the central values for $f_\tau^{\text{incl.}D_s}$ are found to be

$$\begin{aligned} f_\tau^{\text{incl.}D_s} &= 0.782 \pm 0.027 \quad \text{for } \sqrt{s} = 7 \text{ TeV} \\ f_\tau^{\text{incl.}D_s} &= 0.790 \pm 0.025 \quad \text{for } \sqrt{s} = 8 \text{ TeV}. \end{aligned}$$

The uncertainties will be re-addressed in Sect. 11.5.

The uncertainty on the branching fraction $\mathcal{B}(D_s \rightarrow \tau \nu_\tau)$ is a correlated source of uncertainty of $f_\tau^{\text{incl.}D_s}$ and the remaining factors of α . The uncertainty on $\mathcal{B}(D_s \rightarrow \tau \nu_\tau)$ is addressed by calculating three values of $f_\tau^{\text{incl.}D_s}$, once using the nominal values for all input quantities, once increasing $\mathcal{B}(D_s \rightarrow \tau \nu_\tau)$ by 1σ , and once decreasing $\mathcal{B}(D_s \rightarrow \tau \nu_\tau)$ by 1σ . Consequently, three values for α can be computed where the fluctuated value of $\mathcal{B}(D_s \rightarrow \tau \nu_\tau)$ and the correspondingly fluctuated value for $f_\tau^{\text{incl.}D_s}$ are used.

³0.0347 applies for $\sqrt{s} = 7$ TeV. For $\sqrt{s} = 8$ TeV the off-diagonal entry is 0.0248.

The ratios $\frac{\mathcal{B}(D_s \rightarrow \tau \nu_\tau)}{f_\tau^{\text{incl.}D_s}}$ and its uncertainty due to the $\mathcal{B}(D_s \rightarrow \tau \nu_\tau)$ are

$$\begin{aligned} \left(\frac{\mathcal{B}(D_s \rightarrow \tau \nu_\tau)}{f_\tau^{\text{incl.}D_s}} \right)_{7 \text{ TeV}} &= 0.0718^{+0.0022}_{-0.0014} (\text{from } \mathcal{B}(D_s \rightarrow \tau \nu_\tau)) \pm 0.0025 (\text{from before}) \\ \left(\frac{\mathcal{B}(D_s \rightarrow \tau \nu_\tau)}{f_\tau^{\text{incl.}D_s}} \right)_{8 \text{ TeV}} &= 0.0710^{+0.0021}_{-0.0013} (\text{from } \mathcal{B}(D_s \rightarrow \tau \nu_\tau)) \pm 0.0022 (\text{from before}). \end{aligned}$$

The uncertainty due to $\mathcal{B}(D_s \rightarrow \tau \nu_\tau)$ is maximally correlated among the data taking periods.

$D_s \rightarrow \phi(\mu\mu)\pi$ branching fraction: The uncertainty on the branching fraction $\mathcal{B}(D_s \rightarrow \phi\pi)$ appears to be a correlated source of uncertainty of the factor $\mathcal{B}(D_s \rightarrow \phi(\mu\mu)\pi)$ and $f_\tau^{\text{incl.}D_s}$ in the normalisation factor, because the D_s cross section measurement used the $D_s \rightarrow \phi\pi$ decay. It has however been confirmed that independent measurements of $\mathcal{B}(D_s \rightarrow \phi\pi)$ have been used in [129] and Sect. 11.1.

11.3 Efficiency ratio

The ratio of efficiencies is split into three factors. The probability that the decay of a D_s meson or of a τ lepton, which is produced at a collision at LHCb is within the detector acceptance, ε^{acc} ; the probability that a decay in the detector acceptance can be reconstructed and selected, ε^{rec} ; and the probability that a reconstructible decay fulfilling the selection requirements is accepted by the trigger, $\varepsilon^{\text{trig}}$.

11.3.1 Acceptance

The acceptance – the fraction of $\tau \rightarrow \mu\mu\mu$ decays where all three final state particle trajectories are within the sensitive region of the LHCb detector, and correspondingly for $D_s \rightarrow \Phi(\mu\mu)\pi$ – is evaluated in simulated events and the uncertainty includes the statistical uncertainty from the simulated sample size and is listed in Appendix E for the individual production modes. Considering the uncertainty of the production fractions, the acceptance ratio averages to

$$\begin{aligned} \left(\varepsilon_\tau^{\text{acc}} / \varepsilon_{D_s}^{\text{acc}} \right) &= 0.803 \pm 0.043 \quad \text{at 7 TeV} \\ \left(\varepsilon_\tau^{\text{acc}} / \varepsilon_{D_s}^{\text{acc}} \right) &= 0.799 \pm 0.038 \quad \text{at 8 TeV.} \end{aligned}$$

11.3.2 Reconstruction and selection

The combined reconstruction and selection efficiency, excluding the selection cuts on $\mathcal{M}_{3\text{body}}$ and \mathcal{M}_{PID} selection, are determined using simulated events. Three corrections are applied to the efficiency ratio due to a possibly incorrect description of the track reconstruction efficiency, the muon pre-selection efficiency, and the ϕ resonance description in the simulation. These corrections are described in the following paragraphs. Before applying the corrections, the efficiency ratios are

$$\begin{aligned} \left(\varepsilon_\tau^{\text{rec}} / \varepsilon_{D_s}^{\text{rec}} \right) &= 1.381 \pm 0.006 \quad \text{at 7 TeV} \\ \left(\varepsilon_\tau^{\text{rec}} / \varepsilon_{D_s}^{\text{rec}} \right) &= 1.384 \pm 0.006 \quad \text{at 8 TeV.} \end{aligned}$$

The uncertainty stems from the association of reconstructed objects to simulated objects⁴: The reconstructed $\tau \rightarrow \mu\mu\mu$ candidates in simulated events contain a small amount of background from the underlying events. Therefore counting the number of selected candidates in the simulation cannot lead to a correct efficiency determination.

The invariant mass spectrum of the simulated events is fitted with the same method as in Sect. 11.1 to determine the $\tau \rightarrow \mu\mu\mu$ and $D_s \rightarrow \Phi(\mu\mu)\pi$ yield in the simulated events.

Track reconstruction efficiency correction

The ratio of the track reconstruction efficiency in simulated events and real events is measured as a function of the track momentum and pseudorapidity with the tag-and-probe method (see Appendix C.1). This correction factor is applied to every simulated $\tau \rightarrow \mu\mu\mu$ and $D_s \rightarrow \phi(\mu\mu)\pi$ decay. This results in a correction factor

$$c^{\text{track}} := \frac{\frac{\varepsilon_{\tau}^{\text{track}}(\text{data})}{\varepsilon_{D_s}^{\text{track}}(\text{data})}}{\frac{\varepsilon_{\tau}^{\text{track}}(\text{sim})}{\varepsilon_{D_s}^{\text{track}}(\text{sim})}},$$

to correct the ratio of reconstruction efficiencies observed in the simulation to the expected ratio of reconstruction efficiencies in data:

$$\begin{aligned} c^{\text{track}} &= 0.997 \pm 0.009 \quad \text{for 2011 data} \\ c^{\text{track}} &= 0.996 \pm 0.009 \quad \text{for 2012 data} \end{aligned}$$

where the uncertainty is the statistical uncertainty from [87]. The uncertainty on the hadronic interaction length for the pion in the $D_s \rightarrow \phi\pi$ decay contributes an additional uncertainty of 0.026 to these factors.

The measurement of the track reconstruction efficiency has not been performed within the scope of the present analysis. The efficiency ratio (real data over simulation) is taken from [87].

Muon identification efficiency correction

The performance of \mathcal{M}_{PID} is measured in Sect. 10.3. The muon pre-selection is addressed here. The efficiency of the pre-selection is measured by reconstructing $J/\psi \rightarrow \mu\mu$ decays in data without imposing muon identification requirements on one of the muons (a tag-and-probe method as in Appendix C.1). The “success rate” of the muon pre-selection on this sample is then pre-selection efficiency. To account for possible background candidates in the $J/\psi \rightarrow \mu\mu$ sample, the invariant mass spectrum is fitted before and after applying the muon pre-selection – which yields the number of true $J/\psi \rightarrow \mu\mu$ decays.

The measurement has been conducted by the muon identification group [93]. The ratio of the measured efficiency in data and in simulated events is used as a per-track correction factor to the reconstruction efficiency observed in simulated decays. In the case of $\tau \rightarrow \mu\mu\mu$ the correction is applied for all three tracks, for $D_s \rightarrow \phi(\mu\mu)\pi$ it is applied for the two muon tracks.

⁴The explanation is written for $\tau \rightarrow \mu\mu\mu$ but applies to $D_s \rightarrow \phi\pi$ in the same manner.

The correction to the τ -to- D_s efficiency ratio is

$$c^{\mu \text{ id}} := \frac{\frac{\varepsilon_{\tau}^{\mu \text{ id}}(\text{data})}{\varepsilon_{D_s}^{\mu \text{ id}}(\text{data})}}{\frac{\varepsilon_{\tau}^{\mu \text{ id}}(\text{sim})}{\varepsilon_{D_s}^{\mu \text{ id}}(\text{sim})}}.$$

It is found to be

$$\begin{aligned} c^{\mu \text{ id}} &= 0.973 \pm 0.003 \pm 0.026 \quad \text{for 2011 data} \\ c^{\mu \text{ id}} &= 1.007 \pm 0.002 \pm 0.020 \quad \text{for 2012 data.} \end{aligned}$$

where the first uncertainty is statistical and the second is systematic. The systematic uncertainty is the difference between the actual efficiency of the muon pre-selection in simulated events (in simulated events, it is known if a track is a muon) and the efficiency measured with the tag-and-probe method.

Correction of the simulated ϕ resonance

The efficiency of the $\phi \rightarrow \mu\mu$ selection in the $D_s \rightarrow \phi(\mu\mu)\pi$ decay, based on the mass cut ($\pm 20 \text{ MeV}/c^2$ around the nominal value), is not correctly simulated because the $\phi \rightarrow \mu\mu$ resonance is not simulated according to the $\phi \rightarrow \mu\mu$ line shape. The simulated line shape is truncated at $988 \text{ MeV}/c^2$ and $1085 \text{ MeV}/c^2$ and includes a turn-on at the lower end, as it would be expected for a two kaon final state. The integral of a non-relativistic Breit-Wigner for this process⁵ outside of the simulated range is computed to be 0.028. The integral of the simulated line shape within $\pm 20 \text{ MeV}/c^2$ of the ϕ mass is therefore 0.934 instead of the simulated value of 0.962. The $D_s \rightarrow \phi(\mu\mu)\pi$ selection efficiency is therefore decreased by $\sim 3\%$ with respect to the simulation.

It may seem arguable whether the ϕ line shape extends into the range between twice the strange quark mass and twice the muon mass. Additionally, the range of small invariant muon masses is subject to interference with other resonances. As an alternative line shape model, the non-relativistic Breit-Wigner is truncated at the kaon threshold without the kinematic turn-on. This leads to a smaller correction of $\sim 1\%$.

The correction to the simulated efficiency of the di-muon mass requirement is therefore taken to be 0.98 ± 0.01 , the mean of the two corrections and either difference to the mean as uncertainty.

11.3.3 Trigger

The trigger efficiencies are determined in simulated events. To evaluate the uncertainty, the TISTOS method (Appendix C.2) is applied to $D_s \rightarrow \phi(\mu\mu)\pi$ in simulated and real events. The statistical uncertainty of this method is large. The difference $\varepsilon_{D_s}(\text{TISTOS, data}) - \varepsilon_{D_s}(\text{TISTOS, simulation})$ of 5.2% is significant due to the poor sensitivity of the TISTOS method. As a conservative estimate, these 5.2% are assigned as uncertainty to the trigger efficiency ratio.

$$\begin{aligned} \left(\varepsilon_{\tau}^{\text{trig}} / \varepsilon_{D_s}^{\text{trig}} \right) &= 0.6593 \pm 0.0058 \quad \text{at 7 TeV} \\ \left(\varepsilon_{\tau}^{\text{trig}} / \varepsilon_{D_s}^{\text{trig}} \right) &= 0.525 \pm 0.040 \quad \text{at 8 TeV} \end{aligned}$$

⁵I. e. assuming that the ϕ resonance extends below twice the strange quark mass!

The efficiency ratio has been computed by Jonathan Harrison in [38].

11.4 Total normalisation factor α

The full normalisation factor is given in Tab. 11.1. Those sources of uncertainty of the normalisation factor which are uncorrelated to the distribution of $\tau \rightarrow \mu\mu\mu$ decays over the analysis bins, are added in quadrature and called $\sigma_{\alpha\text{-normalisation}}$. These are

- $\mathcal{B}(D_s \rightarrow \phi(\mu\mu)\pi)$
- the number of observed $D_s \rightarrow \phi(\mu\mu)\pi$ decays
- the correction to the track reconstruction efficiency
- the correction to the first muon pre-selection
- the efficiency ratio, except for the uncertainties which are correlated to the signal calibration, as listed in the following section.

The major difference between the two data taking periods are the integrated luminosities, which leads to a larger $N_{\text{obs}}(D_s \rightarrow \phi(\mu\mu)\pi)$ in 2012.

Table 11.1: Factors of the normalisation factor. Uncertainties marked with * are correlated to the likelihood calibration and *not* considered in the uncertainty of the product, $\sigma_{\alpha\text{-normalisation}}$. They are re-addressed in Sect. 11.5. The second column indicates if the quantity is in the numerator or denominator of α .

		2011	2012
$f_{\tau}^{\text{incl.}D_s}$	num	$0.782 \pm 0.027^*$	$0.790 \pm 0.025^*$
$\mathcal{B}(D_s \rightarrow \phi(\mu\mu)\pi)$	num	$(1.29 \pm 0.14) \times 10^{-5}$	
$\mathcal{B}(D_s \rightarrow \tau\nu_{\tau})$	den	$0.0561 \pm 0.0024^*$	
$\varepsilon_{D_s}/\varepsilon_{\tau}$	num	0.574 ± 0.039	0.47 ± 0.11
likelihood calibrations (mass, \mathcal{M}_{PID} , $\mathcal{M}_{\text{3body}}$)	den	$0.68 \pm 0.02^*$	$0.49 \pm 0.02^*$
$N_{\text{obs}}(D_s \rightarrow \phi(\mu\mu)\pi)$	den	$(2.321 \pm 0.044) \times 10^4$	$(5.213 \pm 0.070) \times 10^4$
$\alpha \pm \sigma_{\alpha\text{-normalisation}}$		$(5.36 \pm 0.69) \times 10^{-9}$	$(3.32 \pm 0.51) \times 10^{-9}$

11.5 Correlated uncertainties of likelihood calibration and normalisation

Some sources of uncertainties affect the calibrated likelihoods as well as the normalisation. The most simple one is the invariant mass shape. A change of the invariant mass shape results in a changed distribution of the signal candidates over the mass bins, but also changes the selection efficiency of $|m - m_{\tau}| < 20 \text{ MeV}/c^2$. As described in Sect. 10.4, the three nuisance parameters (NP) of the invariant mass shape are μ , σ_1 , and σ_2 . For variations of each of them by one standard deviation, a “fluctuated”

normalisation factor is computed:

$$\begin{aligned}
 \alpha_n &:= \alpha(\text{NP at their nominal values}) \\
 \alpha_{\mu+} &:= \alpha(\mu \text{ increased by } 1\sigma) \\
 \alpha_{\mu-} &:= \alpha(\mu \text{ decreased by } 1\sigma) \\
 &\vdots
 \end{aligned}$$

and correspondingly for $\alpha_{\sigma_1+}, \alpha_{\sigma_1-}, \alpha_{\sigma_2+}, \alpha_{\sigma_2-}$. The normalisation factor in a pseudo-experiment (PE) with nuisance parameters (NP) μ, σ_1 , and σ_2 is then computed in first order approximation from the *relative* changes of α

$$\begin{aligned}
 \delta_{\mu+} &:= (\alpha_{\mu+} - \alpha_n) / \alpha_n \\
 \delta_{\mu-} &:= (\alpha_n - \alpha_{\mu-}) / \alpha_n \\
 &\vdots
 \end{aligned}$$

and correspondingly for $\delta_{\sigma_1+}, \delta_{\sigma_1-}, \delta_{\sigma_2+}, \delta_{\sigma_2-}$. For each NP a Gaussian (mean at 0 and width of 1) random number ϑ is generated and the randomised normalisation factor is:

$$\begin{aligned}
 \alpha_{\text{PE}} &:= \alpha_n \cdot \left(1 + \vartheta_\mu \cdot \begin{cases} \delta_{\mu+} & \text{if } \vartheta_\mu > 0 \\ \delta_{\mu-} & \text{if } \vartheta_\mu < 0 \end{cases} \right. \\
 &\quad + \vartheta_{\sigma_1} \cdot \begin{cases} \delta_{\sigma_1+} & \text{if } \vartheta_{\sigma_1} > 0 \\ \delta_{\sigma_1-} & \text{if } \vartheta_{\sigma_1} < 0 \end{cases} \\
 &\quad \left. + \vartheta_{\sigma_2} \cdot \begin{cases} \delta_{\sigma_2+} & \text{if } \vartheta_{\sigma_2} > 0 \\ \delta_{\sigma_2-} & \text{if } \vartheta_{\sigma_2} < 0 \end{cases} \right) \quad (11.1)
 \end{aligned}$$

For the τ production fractions a similar method is applied: For each principle component, PC_i , of the production fraction covariance matrix, a new simulated $\tau \rightarrow \mu\mu\mu$ calibration sample is generated (see Sect. 9.2), where the production fractions are changed by one standard deviation of PC_i . With this altered calibration sample, the calibration of \mathcal{M}_{PID} and $\mathcal{M}_{\text{3body}}$ are repeated and an altered distribution of $\tau \rightarrow \mu\mu\mu$ over the multivariate bins is obtained. The fraction $f_\tau^{\text{incl.D}_s}$ is also recomputed for a 1σ variation of PC_i , and the relative change of α is determined, called δ_{PC_i} .

In the pseudo-experiment generation, a Gaussian random number ϑ_i is generated for each of the principle components PC_i . In this pseudo-experiment, α is then

$$\alpha_{\text{PE}} := \alpha_n \left(1 + \sum_{i \text{ NP}} \vartheta_i \cdot \begin{cases} \delta_{i+} & \text{if } \vartheta_i > 0 \\ \delta_{i-} & \text{if } \vartheta_i < 0 \end{cases} \right), \quad (11.2)$$

where the sum ranges over μ, σ_1, σ_2 , and the PC_i .

The number of $\tau \rightarrow \mu\mu\mu$ decays in each analysis bin s_i in the pseudo-experiment is, in the same manner, the interpolation between the nominal fraction, $s_i(\text{nom})$, and the varied fraction, $s_i(\text{PC}_i)$, with the *same* parameter ϑ_i :

$$s_i = s_i(\text{nom}) \sum_i \vartheta_i \left(s_i(\text{PC}_i) - s_i(\text{nom}) \right).$$

11.5. Correlated uncertainties of likelihood calibration and normalisation

Table 11.2: List of nuisance parameters which influence both, the distribution of a signal over the analysis bins and the normalisation factor. The relative change of the normalisation factor under 1σ fluctuations of the parameter are assumed to be maximally correlated among the two data taking periods (except for the the \mathcal{M} selection requirements).

nuisance parameter	$\delta+$	$\delta-$	$\delta+$	$\delta-$
	2011 [%]		2012 [%]	
mass parameter μ	0.032	0.037	0.035	0.040
mass parameter σ_1	0.071	0.039	0.079	0.084
mass parameter σ_2	0.38	0.43	0.061	0.071
$\mathcal{B}(D_s \rightarrow \tau\nu_\tau)$	5.0	5.5	3.0	3.6
τ production fraction, first principle component	0.14	2.6	0.44	1.0
τ production fraction, second principle component	1.9	1.8	1.5	2.2
τ production fraction, third principle component	0.73	0.41	1.2	0.80
τ production fraction, fourth principle component	0.093	0.083	0.20	0.25
$\mathcal{M}_{3\text{body}} > 0.30$ (2011 data)	1.3	1.4		
$\mathcal{M}_{3\text{body}} > 0.15$ (2012 data)			1.7	1.8
$\mathcal{M}_{\text{PID}} > 0.15$ (2011 data)	3.1	3.1		
$\mathcal{M}_{\text{PID}} > 0.35$ (2012 data)			3.3	3.3

The uncertainty on $\mathcal{B}(D_s \rightarrow \tau\nu_\tau)$ is treated in the same way as the PC_i . The relative changes of α for all nuisance parameters are listed in Tab. 11.2. As stated above, the principle components listed in Tabs. E.6 and E.7 have the same physics behaviour and are assumed to be maximally correlated for the different data taking periods, i. e. in each pseudo-experiment the *same* value ϑ_i is used for both data taking periods.

Adding the uncertainties which do not alter the distribution of $\tau \rightarrow \mu\mu\mu$ decays over the likelihood bins (i. e. the uncertainty quoted in Tab. 11.1) to Eq. 11.2 leads to

$$\alpha_{\text{PE}} := \alpha_n \left(1 + \sum_{i \text{ NP}} \vartheta_i \cdot \begin{cases} \delta_{i+} & \text{if } \vartheta_i > 0 \\ \delta_{i-} & \text{if } \vartheta_i < 0 \end{cases} \right) + \vartheta_{\text{normalisation}} \sigma_{\alpha\text{-normalisation}} \cdot \quad (11.3)$$

Limits for the $\tau \rightarrow \mu\mu\mu$ decay

12.1 Inspection of the m_τ mass region

The invariant mass spectrum is fitted in the invariant mass sidebands. The integral of the fitted probability density function in the signal region is the expected background yield in the signal region. The fit function describes two background processes, the $D^+ \rightarrow K^-\pi^+\pi^+$ background and the combinatorial spectrum. All other processes are sufficiently suppressed by not using the lowest bin of \mathcal{M}_{PID} .

The combinatorial background is described by an exponential function.

The $D^+ \rightarrow K^-\pi^+\pi^+$ contribution is described by a Crystal Ball function. As an example, the bin with $0.7 < \mathcal{M}_{\text{3body}} < 0.85$ and $\mathcal{M}_{\text{PID}} < 0.15$ in the 2011 data is shown in Fig. 12.1. This bin is not used in the analysis but the shape of the $D^+ \rightarrow K^-\pi^+\pi^+$ is best visible.

The numbers of expected background events in the different multivariate likelihood bins are listed in Tab. 12.1 and 12.2, where the invariant mass bins are integrated out. The uncertainties are the statistical uncertainties from the fit to the sidebands. The highest likelihood bin¹ for each data taking period is shown in Fig. 12.2 for both data taking periods. The binning in the figure is chosen to coincide with the mass bin boundaries and the border of the outer and inner sidebands. The statistical uncertainty on the fit result is visualised as the green band in the fit region. Within

¹See the footnote on page 126 for an explanation of this tautological term.

Table 12.1: Expected and observed background event yield in the multivariate likelihood bins for 2011 data.

