
Unterschrift BetreuerIn



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DIPLOMARBEIT

Estimation of the background to the $H \rightarrow \tau\tau$ signal due to hadronic jets misidentified as τ lepton decays for the CMS experiment

ausgeführt am Institut für Hochenergiephysik
der Technischen Universität Wien

unter der Anleitung von
Univ. Prof. Dipl.-Phys. Dr. rer. nat. Jochen Schieck
und
Dr. Martin Flechl

durch

Florian Spreitzer

Matthias-Ernst-Pista-Gasse 23, 1210 Wien

Wien, 24. Juli 2019

Unterschrift StudentIn



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Abstract

Detailed research of the properties of the Higgs boson found at CERN is an important task to determine compliance with values predicted by the standard model of particle physics for observables such as its coupling strength to fermions. The decay of the Higgs boson into two tau leptons represents a very promising process for such measurements.

In this thesis the heavily data driven fake factor method for estimation of the background arising from jets initiated by quarks or gluons being misidentified as hadronically decaying tau leptons is analyzed and extended. The effect of using $Z \rightarrow \tau\tau$ samples created with the τ embedding technique is reviewed. Several updates to the fake factor method to increase its performance and robustness are implemented, tested and discussed. Results of the application of the fake factor method to proton-proton collision data collected by the CMS experiment in 2016 and 2017 and its performance are presented. The extended method is used in recent preliminary CMS measurements of Higgs boson properties and for the upcoming legacy publications with the CMS data set recorded from 2016 to 2018.

Kurzfassung

Die Bestimmung der genauen Eigenschaften des im Jahre 2012 am CERN entdeckten Higgs-Bosons ist eine wichtige Aufgabe, um die Übereinstimmung mit den durch das Standardmodell der Teilchenphysik vorhergesagten Werten wie beispielsweise der Kopplungsstärke zu Fermionen zu untersuchen. Der Zerfall des Higgs-Bosons zu zwei Tau-Leptonen stellt einen dafür sehr gut geeigneten Prozess dar.

Das Ziel dieser Arbeit ist es, die datenbasierte Fake-Factor-Methode zur Abschätzung des Anteils von Hintergrundereignissen beim Zerfall des Higgs-Bosons in zwei Tauonen, welcher durch die fälschliche Identifikation von durch Quarks oder Gluonen initiierten Jets als hadronisch zerfallende Tauonen zustandekommt, zu analysieren und zu erweitern. Die Methode wurde erweitert um die Verwendung von $Z \rightarrow \tau\tau$ Datensätzen, erstellt mit der τ -Embedding Methode, zu verwenden und die daraus resultierenden Ergebnisse wurden untersucht. Verschiedene Ansätze zur Erhöhung der Genauigkeit und Robustheit der Fake-Factor-Methode wurden behandelt und ausgewertet. Die Fake-Factor-Methode wurde auf am CMS-Experiment in den Jahren 2016 und 2017 gesammelte Kollisionsdaten angewendet und die entsprechenden Ergebnisse vorgestellt. Die erweiterte Fake-Factor-Methode wird in aktuellen vorläufigen CMS Messungen zu den Eigenschaften des Higgs-Bosons sowie in den bevorstehenden Publikationen über die Ergebnisse aus den Messungen aus dem Zeitraum 2016 bis 2018 verwendet.

Contents

1. Introduction	1
2. The Standard Model of Particle Physics	3
2.1. Elementary Particles	3
2.1.1. Leptons	3
2.1.2. Quarks	3
2.1.3. Gauge Bosons	4
2.2. Fundamental Interactions of the SM	5
2.2.1. Electroweak Interaction	5
2.2.2. Strong Interaction	6
2.3. The Higgs Mechanism	6
2.3.1. Spontaneous Symmetry Breaking	6
2.3.2. Higgs Production Modes	8
2.3.3. Higgs Decay Modes	9
2.3.4. Properties of the Higgs Boson	10
2.4. The Tau Particle	10
2.4.1. τ Decay Modes	11
2.5. Physics beyond the Standard Model	12
2.5.1. Unexplained Phenomena and Discrepancies	12
2.5.2. Minimal Supersymmetric Extension of the Standard Model (MSSM)	13
2.5.3. Theoretical Motivations of the MSSM	14
3. The Large Hadron Collider	17
3.1. The CMS Experiment	18
3.1.1. Superconducting Solenoid	19
3.1.2. Inner Tracker	19
3.1.3. Electromagnetic Calorimeter	19
3.1.4. Hadron Calorimeter	20
3.1.5. Muon System	21
3.1.6. Trigger System	22

4. Event Reconstruction and Simulation	25
4.1. Particle Reconstruction	25
4.1.1. Muons	25
4.1.2. Electrons	26
4.1.3. Hadronic Tau Decays	26
4.1.4. Jets	28
4.1.5. Missing Transverse Energy	28
4.2. Simulation	30
5. Summary of the Measurement of Higgs boson production and decay to the $\tau\tau$ final state	31
5.1. Event categorization	31
5.2. Results	32
6. The Fake Factor Method	35
6.1. Calculation	35
6.1.1. QCD Fake Factors	36
6.1.2. W+jets Fake Factors	38
6.1.3. $t\bar{t}$ Fake Factors	38
6.1.4. Fake Factor Weights	42
6.2. Corrections	42
6.2.1. QCD fake factor corrections	42
6.2.2. W+jets fake factor corrections	44
6.2.3. $t\bar{t}$ fake factor corrections	45
6.3. Uncertainties	47
7. Extension of the fake factor method for the $H \rightarrow \tau\tau$ analysis	49
7.1. Determination of genuine $\tau\tau$ Backgrounds via Embedding	49
7.2. Decay Mode Binning	53
7.3. Separation of Corrections into 0-jet/1-jet Categories	56
7.4. Application of the Fake Factor Method to 2016 and 2017 Data	60
8. Conclusions	63
A. Acronyms	65
B. Software Instructions	67
C. Performance Plots	69
Bibliography	77

1. Introduction

The search for the last missing building block of the standard model of particle physics, the Higgs boson, whose existence was predicted in 1964 [1], took almost 50 years and required enormous technical effort. But while the existence of a Higgs boson with a mass of 125 GeV has been confirmed in 2012 at the CERN LHC [2, 3], the largest particle accelerator in existence, it is important to further investigate its kinematic and coupling properties to either confirm conformity with predictions from the standard model or to detect discrepancies which could further support one of the many approaches to extend the standard model to a more fundamental theory to overcome its several known shortcomings. So far no significant deviations from the expected properties have been found. In extended theories such as supersymmetry, a difference for the coupling of the Higgs boson to fermions, for example the tau lepton, to the standard model predictions is expected. The different Higgs decay modes provide important information about the respective fermion couplings.