\mathcal{M}_{PID} bin	$\mathcal{M}_{\text{3body}}$ bin	$N(\text{expected})$	$N(\text{observed})$	fit quality χ^2/ndf
0.15–0.4	0.3–0.7	212.4 ± 0.2	200	1.6
0.15–0.4	0.7–0.85	24.8 ± 2.7	20	0.4
0.15–0.4	0.85–1	21.1 ± 1.8	16	0.7
0.4–0.8	0.3–0.7	71.1 ± 3.1	76	0.8
0.4–0.8	0.7–0.85	12.6 ± 1.3	14	0.4
0.4–0.8	0.85–1	5.6 ± 0.9	7	1.7
0.8–1	0.3–0.7	6.3 ± 0.9	7	1.5
0.8–1	0.7–0.85	1.8 ± 0.5	4	2.1
0.8–1	0.85–1	1.1 ± 0.4	1	2.2

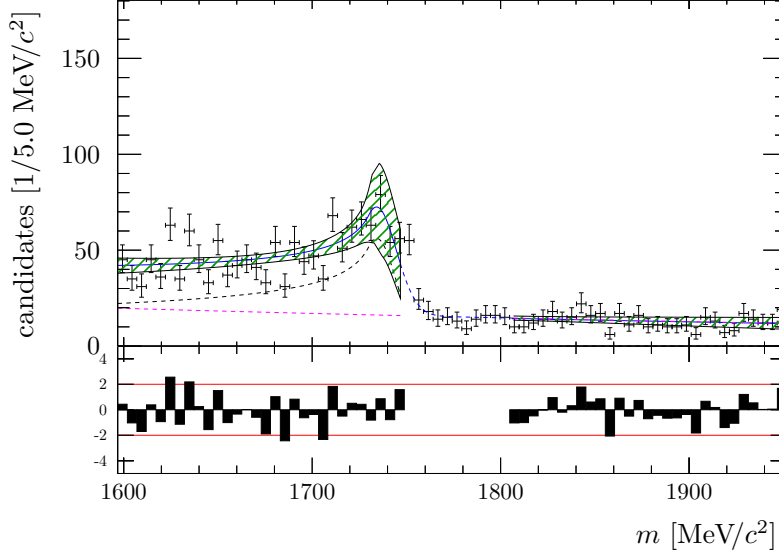
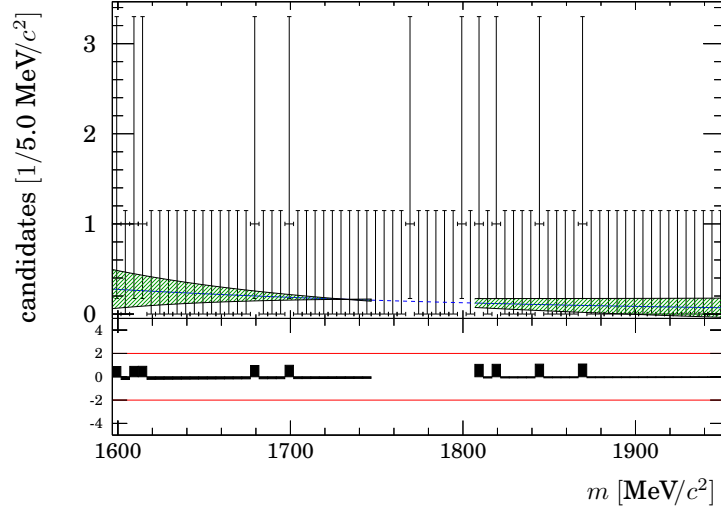


Figure 12.1: The 2011 data with $0.7 < \mathcal{M}_{3\text{body}} < 0.85$ and $\mathcal{M}_{\text{PID}} < 0.15$, fitted with the background fit model: a combinatorial contribution in dashed magenta and a Crystal ball function for the $D^+ \rightarrow K^- \pi^+ \pi^+$ in dashed black. The combined fit function is shown in solid blue, all three curves are drawn in the fit range. The interpolated combined fit function is shown in dashed blue in the signal region. The green band indicates the 1σ uncertainty band on the fitted function. The lower part of the figure shows the pulls – the difference between the (integrated) fit function and the observed event count per bin divided by the statistical uncertainty – red lines indicate 2σ deviations.

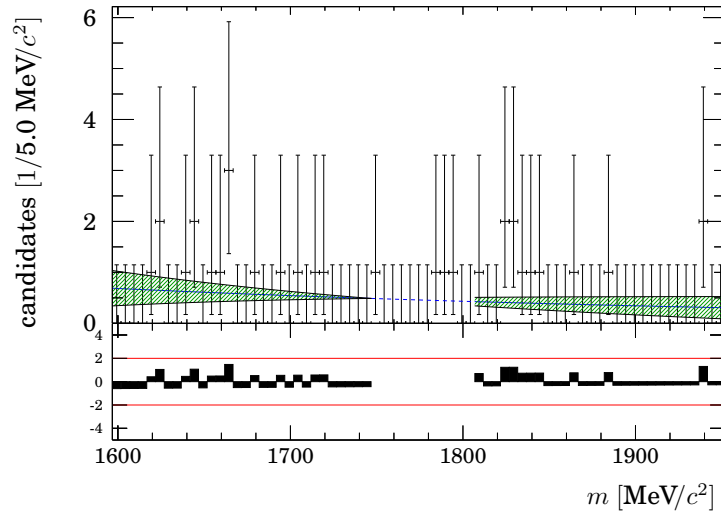
the range which is not considered in the fit, i. e. inner sidebands and signal region, the fit result is shown as dashed line. The inner sidebands are the outermost two bins on each side of that range, the innermost 8 bins are the signal region.

The exponential model for the combinatorial background is found compatible with a linear function, which has been tested by repeating the fit to the sidebands with a linear function instead of the exponential function. In a second test, the range in which the data is fitted has been increased to include the inner sidebands, too. The expected background yields found with the increased fit region are compatible with the nominal fit range, while a bias towards lower values due to the optimisation procedure cannot be ruled out. No systematic uncertainty is assigned due to these effects.

The uncertainties on the expected background yields are assumed uncorrelated for different multivariate likelihood bins and maximally correlated for different invariant mass bins in the same multivariate likelihood bin.



(a) 2011



(b) 2012

Figure 12.2: The invariant mass spectra for the candidates in the highest multivariate likelihood bins for both data taking periods. The fit function is shown in solid blue in the fit range. The interpolated fit function is shown in dashed blue in the signal region. The green band indicates the 1σ uncertainty band on the fitted function. The lower part of the figures shows the pulls – the difference between the (integrated) fit function and the observed event count per bin divided by the statistical uncertainty – red lines indicate 2σ deviations.

Table 12.2: Expected and observed background event yield in the multivariate likelihood bins for 2012 data.

\mathcal{M}_{PID} bin	$\mathcal{M}_{\text{3body}}$ bin	$N(\text{expected})$	$N(\text{observed})$	fit quality χ^2/ndf
0.35 – 0.65	0.15 – 0.35	336.1 ± 7.3	349	1.3
0.35 – 0.65	0.35 – 0.7	122.6 ± 4.1	115	1.5
0.35 – 0.65	0.7 – 1	37.4 ± 2.3	33	1.0
0.65 – 0.7	0.15 – 0.35	28.7 ± 2.0	25	0.6
0.65 – 0.7	0.35 – 0.7	6.8 ± 2.7	8	0.5
0.65 – 0.7	0.7 – 1	2.4 ± 1.4	4	1.4
0.7 – 0.85	0.15 – 0.35	50.1 ± 2.6	54	0.7
0.7 – 0.85	0.35 – 0.7	25.5 ± 1.9	18	0.9
0.7 – 0.85	0.7 – 1	10.7 ± 1.3	5	1.5
0.85 – 1	0.15 – 0.35	11.0 ± 1.2	8	0.7
0.85 – 1	0.35 – 0.7	7.6 ± 1.0	6	0.2
0.85 – 1	0.7 – 1	3.7 ± 0.7	3	1.8

In the pseudo-experiment generation, the uncertainties on the expected yield of background events is considered as nuisance parameter. I. e. for each likelihood bin i , a Gaussian random number ϑ_i is generated. Assume, that the estimation of the expected number of background events in the eight mass bins are $b_{i,j} \pm \sigma_{b,i,j}$. In this pseudo experiment the background hypothesis is then set to a mean number of $b_{i,j} + \vartheta_i \cdot \sigma_{b,i,j}$ in likelihood bin i and mass bin j . The truncation, mentioned in Sect. 3.6, is applied by repeating the random number generation if the result is negative. The number of observed background events in the pseudo-experiment in each bin is then a Poissonian random number with mean $b_{i,j} + \vartheta_i \cdot \sigma_{b,i,j}$.

12.2 Upper limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$

The numbers of expected events from the background fit and the numbers of observed events in the different likelihood bins are given in Tabs. 12.1 and 12.2. The uncertainties are purely the uncertainties on the estimated mean of the event count, an additional Poissonian fluctuation is expected. The background-only expectation is in agreement with the observation, especially considering Poissonian fluctuations of the observed events around the expectation.

The number of expected signal decays for a given branching fraction hypothesis is computed using the normalisation factor from Chap. 11 with

$$N(\tau \rightarrow \mu\mu\mu) = \frac{\mathcal{B}(\tau \rightarrow \mu\mu\mu)}{\alpha}.$$

The distribution of the signal over the analysis bins, as determined in Chap. 10.

The CLs method is applied as explained in Sect. 3.7 to determine the sensitivity: For branching fraction hypotheses \mathcal{B} from 1×10^{-8} to 5×10^{-6} the test statistics Q is defined as

$$Q_{\text{obs}} = -2 \ln \frac{\prod_i \frac{(s_i + b_i)^{n_i}}{n_i!} e^{-(s_i + b_i)}}{\prod_i \frac{(b_i)^{n_i}}{n_i!} e^{-b_i}},$$

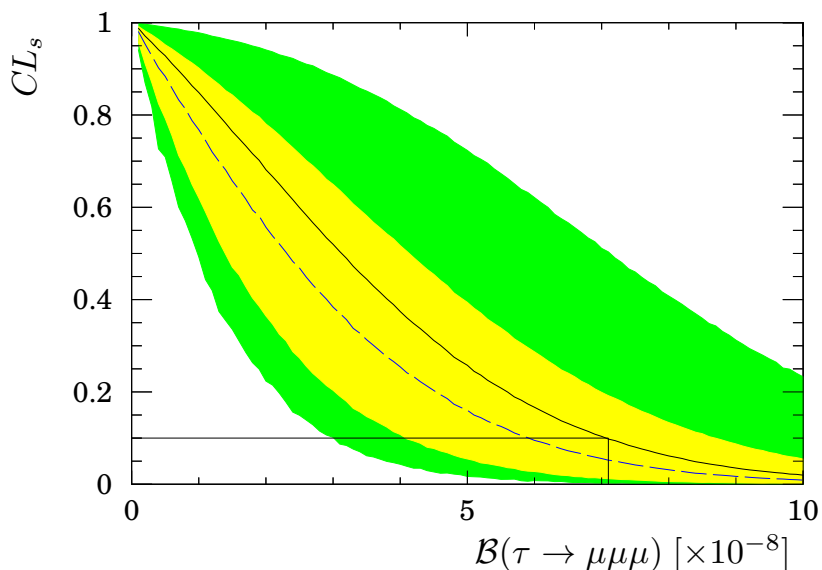


Figure 12.3: CL_s as a function of the assumed branching fraction. The solid black line is the observed CL_s value. The blue dashed line, the yellow band, and the green band are the median, the central 68% coverage interval, and the central 95% coverage interval of the CL_s distribution under the background hypothesis. The observed CL_s is well within the expected range, i.e. the data is compatible with the background hypothesis. At $\mathcal{B}(\tau \rightarrow \mu\mu\mu) = 7.1 \times 10^{-8}$, the observed CL_s value drops below 0.1, i.e. branching fractions above 7.1×10^{-8} are excluded at 90% confidence level.

where the product ranges over all bins. Nuisance parameters do not occur in the test statistics, i.e. they are not floating in the test statistics but fixed to their nominal values instead. The expected Q distribution for both hypotheses (only background or background with the presence of an additional signal of \mathcal{B}) is determined with pseudo-experiments. In each pseudo-experiment, a Gaussian random number ϑ_j is generated for each nuisance parameter to determine the value of the nuisance parameter in the pseudo-experiment, i.e. the nuisance parameters are not fixed to their nominal values in the pseudo-experiment generation but randomised instead.

The expected and observed CL_s values are then used to decide if the branching fraction hypothesis is expected to be rejected (expected $CL_s < 0.1$) and if the branching fraction hypothesis is actually rejected (observed $CL_s < 0.1$).

At 90% confidence level, the exclusion limit is expected to be within 4.1×10^{-8} and 8.7×10^{-8} , with a central value of 5.9×10^{-8} . It is observed that $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is smaller than 7.1×10^{-8} at 90% confidence level, well in agreement with the expectation. Figure 12.3 shows the observed CL_s values as a function of $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$, as well as the range in which CL_s is expected.

The nuisance parameters, which are randomised in the pseudo experiment generation, are listed in Tab. 12.3.

Table 12.3: Nuisance parameters (NP) of the pseudo experiment generation. The categories are those used in Tab. 12.4.

nuisance parameter (NP)	N parameters 2011/2012	reference	category
invariant mass shape $(\mu, \sigma_1, \sigma_2)$	3	Sect. 10.1	signal NP
$\mathcal{M}_{3\text{body}}$	3	Sect. 10.2	signal NP
\mathcal{M}_{PID}	3/4	Sect. 10.3	signal NP
τ production fractions	4	Sect. 11.2	signal NP
$\mathcal{B}(D_s \rightarrow \tau\nu)$	1	Sect. 11.2	signal NP
normalisation	1	Sect. 11.4	α -normalisation NP
background expectation	9/12	Sect. 12.1	background NP

12.3 Influence of nuisance parameters

If all nuisance parameters are fixed to their nominal values in the pseudo-experiment generation the expected sensitivity “improves” to 5.6×10^{-8} . Only some of the nuisance parameters are randomised in each scenario in Tab. 12.4, the other nuisance parameters remain fixed to their nominal values, i. e. ϑ_j is set to zero in all pseudo-experiments instead of to a Gaussian random number. The three scenarios are

background fluctuating only the expected number of background events in the pseudo-experiment generation,

signal fluctuating only the signal distribution over the analysis bins in the pseudo-experiment generation,

α -normalisation fluctuating only the the normalisation factor, for the sources of systematic uncertainty on the normalisation factor which are uncorrelated to the signal distribution.

The nuisance parameters of the signal distribution and the normalisation do not alter the expected sensitivity significantly. The uncertainty on the background estimation degrades the sensitivity. There is no need to improve the likelihood calibration methods beyond its current accuracy, but the analysis would profit from an improved background estimation.

12.4 Review of the binned analysis procedure

The searches for $\tau \rightarrow \mu\mu\mu$ in [3, 167] appear simpler than the present work by avoiding a binned event classification. Both analyses apply an event selection without event classification. The present analysis can thus not be directly compared to these analyses in terms of numbers of background events, produced τ leptons, and selection efficiencies: each of them is a single number in [3] and [167] but 21 numbers in the present work. A curious analyst is therefore interested in the sensitivity gain by performing the presented binned search, whilst complicating the comparison to other results.

For simplicity, the nuisance parameters are fixed to their nominal values in the following studies. Furthermore, only the binning in the multivariate classifiers is

Table 12.4: Influence of the nuisance parameters (NP) on the sensitivity quantified by the expected exclusion limit. The sensitivities for the individual data taking periods are given as well, the influence of separate nuisance parameter categories is not evaluated for the separate data taking periods.

randomised nuisance parameter (NP)	sensitivity both years [10^{-8}]	sensitivity 2011 [10^{-8}]	sensitivity 2012 [10^{-8}]
no NP randomised	5.5	8.9	7.2
only signal NP randomised	5.6	-	-
only α -normalisation NP randomised	5.6	-	-
only background NP randomised	5.8	-	-
all NP randomised	5.9	9.4	8.2
observed	7.1	14.5	7.3

investigated, the nominal binning in the invariant mass is used. The binned analysis of the data taken in 2012 has an expected sensitivity of 7.2×10^{-8} , see Tab. 12.4.

The highest likelihood bin is the bin with the highest (multivariate) likelihood values². As illustrated in Fig. 12.4, the highest $\mathcal{M}_{3\text{body}}$ bin is the bin with the highest $\mathcal{M}_{3\text{body}}$ values; for the 2011 data this is the bin $0.85 < \mathcal{M}_{3\text{body}} \leq 1.0$; for the 2012 data this is the bin $0.70 < \mathcal{M}_{3\text{body}} \leq 1.0$. Correspondingly, the highest \mathcal{M}_{PID} bin is the bin $0.80 < \mathcal{M}_{\text{PID}} \leq 1.0$ for 2011 and $0.85 < \mathcal{M}_{\text{PID}} \leq 1.0$ for 2012. The highest bin in the two-dimensional likelihood binning is consequently the bin with

$$0.85 < \mathcal{M}_{3\text{body}} \leq 1.0 \quad \text{and} \quad 0.80 < \mathcal{M}_{\text{PID}} \leq 1.0$$

²This tautology is a definition and not a finding of statistics.

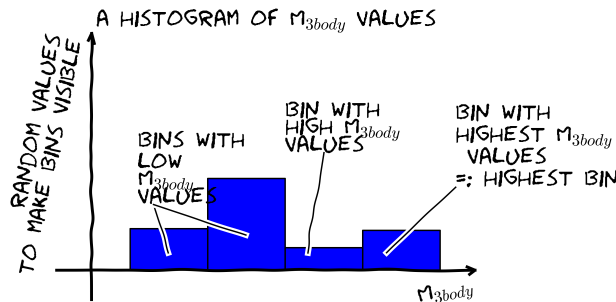


Figure 12.4: Illustration of the term “highest bin” for a binned quantity. The example shows a binning in $\mathcal{M}_{3\text{body}}$ – $\mathcal{M}_{3\text{body}}$ is shown on the x -axis – Whether a bin is a “high” $\mathcal{M}_{3\text{body}}$ bin is thus independent of the y value – these are random numbers here to achieve a visibility of the bins. The $\mathcal{M}_{3\text{body}}$ values increase from left to right, the highest $\mathcal{M}_{3\text{body}}$ values are therefore in the bin which is furthest to the right.

12. LIMITS FOR THE $\tau \rightarrow \mu\mu\mu$ DECAY

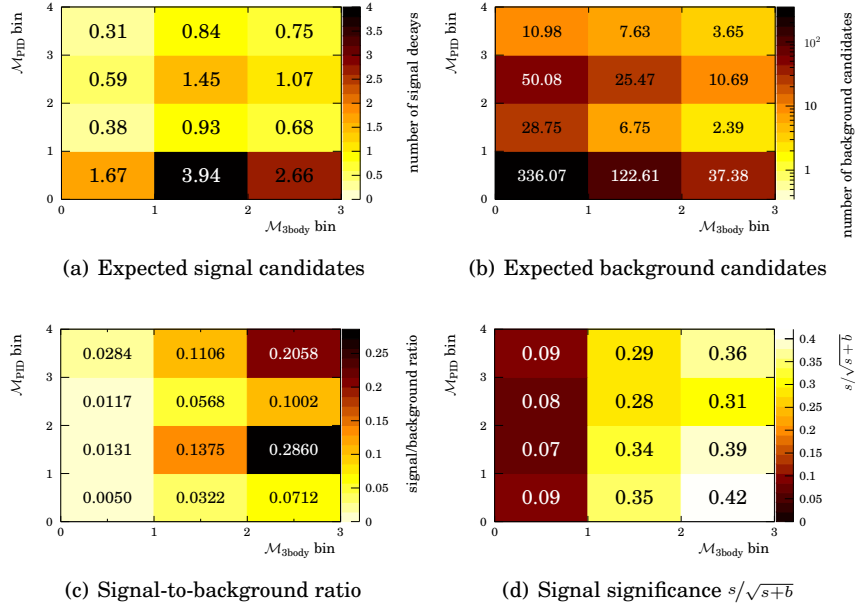


Figure 12.5: Quantification of analysing only a single analysis bin for the 2012 data sample. The signal hypothesis in (a), (c), and (d) is the expected sensitivity of the nominal analysis with fixed nuisance parameters. (a): shows the expected number of signal candidates, the projections are shown in Figs. 10.5 (b) and 10.8 (b), up to an overall scaling due to the signal hypothesis. The pattern is a consequence of the bin widths. (b): the number of expected background candidates, as listed in Tab. 12.2, (c)&(d): the signal to background ratio and the signal significance resulting from (a) and (b).

for 2011, and

$$0.70 < \mathcal{M}_{3\text{body}} \leq 1.0 \quad \text{and} \quad 0.85 < \mathcal{M}_{\text{PID}} \leq 1.0$$

for 2012. The highest bin is commonly regarded as most interesting and relevant, and therefore picked as example illustration (e. g. [168]). In the work of [127], it is actually shown that the highest likelihood bin of the $B_s^0 \rightarrow \mu\mu$ analysis is also the best bin (defined as below). The expected exclusion limit using only the highest likelihood bin in 2012 is 2.5×10^{-7} , as indicated by the colour of the top right bin in Fig. 12.6.

The best bin is defined to be the analysis bin which provides the best sensitivity if the decay $\tau \rightarrow \mu\mu\mu$ is sought for in this bin alone. The best bin is not necessarily the highest bin³.

For each analysis bin, the intermediate results of the nominal analysis are re-used: the background estimate from Sect. 12.1 for that bin is used as background

³Since the binning is done in a multivariate likelihood, “highest bin” is the natural abbreviation for “the bin which is the highest on the likelihood axis” or “the bin which comprises the highest likelihood values”.

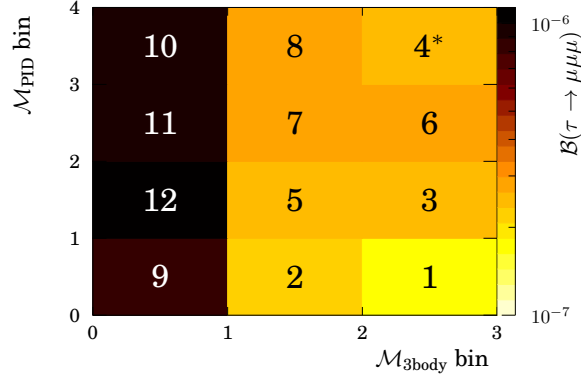


Figure 12.6: The expected sensitivity on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ when analysing only a single analysis bin of the 2012 data sample. The colour indicates the expected limit, the numbers indicate the ranking of the bins, 1 being the bin with the best sensitivity; this best bin is on the bottom right, with $0.7 < \mathcal{M}_{3\text{body}} \leq 1.0$ and $0.35 < \mathcal{M}_{\text{PID}} < 0.65$. This bin has also the best signal significance. The ranking of signal significances is similar but not identical to the sensitivity ranking. The highest likelihood bin is the bin on the top right, marked by *.

prediction. The number of expected signal decays (for a given $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ hypothesis) is determined with the normalisation factor (Chap. 11) and the fraction of signal decays in that bin, given by the calibration (Chap. 10).

As normalisation factor for the single bin analyses, the nominal normalisation factor, corrected by the fraction of the signal decays in that bin, is used. The number of expected signal candidates for $\mathcal{B}(\tau \rightarrow \mu\mu\mu) = 5.0 \times 10^{-8}$, the number of expected background candidates are shown in Fig. 12.5. The signal-to-background ratio as well as $s/\sqrt{s+b}$ are shown in Fig. 12.5 along.