The main goal of this work is to extend the data driven fake factor method which is used to estimate the contribution of jets being misidentified as hadronically decaying τ particle to the signal background in the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels. In Chap. 2, a short summary of the standard model of particle physics is presented. The Higgs mechanism resulting from spontaneous symmetry breaking as well as Higgs boson properties, production and decay modes are discussed. The design and structure of the LHC with a strong emphasis on the setup of the CMS experiment are shown in Chap. 3. Closely related to the former chapter is Chap. 4 in which the reconstruction of observed data within the experimental setup is discussed. Chap. 5 introduces an outline of the analysis obtaining measurements of the SM Higgs production with subsequent decay to two τ leptons cross section. The approach of event categorization based on a neural network is discussed and measured results are presented. In Chap. 6 and Chap. 7 the revision and extension of the fake factor method is discussed. Several approaches to make it more performant and robust are presented and the method is applied to two datasets of pp collision with a center-of-mass energy of 13 TeV collected in the 2016 and 2017 LHC run corresponding to a total integrated luminosity of 77.4 fb^{-1} and the performance of the estimation is discussed.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

2. The Standard Model of Particle Physics

The standard model (SM) is a theory describing all known elementary particles and their interaction via the strong, weak and electromagnetic force. Being a quantum field theory each fundamental particle is described as discrete quantum in a global field. It is one of the most tested and robust theories in physics. However, it is known to be incomplete for several reasons, one of them being the exclusion of general relativity for the description of particle interactions. The SM also does not provide a mechanism to explain the existence of Dark Matter, which is thought to constitute approximately 85% of matter in the universe.

2.1. Elementary Particles

2.1.1. Leptons

In the standard model there are six leptons which are fermions with a spin of $\frac{1}{2}$ and can be grouped into three generations. Additionally there is an anti-particle for each lepton which shares the same mass and spin but has opposite additive quantum numbers, like the electric charge or the magnetic moment respective to the spin. The general qualities of the leptons are listed in Table 2.1. Muons and tau leptons differ from the electron in mass and lifetime but share the same electric charge $e = -e_0 = -1.602 \cdot 10^{-19} C$. In the formalism of the standard model neutrinos are considered massless, however, neutrino oscillation observations have confirmed a non-zero neutrino mass[4].

2.1.2. Quarks

Quarks are fundamental particles with a spin of $\frac{1}{2}$ and an electric charge of either $+\frac{2}{3}e_0$ or $-\frac{1}{3}e_0$, their antiparticles carry the opposite electric charge. The amount of quarks of a specific type is conserved when only the strong interaction is taken into account, however, there can be transitions between the different quark types via the weak interaction. Quarks are the only elementary particles in the standard

Name	Symbol	Mass (MeV)	Electric Charge (e_0)	Lifetime [s]
Electron	e	0.511	-1	∞
Electron neutrino	ν_e	$<2 \cdot 10^{-6}$	0	∞
Muon	μ	105.66	-1	$2.197 \cdot 10^{-6}$
Muon neutrino	ν_μ	<0.17	0	∞
Tau lepton	τ	1777	-1	$2.9 \cdot 10^{-13}$
Tau neutrino	ν_τ	<15.5	0	∞

Table 2.1.: A summary of lepton properties[4].

model interacting with all four fundamental forces and with electric charges that are not multiples of the elementary charge e_0 .

Name	Symbol	Mass	Electric Charge (e_0)
Up	u	$2.3^{+0.7}_{-0.5}$ MeV	$+\frac{2}{3}$
Down	d	$4.8^{+0.7}_{-0.3}$ MeV	$-\frac{1}{3}$
Charm	c	1.275 ± 0.025 GeV	$+\frac{2}{3}$
Strange	s	95 ± 5 MeV	$-\frac{1}{3}$
Top	t	1.275 ± 0.025 GeV	$+\frac{2}{3}$
Bottom	b	4.18 ± 0.03 GeV	$-\frac{1}{3}$

Table 2.2.: A summary of quark properties[4].

There are six different types or flavors of quarks, their different properties are shown in Table 2.2. Experimental analysis of quarks is a difficult task, since there has never been an observation of an isolated quark. They can appear as either mesons, which are systems consisting of a quark and an anti-quark, or as baryons, systems consisting of an odd number of valence quarks (at least three). More exotic configurations include the tetraquark (two quarks and two anti-quarks) and the pentaquark (four quarks and one anti-quark) states.

2.1.3. Gauge Bosons

The interaction between elementary particles described by a gauge theory takes place via the exchange of gauge bosons.

The $SU(3) \times SU(2) \times U(1)$ gauge group of the standard model implies the existence of several quanta with spin 1 [5]. The quantum chromodynamics (QCD) group $SU(3)$ describes eight generators, adding eight gauge bosons called gluons which carry the strong interaction. Based on the $SU(2) \times U(1)$ group the unified electroweak interaction requires four bosons as force carriers, which are found to be the photon, W^\pm and Z^0 bosons. Photons and gluons are massless and can act

on an unlimited distance while the range of the massive W^\pm and Z^0 bosons is limited by the uncertainty principle to about 10^{-3}fm .

Since a direct mass term for gauge bosons in the Lagrange function is not gauge invariant, local gauge symmetry demands massless gauge bosons and long-ranged interactions. Observation shows that the interaction is not long-ranged and that the W^\pm and Z^0 bosons actually carry mass as a result of the local symmetry breaking known as the Higgs mechanism.

2.2. Fundamental Interactions of the SM

Three of the four fundamental forces in our universe can be described by the SM. Gravity, while being the most dominant force at macroscopic scales, is not included in the SM. Compared to the other fundamental interactions the coupling constant of gravity is the smallest one by many orders of magnitude and it can be neglected for typical high energy particle physics processes.