The resulting “single bin sensitivities” are illustrated in Fig. 12.6. It is visible, that the most sensitive bin is not the bin with the highest signal likelihood, marked with a * in Fig. 12.6. This best analysis bin alone achieves an expected sensitivity of 1.8×10^{-7} . Furthermore, comparing Fig. 12.6 to Fig. 12.5 (c) and (d), it becomes visible as well that s/b is a poor measure of the sensitivity of a search while the best $s/\sqrt{s+b}$ bin has the best sensitivity. It must still be noticed that $s/\sqrt{s+b}$ does not correctly reproduce the sensitivity ordering: the second largest $s/\sqrt{s+b}$ value is 0.39 in the highest $\mathcal{M}_{3\text{body}}$ and second \mathcal{M}_{PID} bin, although this bin has only the third best sensitivity, indicated in Fig. 12.6.

The best bin is $0.7 < \mathcal{M}_{3\text{body}} \leq 1.0 \wedge 0.35 < \mathcal{M}_{\text{PID}} \leq 0.65$. This, however would not be chosen as a selection for an unbinned analysis, since the requirement $\mathcal{M}_{\text{PID}} \leq 0.65$ is non-physical. A remedy is the best *possible* bin, as described in the following paragraph.

The best possible bin is an attempt to emulate an unbinned analysis which uses both multivariate classifiers. The selection requirement $\mathcal{M}_{3\text{body}} > x \wedge \mathcal{M}_{\text{PID}} > y$ is

12. LIMITS FOR THE $\tau \rightarrow \mu\mu\mu$ DECAY

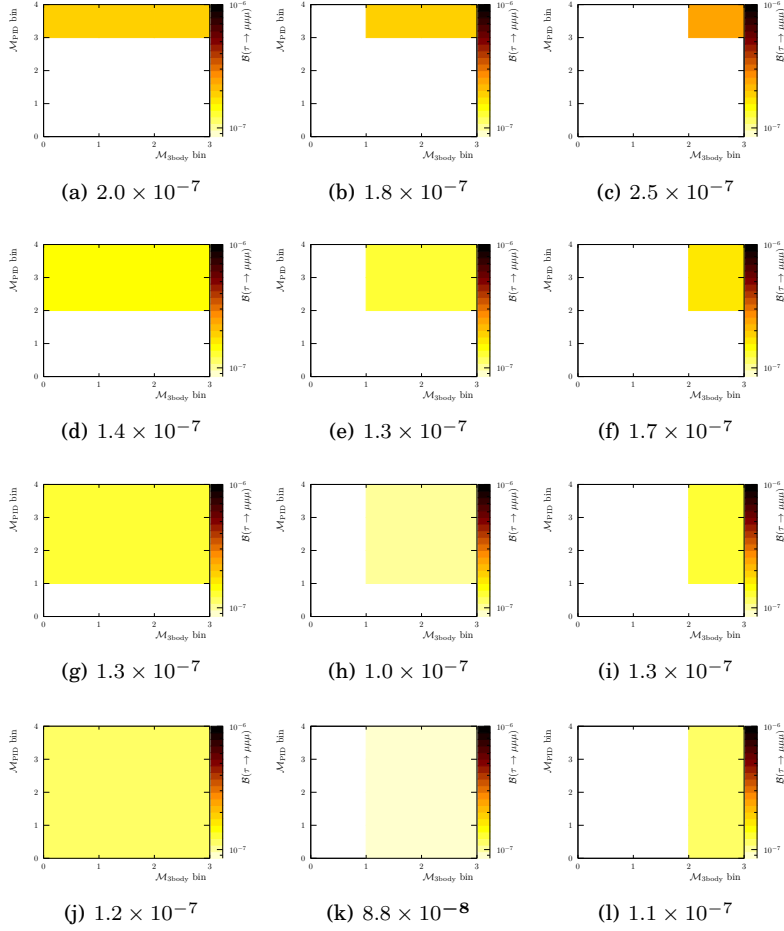


Figure 12.7: Expected sensitivities for different single-bin scenarios. The colour indicates the range in $\mathcal{M}_{3\text{body}}$ and \mathcal{M}_{PID} which is analysed. The same colour scale as in Fig. 12.7 is used in all figures.

emulated by adding bins. Then the sensitivity for an analysis using this selection is determined by applying the CLs method with one analysis bin.

The resulting “merged bin sensitivities” are illustrated in Fig. 12.7, where the coloured area shows which bins are merged and the colour indicates the sensitivity. The best possible bin reaches an expected sensitivity of 8.8×10^{-8} , resulting from 47.5 expected background events and one expected signal decay at a branching fraction hypothesis of 1.2×10^{-8} .

Comparing this sensitivity of 8.8×10^{-8} to the sensitivity of the nominal analysis with the 2012 data, 7.2×10^{-8} , shows that the binned analysis gains approximately 18% in sensitivity. Due to the limitation of only testing selections at bin boundaries, 18% is an upper estimate for the improvement.

The best single measurement by the Belle collaboration [3] expects 0.13 ± 0.06

background events, which is to be compared to 216.7 ± 6.2 expected background events in the best possible bin (summing the expected numbers of background events from Tab. 12.2). The number of expected signal events in [3] for a branching fraction of $\mathcal{B}(\tau \rightarrow \mu\mu\mu) = 5.0 \times 10^{-8}$ is 2.1, which is to be compared 12.3 in the best possible bin for the same branching fraction hypothesis (summing the merged bin contents in Fig. 12.5 (a)).

12.5 Comparison to other measurements of $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$

Comparison to the published result of LHCb

The sensitivity of the present analysis is similar to the sensitivity of the best previous experiments, [3, 167]. A study by [169] of the data used in [4] showed that the sensitivity of LHCb is expected to improve with

$$\text{Limit} \left(\mathcal{B}(\tau \rightarrow \mu\mu\mu) \right) \propto \frac{\sqrt{\sigma_{\text{bkg}} \cdot \mathcal{L}}}{\sigma_{\tau} \cdot \mathcal{L}} \approx \frac{1}{\sqrt{\sigma \mathcal{L}}} \quad (12.1)$$

if the analysis remains unchanged. LHCb will continue operations in 2015 and the inclusive τ production cross section is expected to increase due to a higher centre-of-mass energy of $\sqrt{s} = 13$ TeV. Since τ leptons are produced in bottom and charm decays, and since the background processes are also bottom and charm decays, the cross section for background will increase in the same way. A correction to the latter assumption might be necessary because the cross section for combinatoric background candidates can receive an enhancement due to a higher detector occupancy (the particle density is increased at a higher centre-of-mass energy). Additionally, the detector occupancy depends on the number of primary collisions at a bunch crossing (pile-up) and the detector's capabilities to distinguish consecutive bunch crossings (spill-over; it is planned to operate the LHC at a 25 ns bunch spacing in the future).

A small deviation from Eq. 12.1 can be expected from an updated classifier binning, as shown in Appendix A, the optimal binning is a function of the integrated luminosity.

This work analysed the same data, with the same event selection as the analysis published by the LHCb collaboration in [5]. The present work finds

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < (5.9_{-1.8}^{+2.8}) \times 10^{-8} \quad \text{expected}$$

while [5] reports

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < (5.0_{-1.6}^{+2.4}) \times 10^{-8} \quad \text{expected}$$

at 90% confidence level. The differences between the published result and this analysis shall shortly be listed here. A summary is given in Tab. 12.5.

τ production: The τ production fractions have been re-computed here with respect to [5]. The numeric values differ slightly due to different rounding precisions. The differences are much smaller than the assigned uncertainties. The uncertainties in the present work are evaluated by means of the principle components of the covariance matrix. In the analysis in [5] the uncertainties of the τ production fractions are considered uncorrelated⁴. The method of principle components

⁴up to the overall normalisation $\sum f = 1$.

is more rigorous. However, given that the signal nuisance parameter do not alter the sensitivity in either this work or [5], the τ production cannot account for the different results.

The τ production fractions enter the computation of the expected and the observed limit in the same manner.

Binning: The present work used the observed event count in the inner sidebands in the optimisation of the bin boundaries. In [5] a parametrisation of the distribution of background events in both classifiers has been developed; based on the work by [127]. Ultimately, both methods are limited by the small event count in the inner sidebands for high likelihood values: the method presented here is sensitive to statistical fluctuations. This is acceptable because the sensitivity does not depend strongly on deviations from the optimal boundary position, as shown in Appendix A. The method in [5] requires a parametrisation of the $\mathcal{M}_{3\text{body}}$ and \mathcal{M}_{PID} distribution for the background events in the inner sidebands and this parametrisation needs to be developed on same small event count in the inner sidebands with high likelihood values.

Any effect of the different binning optimisation enter the computation of the expected and the observed limit in the same manner.

$\mathcal{M}_{3\text{body}}$ technique: An ensemble selection [9, 10] of multivariate classifiers, based on the TMVA classifiers developed within this work, is used in [5]. Twice as many simulated events of signal and background have been available for its training. The improvement is estimated to be around 6 % in the final sensitivity, [8], combined with the binning. This 6 % could only be evaluated once the $\tau \rightarrow \mu\mu\mu$ sidebands (the invariant mass region in which the background fit is performed) have been unblinded.

Table 12.5: Breakdown of differences between [5] and the present work and their impact on the expected exclusion limit, except for values marked with *, which contributes only to the difference of the observed exclusion limits. Values marked with † are only known after the unblinding of the calibration data, values marked with ‡ are only known after the unblinding of the signal region, values marked with ✕ are only known after the unblinding of the background region.

	sensitivity difference
τ production	$\pm 0\%$
treatment of the uncertainties of the τ production	$\pm 0\%$
$\mathcal{M}_{3\text{body}}$ classifier and binning prior / after background unblinding	$\mathcal{O}(5\%) / 6\% \text{✕}$
$\mathcal{M}_{3\text{body}}$ calibration i. e. difference in data/simulation differences	$11\% \text{†✕}$
$\mathcal{B}(D_s \rightarrow \phi(\mu\mu)\pi)$	-2%
Poissonian fluctuation of observed background around expected background	$22\% \text{*†✕}$
Σ expected / observed	$15\% \text{†✕} / 35\% \text{*†✕}$

12.5. Comparison to other measurements of $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$

The event classification with $\mathcal{M}_{3\text{body}}$ affects the computation of the expected and the observed limit in the same manner.

Normalisation: The normalisation in [5] uses the $D_s \rightarrow \phi(\mu\mu)\pi$ branching fraction measurement from [170], this compares to the present as follows:

$$\begin{aligned} \mathcal{B}(D_s \rightarrow \phi(\mu\mu)\pi) &= (1.29 \pm 0.14) \times 10^{-5} && \text{present work} \\ \mathcal{B}(D_s \rightarrow \phi(\mu\mu)\pi) &= (1.32 \pm 0.10) \times 10^{-5} && [5]. \end{aligned}$$

This factor in the normalisation factor α applies both, to the expected and the observed limit, linearly.

$\mathcal{M}_{3\text{body}}$ calibration: In the classifier development, based on simulated events, the performance of $\mathcal{M}_{3\text{body}}$ in the present work and in [5] are rather similar (6%, see above). The difference between the expected exclusion limits of [5] and the present analysis is partially due to this. The remaining difference is introduced by the calibration – the difference between the $\mathcal{M}_{3\text{body}}$ performance on $D_s \rightarrow \phi(\mu\mu)\pi$ decays in simulated and real events is larger in the present work than in [5]. It is unknown why one of the two $\mathcal{M}_{3\text{body}}$ versions is more vulnerable to data/simulation differences.

The calibration is an unblinding of the calibration data. The classifier choice must not be done due to the calibrated performance since changing the analysis after the unblinding is forbidden.

The calibration is used in the computation of the expected and the observed limit.

In the present analysis and in [5], the difference between the expected and observed limits are well compatible with Poissonian fluctuations. The present analysis observes a $\mathcal{O}(0.4\sigma)$ upward fluctuation:

$$(5.9^{+2.8}_{-1.8}) \times 10^{-8}(\text{exp.}) \quad \text{vs.} \quad 7.1 \times 10^{-8}(\text{obs.}),$$

while [5] observes a $\mathcal{O}(0.2\sigma)$ downward fluctuation:

$$(5.0^{+2.4}_{-1.6}) \times 10^{-8}(\text{exp.}) \quad \text{vs.} \quad 4.6 \times 10^{-8}(\text{obs.}).$$

The uncertainties on the expected limits here are the 1σ range of the Poissonian fluctuations of the event counts.

Since both analyses analysed the same LHCb data, the overlap of events selected by both versions of $\mathcal{M}_{3\text{body}}$ is determined for cut values x with efficiencies of 20%, 40%, 60%, and 80%. For all four efficiencies, the overlap is only around 60%, i. e.

$$\frac{\varepsilon(\mathcal{M}_{3\text{body,here}} > x \wedge \mathcal{M}_{3\text{body,[5]}} > x)}{\varepsilon(\mathcal{M}_{3\text{body,here}} > x_{\text{here}})} \approx 0.6.$$

Observing a mild upward fluctuation in the present work, and a mild downward fluctuation in [5] at a correlation coefficient of 0.6 is therefore possible.

Combination of the LHCb result with previous measurements

The combination of the Belle result, [3], and the BaBar result, [167], with the LHCb publication [5] has been reported to be

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < 1.2 \times 10^{-8} \quad [171].$$

The LHCb measurement, [5], leads to a small improvement in this combination. Without the LHCb data, the combination of Belle and BaBar is

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < 1.4 \times 10^{-8} \quad [11].$$

Model dependence

The exclusion limit on the branching fraction $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ determined in the previous chapter assumes that the 3-body phase space is flat and the distribution over the Dalitz plane is thus constant. Various models for physics beyond the Standard Model do not fulfil this assumption. Exclusion limits for phase space distributions in effective theories are determined in this chapter. The analysis remains unchanged, only the efficiency ratio in the normalisation factor and the distribution of $\tau \rightarrow \mu\mu\mu$ over the analysis bins are reevaluated.

13.1 Separation in effective phase space distributions

The model independent analysis of $\tau \rightarrow \mu\mu\mu$ by [72], presented in Sect. 1.3, finds that three effective operators need to be taken into account. Allowing for an overall normalisation \mathcal{N} , any effective Hamiltonian can be written as linear combination of the three operators,

$$H_{\text{eff}} = \mathcal{N} \left(c_1 H_{\text{eff}}^{(LL)(LL)} + c_2 H_{\text{eff}}^{(LL)(RR)} + c_3 H_{\text{eff}}^{\text{rad}} \right).$$

The phase space distribution of the decay $\tau \rightarrow \mu\mu\mu$ for the full Hamiltonian is then a linear combination of the phase space distributions for the individual operators and their interferences (\mathcal{N}' is again a normalisation constant to ensure $\int \varrho = 1$)

$$\begin{aligned} \varrho(m_{--}^2, m_{+-}^2) = \mathcal{N}' \left(& |c_1|^2 \underbrace{\varrho_V^{(LL)(LL)}(m_{--}^2, m_{+-}^2)}_{=: \varrho_1} \right. \\ & + |c_2|^2 \underbrace{\varrho_V^{(LL)(RR)}(m_{--}^2, m_{+-}^2)}_{=: \varrho_2} \\ & + |c_3|^2 \underbrace{\varrho_{\text{rad}}^{(LR)}(m_{--}^2, m_{+-}^2)}_{=: \varrho_3} \\ & + 2 \underbrace{|c_1 c_3| \cos \arg(c_1/c_3)}_{=: |c_4|^2} \underbrace{\varrho_{\text{int}}^{(LL)(LL)}(m_{--}^2, m_{+-}^2)}_{=: \varrho_4} \\ & + 2 \underbrace{|c_2 c_3| \cos \arg(c_2/c_3)}_{=: |c_5|^2} \underbrace{\varrho_{\text{int}}^{(LL)(RR)}(m_{--}^2, m_{+-}^2)}_{=: \varrho_5} \\ & \left. + 2 |c_1 c_2| \cos \arg(c_1/c_2) \underbrace{\varrho_{\text{int}}^{((LL)(LL)) \leftrightarrow ((LL)(RR))}}_{\approx 0}(m_{--}^2, m_{+-}^2) \right). \end{aligned}$$

With the definitions of ϱ_i , $|c_4|^2$, and $|c_5|^2$, this reads in short:

$$\varrho(m_{--}^2, m_{+-}^2) = \mathcal{N}' \sum_{i=1}^5 |c_i|^2 \varrho_i(m_{--}^2, m_{+-}^2).$$

As stated in Sect. 1.3, the interference between the vector operators $H_{\text{eff}}^{(LL)(LL)}$ and $H_{\text{eff}}^{(LL)(RR)}$ can be neglected because the interference of different helicity final states is suppressed by m_μ^2/m_τ^2 .

13.2 Efficiency dependence

The following paragraphs are *just* a mathematically more rigorous formulation of the efficiency determination method in Sect. 11.3.

The efficiency which is used in the nominal analysis, ε_{nom} is a phase space averaged value:

$$\varepsilon_{\text{nom}} = \iint dm_{--}^2 dm_{+-}^2 \varepsilon(m_{--}^2, m_{+-}^2) \varrho_{\text{nom}}(m_{--}^2, m_{+-}^2)$$

$$\text{where } \varrho_{\text{nom}}(m_{--}^2, m_{+-}^2) = \frac{1}{V_{\text{Dalitz}}} = \text{const.}$$

The integration is carried out by Monte Carlo integration. The integrand in the Monte Carlo integration is the observed phase space distribution of $\tau \rightarrow \mu\mu\mu$ decays in the sample of simulated $\tau \rightarrow \mu\mu\mu$ decays after applying the reconstruction, event selection, and trigger requirement:

$$\varepsilon_{\text{nom}} = \frac{1}{N_{\text{sample}}} \sum_{\text{sample}} \frac{1}{cV_{\text{Dalitz}}}, \quad (13.1)$$

where V_{Dalitz} is the volume of the kinematically allowed phase space, and c is a constant to make the summand dimensionless. The value of c is not determined as it drops out later.

For the above linear combination of phase space distributions, the efficiency integral becomes

$$\varepsilon(H_{\text{eff}}) = \iint dm_{--}^2 dm_{+-}^2 \varepsilon(m_{--}^2, m_{+-}^2) \sum_{i=1}^5 |c_i|^2 \varrho_i(m_{--}^2, m_{+-}^2).$$

The integral can be rearranged¹ to

$$\begin{aligned} \varepsilon(H_{\text{eff}}) &= \sum_{i=1}^5 |c_i|^2 \underbrace{\iint dm_{--}^2 dm_{+-}^2 \varepsilon(m_{--}^2, m_{+-}^2) \varrho_i(m_{--}^2, m_{+-}^2)}_{=:\varepsilon_i} \\ &= \sum_{i=1}^5 |c_i|^2 \varepsilon_i. \end{aligned}$$

¹[172]

Table 13.1: Efficiency ratios $\varepsilon_i/\varepsilon_{\text{nom}}$ for different phase space distributions relative to a flat phase space distribution. The efficiency hereby includes acceptance, trigger, and event selection, except for the requirements on the minimum multivariate classifier response. The uncertainties are the statistical uncertainty due to the finite size of the simulated sample. The values have been determined by [38].

	2011 data	2012 data
$\varrho_V^{(LL)(LL)}$	1.0049 ± 0.0019	1.0060 ± 0.0016
$\varrho_V^{(LL)(RR)}$	1.0772 ± 0.0010	1.0680 ± 0.0008
$\varrho_{\text{rad}}^{(LR)}$	0.6318 ± 0.0017	0.6422 ± 0.0014
$\varrho_{\text{int}}^{(LL)(LL)}$	0.9193 ± 0.0027	0.9280 ± 0.0023
$\varrho_{\text{int}}^{(LL)(RR)}$	1.0124 ± 0.0013	1.0089 ± 0.0010

This transformation is only possible when expressing the full phase space distribution as a linear combination of the five phase space distributions ϱ_i . Due to the interference terms, the expansion cannot be done for the three effective Hamiltonians.

The efficiency is expected to depend strongly on the phase space distribution due to the veto against $\eta \rightarrow \mu\mu\gamma$ decays. Requiring the simulated decays to be within the detector acceptance, pass the trigger, be reconstructed, and fulfil the selection requirements immediately yields the dependence of the full efficiency as a function of (m_{--}^2, m_{+-}^2) as the observed phase space distribution, shown in Fig. 13.1. The impact of the ϕ veto and the η veto is visible in Fig. 13.1, the double reconstruction veto is hardly visible because the requirement $m_{--} > 250 \text{ MeV}/c^2$ is very close to the edge of the Dalitz plot.

The efficiency for each of the phase space distributions is determined using Monte Carlo integration with the simulated $\tau \rightarrow \mu\mu\mu$ decays, again.

$$\varepsilon_i = \frac{1}{N_{\text{sample}}} \sum_{\text{sample}} \frac{\varrho_i(m_{--}^2, m_{+-}^2)}{c},$$

with the same constant c as in Eq. 13.1. The only difference with respect to Eq. 13.1 is that instead of a constant $1/V_{\text{Dalitz}}$ a non-constant phase space distribution $\varrho(m_{--}^2, m_{+-}^2)$ is summed/integrated.

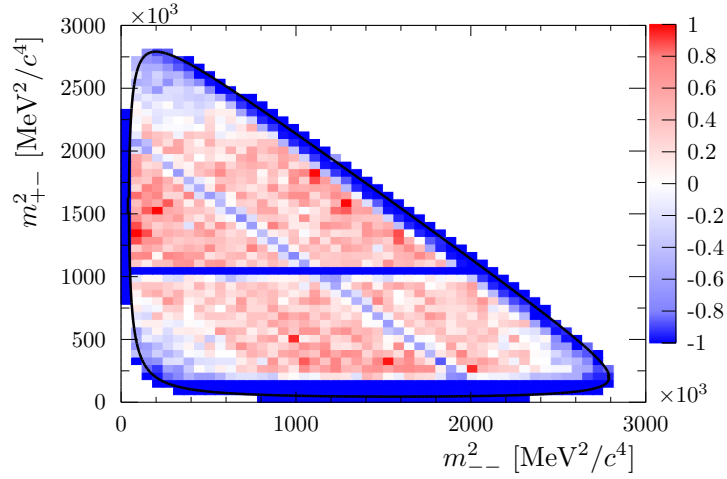
It follows

$$\frac{\varepsilon_i}{\varepsilon_{\text{nom}}} = \frac{1}{N_{\text{sample}}} V_{\text{Dalitz}} \sum_{\text{sample}} \varrho_i(m_{--}^2, m_{+-}^2) \quad \forall 1 \leq i \leq 5$$

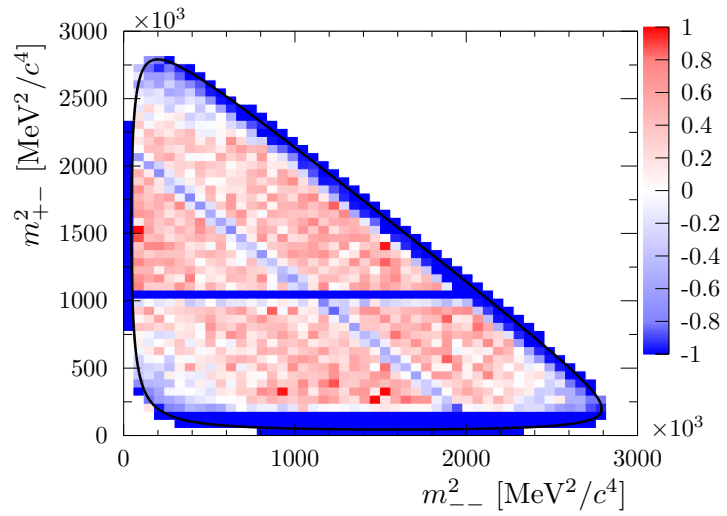
These five sums have been evaluated by Jonathan Harrison for each data taking period separately. The results, $\varepsilon_i/\varepsilon_{\text{nom}}$, are summarised in Tab. 13.1. As expected, the most striking model dependence is found for the radiative operator where the dominant contribution is the η veto.