2.2.1. Electroweak Interaction

Electromagnetism as described by the Maxwell equations visible in Eq. (2.1) and Eq. (2.2) is a classical field theory.

$$\partial_\mu F^{\mu\nu} = J^\nu, \quad (2.1)$$

$$\partial^\lambda F^{\mu\nu} + \partial^\nu F^{\lambda\mu} + \partial^\mu F^{\nu\lambda} = 0, \quad (2.2)$$

here in covariant form. In this notation the Maxwell equations consist of an inhomogeneous equation connecting the electromagnetic tensor $F^{\mu\nu}$ defined as the exterior derivative of the electromagnetic four-potential $A^\mu = (\phi, \vec{A})$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (2.3)$$

with an external current $J^\mu = (\rho, \vec{J})$ on the one hand, and a homogenous equation on the other hand. The covariant derivatives $\partial_\mu = \frac{\partial}{\partial x^\mu}$ are used. Interactions based on electromagnetism are defined by the Lorentz force and mediated by photons.

Compared to the interactions of gravity and electromagnetism, the weak force is very limited in its effective range to about 10^{-3} fm due to the short lifetime of the particles mediating it, the W^\pm and Z bosons. It is the mechanism behind many subatomic interactions and dictates processes like radioactive β decays of atoms. It is the only one of the four interactions known to violate parity (P) symmetry and also the only one violating charge-parity (CP) symmetry.

The SM describes the electromagnetic and the weak force as different manifestations of the same force at different energy levels unified under an $SU(2)\times U(1)$

gauge group. The unification energy level lies at around 246 GeV, equal to the vacuum expectation value of the Higgs field. Glashow, Weinberg and Salam were awarded the Nobel Prize in 1979 for their contributions to this theory[6, 7].

2.2.2. Strong Interaction

The dynamics of quarks interacting via a gluon field is described by quantum chromodynamics (QCD). This theory is based on a SU(3) symmetry group, which acts in the so-called color space. The force between quarks is mediated by eight gluons which additionally participate in the strong interaction by carrying a color charge themselves, which complicates the analysis of QCD. The strong force acts on a very small scale. Hadrons with a dimension of less than 1 fm are bound states of quarks, held together by gluons mediating the strong force. On a larger scale of 1-3 fm mesons carry the force between baryons, forming nuclei. Beyond that range the strong interaction has no effect. Due to an effect called *quark confinement* no particle carrying a color charge is observed as a free particle, only in a bound state with a neutral total color charge. This is a consequence of the strong force not diminishing in strength with the distance between quarks but remaining constant after a limiting distance. With increasing distance, the creation of a quark-antiquark pair matching to the original one out of the vacuum is energetically favorable, hence no individual quark is observed.

2.3. The Higgs Mechanism

The Higgs mechanism introduced 1964 by Peter Higgs[1] and its electroweak symmetry breaking at low energies is essential for explaining the masses of the electroweak gauge bosons. In 2012 the CMS and ATLAS experiments at the LHC at CERN announced the discovery of a new boson compatible with theoretical predictions of a massive neutral boson, which was the last missing particle of the SM[2, 3].

2.3.1. Spontaneous Symmetry Breaking

It is assumed that a new field Φ , an SU(2) doublet called the Higgs field is breaking the SU(2)×U(1) gauge symmetry of the electroweak interaction with a non-zero vacuum expectation value. In its most general form Φ is of the form

$$\Phi = \begin{pmatrix} \Phi_A \\ \Phi_B \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.4)$$

The Lagrangian density describing such a field can be written as

$$\mathcal{L} = (\partial_\mu \Phi^\dagger) \partial^\mu \Phi - V(\Phi^\dagger \Phi) \quad (2.5)$$

consisting of a general potential term $V(\Phi^\dagger \Phi)$ and a kinetic term, with Φ^\dagger being the hermitian conjugate of Φ . Based on the work of Yang and Mills [8] on the generalization of the invariance to local $U(1) \times SU(2)$ gauge transformations of the form

$$\Phi \rightarrow \Phi' = e^{-i\theta} \mathcal{U} \Phi \quad (2.6)$$

with \mathcal{U} being an arbitrary element of $SU(2)$ fulfilling $\mathcal{U}^\dagger \mathcal{U} = \mathcal{U} \mathcal{U}^\dagger = 1$, one can define a Lagrangian density

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger D^\mu \Phi - V(\Phi^\dagger \Phi) \quad (2.7)$$

with the newly defined covariant derivatives

$$D_\mu = \left(\partial_\mu + i \frac{g_1}{2} B_\mu + i \frac{g_2}{2} \mathbf{W}_\mu \right) \quad (2.8)$$

which fulfill the relation

$$D'_\mu \Phi' = e^{-i\theta} \mathcal{U} D_\mu \Phi. \quad (2.9)$$

The gauge field B_μ was introduced to ensure the invariance regarding local $U(1)$ gauge transformations, the vector gauge field \mathbf{W}_μ adds invariance to local $SU(2)$ gauge transformations. Their respective transformation characteristics are defined in Eq. (2.10) and Eq. (2.11).

$$B_\mu(x) \rightarrow B'_\mu(x) = B_\mu(x) + \frac{2}{g_1} \partial_\mu \theta(x) \quad (2.10)$$

$$\mathbf{W}_\mu(x) \rightarrow \mathbf{W}_\mu(x)' = \mathcal{U} \mathbf{W}_\mu(x) \mathcal{U}^\dagger + i \frac{2}{g_2} (\partial_\mu \mathcal{U}(x)) \mathcal{U}^\dagger(x) \quad (2.11)$$

The parameters g_1 and g_2 have to be determined experimentally. The Lagrangian defined in Eq. (2.7) is now gauge invariant in respect to local $U(1) \times SU(2)$ gauge transformations and in the case of B_μ and \mathbf{W}_μ transforming like covariant 4-vectors also Lorentz invariant. Using the definition of Φ with the most general, non-trivial and renormalizable potential term $V(\Phi^\dagger \Phi) = \frac{m^2}{2\phi_0^2} (\Phi^\dagger \Phi - \phi_0^2)^2$, it can be written as

$$V(\Phi^\dagger \Phi) = \frac{m^2}{2\phi_0^2} (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 - \phi_0^2)^2, \quad (2.12)$$

resulting in a degenerate ground state. Without the loss of generalization one can choose a ground state with $\Phi_A = 0$ and Φ_B real and expand a general state around this ground state $(0, \phi_0)^T$ with a real function $h(x)$.

$$\Phi = \begin{pmatrix} 0 \\ \phi_0 + \frac{1}{\sqrt{2}} h(x) \end{pmatrix} \quad (2.13)$$

The $U(1)$ symmetry in this gauge condition is unchanged however $SU(2)$ symmetry is broken. Using this state in the Lagrangian in Eq. (2.7) one gets a system describing a massive neutral boson field $h(x)$, a massive neutral vector field $Z_\mu(x)$ and a pair of massive charged vector boson fields $W_\mu^\pm(x)$ and the electromagnetic massless photon field A_μ . For this notation the definitions

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2), \quad (2.14)$$

as well as

$$Z_\mu = W_\mu^3 \cos(\theta_W) - B_\mu \sin(\theta_W), A_\mu = W_\mu^3 \sin(\theta_W) + B_\mu \cos(\theta_W) \quad (2.15)$$

are used. The Weinberg angle θ_W is defined via the coupling constants g_1, g_2 in Eq.(2.16).

$$\cos(\theta_W) = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}, \sin(\theta_W) = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \quad (2.16)$$

This process describes the masses of gauge bosons resulting from spontaneous symmetry breaking of the electroweak $SU(2) \times U(1)$ gauge symmetry. In this formulation fermion mass terms are described by Yukawa couplings of the form

$$\mathcal{L}_{Yukawa} = -g(\Psi_R^\dagger \Phi \Psi_L + \Psi_L^\dagger \Phi \Psi_R) \quad (2.17)$$

where the Higgs doublet Φ couples to the left-handed (Ψ_L) and right-handed (Ψ_R) components of the fermionic spin doublet.

2.3.2. Higgs Production Modes

The dominant production modes for Higgs bosons in hadron colliders are based on the strong coupling to heavy particles, namely the W and Z vector bosons, the top quark and the bottom quark. The most common process is the production via gluon-gluon fusion. Here, the Higgs boson does not couple directly to gluons but via a triangular quark loop. Due to the mass dependence of the Higgs-fermion couplings only the heavy top and bottom quarks contribute non-negligibly to this production mode. By a factor of 10 less common, however, very significant due to its distinctive signature is the process of vector boson fusion. Here the Higgs boson is accompanied by two quarks forming hadronic jets. A fermion colliding with its antiparticle can form a virtual Z boson which, given it carries enough energy, can emit a Higgs boson. This process is called Higgs strahlung and is the third-most likely production mode at the LHC. A fourth possibility is the associated production with top or bottom quarks. Evidence for $t\bar{t}H$ production was published by the ATLAS and CMS collaborations in 2018 [9, 10]. Feynman diagrams of these main production modes are shown in Fig. 2.1. Other processes, including pairwise Higgs production are suppressed with a much smaller production cross section.

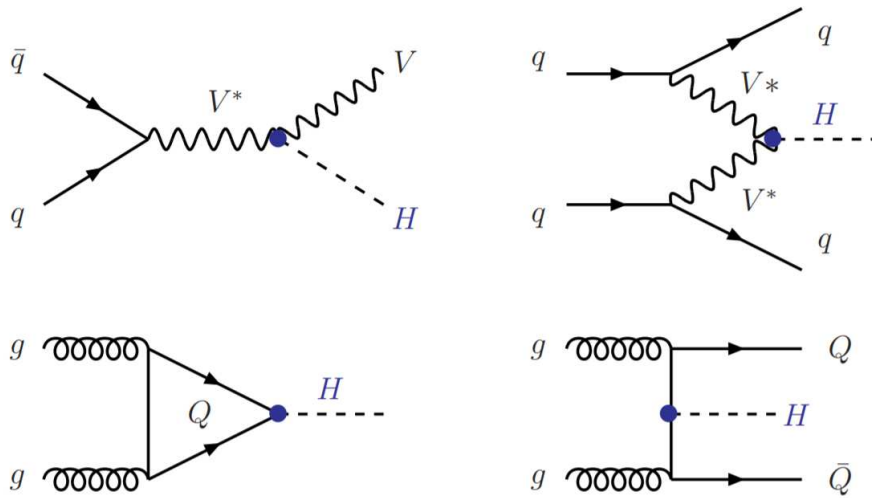


Figure 2.1.: The four most dominant production modes of the Higgs boson at the LHC [11]. Top left: Higgs Strahlung, a vector boson (V) with enough energy can emit a Higgs boson, top right: vector boson fusion, bottom left: gluon-gluon fusion via a quark (Q) loop, bottom right: associated production with heavy quarks.

2.3.3. Higgs Decay Modes

The decay modes of the Higgs boson are numerous since with a mass of 125 GeV it is the second most massive particle in the SM and couples to all other massive particles. Due to the dependence of the coupling strength of the Higgs boson to gauge bosons and fermions on their masses one can determine the respective decay width after measuring those masses. So generally it is more likely to decay into heavier particles, with the production of a $b\bar{b}$ quark pair being the most likely one of all decay modes with a predicted branching ratio of about 57.7% [12]. Although the probability for this decay mode is very high, a large number of background processes mimicking the measurable outcome products make it hard to extract signal events. The observation of the $H \rightarrow b\bar{b}$ process was published in 2018 by the ATLAS and CMS collaborations [13, 14].

The second highest branching ratio of 21.5% is due to the decay into a pair of W bosons, the second most common fermionic decay mode is $H \rightarrow \tau\tau$ with a predicted branching ratio of 6.32%, already about an order of magnitude smaller than $H \rightarrow b\bar{b}$ [12]. Predicted branching ratios in dependence of the Higgs mass in a region around 125 GeV is included in Fig. 2.2.