13.3 Signal distribution

The signal distributions over the analysis bins for the different phase space distributions are obtained as in Chap. 10 for the nominal analysis, weighting the events in



(a) 2011



(b) 2012

Figure 13.1: The relative deviation from the average efficiency as a function of m_{--}^2 and m_{+-}^2 . The efficiency in bins which are partially outside of the kinematically allowed region is low due to the η veto or veto of double reconstructed tracks. The fraction of those bins which is outside of the allowed phase space is taken into account in the computation, i. e. the visualised efficiency in those bins is not biased to small values due to the binning.

the simulated $\tau \rightarrow \mu\mu\mu$ calibration sample (see Sect. 9.2) with $\sum |c_i|^2 \varrho_i(m_{--}^2, m_{+-}^2)$.

13.4 Sensitivities

The determination of the expected and observed exclusion limit for the branching fraction $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ is repeated with the CLs method as in Chap. 12. As state above, the following modifications are made to account for the altered phase space distribution:

- For simplicity, the nuisance parameters related to the distribution of $\tau \rightarrow \mu\mu\mu$ over the analysis bins are fixed to their nominal values.
- The nominal normalisation factor is corrected by the aforementioned efficiency ratio

$$\alpha \rightarrow \alpha \cdot \left(\sum |c_i|^2 \frac{\varepsilon_i}{\varepsilon_{\text{nom}}} \right)^{-1}$$

- The signal distribution over the analysis bins for the weighted calibration sample is used instead of the un-weighted calibration sample.

The expected and observed exclusion limits for different phase space distributions are summarised in Tab. 13.2. The differences with respect to the nominal analysis, are close to the efficiency ratios quoted in Tab. 13.1, i. e. the dominant dependence on the decay kinematics is introduced by the efficiency.

Since no combination of complex numbers c_1, c_2, c_3 can lead to $|c_1|^2 = |c_2|^2 = |c_3|^2 = |c_4|^2 = 0 \neq |c_5|^2$, there cannot be any new physics model with a phase space distribution of ϱ_5 (and correspondingly for ϱ_4). The “exclusion limits” for ϱ_4 and ϱ_5 are necessary for the following parametrisation.

The upper limit on a branching fraction in Eq. 3.4 is

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < \text{Limit}(\tau \rightarrow \mu\mu\mu) \propto \frac{1}{\varepsilon}.$$

The other terms are not phase-space-distribution dependent. Above it is found that $\varepsilon(H_{\text{eff}}) = \sum |c_i|^2 \varepsilon_i$. This means, that when the Hamiltonian of a new physics model is expanded as a linear combination of the effective operators

$$H_{\text{eff}} = \mathcal{N} \left(c_1 H_{\text{eff}}^{(LL)(LL)} + c_2 H_{\text{eff}}^{(LL)(RR)} + c_3 H_{\text{eff}}^{\text{rad}} \right)$$

and for each of the phase space distributions an exclusion limit is known

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu, \varrho(m_{--}^2, m_{+-}^2) = \varrho_i) < \text{Limit}(\tau \rightarrow \mu\mu\mu, \varrho(m_{--}^2, m_{+-}^2) = \varrho_i) =: L_i \quad \forall 1 \leq i \leq 5,$$

then the limit on the branching fraction for the phase space distribution of the new physics model is

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu, \text{via } H_{\text{eff}}) \lesssim \left(\sum_{i=1}^5 \frac{|c_i|^2}{L_i} \right)^{-1} \quad (13.2)$$

Equation 13.2 is for analyses with a single analysis bin strictly valid. For the present analysis it holds approximately, due to the a possible influence of the phase

Table 13.2: Expected and observed exclusion limits for $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ for different phase space distribution of the decay. The nuisance parameters labelled as “signal NP” in Tab. 12.3 are fixed to their nominal values. All numbers are given at 90% confidence level. There are four categories of phase space distributions, separated by horizontal lines. Firstly, the expected and observed exclusion limit for the nominal analysis. Secondly, the exclusion limits for pure contributions from one of the effective operators. Thirdly, the exclusion limits for the unphysical phase space distributions of the interference terms. Fourthly, exclusion limits for simple linear combinations of effective operators. For the latter ones, the exclusion limit for the parametrisation in Eq. 13.2 is given, too.

distribution	exclusion limit		
	expected [10^{-8}]	observed [10^{-8}]	parametric observed [10^{-8}]
$\varrho = \text{const}$	5.8	7.0	-
$\varrho_1 := \varrho_V^{(LL)(LL)}$	5.8	$L_1 := 6.9$	-
$\varrho_2 := \varrho_V^{(LL)(RR)}$	5.4	$L_2 := 6.5$	-
$\varrho_3 := \varrho_{\text{rad}}^{(LR)}$	9.1	$L_3 := 10.8$	-
$\varrho_4 := \varrho_{\text{int}}^{(LL)(LL)}$	6.3	$L_4 := 7.5$	-
$\varrho_5 := \varrho_{\text{int}}^{(LL)(RR)}$	5.8	$L_5 := 6.9$	-
$\frac{1}{2} \cdot \left(\varrho_V^{(LL)(LL)} + \varrho_{\text{rad}}^{(LR)} \right)$	7.0	8.4	8.4
$\frac{1}{2} \cdot \left(\varrho_V^{(LL)(RR)} + \varrho_{\text{rad}}^{(LR)} \right)$	6.8	8.1	8.1
$\frac{1}{2} \cdot \left(\varrho_V^{(LL)(RR)} + \varrho_V^{(LL)(RR)} \right)$	5.6	6.7	6.7
$\frac{1}{4} \cdot \left(\varrho_V^{(LL)(LL)} + \varrho_{\text{rad}}^{(LR)} + 2\varrho_{\text{int}}^{(LL)(LL)} \right)$	6.6	7.9	7.9
$\frac{1}{4} \cdot \left(\varrho_V^{(LL)(RR)} + \varrho_{\text{rad}}^{(LR)} + 2\varrho_{\text{int}}^{(LL)(RR)} \right)$	6.3	7.5	7.5

space distribution on the the distribution of the signal over the multivariate likelihood bins. The validity of the approximation can be seen in Tab. 13.2, where the parametric exclusion limit coincides with the exact calculation within the printed rounding precision. For a given new physics model thus only the coefficients $|c_i|^2$ need to be determined to obtain an exclusion limit on $\mathcal{B}(\tau \rightarrow \mu\mu\mu)$ based on the present analysis.

The results in Tab. 13.2 show that the limit on new physics models with a pure vector contribution (ϱ_1, ϱ_2 , or linear combinations like $1/2(\varrho_1 + \varrho_2)$) is stronger than the nominal analysis with the assumption of a flat phase space distribution. The exclusion limit of the nominal analysis is, however, too strong for new physics models with a radiative contribution. The same is observed in [5].

Conclusion

In the Standard Model, the lepton flavour violating decay $\tau \rightarrow \mu\mu\mu$ is expected to occur at unobservable low rates. Models for physics beyond the Standard Model, however, can predict larger rates and they can be constrained by observed limits on the branching fraction.

The search for the lepton flavour violating decay at LHCb has reached a sensitivity similar to previous experiments at e^+e^- colliders. The entire data set taken by LHCb in the first run of the LHC is analysed in this work. In the absence of a signal, it is expected to obtain an exclusion limit between 4.1×10^{-8} and 8.7×10^{-8} . The observed exclusion limit,

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < 7.1 \times 10^{-8}$$

is in the expected range. The sensitivity is reached by applying a multivariate event classification which separates the data into 21 independent search regions.

The application of a classification, instead of a simpler “cut-and-count” technique contributes approximately 18% to the sensitivity. The classification technique developed in this work has been picked up in an ensemble selection within a different multivariate toolkit. Together with a larger sample of simulated events, the publication in [5] expects an exclusion between 3.4×10^{-8} and 7.4×10^{-8} and observes

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < 4.6 \times 10^{-8}.$$

The largest part of the difference of the observed exclusion limits between both analyses is compatible with Poissonian fluctuations of the observed event count around the background expectation, additionally to an improvement of 15% in the expected limit.

The current world combination of $\tau \rightarrow \mu\mu\mu$ searches (including the LHCb publication [5]) is

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < 1.2 \times 10^{-8}.$$

The above exclusion limits are valid for the precise decay sought for, a $\tau \rightarrow \mu\mu\mu$ decay with a flat phase space distribution. Within an effective theory, exclusion limits for other phase space distributions can be computed. The exclusion limits can be as low as 6.9×10^{-8} or as large as 1.1×10^{-7} . The latter for a dominant radiative contribution.

The LHC and LHCb will continue operations in 2014. The sensitivity can be expected to improve with the square root of the τ production, $1/\sqrt{\sigma\mathcal{L}}$.

Toy studies on classifier binning

To illustrate the benefit of using several analysis bins, a toy study is performed. Simulated candidate events have one property \mathcal{M} which can be thought of being a multivariate classifier, which discriminates signal from background. For each event, \mathcal{M} is a number between 0 and 1. Signal events have a flat distribution in \mathcal{M} . Background events have an exponential distribution in \mathcal{M} .

$$P(\mathcal{M}(\text{event}) < v | \text{event is signal}) = v \quad (\text{A.1})$$

$$P(\mathcal{M}(\text{event}) < v | \text{event is background}) = \mathcal{N} \cdot (1 - e^{-5v}) \quad (\text{A.2})$$

$$\text{with a normalisation } \mathcal{N} = \frac{1}{1 - e^{-5}}$$

One of these two assumptions can be achieved by means of a monotonous variable transformation. The other is a choice motivated by [168, Fig. 3].

A.1 Application of a classifier requirement

The CLs method, as described in Sects. 3.5 through 3.7, is used to determine the expected sensitivity of an experiment. The outcome of the study does not depend on α . For simplicity α is set to 10^{-3} prior to the application of a selection. If a selection is applied, α is corrected correspondingly $\alpha \rightarrow \alpha/\varepsilon$. The results do not depend on the choice of α .

A.1.1 Selection requirement vs. no selection requirement

Firstly, it is shown that applying a selection requirement improves the sensitivity of a rare decay search.

To quantify the effect of introducing a selection requirement, ten scenarios are compared for $b = 90$ expected background events. Not applying a requirement, and applying nine different requirements $\mathcal{M} > x$ with equally spaced values from 0.1 to 0.9 for x . The expected upper limit as a function of x shown in figure A.1.

There are two extreme cases. The selection is not restrictive enough, at $\mathcal{M} > 0.0$, where the efficiency is 100% and all background events are selected. The sensitivity improves when increasing x . The other extreme case is at $\mathcal{M} > 0.9$. The efficiency is only 10% and 4 remaining background events are expected to be selected. It is pointed out that the expected upper limit for the extreme case $\mathcal{M} > 0.9$ is worse than the expected upper limit without a selection requirement.

Among the requirements tested, the best sensitivity is obtained requiring $\mathcal{M} > 0.4$ with an efficiency of 60%. The application of a selection requirement improves the

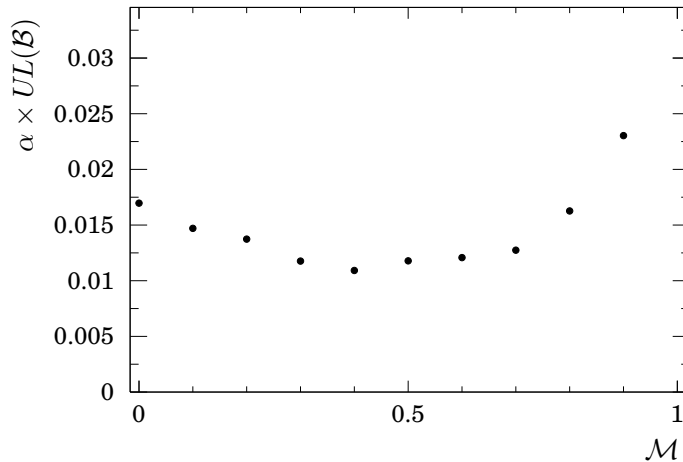


Figure A.1: Expected upper limit of the toy experiment as a function of the selection requirement on the classifier. The most left point, $\mathcal{M} > 0$ corresponds to not applying a selection requirement.

expected upper limit from 0.017 to 0.011. The function looks “nonsmooth”, which is a consequence of the discrete nature of the observable. The nonsmoothness vanishes when choosing larger numbers of expected background events, as shown in Fig. A.2.

A.1.2 Optimal selection for increasing integrated luminosity

When more data is accumulated it is worth restricting the selection further. The number of expected background events is changed to 6 different values from 5 to 30 000. For each of these, the expected upper limit under the background hypothesis is shown as a function of the selection requirement $\mathcal{M} > x$ in Fig. A.2. At high b , the minimum is found at higher values than at low b .

At first sight, the study shown in Fig. A.2 only shows that the selection must be more restrictive when large numbers of background events are expected, and that in these cases the sensitivity degrades.

One can also consider this study as an increase of the integrated luminosity when correcting α correspondingly. The number of expected background events should be linear to the accumulated integrated luminosity. Furthermore $\alpha \propto 1/\mathcal{L}$. The results from Fig. A.2 are summarised in Tab. A.1 with the minimal upper limits from the figure ($UL \times 10^{-3}$) and with a luminosity corrected normalisation factor ($UL \times \alpha(\mathcal{L})$). One finds that the selection must be more restrictive when more data is accumulated.

The optimal selection requirement shows the same nonsmooth behaviour as UL for small values of b , visible as “jump” from 0.4 to 0.3 and back for $b = 30$. Testing for b each integer number from 30 to 90, one finds that the optimal value for x fluctuates between 0.3 and 0.5, with a general trend that the optimal value becomes larger with increasing b . It is visible from Fig. A.2 that the optimal selection requirement of 0.6 or 0.7 for large values of b is not optimal for small values of b and not within the range of the fluctuations observed. The finding that the selection must become restrictive is thus not a result of fluctuations.

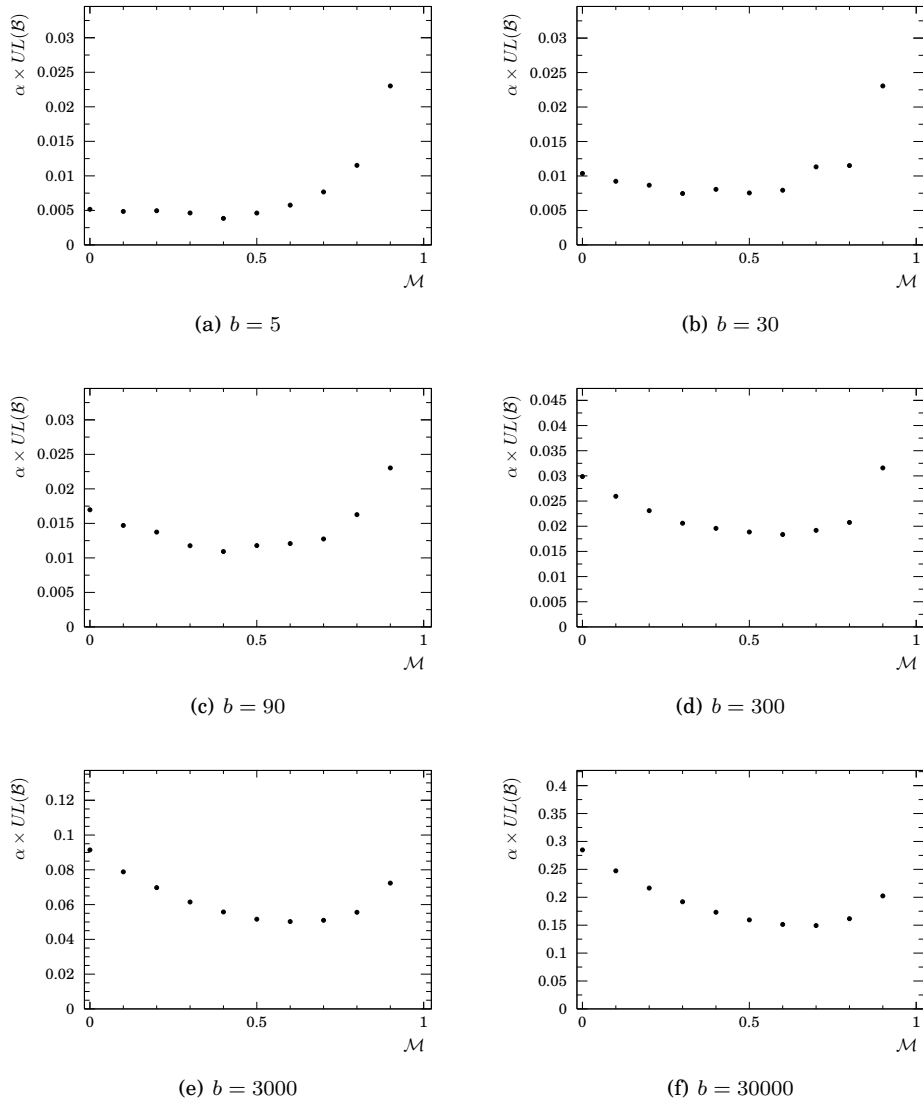


Figure A.2: Expected upper limit for different numbers of expected background events with $\alpha = 10^{-3}$. The ordinate ranges change in the respective subfigures. The ranges of the abscissae are the same in all subfigures.

Table A.1: Expected upper limits for the best selection in Fig. A.2. No uncertainties are given since the numbers are either exact or the dominant uncertainty is rounding to two significant digits.

$\mathcal{L}[\text{a.u.}]$	$\alpha(\mathcal{L})$	b	$UL \times \alpha(\mathcal{L})$	$UL \times 10^{-3}$	best selection $\mathcal{M} > x$
1	1.8×10^{-2}	5	0.069	0.0038	0.4
6	3×10^{-3}	30	0.022	0.0075	0.3
18	1×10^{-3}	90	0.011	0.011	0.4
60	3×10^{-4}	300	0.0055	0.018	0.6
600	3×10^{-5}	3000	0.0015	0.050	0.6
6000	3×10^{-6}	30000	0.00045	0.15	0.7

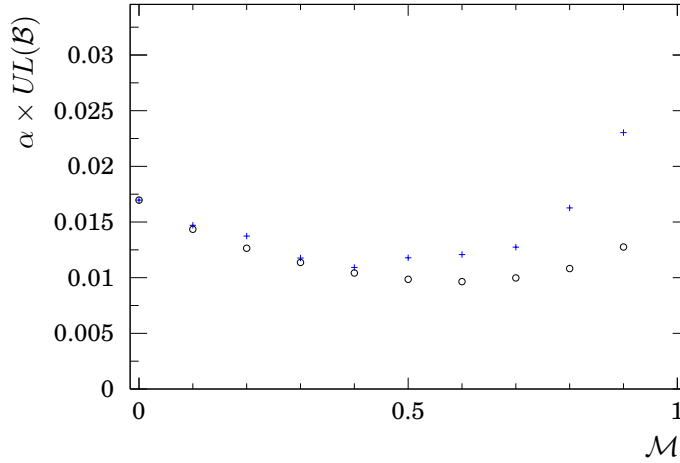


Figure A.3: Comparison of the expected upper limit in an analysis with a selection requirement $\mathcal{M} > x$ (blue hair cross) to an analysis in two bins with $\mathcal{M} > x$ and $\mathcal{M} < x$ respectively (black circle). The data points from the former are identical to those shown in Fig. A.1.

A.2 Application of binning

It is argued in Sect. 3.4 that it is favourable to do an analysis in bins over the application of a selection requirement. The behaviour of the sensitivity of a binned analysis is illuminated in this section.

A.2.1 Two bins compared to a selection requirement

As an example, the same toy setup as before ($b = 90$) is used to compare the application of a selection requirement $\mathcal{M} > x$ to an analysis in two bins, where x is the bin boundary.

The result is shown in Fig. A.3. For each tested value of x , the binned analysis is equally or better performing than the analysis with a selection requirement. In the extreme case $x = 0$, both analyses are identical. In contrast to the findings of the

A.2. Application of binning

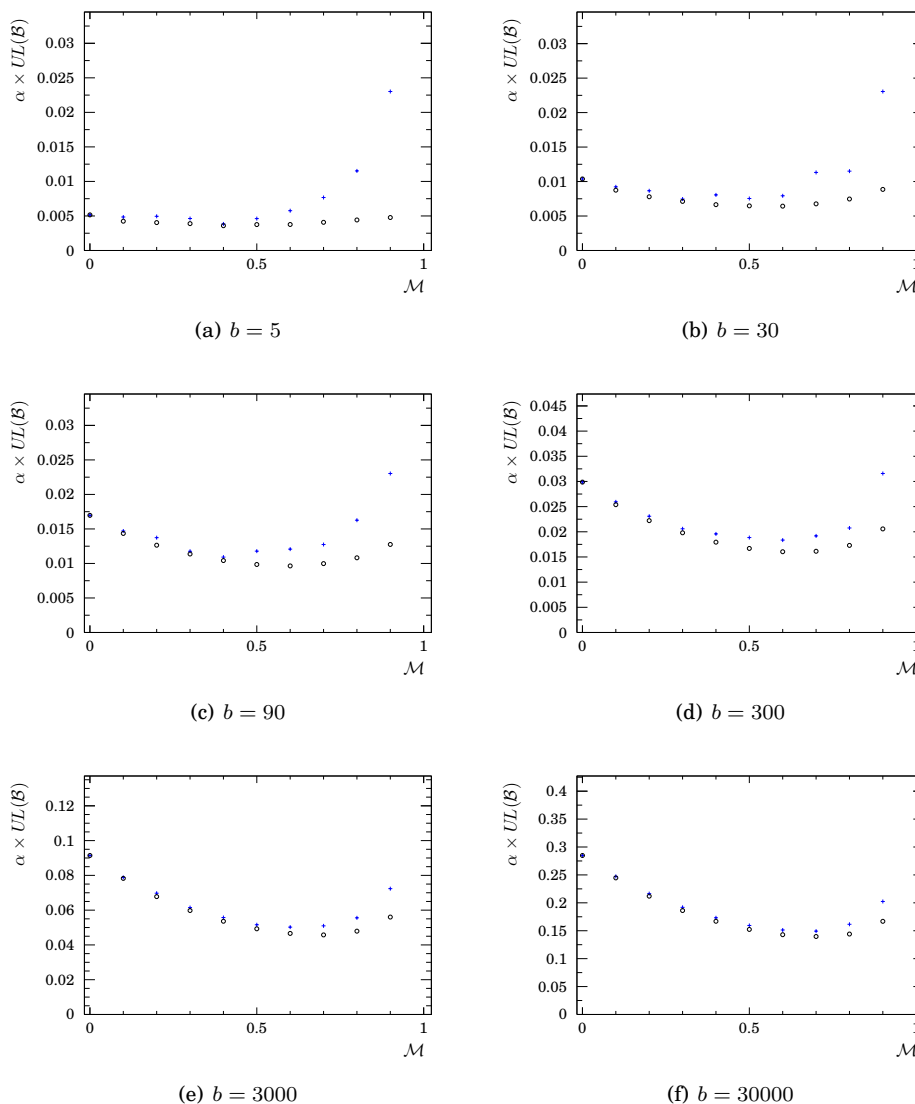


Figure A.4: Comparison of the sensitivity of a binned analysis (black circle) to the sensitivity using a selection requirement (blue hair cross). The ordinate ranges change in the respective subfigures. The ranges of the abscissae are the same in all subfigures.

previous section, the binned analysis cannot perform worse than the extreme case of $x = 0$. As an example, the expected upper limit for $x = 0.9$ in the binned analysis is still better than the analysis with $x = 0$. The change from the optimal selection requirement to the optimal binned analysis improves the expected upper limit from 1.1×10^{-2} to 9.6×10^{-3} . The optimal bin boundary is at $x = 0.6$, which does not coincide with the optimal selection requirement at $\mathcal{M} > 0.4$.

The same comparison, between an analysis with a selection requirement to an

Table A.2: Expected upper small for the best selection and best binning in Fig. A.2. No uncertainties are given since the numbers are either exact or the dominant uncertainty is rounding to two significant digits.

b	best selection	$UL \times 10^{-3}$ (selection)	best bin boundary	$UL \times 10^{-3}$ (bins)	relative improvement
5	0.4	0.0038	0.4	0.0036	6.7%
30	0.3	0.0075	0.6	0.0064	14%
90	0.4	0.011	0.6	0.0096	17%
300	0.6	0.018	0.6	0.016	13%
3000	0.6	0.050	0.7	0.046	8.9%
30000	0.7	0.15	0.7	0.14	6.5%

analysis in two bins, is shown for different values of b in Fig. A.4. The expected upper limit as a function of x in the binned analysis tends to be smoother than in the analyses with a selection requirement. The optimal bin boundaries are situated at larger values than the optimal selection requirement.

The sensitivity improvement for changing from a selection requirement to a binned analysis is different for the tested scenarios as summarised in Tab. A.2. It is suspected that the relative improvements at small values for b are dominated by the fluctuations, mentioned earlier. At large values for b , a trend is visible that the improvement shrinks. It is not investigated if this stems from the step size of possible values for x .

Note

- One might be interested in not only optimising the expected upper limit, but also to keep the 1σ range of the expected upper limit at limited size. Conversely, if the 1σ range increases drastically in the optimisation, the experiment's outcome would be dominated by (un)lucky fluctuations instead of the achieved optimisation. The behaviour of the 1σ range of the expected upper limit for $b = 90$ is shown in Fig. A.5. No increase of the width of the 1σ interval within which the upper limit is expected is observed as a side effect of the optimisation.

A.2.2 Multiple bins

The scenario of using two analysis bins is compared to that with three analysis bins, for $b = 90$. The expected upper limit becomes a function of two bin boundaries, x_1 and x_2 , and is shown in Fig. A.6. The findings are invariant under exchange of x_1 and x_2 . The outer most rows and columns correspond to either of the bin boundaries being 0 or 1, which means that one bin vanishes. The same holds for the diagonal $x_1 = x_2$. The expected upper limit as a function of e.g. x_1 for $x_2 = 0$ is identical to that shown in Fig. A.3.

The minimum is 9.0×10^{-3} at $x_1 = 0.7$ and $x_2 = 0.4$. Neither of the bin boundaries is the optimal bin boundary for the analysis with two bins ($x = 0.6$, yielding an expected upper limit of 9.6×10^{-3}). The analysis with three bins thus improves the sensitivity

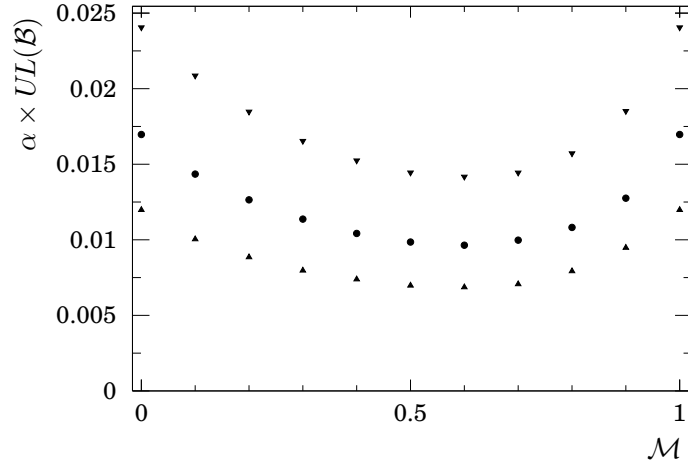


Figure A.5: Expected upper limit and its 1σ range for the toy study using 2 bins as function of the position of the bin boundary. The data points at 1 and 0 are identical as these are the cases where only one bin spans from 0 to 1.

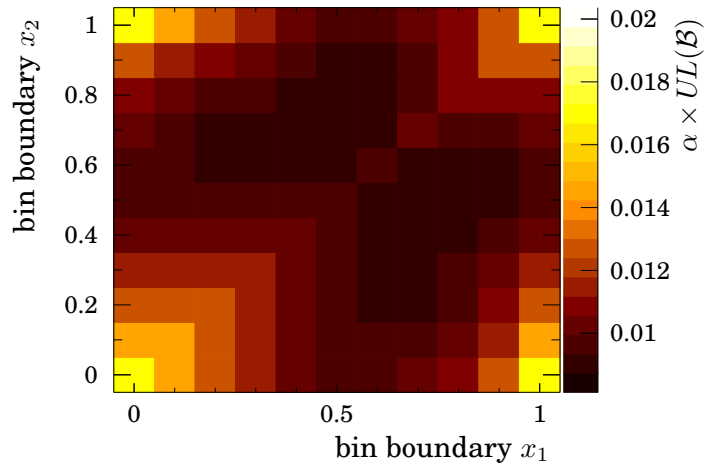


Figure A.6: Expected upper limit for an analysis with three bins as bin boundaries x_1 and x_2 . The cases where $x_1 = x_2$, $x_i = 0$, or $x_i = 1$ are identical to analyses with two bins. Cases where both x_i take pathological values 1 or 0 are identical to not applying binning or a selection requirement.

A. TOY STUDIES ON CLASSIFIER BINNING

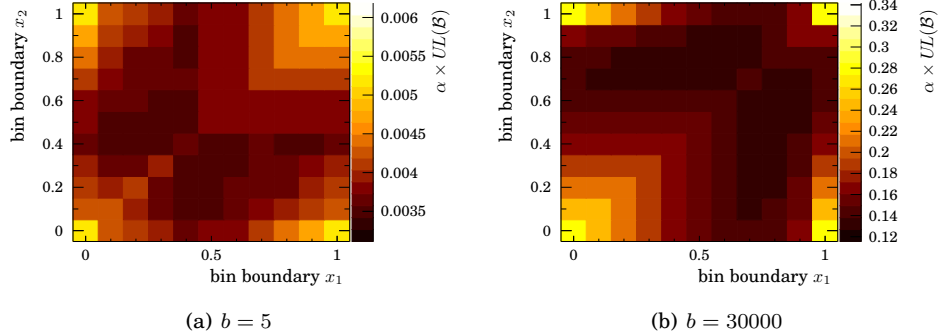


Figure A.7: Expected upper limit for an analysis with three bins as bin boundaries x_1 and x_2 for different numbers of expected background events.

Table A.3: Optimal expected upper limits for an analysis with a selection requirement, two bin, and three bins compared, for different numbers of expected background events. Numerical errors are smaller than the rounding precision.

b	$UL \times 10^{-3}$ (selection)	$UL \times 10^{-3}$ (2 bins)	$UL \times 10^{-3}$ (3 bins)	relative improvement (2 \rightarrow 3 bins)	relative improvement (sel. \rightarrow 3 bins)
30	0.0075	0.0064	0.0060	6.5 %	19 %
90	0.011	0.0096	0.0090	6.4 %	17 %
300	0.018	0.016	0.015	7.5 %	19 %
3000	0.050	0.046	0.042	7.9 %	16 %
30000	0.15	0.14	0.13	8.6 %	14 %

by 6.4 % over the analysis with two bins. The improvement over the application of a selection requirement is 17 %.

For illustrative purposes, Fig. A.7, shows the same for the extreme cases $b = 5$ and $b = 30\,000$. The conclusions of the previous sections hold: The optimal bin boundaries depend on the number of expected background events (the absolute number, independent of the normalisation factor). The bin boundaries need to be put at larger values for higher numbers of background events. The optimal bin boundaries using three bins do not coincide with the optimal bin boundary using two bins. Increasing the number of bins optimally therefore can not be done by splitting one bin into two.

The improvement for increasing the number of bins from two to three increases for larger numbers of expected background events, as summarised in Tab. A.3.

A.2.3 Multi-dimensional binning

In the search for $\tau \rightarrow \mu\mu\mu$ a binning is applied in two dimensions, \mathcal{M}_{PID} and $\mathcal{M}_{3\text{body}}$. The binning can be regarded as two one-dimensional binnings. The simplest approach is to find an optimal binning in the two dimensional range, is to find an optimal binning in \mathcal{M}_{PID} regardless of the distribution of signal and background in $\mathcal{M}_{3\text{body}}$ and

vice versa. Because the *marginal* distributions of the two-dimensional $\mathcal{M}_{\text{PID}} \times \mathcal{M}_{3\text{body}}$ distribution are input to the optimisation, this procedure of finding a two-dimensional binning is called *marginal optimisation* here.

The marginal optimisation will, however, not yield an optimal two-dimensional binning. In the entire $\mathcal{M}_{3\text{body}}$ range, there are more background events expected than in the range of high $\mathcal{M}_{3\text{body}}$ values. Therefore, the \mathcal{M}_{PID} binning which would be found optimal in the high $\mathcal{M}_{3\text{body}}$ range would be different from that found in the entire $\mathcal{M}_{3\text{body}}$ range.

As a hand waving argument, most of the sensitivity comes from high $\mathcal{M}_{3\text{body}}$ range – see Fig. A.3, where the improvement going from a selection to a binning is much smaller than the improvement from introducing a selection. On that scale, the usage of the bin with $\mathcal{M} < x$ is only a small sensitivity gain. Consequently, an optimal \mathcal{M}_{PID} binning must be found in the highest $\mathcal{M}_{3\text{body}}$ bin. This is also done in [127, Sect. 3.2.3]. Optimising the binnings in \mathcal{M}_{PID} and $\mathcal{M}_{3\text{body}}$ introduces a circular reference, since one must be known to find the other. As a consequence, an optimal two-dimensional binning must be found by varying the bin boundaries in both dimensions simultaneously to optimise an analysis with $N_{\text{PID}} \cdot N_{3\text{body}}$ bins; where N_{PID} is the number of bins in \mathcal{M}_{PID} and $N_{3\text{body}}$ is the number of bins in $\mathcal{M}_{3\text{body}}$.

A.3 Figure of merit for classifier binning

In the previous sections, the expected sensitivity of the toy experiment has been studied. It is desired to choose a binning in a real analysis which optimises the sensitivity. The approach from the toy study, trying several bin boundary settings and finding the optimal, is not possible in real applications. The toy study presented in the previous section needed $\mathcal{O}(100)$ CPU weeks for only three bins. The search for $\tau \rightarrow \mu\mu\mu$ used 5 bins in the $\mathcal{M}_{3\text{body}}$ in [4]. The complexity of the problem scales with $\binom{n}{k}$, where n is the number of places where a bin boundary is allowed to be, and k is the number of bin boundaries. The optimisation would thus have occupied the computing cluster of the Heidelberg LHCb group for roughly 300 weeks.

For the analysis, an expected sensitivity thus cannot be computed for each possible binning due to computing constrains. This conclusion is also found in [127], where ΔQ is used as a figure of merit.

A.3.1 Punzi’s figure of merit

A commonly used figure of merit is the signal significance $s/\sqrt{s+b}$. As described in [124] this is not optimal in the case of small statistics. It is found there that the optimal figure of merit is

$$FOM_{\text{Punzi}} := \frac{\varepsilon}{a/2 + \sqrt{b}}, \quad (\text{A.3})$$

where a is “the number of sigmas corresponding to one-sided Gaussian tests at [the desired] significance.”

Note

- Punzi’s figure of merit is independent of the branching fraction which is searched for. It is thus favourable over the signal significance.

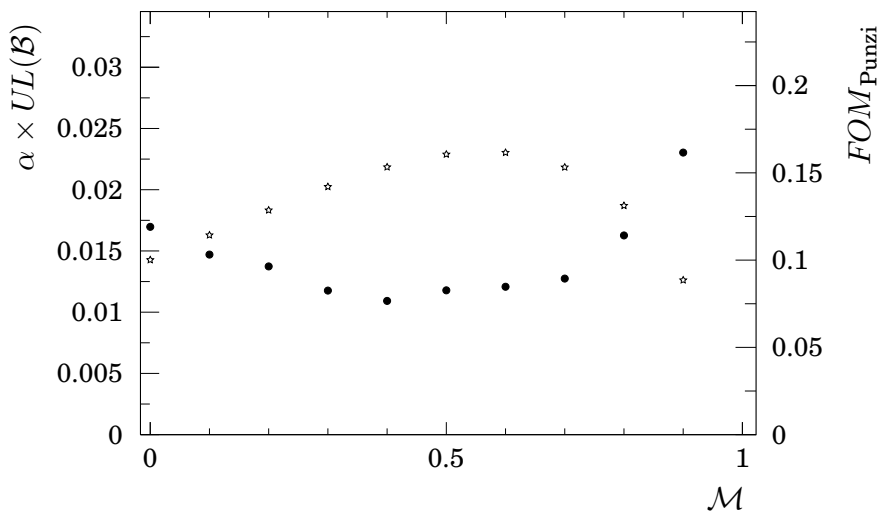


Figure A.8: Comparison of the CLs expected upper limit (at 90 % confidence level) to Punzi’s figure of merit (for 1.28σ) for $b = 90$. The expected upper limits as a function of a selection requirement from Fig. A.1 (black dots, left ordinate axis) behaves similar as Punzi’s figure of merit (open stars, right ordinate axis).

- Punzi’s figure of merit does not have a natural analogue for an analysis in multiple bins. It does not behave additive or multiplicative under random splits, like Q (see appendix B.1).
- For $a = 1.28$, the number of sigmas of a one sided 90 % confidence interval, FOM_{Punzi} agrees moderately well with the findings of Sect. A.2, as illustrated in Fig. A.8. The maximum of FOM_{Punzi} and the minimum of the expected upper limit do not coincide perfectly. It is assumed to be an effect of either the discreteness of event counts, which is not considered in FOM_{Punzi} , or the fact that in the derivation of FOM_{Punzi} the Poisson distribution is approximated by a Gaussian distribution, which does not hold in the toy study for $b = 90$ once a selection requirement is applied (For the optimal selection, 12 background events are expected). The maximum of FOM_{Punzi} and the minimum of the expected upper limit are closer to each other when larger values for b are chosen.

A.3.2 Difference of median test statistics

A figure of merit closely related to the CLs method is ΔQ , as motivated in [127], therein called ΔLQ . ΔQ is the difference between the medians of the Q distribution under the $s + b$ hypothesis and the b hypothesis.

$$\begin{aligned}
Q_{s+b}^{\text{med}} &= -2 \ln \frac{L_{s+b}(x_i = b_i + s_i)}{L_b(x_i = b_i + s_i)} \\
Q_b^{\text{med}} &= -2 \ln \frac{L_{s+b}(x_i = b_i)}{L_b(x_i = b_i)} \\
\Delta Q &= Q_b^{\text{med}} - Q_{s+b}^{\text{med}}
\end{aligned}$$

The computation of ΔQ does not require a complex Monte Carlo integration, as opposed to the computation of CL_s values. As pointed out in [127], the reduction of required computing resources, using ΔQ instead of CL_s as a figure of merit, makes a binning optimisation feasible.

As pointed out by [127], assuming Wilk's theorem [173], one can interpret ΔQ as significance square, at which the b and $s + b$ hypotheses are separated.

For the toy scenario with 2 bins, the expected upper limit as a function of the position of the bin boundary is compared to the behaviour of ΔQ . The comparison is shown in Fig. A.9 for $b = 90$. The minimal expected sensitivity is achieved with a bin boundary at 0.6 while ΔQ is maximal at 0.7 or 0.8, depending on the assumed $s + b$ hypothesis. The $s + b$ hypothesis for $\mathcal{B} = \alpha$ leads to a maximum of ΔQ at higher x than the $s + b$ hypothesis with $\mathcal{B} = \alpha \times UL_{\text{best 2 binned}}$, the best sensitivity for two analysis bins found earlier. The latter is closer to the bin boundary with the best sensitivity. For smaller values of b the discrepancy is larger while from $b = 3000$ onward the toy study didn't find a discrepancy between the true optimum and either of the ΔQ maxima.

It is concluded that ΔQ enables analysts to find the optimal bin boundaries, if the expected event count in all bins remains large enough. This is similar to the observation that FOM_{Punzi} finds the optimal selection criterion only if the expected number of background events which are selected remains large enough to justify the approximation of the Poisson distribution by a Gaussian distribution.

Note

- ΔQ depends on a $s + b$ hypothesis.
- The optimal bin boundaries found using ΔQ are only the true optimal bin boundaries, if enough expected remain in every bin. For practical reasons this must be ensured anyhow, as mentioned already in [127].

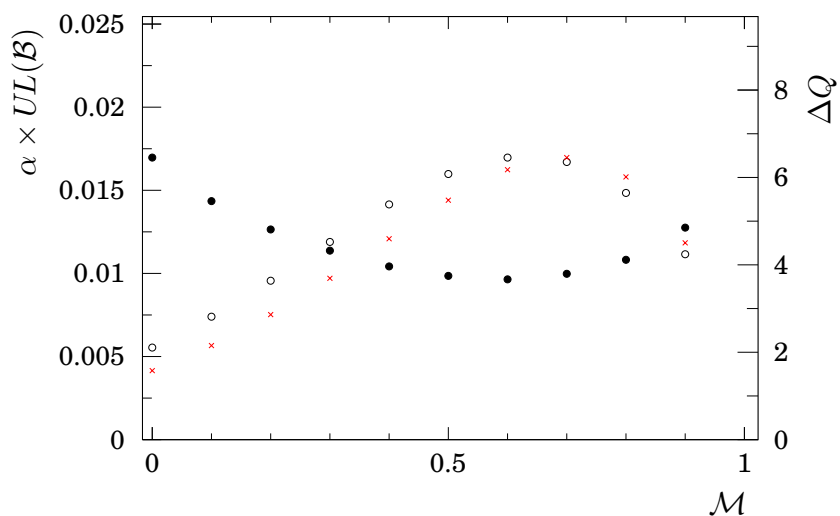


Figure A.9: Comparison of the CLs expected upper limit (at 90% confidence level) to ΔQ for $b = 90$. The expected upper limits as a function of the bin boundary from Fig. A.3 (black dots, left ordinate axis) behaves similar to ΔQ for one expected signal decay (i. e. $\mathcal{B} = \alpha$; red crosses, arbitrary units) and 11 expected signal decays (i. e. the branching fraction equal to the expected upper limit found in Tab. A.2; open circles, right ordinate axis).

Statistics glossary

B.1 Behaviour of Q under random splits

It is claimed in Sect. 3.5 that the test statistics Q is invariant under random splits of an analysis bins. WLOG, consider the case that the full data set is split into two bins. Let s and b be the total numbers of expected signal and background events, respectively, and s_i and b_i for $i = 1, 2$ the corresponding numbers for each of the two bins. It follows that

$$\begin{aligned} s &= s_1 + s_2 \\ b &= b_1 + b_2. \end{aligned}$$

If the split is randomly done the fraction of signal events which get classified in the first bin, s_1/s , is equal to the fraction of background events which get classified in the first bin, b_1/b :

$$\begin{aligned} f &:= \frac{s_1}{s} \\ \Rightarrow s_1 &= fs \\ b_1 &= fb \\ s_2 &= (1-f)s \\ b_2 &= (1-f)b \end{aligned}$$

The definition of λ can immediately be simplified, as most factors appear in the numerator and the denominator:

$$\begin{aligned} \lambda_{2 \text{ bins}} &= \frac{\prod_{i=1}^2 \frac{(s_i+b_i)^{n_i}}{n_i!} e^{-(s_i+b_i)}}{\prod_{i=1}^2 \frac{b_i^{n_i}}{n_i!} e^{-b_i}} \\ &= \frac{\prod_{i=1}^2 (s_i+b_i)^{n_i} e^{-s_i}}{\prod_{i=1}^2 b_i^{n_i}} \\ &= e^{-(s_1+s_2)} \frac{(s_1+b_1)^{n_1}}{b_1^{n_1}} \frac{(s_2+b_2)^{n_2}}{b_2^{n_2}} \\ &= e^{-s} \frac{f^{n_1} (s+b)^{n_1}}{f^{n_1} b^{n_1}} \frac{(1-f)^{n_2} (s+b)^{n_2}}{(1-f)^{n_2} b^{n_2}} \\ &= e^{-s} \frac{(s+b)^{n_1+n_2}}{b^{n_1+n_2}} = e^{-s} \frac{(s+b)^n}{b^n}. \end{aligned}$$

Firstly introducing factors e^{-b} and secondly $1/n!$ in numerator and denominator leads to the desired equality to $\lambda_{1 \text{ bin}}$.

$$\begin{aligned}
 \lambda_{2 \text{ bins}} &= \frac{(s+b)^n e^{-(s+b)}}{b^n e^{-b}} \\
 &= \frac{\frac{(s+b)^n}{n!} e^{-(s+b)}}{\frac{b^n}{n!} e^{-b}} = \lambda_{1 \text{ bin}} \\
 \Rightarrow -2 \ln \lambda_{2 \text{ bins}} &= -2 \ln \lambda_{1 \text{ bin}} \\
 \Rightarrow Q_{2 \text{ bins}} &= Q_{1 \text{ bin}}
 \end{aligned} \tag{B.1}$$

B.2 Standard distribution functions

The following distribution functions are used throughout the thesis.

B.2.1 Gauss distribution

The Gauss distribution (also known as normal distribution) describes a peak at mean position μ and with variance σ^2

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Despite its many mathematical properties, one of the most exploited features of the Gauss distribution throughout this work is its qualitative shape: “Mostly nothing, but around one position it goes up and down again. Nothing strange. Symmetric, simple, only two parameters.”

Since it is one of the simplest distributions imaginable, and its role in the central limit theorem, it is used as “default distribution” whenever nothing is known about a distribution, except for its mean and variance. Therefore the Gaussian distribution is used for the randomisation of nuisance parameters in the CLs method.

Unless otherwise noted, Gaussian distributions throughout this work are assumed to be centred at zero and have unit width, $\mu = 0$ and $\sigma = 1$.