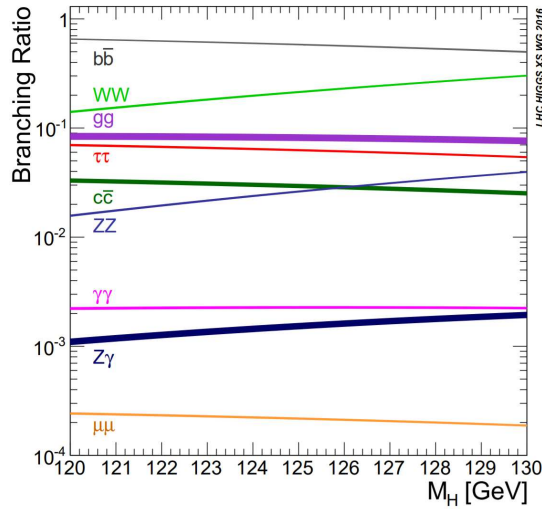


Figure 2.2.: Branching ratios as functions of the Higgs boson mass as predicted by the Higgs Cross Section Working Group [15].

2.3.4. Properties of the Higgs Boson

The mass of the Higgs boson observed at the LHC was measured [16] to be

$$m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})\text{GeV} \quad (2.18)$$

and was the last of the free parameters of the SM to be measured. It carries no electric or color charge and since the Higgs field is a scalar field, the Higgs boson has no spin. The SM predicts the Higgs boson to be CP-even and to be its own antiparticle. Observations did not show deviations from those predictions so far.

2.4. The Tau Particle

The τ -lepton was first discovered by Martin Perl et. al. as an explanation for the observed missing energy distribution in electron-positron annihilation experiments of the form

$$e^+ + e^- \rightarrow e^\pm + \mu^\mp + E_{\text{missing}} \quad (2.19)$$

suggesting that the production and decay of a pair of unknown leptons in a mass range of 1.6 to 2.0 GeV may be responsible [17]. The verification of this hypothesis was complicated by the fact that the missing energy was in the magnitude of the threshold for D meson production, obscuring the τ measurement. Perl was awarded the Nobel Prize in 1995. The τ mass was found to be 1.777 GeV [4]. Due to its

	Decay Mode	Branching Fraction (%)
Leptonic Decays	$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$	17.83
	$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$	17.41
	Σ_{leptonic}	35.24
Hadronic Decays	$\tau^- \rightarrow h^- \nu_\tau$	11.53
	$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	25.95
	$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	9.52
	$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	9.80
	$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$	4.76
	Other	3.23
	Σ_{hadronic}	64.79

Table 2.3.: Branching ratios for leptonic and hadronic τ decays [4] with h^\pm denoting a charged pion or kaon and Σ indicating the sum of all branching ratios in the respective category.

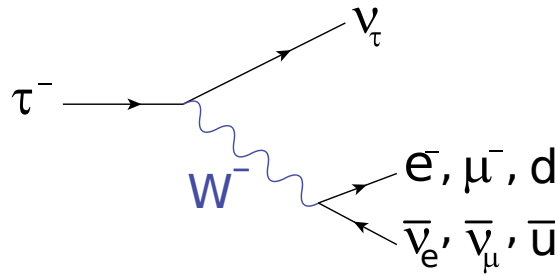


Figure 2.3.: Feynman diagram of the most common τ decays via emission of a W-boson.

very short lifetime of $2.9 \cdot 10^{-13}$ s the τ lepton can not be detected directly in hadron colliders but indirectly by the products into which it decays.

2.4.1. τ Decay Modes

The high mass of the τ particle enables it to decay hadronically, unlike any other lepton, with a branching ratio for hadronic decays of approximately 64.79%. Leptonic decay into $\nu_\tau + \mu(e) + \nu_{\mu(e)}$ is possible with a probability of 17.41%(17.83%). To conserve the tauonic lepton number in the occurring weak decays a tau neutrino is created in every different mode. Branching ratios for different decay modes are shown in Table 2.4, a feynman diagram summarizing the most common τ decays is included in Fig. 2.3.

2.5. Physics beyond the Standard Model

Despite the fact that the SM is one of the best performing and most successful theories in physics, there are some inherent shortcomings that motivate the research of different approaches and theories about our universe, so-called beyond the standard model (BSM) theories. The following section will provide a selection of the SM's limitations arising from the very formulation of the theory as well as currently still unexplained phenomena and experimental results.

2.5.1. Unexplained Phenomena and Discrepancies

Gravity

The SM does not explain gravity and so far could not be unified with the experimentally validated theory of general relativity. Quantum field theories with a hypothetical graviton as the mediator of the gravitational force are not renormalizable.

Hierarchy Problem

The mass of the Higgs boson is subject to corrections due to the presence of virtual particles, mostly virtual top quarks, which are much larger than the Higgs boson mass, requiring very precise fine-tuning of the *bare Higgs boson mass* (see Section 2.5.3) to explain a nonzero Higgs boson mass below the Planck scale.

Dark Matter and Dark Energy

Only about 5% of the energy present in the universe can be explained by the SM [18]. Dark matter is supposed to make up 26% of the total energy and dark energy accounts for about 69%. There is no fundamental particle in the SM fitting as a candidate for the current model of cold dark matter.

Neutrino Mass

In the SM formulation neutrinos which only appear with left-handed chirality are considered massless, since no renormalizable mass term can be added without right-handed partner particles. Experiments however show oscillations between neutrino types which can only be explained if they are massive [19]. Upper bounds of neutrino masses are given in Tab. 2.1.

Matter-Antimatter Asymmetry

There is no explanation within the SM why an asymmetry in favor of matter exists.

2.5.2. Minimal Supersymmetric Extension of the Standard Model (MSSM)

Introduced in 1981 by S. Dimopoulos [20], the MSSM is extending the SM by adding a minimal number of particles and interactions to realize Supersymmetry while preserving local gauge symmetries.

Every SM field gets assigned a supersymmetric partner that a priori only differs by its spin which shows a difference by a unit of $\frac{1}{2}$. However, with only a single Higgs doublet a gauge anomaly would occur [21], hence a second doublet with corresponding superpartners is introduced. Additionally, the second Higgs doublet is required for generating the masses of all fermions. To derive a naming scheme for the new particles the fermionic superpartners of SM bosons are marked with the suffix *-ino* while the bosonic superpartners of SM fermions are discernible by the prefix *s-*.