B.2.2 Crystal Ball distribution

The Crystal Ball function is named after the Crystal Ball Collaboration, which established this function for the parametrisation of the peak in an invariant mass spec-

trum, with a radiative tail (i. e. FSR). The probability density function is given by

$$f(x; \alpha, n, \mu, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{if } \frac{x-\mu}{\sigma} > -\alpha \\ A \cdot \left(B - \frac{x-\mu}{\sigma}\right)^{-n}, & \text{if } \frac{x-\mu}{\sigma} \leq -\alpha \end{cases}$$

with

$$\begin{aligned} A &= \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \\ B &= \frac{n}{|\alpha|} - |\alpha| \\ N &= \frac{1}{\sigma \cdot (C + D)} \\ C &= \frac{n}{|\alpha|} \cdot \frac{1}{n-1} \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \\ D &= \sqrt{\frac{\pi}{2}} \left(1 + \operatorname{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right). \end{aligned}$$

The function consists of a Gaussian main part, and at the threshold $\frac{x-\mu}{\sigma} = -\alpha$, it is replaced by the tail function, which is a power-law, which is designed for continuity of the function and its first derivative at the threshold.

B.2.3 Poisson distribution

The Poisson distribution describes the probability of the occurrence of a number of events in a fixed interval. For a mean number of λ events the probability mass function is

$$f(x) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

The variance is k , which motivates the \sqrt{N} -law of uncertainties of observed event counts. For large λ it can be well approximated by a Gaussian distribution with $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$. The variance of the Poisson distribution is *always* λ . However the approximation with a Gaussian distribution becomes invalid for small values of λ . If this is the case, then the coverage of a $\pm 1\sigma$ interval, i. e. $[\lambda - \sqrt{\lambda}, \lambda + \sqrt{\lambda}]$ is not 68%.

B.2.4 Bernoulli distribution

The Bernoulli distribution describes random variables which can take the values 0 and 1. Its only parameter p is the so-called success rate, i. e. the probability at which the value 1 occurs. The cumulative density function is thus

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x \end{cases}$$

The expected value is p and the variance $p(1-p)$, which is the reason for the commonly used efficiency estimator uncertainty $\sqrt{\varepsilon(1-\varepsilon)/N}$.

B.3 Computation of *UL* in the certain absence of backgrounds

In Poissonian statistics, the relation between the true mean, the observation, and confidence limits for the true mean inferred from the observation can be computed as follows. The probability to observe k events, for a true mean of μ events is

$$P(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}.$$

Consequently the probability of observing no event given a true mean of μ events is

$$P(0; \mu) = e^{-\mu}.$$

To turn this into a conclusive statement, one must fix a confidence level for which the result is given. Generally, it should be fixed before evaluating the data. Searches for $\tau \rightarrow \mu\mu\mu$ at previous experiments use a 90 % confidence level.

Numeric example at 90 % confidence level: Since $P(0; \mu) < 0.1 = 1 - 90\%$ for $\mu > 2.3025$, the probability of observing no event is smaller than 10 % for a true mean larger than 2.3025; which can be expressed as $\mu < 2.3025$ at 90 % confidence level. When no events are observed, the branching fraction thus is

$$\mathcal{B}(\tau \rightarrow \mu\mu\mu) < \alpha \times 2.3025$$

at 90 % confidence level. Corresponding limits can be set for the observations of other numbers of events.

B.4 The sPlot technique

The sPlot technique [164] applies the principle “If you know the background distribution and the full distribution, then you also know the signal distribution.” The sPlot technique is applicable when there is an observable distribution of variables \vec{m}, \vec{x} which is a superposition of N components. The distribution of the variables \vec{x} is of interest and the variables \vec{m} are a mean to distinguish the components. It is required that \vec{x} and \vec{m} are independent for all components:

$$\begin{aligned} f(\vec{x}, \vec{m}) &= \sum_N n_i f^i(\vec{x}, \vec{m}) \\ &= \sum_N n_i f_{\text{interest}}^i(\vec{x}) f_{\text{discriminate}}^i(\vec{m}). \end{aligned}$$

The most common use case is the two components “signal” (s) and “background” (b) with the discriminating variable “invariant mass” (m) and \vec{x} is only one variable x . The components can be isolated by means of the discriminating variable as illustrated in Fig. B.1. The x -distribution of the background component can be observed when restricting to the background region. The assumption of independence implies that the x -distribution in the background region is identical to the distribution of the background component in the signal region, up to the normalisation:

$$f^b(x|m \in \text{signal box}) \propto f^b(x|m \in \text{background region})$$

Further discrimination is done by determining the normalisation of the components in the signal region by means of a fit to the invariant mass distribution

$$f(m) = n_s f^s(m) + n_b f^b(m).$$

Then, the normalisation of both components in the signal box can be determined as integrals

$$\begin{aligned} \tilde{n}_s &= \int_{\text{signal box}} dm n_s f^s(m) \\ \tilde{n}_b &= \int_{\text{signal box}} dm n_b f^b(m). \end{aligned}$$

The contribution of the background component to the full x -distribution is thereby fixed and it can be subtracted from the full x distribution to isolate the signal contribution.

$$\begin{aligned} f^b(x|m \in \text{signal box}) &= \tilde{n}_b \text{Normalise} \left(f^b(x|m \in \text{background region}) \right) \\ f^s(x|m \in \text{signal box}) &= f(x|m \in \text{signal box}) - f^b(x|m \in \text{signal box}). \end{aligned}$$

Up until this point the procedure is effectively replacing $f(x)$ by a weighted distribution

$$g(x) = \int_m dm f(x, m) \cdot \begin{cases} w_s & \text{if } m \in \text{signal box} \\ -w_b & \text{if } m \in \text{background region} \end{cases}$$

where w_s and w_b are positive parameters which need to fulfil the normalisation condition that (expressed for observed distributions) the weights for the signal component sum up to the number of signal events and the weights sum up to zero for the background component:

$$\int_{\text{signal box}} dm w_1 f^s(m) + \int_{\text{background region}} dm -w_2 f^s(m) = N_s \quad (\text{B.2})$$

$$\int_{\text{signal box}} dm w_1 f^b(m) + \int_{\text{background region}} dm -w_2 f^b(m) = 0. \quad (\text{B.3})$$

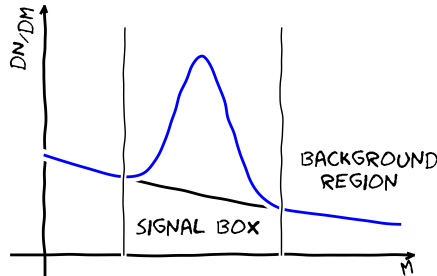


Figure B.1: Illustration for the simple decoupling of two components.

B. STATISTICS GLOSSARY

Several possible choices for the weight functions $w(m)$ fulfil these conditions. To maximise the statistical significance of the observed $f^s(x)$ distribution, the following expression must be minimised:

$$\int dm w(m) f^s(m). \quad (\text{B.4})$$

Minimising Eq. B.4 under the constraints of Eqs. B.2 and B.3 is a Lagrange multiplier problem, solved by:

$$w^s(m) = \frac{V_{ss} f^s(m) + V_{sb} f^b(m)}{N_s f^s(m) + N_b f^b(m)}, \quad (\text{B.5})$$

where the variance matrix V is defined by its inverse

$$V_{ij}^{-1} = \int dm \frac{f^i(m) f^j(m)}{(N_s f^s(m) + N_b f^b(m))^2}.$$

Tag-and-probe and TISTOS

C.1 Tag-and-probe

The tag-and-probe technique is used to measure the efficiency of the track reconstruction or particle identification at LHCb in data.

The particle identification efficiency is measured by selecting $J/\psi \rightarrow \mu\mu$ decays without imposing selection requirements on the particle identification on one of the muons (the *probe*). A fit to the invariant mass spectrum of the J/ψ candidates yields the number of true $J/\psi \rightarrow \mu\mu$ decays in that sample,

$$N_{\text{all}} := N(J/\psi, \text{w/o requiring muon ID}).$$

Any muon identification criterion can then be tested by imposing this additional criterion and determining the number of remaining candidates with a fit to the invariant mass spectrum,

$$N_{\text{good}} := N(J/\psi, \text{w/ requiring muon ID}).$$

The particle identification efficiency is then

$$\varepsilon = \frac{N_{\text{good}}}{N_{\text{all}}}.$$

This method can be applied to any muon identification criterion – the muon pre-selection and requirements $\mathcal{M}_{\text{PID}} > x$.

The track reconstruction efficiency is similarly measured by reconstructing $J/\psi \rightarrow \mu\mu$ by using the standard track reconstruction for one of the muons and an independent reconstruction for the other muon – independent in the tracking detectors. For the track reconstruction efficiency measurement, three complementary implementations exist: reconstructing muons as downstream tracks, reconstructing the muon in the TT and muon stations, and reconstructing the muon in the VELO and the muon stations. These implementations probe the long track reconstruction in the VELO, the full tracking system, and the T stations, respectively.

$$\varepsilon = \frac{N(\text{both muons found as longtrack})}{N(\text{at least one muon found as longtrack})}$$

The kinematic dependence of either efficiency is determined by restricting the J/ψ samples to those candidates where the probe is in a certain kinematic range.

$$\varepsilon(x_1 < p < x_2) = \frac{N(\text{good and } x_1 < p < x_2)}{N(\text{all and } x_1 < p < x_2)}$$

Continuous criteria can be probed more elegantly. Instead of fitting the invariant mass spectrum of $J/\psi \rightarrow \mu\mu$ candidates for any possible selection criterion $\mathcal{M}_{\text{PID}} > x$, the selection variable \mathcal{M}_{PID} is sPlot'ed, $f(\mathcal{M}_{\text{PID}})$. If the sPlot is normalised ($\int f(\mathcal{M}_{\text{PID}})d\mathcal{M}_{\text{PID}} = 1$) then the efficiency can as well be obtained by integrating the sPlot:

$$\varepsilon(\mathcal{M}_{\text{PID}} > x) = \int_x^\infty f(\mathcal{M}_{\text{PID}})d\mathcal{M}_{\text{PID}}.$$

C.2 TISTOS

The trigger efficiency for TOS triggers¹ is measured in data with TIS candidates. The invariant mass spectrum of $D_s \rightarrow \phi(\mu\mu)\pi$ TIS candidates is fitted to obtain the number of $D_s \rightarrow \phi(\mu\mu)\pi$ decays within this sample $N(\text{TIS})$. The TOS efficiency is then probed by restricting the sample further to candidates which are TIS and TOS at the same time:

$$\varepsilon_{\text{trig}} = \frac{N(\text{TIS and TOS})}{N(\text{TIS})}.$$

¹Only TOS candidates are used in this work.

b hadronisation at LHC

In this work, inclusive measurements of b branching fractions at LEP are used. It is assumed that these measurements are valid at the LHC due to similar hadronisation fractions. A short comparison of the hadronisation at LEP and at the LHC is done in this appendix.

D.1 Baryonic hadronisation

The inclusive $b \rightarrow \tau$ branching fraction is given by the sum over all b hadronisations.

$$\mathcal{B}(b \rightarrow \tau) = \sum_{i \in \{\text{b hadrons}\}} \frac{f_i}{\sum_j f_j} \cdot \mathcal{B}(h_i \rightarrow \tau)$$

It is argued in [174] and the references therein, that the semileptonic decay widths for all b mesons are equal. For the ratio of branching fractions it is stated that

$$\frac{\mathcal{B}(B_{u/d} \rightarrow D\ell X)}{\mathcal{B}(\Lambda_b \rightarrow \Lambda_c^+ \ell X)} = \frac{\tau_{B^+} + \tau_{B^0}}{2\tau_{\Lambda_b}} \cdot (1 - \xi),$$

where the ratio of life times is 1.14 ± 0.03 and the chromomagnetic correction ξ is $(4 \pm 2)\%$. Given that the mass differences $m_B - m_D$ and $m_{\Lambda_b} - m_{\Lambda_c^+}$ are about equal, no additional phase space factor is considered here.

The full $b \rightarrow \tau$ branching fraction is thus rewritten

$$\mathcal{B}(b \rightarrow \tau) = \frac{1}{\sum \frac{f_i}{f_d}} \cdot \mathcal{B}(B \rightarrow \tau) \cdot \left(\frac{f_u}{f_d} + \frac{f_d}{f_d} + \frac{f_s}{f_d} + \frac{\mathcal{B}(\Lambda_b \rightarrow \tau)}{\mathcal{B}(B \rightarrow \tau)} \cdot \frac{f_\Lambda}{f_d} \right).$$

Of all the terms on the right side of the equation, only f_Λ/f_d differs at LEP and the LHC. At first order $\mathcal{B}(b \rightarrow \tau)$ at the LHC is therefore

$$\begin{aligned} \mathcal{B}(b \rightarrow \tau)_{\text{LHC}} &= \mathcal{B}(b \rightarrow \tau)_{\text{LEP}} \\ &\times \left(1 + \frac{\left(\frac{f_\Lambda}{f_d}\right)_{\text{LHC}} - \left(\frac{f_\Lambda}{f_d}\right)_{\text{LEP}}}{\sum_{\text{hadrons}} \frac{f_i}{f_d}} \cdot \frac{\sum_{\text{mesons}} \frac{f_i}{f_d}}{\sum_{\text{hadrons}} \frac{f_i}{f_d}} \cdot \left(\frac{\mathcal{B}(\Lambda_b \rightarrow \tau)}{\mathcal{B}(B \rightarrow \tau)} - 1 \right) \right) \end{aligned}$$

Due to isospin symmetry $f_u = f_d$. It is at both, LHC and LEP, found that $f_s/f_d \sim 0.25$. The sum of the meson f_i/f_d is therefore 2.25, neglecting $f_c \sim \mathcal{O}(0.01) \cdot f_d$. The

baryonic fragmentation function at the LHC is in average $f_{\Lambda} \sim 0.4 f_d$, as averaged by eye from [132, Fig. 4], the LEP value from [80, page 1105] is 0.23. Using the above expansion at the LEP hadronisation, it follows that

$$\begin{aligned} \mathcal{B}(b \rightarrow \tau)_{\text{LHC}} &= \mathcal{B}(b \rightarrow \tau)_{\text{LEP}} \\ &\times \left(1 + \frac{0.4 - 0.23}{2.25 + 0.23} \cdot \frac{2.25}{2.25 + 0.23} \cdot \left(\frac{1}{1.14 \cdot (1 - 0.04)} - 1 \right) \right) \\ &= \mathcal{B}(b \rightarrow \tau)_{\text{LEP}} \times (1 + 0.069 \cdot 0.91 \cdot (-0.086)) \\ &= \mathcal{B}(b \rightarrow \tau)_{\text{LEP}} \times 0.99. \end{aligned}$$

The $b \rightarrow \tau$ branching fraction from the inclusive LEP measurements is therefore expected to be correct up to 1 %.

D.2 B_c^+ hadronisation

The B_c^+ production has been measured by LHCb in [175]. It is found that

$$f_c \cdot \mathcal{B}(B_c^+ \rightarrow J/\psi \pi^+) = (0.68 \pm 0.10) \% \times f_u \cdot \mathcal{B}(B^+ \rightarrow J/\psi K^+),$$

which is in agreement with the LEP measurements of $f_c \cdot \mathcal{B}(B_c^+ \rightarrow J/\psi \pi^+) < 8.2 \times 10^{-5}$. The hadronisation into charmed b mesons at LHCb and at the LEP experiments is thus in agreement.

τ production fractions

To create simulated signal samples, the production cross sections for the respective production channels need to be expressed as fractions among the final sample. This computation is in principle a simple division, e. g. $f_{\tau}^{D_s} = \sigma(\tau)_{D_s} / \sum \sigma(\tau)_i$. To correctly assess the uncertainties on the fractions, a more diligent study is necessary. The quantities, from which the cross sections are computed, are the branching fractions

- $\mathcal{B}(D_s \rightarrow \tau \nu_{\tau}) = 0.0543 \pm 0.0031$
- $\mathcal{B}(D^+ \rightarrow \tau \nu_{\tau}) = 0.00102 \pm 0.00009$
- $\mathcal{B}(b \rightarrow \tau X) = 0.0241 \pm 0.0023$
- $\mathcal{B}(b \rightarrow D_s X) = 0.248 \pm 0.037$
- $\mathcal{B}(b \rightarrow D^+ X) = 0.233 \pm 0.017$

and the cross sections

- $\sigma(pp \rightarrow D_s X)$
- $\sigma(pp \rightarrow D^+ X)$
- $\sigma(pp \rightarrow bX)$.

The latter three are correlated. To uncorrelate them, the latter two are expressed as ratios to the D_s cross section

- $\sigma(pp \rightarrow D_s X)$
- $\sigma(pp \rightarrow D^+ X) / \sigma(pp \rightarrow D_s X)$
- $\sigma(pp \rightarrow bX) / \sigma(pp \rightarrow D_s X)$.

The ratio of charm cross sections is quoted already in [129], considering the correlated uncertainties between the two measurements. The ratio is reported to be $1. / (0.305 \pm 0.042) = 3.28 \pm 0.45$. This corresponds to a correlation coefficient of 0.527.

For the ratio of the b cross section over the D_s cross section, it is assumed that from the lists of systematic uncertainty sources of [129] and [137] ([134]), the uncertainty on the luminosity measurement is maximally correlated and the uncertainty on the track finding efficiency per track is the same for both measurements. The b cross section measurement used two reconstructed tracks, the c cross section measurement used three reconstructed tracks. The track finding efficiency is thus not considered

Table E.1: Estimated correlation coefficients between the production cross sections $\sqrt{s} = 7$ TeV.

	D ⁺	D _s	b
D ⁺	1.000	0.527	0.849
D _s		1.000	0.389
b			1.000

 Table E.2: Estimated correlation coefficients between the production cross sections $\sqrt{s} = 8$ TeV.

	D ⁺	D _s	b
D ⁺	1.000	0.527	0.513
D _s		1.000	0.362
b			1.000

for the b cross section uncertainty, and reduced to the single track track-finding uncertainty for the c cross section measurement, while the hadronic interaction uncertainty¹ is kept for all three tracks. The updated values from [87] lead to a relevant relative uncertainty of the c cross section measurement due to track finding of 2.9% instead of 9%.

The maximally correlated uncertainties are subtracted from the quoted uncertainties of the measurements. The uncorrelated uncertainties are then (the uncertainties without the subtraction are given in parentheses):

$$\begin{aligned}\Delta_{\sigma(D_s)} &= 13.6\% (15.7\%) \\ \Delta_{\sigma(b),7\text{ TeV}} &= 10.8\% (16.7\%) \\ \Delta_{\sigma(b),8\text{ TeV}} &= 6.3\% (12.1\%)\end{aligned}$$

The ratio is found to be

$$\begin{aligned}\frac{\sigma(\text{pp} \rightarrow bX)}{\sigma(\text{pp} \rightarrow D_s X)} &= 0.249 \pm 0.041 && \text{for } \sqrt{s} = 7 \text{ TeV} \\ \frac{\sigma(\text{pp} \rightarrow bX)}{\sigma(\text{pp} \rightarrow D_s X)} &= 0.245 \pm 0.042 && \text{for } \sqrt{s} = 8 \text{ TeV},\end{aligned}$$

corresponding to correlation coefficients of 0.389 and 0.362, respectively.

In the same manner the correlation coefficients between the D⁺ cross section and the b cross section is computed. The correlation matrices are shown in Tabs. E.1 and E.2.

The uncertainties of the two ratios $\frac{\sigma(\text{pp} \rightarrow bX)}{\sigma(\text{pp} \rightarrow D_s X)}$ and $\frac{\sigma(\text{pp} \rightarrow D^+ X)}{\sigma(\text{pp} \rightarrow D_s X)}$ are correlated, too. The covariance is determined by error propagation of the correlation matrices in Tabs. E.1 and E.2. The covariance of the two ratios amounts to 0.0347 for $\sqrt{s} = 7$ TeV and 0.0248 for $\sqrt{s} = 8$ TeV.

Once the five fractions are expressed as functions of the 7 input variables, the Jacobi matrix A of the mapping $\vec{x} \mapsto \vec{f}_\tau$ can be computed.

$$\text{Cov}(\vec{f}_\tau) = A \text{Cov}(x_i) A^T.$$

¹See [87] for an explanation of the contribution of the hadronic interaction length to the track finding efficiency

Table E.3: Conversion of the τ production cross sections to the simulated LHCb acceptance.

τ source	cross section ($2 < y < 4.5$)	$\alpha_{4\pi}$	ϵ_{SIM}	cross section (SIM)	fraction
$\sqrt{s} = 7 \text{ TeV}$					
D_s	$(10.7 \pm 1.8) \mu\text{b}$	4.943	9.301 ± 0.033	$(4.92 \pm 0.83) \mu\text{b}$	0.705 ± 0.038
D^+	$(0.66 \pm 0.10) \mu\text{b}$	4.943	9.257 ± 0.047	$(0.30 \pm 0.05) \mu\text{b}$	0.043 ± 0.006
b	$(2.4 \pm 0.4) \mu\text{b}$	5.88	8.024 ± 0.045	$(1.11 \pm 0.19) \mu\text{b}$	0.160 ± 0.024
$b \rightarrow D_s$	$(1.32 \pm 0.36) \mu\text{b}$	5.88	8.088 ± 0.047	$(0.63 \pm 0.17) \mu\text{b}$	0.090 ± 0.017
$b \rightarrow D^+$	$(23 \pm 4) \text{nb}$	5.88	8.07 ± 0.15	$(0.0110 \pm 0.0019) \mu\text{b}$	0.00158 ± 0.00027
incl. D_s	n/a	n/a	n/a	n/a	0.795 ± 0.028
$\sqrt{s} = 8 \text{ TeV}$					
D_s	$(12.2 \pm 2.1) \mu\text{b}$	4.943	9.472 ± 0.028	$(5.71 \pm 0.98) \mu\text{b}$	0.719 ± 0.035
D^+	$(0.75 \pm 0.11) \mu\text{b}$	4.943	9.500 ± 0.027	$(0.35 \pm 0.05) \mu\text{b}$	0.044 ± 0.006
b	$(2.7 \pm 0.4) \mu\text{b}$	5.4	8.349 ± 0.046	$(1.22 \pm 0.18) \mu\text{b}$	0.151 ± 0.023
$b \rightarrow D_s$	$(1.48 \pm 0.37) \mu\text{b}$	5.4	8.392 ± 0.017	$(0.67 \pm 0.17) \mu\text{b}$	0.084 ± 0.015
$b \rightarrow D^+$	$(26 \pm 4) \text{nb}$	5.4	8.320 ± 0.015	$(0.0117 \pm 0.0018) \mu\text{b}$	0.00148 ± 0.00025
incl. D_s	n/a	n/a	n/a	n/a	0.803 ± 0.026

Table E.4: Correlation matrix for the τ production fractions at $\sqrt{s} = 7$ TeV.

	D_s	D^+	b	$b \rightarrow D_s$	$b \rightarrow D^+$	fraction
D_s	1.000	-0.602	-0.896	-0.734	-0.673	0.705 ± 0.038
D^+		1.000	0.491	-0.602	0.818	0.043 ± 0.006
b			1.000	0.384	0.578	0.160 ± 0.024
$b \rightarrow D_s$				1.000	0.376	0.090 ± 0.017
$b \rightarrow D^+$					1.000	0.00158 ± 0.00027

Table E.5: Correlation matrix for the τ production fractions at $\sqrt{s} = 8$ TeV.

	D_s	D^+	b	$b \rightarrow D_s$	$b \rightarrow D^+$	fraction
D_s	1.000	-0.455	-0.890	-0.725	-0.668	0.719 ± 0.035
D^+		1.000	0.307	0.135	0.628	0.044 ± 0.006
b			1.000	0.380	0.573	0.151 ± 0.023
$b \rightarrow D_s$				1.000	0.371	0.084 ± 0.015
$b \rightarrow D^+$					1.000	0.00148 ± 0.00025

Table E.6: Nominal τ production fractions, variations by 1σ for the principle components of the covariance matrix, and for variations of $\mathcal{B}(D_s \rightarrow \tau\nu_\tau)$ by 1σ . For collision energies of 7 TeV.

	nominal					nominal $+1\sigma_{\mathcal{B}(D_s \rightarrow \tau\nu_\tau)}$
D_s	0.705	0.742	0.707	0.707	0.705	0.713
D^+	0.043	0.040	0.043	0.039	0.043	0.041
b	0.160	0.139	0.169	0.162	0.160	0.153
$b \rightarrow D_s$	0.090	0.078	0.079	0.092	0.090	0.091
$b \rightarrow D^+$	0.00158	0.00142	0.00159	0.00144	0.00173	0.00151
incl. D_s	0.795	0.820	0.786	0.799	0.795	0.804

Table E.7: Nominal τ production fractions, variations by 1σ for the principle components of the covariance matrix, and for variations of $\mathcal{B}(D_s \rightarrow \tau\nu_\tau)$ by 1σ . For collision energies of 8 TeV.

	nominal					nominal $+1\sigma_{\mathcal{B}(D_s \rightarrow \tau\nu_\tau)}$
D_s	0.719	0.752	0.721	0.720	0.719	0.727
D^+	0.044	0.042	0.044	0.039	0.044	0.042
b	0.151	0.131	0.159	0.153	0.151	0.144
$b \rightarrow D_s$	0.084	0.073	0.074	0.087	0.085	0.086
$b \rightarrow D^+$	0.00148	0.00133	0.00148	0.00138	0.00164	0.00141
incl. D_s	0.803	0.825	0.795	0.807	0.804	0.813

The computation is done using computer algebra software, the resulting correlation matrix is given in Tabs. E.4 and E.5.

The covariance matrix is diagonalised to obtain the uncorrelated degrees of freedom of the fractions in Tab. E.3. For each degree of freedom, a 1σ change is given in Tabs. E.6 and E.7, they are sorted in magnitude of the eigenvalues of the covariance matrix. Since the D_s branching fraction $\mathcal{B}(D_s \rightarrow \tau\nu_\tau)$ appears in both, the production fractions and the normalisation factor, the uncertainty due to $\mathcal{B}(D_s \rightarrow \tau\nu_\tau)$ is treated separately and set to 0 for the decorrelation of the covariance matrix. The uncertainty on the D_s cross section measurement due to the uncertainty on $\mathcal{B}(D_s \rightarrow \phi\pi)$ is not excluded, because a different measurement is used in [129] compared to this work. The fifth eigenvalue of the covariance matrix is 17 orders of magnitude smaller than the largest. The 1σ change corresponding to this degree of freedom is thus not visible with the printed numbers of digits and omitted in the table.

The mixing fractions resulting from a 1σ change of $\mathcal{B}(D_s \rightarrow \tau\nu_\tau)$ are given in the last column of Tab. E.6 and E.7, respectively.

Surprisingly, and fortunately, the principle components describe the same physics behaviour. (Otherwise the treatment of the correlated uncertainties of the production fractions among the data taking periods would become less trivial.) The first component is an enhancement of prompt D_s , compensated by a reduction of all other modes. The second component is an enhancement of the $b \rightarrow \tau$ production, compensated by the $b \rightarrow D_s \rightarrow \tau$ production. The third component is a reduction of the prompt and

E. τ PRODUCTION FRACTIONS

non-prompt D^+ contribution. The fourth is an enhancement of the non-prompt D^+ contribution.

References

- [1] E. Noether. “Invariant Variation Problems”. In: *Gott.Nachr.* 1918 (1918), pp. 235–257. DOI: 10.1080/00411457108231446. arXiv:physics/0503066 [physics].
- [2] Y. Amhis et al., The Heavy Flavor Averaging Group. “Averages of b-hadron, c-hadron, and τ -lepton properties as of early 2012”. preprint. 2012. arXiv:1207.1158 [hep-ex].
- [3] K. Hayasaka et al., Belle Collaboration. “Search for Lepton Flavor Violating τ Decays into Three Leptons with 719 Million Produced $\tau^+\tau^-$ Pairs”. In: *Phys.Lett.* B687 (2010), pp. 139–143. DOI: 10.1016/j.physletb.2010.03.037. arXiv:1001.3221 [hep-ex].
- [4] R. Aaij et al., LHCb Collaboration. “Searches for violation of lepton flavour and baryon number in tau lepton decays at LHCb”. In: *Phys.Lett.* B724 (2013), pp. 36–45. DOI: 10.1016/j.physletb.2013.05.063. arXiv:1304.4518 [hep-ex].
- [5] R. Aaij et al., LHCb Collaboration. “Search for the lepton flavour violating decay $\tau^- \rightarrow \mu^- \mu^+ \mu^-$ ”. submitted to JHEP. 2014. arXiv:1409.8548 [hep-ex].
- [6] J. Albrecht et al. “Search for the lepton flavour violating decay $\tau^- \rightarrow \mu^+ \mu^- \mu^-$ ”. LHCb Analysis Note LHCb-ANA-2011-100. URL: <https://cds.cern.ch/record/1403053>.
- [7] J. Albrecht et al. “Search for Lepton Flavor Violation and Baryon Number Violation in τ^- decays”. LHCb Analysis Note LHCb-ANA-2012-095. URL: <https://cds.cern.ch/record/1484844>.
- [8] J. Albrecht et al. “Search for the lepton flavour violating decay $\tau \rightarrow \mu^+ \mu^- \mu^-$ ”. LHCb Analysis Note LHCb-ANA-2014-005. URL: <https://cds.cern.ch/record/1645011>.
- [9] R. Caruana et al. “Ensemble Selection from Libraries of Models”. In: *Proceedings of the Twenty-first International Conference on Machine Learning*. ICML ’04. ACM, 2004. DOI: 10.1145/1015330.1015432.
- [10] A. Gulin, I. Kuralenok, and D. Pavlov. “Winning the transfer learning track of Yahoo!’s Learning to Rank Challenge with YetiRank”. In: *JMLR Workshop and Conference Proceedings. Yahoo! Learning to Rank Challenge*. Vol. 14. 2011, p. 63. URL: <http://jmlr.csail.mit.edu/proceedings/papers/v14/>.
- [11] M. Chrzaszcz. private communication. Computed the world-combined LFV exclusion limits in [171]. 2014.

REFERENCES

- [12] The Wikipedia Community. *Wikipedia, The Free Encyclopedia*. [Online]. 2014. URL: <http://en.wikipedia.org/w/index.php>.
- [13] S.L. Glashow. “Partial Symmetries of Weak Interactions”. In: *Nucl.Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.
- [14] A. Salam and J.C. Ward. “Electromagnetic and weak interactions”. In: *Phys.Lett.* 13 (1964), pp. 168–171. DOI: 10.1016/0031-9163(64)90711-5.
- [15] S. Weinberg. “A Model of Leptons”. In: *Phys.Rev.Lett.* 19 (1967). Inaccessible to the general public, pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.
- [16] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. “Advantages of the Color Octet Gluon Picture”. In: *Phys.Lett.* B47 (1973). Inaccessible to the general public, pp. 365–368. DOI: 10.1016/0370-2693(73)90625-4.
- [17] D.J. Gross and F. Wilczek. “Ultraviolet Behavior of Nonabelian Gauge Theories”. In: *Phys.Rev.Lett.* 30 (1973). Inaccessible to the general public, pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.
- [18] S. Weinberg. “Nonabelian Gauge Theories of the Strong Interactions”. In: *Phys.Rev.Lett.* 31 (1973). Inaccessible to the general public, pp. 494–497. DOI: 10.1103/PhysRevLett.31.494.
- [19] P.W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys.Rev.Lett.* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.
- [20] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys.Rev.Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.
- [21] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys.Rev.Lett.* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.
- [22] G. Aad et al., ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys.Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv:1207.7214 [hep-ex].
- [23] S. Chatrchyan et al., CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys.Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv:1207.7235 [hep-ex].
- [24] A.D. Sakharov. “Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe”. In: *Pisma Zh.Eksp.Teor.Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.
- [25] P. Huet and E. Sather. “Electroweak baryogenesis and standard model CP violation”. In: *Phys.Rev.* D51 (1995), pp. 379–394. DOI: 10.1103/PhysRevD.51.379. arXiv:hep-ph/9404302 [hep-ph].
- [26] I. [Hrsg.] Brock, ed. *Physics at the Terascale*. eng. 1. Aufl. Weinheim: Wiley-VCH, 2011, XXXVIII, 476 S. ISBN: 978-3-527-41001-9 ; 3-527-41001-5.
- [27] N. Arkani-Hamed, S. Dimopoulos, and G.R. Dvali. “The Hierarchy problem and new dimensions at a millimeter”. In: *Phys.Lett.* B429 (1998), pp. 263–272. DOI: 10.1016/S0370-2693(98)00466-3. arXiv:hep-ph/9803315 [hep-ph].

-
- [28] M. Drees. “An Introduction to supersymmetry”. In: *Lectures given at: Inauguration Conference of the Asia Pacific Center for Theoretical Physics (APCTP)*. 1996. arXiv:hep-ph/9611409 [hep-ph].
- [29] K. Eguchi et al., KamLAND Collaboration. “First results from KamLAND: Evidence for reactor anti-neutrino disappearance”. In: *Phys.Rev.Lett.* 90 (2003), p. 021802. DOI: 10.1103/PhysRevLett.90.021802. arXiv:hep-ex/0212021 [hep-ex].
- [30] Q.R. Ahmad et al., SNO Collaboration. “Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory”. In: *Phys.Rev.Lett.* 87 (2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301. arXiv:nucl-ex/0106015 [nucl-ex].
- [31] Y. Fukuda et al., Super-Kamiokande Collaboration. “Evidence for oscillation of atmospheric neutrinos”. In: *Phys.Rev.Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv:hep-ex/9807003 [hep-ex].
- [32] Z. Maki, M. Nakagawa, and S. Sakata. “Remarks on the unified model of elementary particles”. In: *Prog.Theor.Phys.* 28 (1962), pp. 870–880. DOI: 10.1143/PTP.28.870.
- [33] M. Giffels. “Study of the Sensitivity of CMS to the Lepton Flavour Violating Neutrinoless τ Decay $\tau \rightarrow \mu\mu\mu$ ”. PhD thesis. RWTH Aachen University, 2009. URL: <https://cds.cern.ch/record/1292236>.
- [34] J.H.J. von Loeben. “Calibration of the ATLAS Precision Muon Chambers and Study of the Decay $\tau \rightarrow \mu\mu\mu$ at the Large Hadron Collider”. CERN-THESIS-2010-099. PhD thesis. Munich, Max Planck Inst., 2010. URL: <https://cds.cern.ch/record/1283473>.
- [35] L. Gavardi. “Search for lepton flavour violation in τ decays at the LHCb experiment”. CERN-THESIS-2013-259. Master thesis. Milan Bicocca U., 2013. URL: <https://cds.cern.ch/record/1645251>.
- [36] M. Giffels et al. “The Lepton-flavour violating decay $\tau \rightarrow \mu\mu\bar{\mu}$ at the CERN LHC”. In: *Phys.Rev.* D77 (2008), p. 073010. DOI: 10.1103/PhysRevD.77.073010. arXiv:0802.0049 [hep-ph].
- [37] J.L. Hewett, ed. *The Discovery potential of a Super B Factory*. Proceedings, SLAC Workshops, Stanford, USA, 2003. 2004. arXiv:hep-ph/0503261 [hep-ph].
- [38] J.R. Harrison. “Radiation damage studies in the LHCb VELO detector and searches for lepton flavour and baryon number violating tau decays”. CERN-THESIS-2014-068. PhD thesis. University of Manchester, 2014. URL: <https://cds.cern.ch/record/1712972>.
- [39] P. Paradisi. “LFV searches”. In: [176]. Ed. by J.R. Ellis et al. First workshop. Resulted in [84]. 2011. URL: <http://indico.cern.ch/event/155764/>.
- [40] M. Blanke et al. “Charged Lepton Flavour Violation and $(g-2)_\mu$ in the Lightest Higgs Model with T-Parity: A Clear Distinction from Supersymmetry”. In: *JHEP* 0705 (2007). As pointed out in the note added to the latest version, this reference neglects the contribution of the Z penguin. An updates relevant for $\tau \rightarrow \mu\mu\mu$ are available in [49]., p. 013. DOI: 10.1088/1126-6708/2007/05/013. arXiv:hep-ph/0702136 [hep-ph].

REFERENCES

- [41] M. Hirsch, F. Staub, and A. Vicente. “Enhancing $l_i \rightarrow 3l_j$ with the Z^0 -penguin”. In: *Phys.Rev.* D85 (2012), p. 113013. DOI: 10.1103/PhysRevD.85.113013. arXiv:1202.1825 [hep-ph].
- [42] T. Goto, Y. Okada, and Y. Yamamoto. “Ultraviolet divergences of flavor changing amplitudes in the littlest Higgs model with T-parity”. In: *Phys.Lett.* B670 (2009), pp. 378–382. DOI: 10.1016/j.physletb.2008.11.022. arXiv:0809.4753 [hep-ph].
- [43] F. del Aguila, J.I. Illana, and M.D. Jenkins. “Precise limits from lepton flavour violating processes on the Littlest Higgs model with T-parity”. In: *JHEP* 0901 (2009), p. 080. DOI: 10.1088/1126-6708/2009/01/080. arXiv:0811.2891 [hep-ph].
- [44] A. Vicente. “Lepton Flavor Violating Decays”. In: [177]. Ed. by J.R. Ellis et al. 4th workshop. 2014. URL: <https://indico.cern.ch/event/324660/>.
- [45] J.R. Ellis et al. “A New parametrization of the seesaw mechanism and applications in supersymmetric models”. In: *Phys.Rev.* D66 (2002), p. 115013. DOI: 10.1103/PhysRevD.66.115013. arXiv:hep-ph/0206110 [hep-ph].
- [46] A. Dedes, J.R. Ellis, and M. Raidal. “Higgs mediated $B_{s,d}^0 \rightarrow \mu\tau, e\tau$ and $\tau \rightarrow 3\mu, e\mu\mu$ decays in supersymmetric seesaw models”. In: *Phys.Lett.* B549 (2002), pp. 159–169. DOI: 10.1016/S0370-2693(02)02900-3. arXiv:hep-ph/0209207 [hep-ph].
- [47] K.S. Babu and C. Kolda. “Higgs mediated $\tau \rightarrow 3\mu$ in the supersymmetric seesaw model”. In: *Phys.Rev.Lett.* 89 (2002), p. 241802. DOI: 10.1103/PhysRevLett.89.241802. arXiv:hep-ph/0206310 [hep-ph].
- [48] H.K. Dreiner et al. “New bounds on trilinear R -parity violation from lepton flavor violating observables”. In: *Phys.Rev.* D86 (2012), p. 015003. DOI: 10.1103/PhysRevD.86.015003. arXiv:1204.5925 [hep-ph].
- [49] M. Blanke et al. “FCNC Processes in the Littlest Higgs Model with T-Parity: an Update”. In: *Acta Phys.Polon.* B41 (2010), pp. 657–683. arXiv:0906.5454 [hep-ph].
- [50] A.G. Akeroyd, M. Aoki, and Y. Okada. “Lepton Flavour Violating tau Decays in the Left-Right Symmetric Model”. In: *Phys.Rev.* D76 (2007), p. 013004. DOI: 10.1103/PhysRevD.76.013004. arXiv:hep-ph/0610344 [hep-ph].
- [51] A.J. Buras et al. “Lepton Flavour Violation in the Presence of a Fourth Generation of Quarks and Leptons”. In: *JHEP* 1009 (2010), p. 104. DOI: 10.1007/JHEP09(2010)104. arXiv:1006.5356 [hep-ph].
- [52] C. Yue, Y. Zhang, and L. Liu. “Nonuniversal gauge bosons Z' and lepton flavor violation tau decays”. In: *Phys.Lett.* B547 (2002), pp. 252–256. DOI: 10.1016/S0370-2693(02)02781-8. arXiv:hep-ph/0209291 [hep-ph].
- [53] A. Hektor, Y. Kajiyama, and K. Kannike. “Muon Anomalous Magnetic Moment and Lepton Flavor Violating Tau Decay in Unparticle Physics”. In: *Phys.Rev.* D78 (2008), p. 053008. DOI: 10.1103/PhysRevD.78.053008. arXiv:0802.4015 [hep-ph].
- [54] E. Arganda and M.J. Herrero. “Testing supersymmetry with lepton flavor violating τ and μ decays”. In: *Phys.Rev.* D73 (2006), p. 055003. DOI: 10.1103/PhysRevD.73.055003. arXiv:hep-ph/0510405 [hep-ph].

-
- [55] P. Paradisi. “Higgs-mediated $\tau \rightarrow \mu$ and $\tau \rightarrow e$ transitions in II Higgs doublet model and supersymmetry”. In: *JHEP* 0602 (2006), p. 050. DOI: 10.1088/1126-6708/2006/02/050. arXiv:hep-ph/0508054 [hep-ph].
- [56] R. Barbier et al. “ R -parity violating supersymmetry”. In: *Phys.Rept.* 420 (2005), pp. 1–202. DOI: 10.1016/j.physrep.2005.08.006. arXiv:hep-ph/0406039 [hep-ph].
- [57] T. Goto, Y. Okada, and Y. Yamamoto. “Tau and muon lepton flavor violations in the littlest Higgs model with T-parity”. In: *Phys.Rev.* D83 (2011), p. 053011. DOI: 10.1103/PhysRevD.83.053011. arXiv:1012.4385 [hep-ph].
- [58] N. Arkani-Hamed, A.G. Cohen, and H. Georgi. “Electroweak symmetry breaking from dimensional deconstruction”. In: *Phys.Lett.* B513 (2001), pp. 232–240. DOI: 10.1016/S0370-2693(01)00741-9. arXiv:hep-ph/0105239 [hep-ph].
- [59] N. Arkani-Hamed et al. “Phenomenology of electroweak symmetry breaking from theory space”. In: *JHEP* 0208 (2002), p. 020. arXiv:hep-ph/0202089 [hep-ph].
- [60] N. Arkani-Hamed et al. “The Minimal moose for a little Higgs”. In: *JHEP* 0208 (2002), p. 021. DOI: 10.1088/1126-6708/2002/08/021. arXiv:hep-ph/0206020 [hep-ph].
- [61] H.-C. Cheng and I. Low. “TeV symmetry and the little hierarchy problem”. In: *JHEP* 0309 (2003), p. 051. DOI: 10.1088/1126-6708/2003/09/051. arXiv:hep-ph/0308199 [hep-ph].
- [62] H.-C. Cheng and I. Low. “Little hierarchy, little Higgses, and a little symmetry”. In: *JHEP* 0408 (2004), p. 061. DOI: 10.1088/1126-6708/2004/08/061. arXiv:hep-ph/0405243 [hep-ph].
- [63] J. Hubisz et al. “Electroweak precision constraints on the littlest Higgs model with T parity”. In: *JHEP* 0601 (2006), p. 135. DOI: 10.1088/1126-6708/2006/01/135. arXiv:hep-ph/0506042 [hep-ph].
- [64] J.N. Esteves et al. “LHC and lepton flavour violation phenomenology of a left-right extension of the MSSM”. In: *JHEP* 1012 (2010), p. 077. DOI: 10.1007/JHEP12(2010)077. arXiv:1011.0348 [hep-ph].
- [65] A.J. Buras et al. “Patterns of Flavour Violation in the Presence of a Fourth Generation of Quarks and Leptons”. In: *JHEP* 1009 (2010), p. 106. DOI: 10.1007/JHEP09(2010)106. arXiv:1002.2126 [hep-ph].
- [66] R. Aaij et al., LHCb Collaboration. “Search for the lepton-flavor violating decays $B_s^0 \rightarrow e^\pm \mu^\mp$ and $B^0 \rightarrow e^\pm \mu^\mp$ ”. In: *Phys.Rev.Lett.* 111.14 (2013), p. 141801. DOI: 10.1103/PhysRevLett.111.141801. arXiv:1307.4889 [hep-ex].
- [67] A. Lenz. “Constraints on a fourth generation of fermions from Higgs Boson searches”. In: *Adv.High Energy Phys.* 2013 (2013), p. 910275. DOI: 10.1155/2013/910275.
- [68] W. Altmannshofer. “New Physics Interpretations of the $B \rightarrow K^* \mu^+ \mu^-$ Anomaly”. In: *49th Rencontres de Moriond on Electroweak Interactions and Unified Theories.* 2014. arXiv:1405.5182 [hep-ph].
- [69] R. Gauld, F. Goertz, and U. Haisch. “An explicit Z' -boson explanation of the $B \rightarrow K^* \mu^+ \mu^-$ anomaly”. In: *JHEP* 1401 (2014), p. 069. DOI: 10.1007/JHEP01(2014)069. arXiv:1310.1082 [hep-ph].

REFERENCES

- [70] T. Banks and A. Zaks. “On the Phase Structure of Vector-Like Gauge Theories with Massless Fermions”. In: *Nucl.Phys.* B196 (1982). Inaccessible to the general public, p. 189. DOI: 10.1016/0550-3213(82)90035-9.
- [71] H. Georgi. “Unparticle physics”. In: *Phys.Rev.Lett.* 98 (2007), p. 221601. DOI: 10.1103/PhysRevLett.98.221601. arXiv:hep-ph/0703260 [hep-ph].
- [72] B.M. Dassingier et al. “Model-independent analysis of lepton flavour violating tau decays”. In: *JHEP* 0710 (2007), p. 039. DOI: 10.1088/1126-6708/2007/10/039. arXiv:0707.0988 [hep-ph].
- [73] A. Celis, V. Cirigliano, and E. Passemar. “Disentangling new physics contributions in lepton flavour violating tau decays”. In: *37th International Conference on High Energy Physics (ICHEP 2014)*. 2014. arXiv:1409.4439 [hep-ph].
- [74] N. Agafonova et al., OPERA Collaboration. “Observation of ν_τ appearance in the CNGS beam with the OPERA experiment”. preprint. 2014. arXiv:1407.3513 [hep-ex].
- [75] K. Abe et al., Super-Kamiokande Collaboration. “A Measurement of the Appearance of Atmospheric Tau Neutrinos by Super-Kamiokande”. In: *Phys.Rev.Lett.* 110 (2013), p. 181802. DOI: 10.1103/PhysRevLett.110.181802. arXiv:1206.0328 [hep-ex].
- [76] B. Aubert et al., BaBar Collaboration. “Searches for Lepton Flavor Violation in the Decays $\tau^\pm \rightarrow e^\pm \gamma$ and $\tau^\pm \rightarrow \mu^\pm \gamma$ ”. In: *Phys.Rev.Lett.* 104 (2010), p. 021802. DOI: 10.1103/PhysRevLett.104.021802. arXiv:0908.2381 [hep-ex].
- [77] K. Hayasaka et al., Belle Collaboration. “New search for $\tau \rightarrow \mu \gamma$ and $\tau \rightarrow e \gamma$ decays at Belle”. In: *Phys.Lett.* B666 (2008), pp. 16–22. DOI: 10.1016/j.physletb.2008.06.056. arXiv:0705.0650 [hep-ex].
- [78] K. Hayasaka. “Lepton Flavor Violating tau decays at Belle”. In: *The 13th International Workshop on Tau Lepton Physics (TAU2014)*. 2014. URL: <https://indico.cern.ch/event/300387/session/16/contribution/40>.
- [79] J.P. Lees et al., Babar Collaboration. “A search for the decay modes $B^\pm \rightarrow h^\pm \tau \ell$ ”. In: *Phys.Rev.* D86 (2012), p. 012004. DOI: 10.1103/PhysRevD.86.012004. arXiv:1204.2852 [hep-ex].
- [80] J. Beringer et al., Particle Data Group. “Review of Particle Physics (RPP)”. In: *Phys.Rev.* D86 (2012), p. 010001. DOI: 10.1103/PhysRevD.86.010001.
- [81] The CMS Collaboration. “Search for Lepton Flavour Violating Decays of the Higgs Boson”. CMS Physics Analysis Summary CMS-PAS-HIG-14-005. 2014. URL: <http://cds.cern.ch/record/1740976>.
- [82] A.A. Alves Jr. et al., LHCb Collaboration. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.
- [83] B. Adeva et al., LHCb Collaboration. “Roadmap for selected key measurements of LHCb”. preprint. 2009. arXiv:0912.4179 [hep-ex].
- [84] R. Aaij et al., LHCb Collaboration. “Implications of LHCb measurements and future prospects”. In: *Eur.Phys.J.* C73 (2013), p. 2373. DOI: 10.1140/epjc/s10052-013-2373-2. arXiv:1208.3355 [hep-ex].
- [85] R. Aaij et al., LHCb Collaboration. “LHCb Detector Performance”. in preparation. LHCb-DP-2014-002. 2014.