For an exact symmetry the assumption of superpartners only differing by a spin of $\frac{1}{2}$ would be correct, however, the fact that no SUSY particles have been found in the mass regions of their SM counterparts shows that if SUSY exists, it is a broken symmetry and the masses of superpartners exceed the masses of corresponding SM particles by a large margin.

The charges of all fields except for the new Higgs doublet are known from the SM so the gauge interactions are already determined. This means that in the MSSM the symmetry group $SU(2) \times U(1)$ must be broken to the electromagnetic symmetry $U(1)$ as well. Again, non-vanishing vacuum expectation values of the Higgs fields are the reason. As in the SM, only three degrees of freedom of the Higgs fields are absorbed by gauge bosons, this leads to overall five Higgs particles in the MSSM:

- h^0 , a scalar particle.
- H^0 , a scalar particle, heavier than h^0 .
- A^0 , a heavy pseudoscalar particle.
- H^+, H^- , a pair of charged Higgs bosons.

The MSSM introduces 105 new free parameters in addition to the 18 parameters of the SM. For comparison with experimental results models with fewer free parameters are used. For example, using the assumptions

- no new source of CP-violation
- no flavor changing neutral currents
- first and second generation universality

Name	Gauge eigenstate	Spin
Sleptons	$\tilde{\nu}_e^L, \tilde{e}^L, \tilde{e}^R$	0
	$\tilde{\nu}_\mu^L, \tilde{\mu}^L, \tilde{\mu}^R$	0
	$\tilde{\nu}_\tau^L, \tilde{\tau}^L, \tilde{\tau}^R$	0
Squarks	$\tilde{u}^L, \tilde{u}^R, \tilde{d}^L, \tilde{d}^R$	0
	$\tilde{c}^L, \tilde{c}^R, \tilde{s}^L, \tilde{s}^R$	0
	$\tilde{t}^L, \tilde{t}^R, \tilde{b}^L, \tilde{b}^R$	0
Neutralinos	$\tilde{B}^0, \tilde{W}^0, \tilde{H}_u^0, \tilde{H}_d^0$	$\frac{1}{2}$
Charginos	$\tilde{W}^\pm, \tilde{H}_u^\pm, \tilde{H}_d^\pm$	$\frac{1}{2}$
Gluino	\tilde{g}	$\frac{1}{2}$

Table 2.4.: Supersymmetric partners of SM fields.

the phenomenological MSSM reduces the number of additional free parameters to 19 [22].

2.5.3. Theoretical Motivations of the MSSM

Naturalness In the SM with a Potential

$$V = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (2.20)$$

the Higgs mass is given by the expression

$$m_h^2 = 2\mu^2 + \delta m_h^2. \quad (2.21)$$

The term δm_h^2 is the radiative correction and is divergent of the order Λ^2 with Λ being the energy cut-off to the divergent loop integrals. One could presume the *natural* mass of the Higgs to be in the same order of magnitude as its correction, however with an experimentally found mass of 125 GeV it is smaller than the Planck-Scale Λ by a factor 10^{-16} . In unbroken supersymmetric models a second correction term with opposite sign cancels out the term in question, broken supersymmetry, however, again gives room for a quadratic divergence. Solving the naturalness problem regarding the quadratically divergent corrections of the Higgs mass was one of the main reasons for the introduction of the MSSM.

Dark Matter and Proton Stability Baryon and lepton number conservation, a concept that has been tested very rigorously, is no longer inherent to the MSSM due to several, not a priori vanishing, renormalizable couplings in the theory. However, violation of this principle would lead to very short proton life times,

while the proton lifetime is constrained to be at least of the order of $10^{33} - 10^{34}$ years, depending on the decay channel [23]. The introduction of R-parity, a \mathbb{Z}_2 symmetry in the MSSM conserves baryon and lepton numbers, as in the SM. R-parity is defined as

$$P_R = (-1)^{3B+L+2s}, \quad (2.22)$$

with the baryon number B , the lepton number L and the spin s . The R-parity of SM particles is +1, supersymmetric particles have an R-parity of -1 . With preservation of R-parity the lightest supersymmetric particle has to be stable. As a weakly interacting massive particle (WIMP), it would make a good dark matter candidate. In most supersymmetric models the neutralino is a candidate for the lightest supersymmetric particle, but a sufficiently light gravitino, the superpartner of the hypothetical graviton, could be a good candidate as well.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

3. The Large Hadron Collider

The Large Hadron Collider project (LHC) was approved by the CERN Council in 1994, first collisions were achieved in 2010. It is built at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland and with a diameter of 27 km and collision energies of 13 TeV it is the world's largest and most powerful particle accelerator. To get the protons up to high velocities, a series of different accelerators is used in combination. At first the source hydrogen atoms are stripped of their electrons to yield protons that are accelerated in the linear accelerator Linac-2 to an energy of 50 MeV. The resulting beam is then sped up further by the Proton Synchrotron Booster to 1.4 GeV. The Proton Synchrotron, which has been operative since 1959, long before the commission of the LHC, and was once the most powerful particle accelerator worldwide, pushes the proton beam to an energy of 25 GeV before the beam is sent to the Super Proton Synchrotron and accelerated to 450 GeV. After that the protons are injected into the two LHC beam pipes (one running clockwise and one counter-clockwise) in which they reach their maximum energy of 6.5 TeV after 20 minutes. The two proton beams are colliding with a center of mass energy of 13 TeV at four different collision points, around which the following seven experiments are positioned:

- ALICE (A Large Ion Collider Experiment)
- ATLAS (A Toroidal LHC Apparatus)
- CMS (Compact Muon Solenoid)
- LHCb (Large Hadron Collider beauty)
- LHCf (Large Hadron Collider forward)
- TOTEM (TOTAl Elastic and diffractive cross section Measurement)
- MoEDAL (Monopole and Exotics Detector At the LHC)

In 2012 one of the main goals of the LHC could be achieved by the independent discovery by the CMS and ATLAS experiments of a Higgs boson with a mass of 125 GeV [3][2]. A map of the LHC with the location of the four biggest experiments (ATLAS, CMS, ALICE, LHCb) at the four collision points is shown in Fig. 3.1. The TOTEM, LHCf and MoEDAL experiments are built near the CMS, ATLAS and LHCb collision points, respectively.