- [86] R. Aaij et al., LHCb trigger group. “The LHCb Trigger and its Performance in 2011”. In: *JINST* 8 (2013), P04022. DOI: 10.1088/1748-0221/8/04/P04022. arXiv:1211.3055 [hep-ex].
- [87] R. Aaij et al., LHCb Collaboration. “Measurement of the track reconstruction efficiency at LHCb”. submitted to JINST. LHCb-DP-2013-002. 2014. arXiv:1408.1251 [hep-ex].
- [88] R. Aaij et al., LHCb VELO Group. “Performance of the LHCb Vertex Locator”. In: *JINST* 9 (2014), p. 09007. DOI: 10.1088/1748-0221/9/09/P09007. arXiv:1405.7808 [physics.ins-det].
- [89] C. Abellan Beteta et al., LHCb Calorimeters Group. “Performance of the LHCb Calorimeters”. in preparation. LHCb-DP-2013-004. 2014.
- [90] M. Adinolfi et al., LHCb RICH Group. “Performance of the LHCb RICH detector at the LHC”. In: *Eur.Phys.J. C* 73 (2013), p. 2431. DOI: 10.1140/epjc/s10052-013-2431-9. arXiv:1211.6759 [physics.ins-det].
- [91] A.A. Alves Jr. et al. “Performance of the LHCb muon system”. In: *JINST* 8 (2013), P02022. DOI: 10.1088/1748-0221/8/02/P02022. arXiv:1211.1346 [physics.ins-det].
- [92] J. Albrecht et al., LHCb HLT project. “Performance of the LHCb High Level Trigger in 2012”. In: *J.Phys.Conf.Ser.* 513 (2014), p. 012001. DOI: 10.1088/1742-6596/513/1/012001. arXiv:1310.8544 [hep-ex].
- [93] F. Archilli et al. “Performance of the muon identification at LHCb”. In: *JINST* 8 (2013), P10020. DOI: 10.1088/1748-0221/8/10/P10020. arXiv:1306.0249 [physics.ins-det].
- [94] R. Arink et al., LHCb Outer Tracker group. “Performance of the LHCb Outer Tracker”. In: *JINST* 9 (2014), p. 01002. DOI: 10.1088/1748-0221/9/01/P01002. arXiv:1311.3893 [physics.ins-det].
- [95] E. Aslanides et al. “The Level-0 muon trigger for the LHCb experiment”. In: *Nucl.Instrum.Meth.* A579 (2007), pp. 989–1004. DOI: 10.1016/j.nima.2007.06.022. arXiv:0705.0310 [physics.ins-det].
- [96] M. Vesterinen, on behalf of the LHCb Collaboration. “Considerations on the LHCb dipole magnet polarity reversal”. LHCb Public Note LHCb-PUB-2014-006. 2014. URL: <https://cds.cern.ch/record/1642153>.
- [97] A. Kozlinskiy. “Outer Tracker calibration and open charm production cross section measurement at LHCb”. CERN-THESIS-2012-338. PhD thesis. Vrije Universiteit Amsterdam, 2012. URL: <https://cds.cern.ch/record/1594180>.
- [98] M. De Cian. “Track Reconstruction Efficiency and Analysis of $B^0 \rightarrow K^* \mu^+ \mu^-$ at the LHCb Experiment”. CERN-THESIS-2013-145. PhD thesis. Universität Zürich, 2013. URL: <https://cds.cern.ch/record/1605179>.
- [99] F. Dordei. “Measurement of beauty hadron lifetimes (tbc)”. in preparation. PhD thesis. Heidelberg University, 2014. URL: <http://www.physi.uni-heidelberg.de/Forschung/he/LHCb/finishedTheses/>.
- [100] P.V.C. Hough. “Machine Analysis Of Bubble Chamber Pictures”. In: *2nd International Conference on High-Energy Accelerators (HEACC 59)*. Vol. C590914. 1959, pp. 554–558. URL: http://inspirehep.net/record/919922/files/HEACC59_598-602.pdf.

REFERENCES

- [101] R.O. Duda and P.E. Hart. “Use of the Hough transformation to detect lines and curves in pictures”. In: *Commun. ACM* 15.1 (1972). Inaccessible to the general public, pp. 11–15. DOI: 10.1145/361237.361242.
- [102] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nucl.Instrum.Meth.* A262.2-3 (1987). Inaccessible to the general public, pp. 444–450. ISSN: 0168-9002. DOI: 10.1016/0168-9002(87)90887-4.
- [103] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1 (1951), p. 79. URL: <http://www.jstor.org/stable/2236703>.
- [104] M. Needham. “Clone Track Identification using the Kullback-Liebler Distance”. LHCb Note LHCb-2008-002 (public). 2008. URL: <https://cds.cern.ch/record/1082460>.
- [105] M. Alemi et al., TA2 Collaboration. “First operation of a hybrid photon detector prototype with electrostatic cross-focussing and integrated silicon pixel read-out”. In: *Nucl.Instrum.Meth.* A449 (2000). Inaccessible to the general public, pp. 48–59. DOI: 10.1016/S0168-9002(99)01448-5.
- [106] G. Bencivenni et al. “A triple-GEM detector with pad readout for the inner region of the first LHCb muon station”. LHCb Note LHCb-2001-051. 2001. URL: <https://cds.cern.ch/record/691604>.
- [107] M. Frank et al. “Deferred High Level Trigger in LHCb: A Boost to CPU Resource Utilization”. In: *J.Phys.Conf.Ser.* 513.1 (2014), p. 012006. DOI: 10.1088/1742-6596/513/1/012006.
- [108] S. Tolk et al. “Data driven trigger efficiency determination at LHCb”. LHCb Public Note LHCb-PUB-2014-039. 2014. URL: <http://cds.cern.ch/record/1701134>.
- [109] T. Sjöstrand, S. Mrenna, and P. Skands. “PYTHIA 6.4 physics and manual”. In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv:hep-ph/0603175 [hep-ph].
- [110] T. Sjöstrand, S. Mrenna, and P. Skands. “A brief introduction to PYTHIA 8.1”. In: *Comput.Phys.Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv:0710.3820 [hep-ph].
- [111] I. Belyaev et al. “Handling of the generation of primary events in Gauss, the LHCb simulation framework”. In: *Nuclear Science Symposium Conference Record (NSS/MIC) IEEE* (2010). Inaccessible to the general public, p. 1155. DOI: 10.1109/NSSMIC.2010.5873949.
- [112] D.J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl.Instrum.Meth.* A462 (2001). Inaccessible to the general public, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.
- [113] P. Golonka and Z. Was. “PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays”. In: *Eur.Phys.J.* C45 (2006), pp. 97–107. DOI: 10.1140/epjc/s2005-02396-4. arXiv:hep-ph/0506026 [hep-ph].
- [114] J. Allison et al., Geant4 Collaboration. “Geant4 developments and applications”. In: *IEEE Trans.Nucl.Sci.* 53 (2006). Inaccessible to the general public, p. 270. DOI: 10.1109/TNS.2006.869826.

- [115] S. Agostinelli et al., Geant4 Collaboration. “Geant4: a simulation toolkit”. In: *Nucl.Instrum.Meth.* A506 (2003). Inaccessible to the general public, p. 250. DOI: 10.1016/S0168-9002(03)01368-8.
- [116] M. Clemencic et al. “The LHCb simulation application, Gauss: Design, evolution and experience”. In: *J.Phys.Conf.Ser.* 331 (2011), p. 032023. DOI: 10.1088/1742-6596/331/3/032023.
- [117] A.L. Read. “Presentation of search results: The CL(s) technique”. In: *J.Phys.* G28 (2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.
- [118] T. Junk. *Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties*. CDF public note CDF/DOC/STATISTICS/PUBLIC/8128. 2012. URL: http://www-cdf.fnal.gov/~trj/mclimit/mclimit_csm.pdf.
- [119] A. Read. “Statistics for Theorists”. In: *Higgs Days at Santander. Theory meets Experiment*. Ed. by S. Heinemeyer. 2011. URL: <http://indico.ifca.es/indico/contributionDisplay.py?contribId=14&sessionId=10&confId=378>.
- [120] A. Rogozhnikov et al. “New approaches for boosting to uniformity”. preprint. 2014. arXiv:1410.4140 [hep-ex].
- [121] H.B. Prosper. “Multivariate methods in particle physics: Today and tomorrow”. In: *PoS ACAT08* (2008), p. 010. URL: http://inspirehep.net/record/858299/files/ACAT08_010.pdf.
- [122] H. Voss. “Multivariate Data Analysis in HEP. Successes, challenges and future outlook.” In: *16th International workshop on Advanced Computing and Analysis Techniques in physics research (ACAT)*. 2014. URL: <https://indico.cern.ch/event/258092/session/0/contribution/124>.
- [123] P. Baldi, P. Sadowski, and D. Whiteson. “Enhanced Higgs to $\tau^+\tau^-$ Searches with Deep Learning”. preprint. 2014. arXiv:1410.3469 [hep-ph].
- [124] G. Punzi. “Sensitivity of searches for new signals and its optimization”. In: *eConf C030908* (2003), MODT002. arXiv:physics/0308063 [physics].
- [125] G. Ranucci. “The Profile likelihood ratio and the look elsewhere effect in high energy physics”. In: *Nucl.Instrum.Meth.* A661 (2012), pp. 77–85. DOI: 10.1016/j.nima.2011.09.047. arXiv:1201.4604 [physics.data-an].
- [126] J. Neyman and E.S. Pearson. “On the Problem of the Most Efficient Tests of Statistical Hypotheses”. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009. URL: <http://rsta.royalsocietypublishing.org/content/231/694-706/289.full.pdf+html>.
- [127] M. Perrin-Terrin. “Searches for B meson decays to purely leptonic final states”. CERN-THESIS-2013-249. PhD thesis. Aix-Marseille U., 2013. URL: <https://cds.cern.ch/record/1642541>.
- [128] T. Junk. “Confidence Level Computation for Combining Searches with Small Statistics”. In: *Nucl.Instrum.Meth.* A434 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv:hep-ex/9902006.

REFERENCES

- [129] R. Aaij et al., LHCb Collaboration. “Prompt charm production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Nucl.Phys.* B871 (2013), pp. 1–20. DOI: 10.1016/j.nuclphysb.2013.02.010. arXiv:1302.2864 [hep-ex].
- [130] B.I. Eisenstein et al., CLEO Collaboration. “Precision Measurement of $\mathcal{B}(D^+ \rightarrow \mu^+\nu)$ and the Pseudoscalar Decay Constant f_{D^+} ”. In: *Phys.Rev.* D78 (2008), p. 052003. DOI: 10.1103/PhysRevD.78.052003. arXiv:0806.2112 [hep-ex].
- [131] R. Aaij et al., LHCb Collaboration. “Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics”. In: *JHEP* 1304 (2013), p. 001. DOI: 10.1007/JHEP04(2013)001. arXiv:1301.5286 [hep-ex].
- [132] R. Aaij et al., LHCb Collaboration. “Study of the kinematic dependences of Λ_b production in pp collisions and a measurement of the $\Lambda_b \rightarrow \Lambda_c^+\pi^-$ branching fraction”. In: *JHEP* 1408 (2014), p. 143. DOI: 10.1007/JHEP08(2014)143. arXiv:1405.6842 [hep-ex].
- [133] R. Aaij et al., LHCb Collaboration. “Measurement of J/ψ production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Eur.Phys.J.* C71 (2011), p. 1645. DOI: 10.1140/epjc/s10052-011-1645-y. arXiv:1103.0423 [hep-ex].
- [134] R. Aaij et al., LHCb Collaboration. “Production of J/ψ and Υ mesons in pp collisions at $\sqrt{s} = 8$ TeV”. In: *JHEP* 1306 (2013), p. 064. DOI: 10.1007/JHEP06(2013)064. arXiv:1304.6977 [hep-ex].
- [135] M. Ablikim et al., BES Collaboration. “Improved measurement of $\psi(2S)$ decays into $\tau^+\tau^-$ ”. In: *Phys.Rev.* D74 (2006), p. 112003. DOI: 10.1103/PhysRevD.74.112003.
- [136] R. Aaij et al., LHCb Collaboration. “Measurement of $\psi(2S)$ meson production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Eur.Phys.J.* C72 (2012), p. 2100. DOI: 10.1140/epjc/s10052-012-2100-4. arXiv:1204.1258 [hep-ex].
- [137] R. Aaij et al., LHCb Collaboration. “Measurement of Υ production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Eur.Phys.J.* C72 (2012), p. 2025. DOI: 10.1140/epjc/s10052-012-2025-y. arXiv:1202.6579 [hep-ex].
- [138] R. Aaij et al., LHCb Collaboration. “Inclusive low mass Drell-Yan production in the forward region at $\sqrt{s} = 7$ TeV”. LHCb Conference Note LHCb-CONF-2012-013. 2012. URL: <https://cdsweb.cern.ch/record/1434424>.
- [139] M.B. Voloshin. “The Onset of $e^+e^- \rightarrow \tau^+\tau^-$ at threshold revisited”. In: *Phys.Lett.* B556 (2003), pp. 153–162. DOI: 10.1016/S0370-2693(03)00125-4. arXiv:hep-ph/0212207 [hep-ph].
- [140] R. Aaij et al., LHCb Collaboration. “Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV”. In: *JHEP* 1206 (2012), p. 058. DOI: 10.1007/JHEP06(2012)058. arXiv:1204.1620 [hep-ex].
- [141] R. Aaij et al., LHCb Collaboration. “Measurement of the forward W boson cross-section in pp collisions at $\sqrt{s} = 7$ TeV”. submitted to JHEP. LHCb-PAPER-2014-033. 2014. arXiv:1408.4354 [hep-ex].
- [142] S. Bifane. “Measurement of the W production cross-section in pp collisions at $\sqrt{s} = 7$ TeV with the 2011 data”. LHCb Analysis Note LHCb-ANA-2014-049, draft 4.0 (2nd June 2014). URL: <https://cds.cern.ch/record/1694754>.
- [143] J. Harrison. “The inclusive τ cross section at LHCb at $\sqrt{s} = 7$ TeV”. LHCb Internal Note LHCb-INT-2011-041. URL: <https://cds.cern.ch/record/1379468>.

- [144] A. Keune. “Reconstruction of the Tau Lepton and the study of $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ at LHCb”. PhD thesis. École Polytechnique Fédérale de Lausanne, 2012. URL: <http://lphe.epfl.ch/publications/theses-en.php#theses>.
- [145] A. Powell et al. “Particle identification at LHCb”. In: *PoS ICHEP2010* (2010), p. 020. URL: <https://cdsweb.cern.ch/record/1322666>.
- [146] P.F. Harrison. “Blind analysis”. In: *J.Phys. G28* (2002), pp. 2679–2692. DOI: 10.1088/0954-3899/28/10/312.
- [147] R. Aaij et al., LHCb Collaboration. “Search for the lepton flavour violating decay $\tau^- \rightarrow \mu^+\mu^-\mu^-$ ”. LHCb Conference Note LHCb-CONF-2012-015. 2012. URL: <https://cds.cern.ch/record/1434456>.
- [148] M. Jaffe. CERN Summer Student project. 2011.
- [149] M. Meißner. “Studie zum Nachweis des leptonzahlverletzenden Zerfalls $\tau \rightarrow \mu\mu\mu$ mit dem LHCb Experiment”. diploma thesis. Heidelberg University, 2009. URL: <http://www.physi.uni-heidelberg.de/Forschung/he/LHCb/finishedTheses/>.
- [150] R. Aaij and J. Albrecht. “Muon triggers in the High Level Trigger of LHCb”. LHCb Public Note LHCb-PUB-2011-017. 2011. URL: <http://cds.cern.ch/record/1384386>.
- [151] M. Feindt. *A Neural Bayesian Estimator for Conditional Probability Densities*. 2004. arXiv:physics/0402093.
- [152] M. Feindt and U. Kerzel. “The NeuroBayes neural network package”. In: *Nucl.Instrum.Meth. A559* (2006). Inaccessible to the general public, pp. 190–194. DOI: 10.1016/j.nima.2005.11.166.
- [153] A. Hoecker et al. “TMVA: Toolkit for Multivariate Data Analysis”. In: *PoS ACAT* (2007), p. 040. arXiv:physics/0703039.
- [154] G. Krocker. “Development and calibration of a same side kaon tagging algorithm and measurement of the B_s^0 - \bar{B}_s^0 oscillation frequency Δm_s at the LHCb experiment”. CERN-THESIS-2013-213. PhD thesis. Heidelberg University, 2013. URL: <https://cds.cern.ch/record/1631104>.
- [155] C.R. Jones. private communication. 2013.
- [156] L. Breiman et al. *Classification and regression trees*. Belmont, California, USA: Wadsworth international group, 1984.
- [157] R.E. Schapire and Y. Freund. “A decision-theoretic generalization of on-line learning and an application to boosting”. In: *Jour.Comp.Syst.Sc.* 55 (1997), p. 119. DOI: 10.1006/jcss.1997.1504.
- [158] B.P. Roe et al. “Boosted decision trees as an alternative to artificial neural networks for particle identification”. In: *Nucl.Instrum.Meth. A543* (2005), pp. 577–584. DOI: 10.1016/j.nima.2004.12.018. arXiv:physics/0408124 [physics].
- [159] H. Voss. “Multivariate Analysis Techniques”. In: *Terascale Statistics Tools School*. Ed. by O. Behnke, C. Kleinwort, and S. Schmitt. 2008. URL: <https://indico.desy.de/conferenceDisplay.py?confId=1149>.
- [160] H. Voss. “Multivariate Data Analysis and Machine Learning in High Energy Physics”. In: *Lecture: Graduierten-Kolleg Freiburg*. 2009. URL: <http://tmva.sourceforge.net/talks.shtml>.

REFERENCES

- [161] H. Voss. private communication. 2011.
- [162] A. Abulencia et al., CDF Collaboration. “Search for $B_s^0 \rightarrow \mu^+\mu^-$ and $B_d^0 \rightarrow \mu^+\mu^-$ decays in $p\bar{p}$ collisions with CDF II”. In: *Phys.Rev.Lett.* 95 (2005), p. 221805. DOI: 10.1103/PhysRevLett.95.221805. arXiv:hep-ex/0508036 [hep-ex].
- [163] D. Casadei. “Estimating the selection efficiency”. In: *JINST* 7 (2012), p. 8021. DOI: 10.1088/1748-0221/7/08/P08021. arXiv:0908.0130 [physics.data-an].
- [164] M. Pivk and F.R. Le Diberder. “sPlot: a statistical tool to unfold data distributions”. In: *Nucl.Instrum.Meth.* A555 (2005), p. 356. DOI: 10.1016/j.nima.2005.08.106. arXiv:physics/0402083.
- [165] R. Aaij et al., LHCb Collaboration. “Search for $D_{(s)}^+ \rightarrow \pi^+\mu^+\mu^-$ and $D_{(s)}^+ \rightarrow \pi^-\mu^+\mu^+$ decays”. In: *Phys.Lett.* B724 (2013), pp. 203–212. DOI: 10.1016/j.physletb.2013.06.010. arXiv:1304.6365 [hep-ex].
- [166] C. Barschel. private communication. 2014.
- [167] J.P. Lees et al., BaBar Collaboration. “Limits on τ Lepton-Flavor Violating Decays in three charged leptons”. In: *Phys.Rev.* D81 (2010), p. 111101. DOI: 10.1103/PhysRevD.81.111101. arXiv:1002.4550 [hep-ex].
- [168] R. Aaij et al., LHCb Collaboration. “Search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ ”. In: *Phys.Lett.* B708 (2012), pp. 55–67. DOI: 10.1016/j.physletb.2012.01.038. arXiv:1112.1600 [hep-ex].
- [169] D. Martinez Santos. private communication. 2012.
- [170] P. del Amo Sanchez et al., BaBar Collaboration. “Dalitz plot analysis of $D_s^+ \rightarrow K^+K^-\pi^+$ ”. In: *Phys.Rev.* D83 (2011), p. 052001. DOI: 10.1103/PhysRevD.83.052001. arXiv:1011.4190 [hep-ex].
- [171] A. Lusiani. “HFAG-Tau report with theory introduction for $|V_{us}|$ ”. In: *The 13th International Workshop on Tau Lepton Physics (TAU2014)*. 2014. URL: <https://indico.cern.ch/event/300387/session/6/contribution/12>.
- [172] Otto Forster. *Analysis 1. Differential- und Integralrechnung einer Veränderlichen*. ger. 2013. DOI: 10.1007/978-3-658-00317-3.
- [173] S.S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *The Annals of Mathematical Statistics* 9.1 (Mar. 1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.
- [174] R. Aaij et al., LHCb Collaboration. “Measurement of b-hadron production fractions in 7 TeV pp collisions”. In: *Phys.Rev.* D85 (2012), p. 032008. DOI: 10.1103/PhysRevD.85.032008. arXiv:1111.2357 [hep-ex].
- [175] R. Aaij et al., LHCb Collaboration. “Measurements of B_c^+ production and mass with the $B_c^+ \rightarrow J/\psi\pi^+$ decay”. In: *Phys.Rev.Lett.* 109 (2012), p. 232001. DOI: 10.1103/PhysRevLett.109.232001. arXiv:1209.5634 [hep-ex].
- [176] J.R. Ellis et al., eds. *Implications of LHCb measurements and future prospects workshop*. First workshop. Resulted in [84]. 2011. URL: <http://indico.cern.ch/event/155764/>.
- [177] J.R. Ellis et al., eds. *Implications of LHCb measurements and future prospects workshop*. 4th workshop. 2014. URL: <https://indico.cern.ch/event/324660/>.