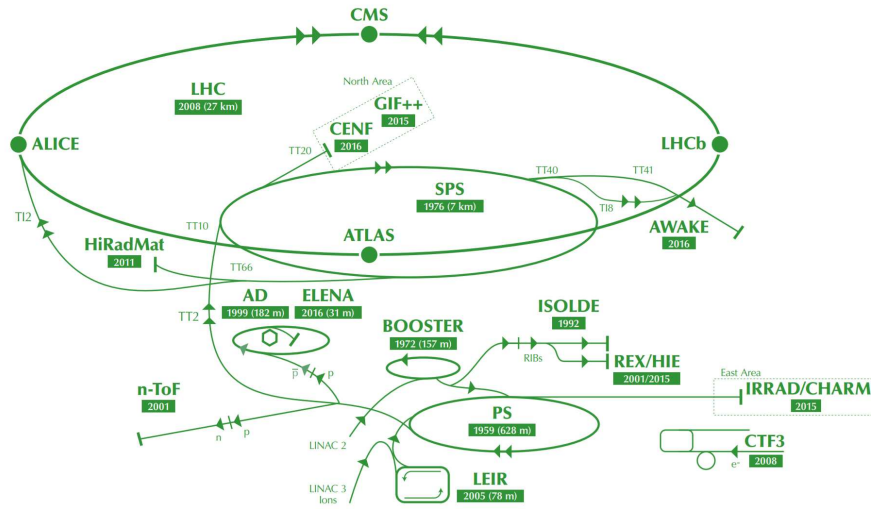


Figure 3.1.: Layout of the LHC accelerator complex [24]

3.1. The CMS Experiment

The Compact Muon Solenoid (CMS) experiment is one of the four big detectors located 100 m below the ground at the LHC at CERN. The multi-purpose detector is collecting data from proton-proton collisions since 2010 and was able to confirm the existence of the SM Higgs boson in 2012 [3]. The eponymous superconducting solenoid generates a high magnetic field of 4 Tesla to bend the track of charged particles for precise measurements of their momentum. To describe the paths of particles resulting from pp-collisions, a special coordinate system with its origin in the designated interaction point is defined. The x-axis is defined as pointing radially to the center of the LHC, the y-axis points vertically upwards to the surface, the z-axis is pointing in the beam direction so that a right handed coordinate system is formed. Starting from this Cartesian system usually a spherical system is used that is defined as follows. The azimuth angle ϕ is measured in the x-y-plane starting from the x-axis with the radius r indicating the distance from the origin while the polar angle θ is measured in the r-z-plane starting from the z-axis. The pseudorapidity η is used instead of the polar angle since the resulting particle flux as a function of η is characteristically nearly constant,

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]. \quad (3.1)$$

3.1.1. Superconducting Solenoid

The central device around which the CMS experiment is built is the superconducting solenoid, designed to reach a magnetic field of 4 Tesla in a free bore with a diameter of 6 meters and a length of 12.5 meters. With a nominal current of 19.14 kA and an inductance of 14.2 H an energy of 2.6 GJ is stored in the magnetic field. The system is cooled to a target temperature of 4.5 K via a helium refrigeration plant. The solenoid's winding, which is made from a stabilized reinforced NbTi conductor, is composed of 4 layers, due to the high number of ampere-turns required to generate the desired field strength.

3.1.2. Inner Tracker

The innermost tracking system of CMS surrounds the interaction point of the particle beam and has a length of 5.8 m and a diameter of 2.5 m. In its whole volume a homogenous magnetic field is provided by the superconducting solenoid. The pixel and silicon strip detectors of the tracker are responsible for measuring the trajectories of charged particles resulting from primary collisions, as well as reconstructing secondary vertices. At a luminosity of $2 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, twice the original design luminosity, bunch crossings happen every 25 nanoseconds with up to 60 proton-proton collisions per bunch crossing. Each collision produces up to 1000 particles crossing through the detector. The detector design is optimized to provide high granularity and fast response while minimizing the amount of material used, in order to limit unwanted multiple scattering. The vast number of particles traversing the tracker system deploy very high energies in its components, so another big challenge was to develop detector components able to function under that constant stress for its designed lifetime of 10 years.

3.1.3. Electromagnetic Calorimeter

Around the inner tracking system lies the electromagnetic calorimeter (ECAL). The capability of the avalanche photodiodes in the barrel and the vacuum phototriodes in the endcaps to detect the decay of the Higgs boson into two photons is enhanced by the high energy resolution of the homogeneous crystal calorimeter. Its central barrel part consists of 61200 lead tungstate crystals (PbWO_4). Both of the endcaps contain 7324 crystals. The preshower detector in front of the endcaps is primarily designed to identify neutral pions. The layout of the ECAL is shown in Fig. 3.2.

At energies above 500 GeV shower leakage from the rear of the calorimeter starts to play a significant role. Below those energies the energy resolution can be

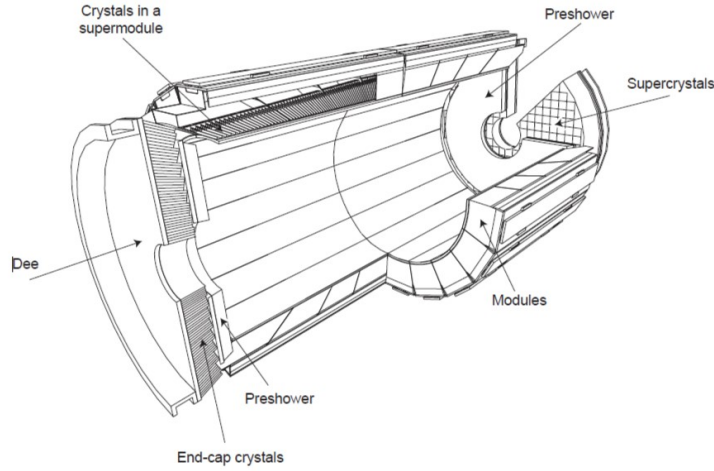


Figure 3.2.: ECAL layout of crystal modules, supermodules and endcaps [25].

parametrized as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (3.2)$$

where S stands for stochastic contributions, N is the noise term and C a constant term. Measurements during beam tests showed typical values of $S=2.8\% \text{ GeV}^{\frac{1}{2}}$, $N=0.12 \text{ GeV}$ and $C=0.30\%$ [25], the resulting energy resolution is shown in Fig. 3.3.

3.1.4. Hadron Calorimeter

The Hadron Calorimeter (HCAL) primarily measures the energy of hadrons, but it is also important for indirectly measuring non-interacting particles such as neutrinos by reconstructing the missing energy and momentum of a proton-proton collision. The HCAL is divided into four subsystems. The hadron barrel (HB), consisting of two half-barrel sections, is detecting particles in the pseudorapidity range of $|\eta| < 1.3$. It is located between the ECAL and the solenoid which represent the radial boundaries of $1.77 \text{ m} < R < 2.95 \text{ m}$. To provide additional material that can be used to absorb the hadronic shower an outer hadron calorimeter (HO) is placed outside the solenoid. The HB is closed off by the hadronic endcaps (HE) which extend the covered pseudorapidity range to $|\eta| < 3$. Beyond that range the forward hadron calorimeters (HF) extends the detectable range up to $|\eta| < 5.2$. It is used to measure the forward energy flux with on average 760 GeV being deposited into the two forward calorimeters per proton-proton interaction. With this

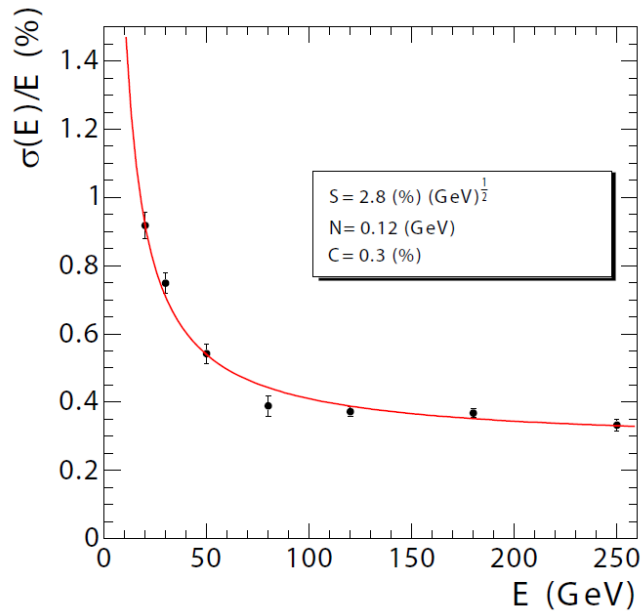


Figure 3.3.: Energy Resolution of the electromagnetic calorimeter as measured from a beam test [25].

luminosity information the LHC's performance can be monitored in real time and other measurements can be normalized.

3.1.5. Muon System

The ability to precisely detect muons and measure their properties is essential for recognizing many different interesting processes in the Standard Model and other models. The design goal is to accurately identify muons, measure their momentum and apply triggers to decide whether to record an event..

The muon system, separated in a cylindrical barrel section and two endcaps, consists of 25000 m² of detection planes, using 3 types of gaseous particle detectors. In the barrel detector a total of 250 drift chambers form four concentric cylinders around the beam line, covering a pseudorapidity range of $|\eta| < 1.2$. At the endcaps 468 cathode strip chambers (CSCs) arranged in several groups cover the pseudorapidity region $0.9 < |\eta| < 2.4$ and thus overlap with the barrel section. Both precise muon measurements and muon triggering can be done in a CSC. The layout of the CSC groups and other components of the muon system is given in Fig. 3.4.

A total amount of 480 rectangular resistive plate chambers (RPCs) form six

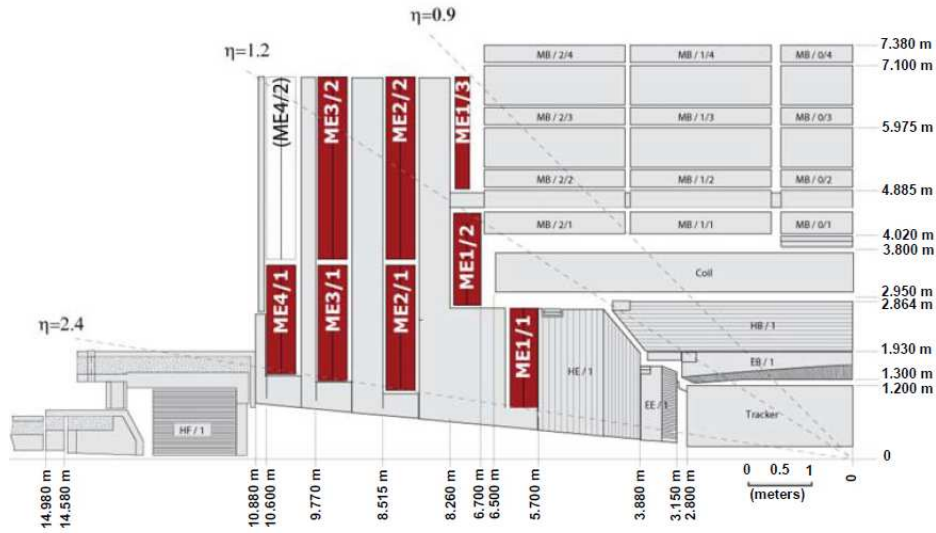


Figure 3.4.: Quarter-view of the layout of the CMS muon system [25]. ME denotes the endcap muon system CSCs, MB stands for the detectors in the muon barrel section. Also visible are the HCAL barrel section (HB), the HCAL endcaps (HE) and the forward hadronic calorimeter (HF), as well as the ECAL barrel section (EB), the ECAL endcaps (EE) and the inner tracker.

coaxial cylinders around the beam axis. Due to their capability of tagging an ionising event with a higher frequency than the rate of LHC bunch crossings they are able to clearly identify the bunch crossing a muon track is associated to.

By combining the data from the muon system and the inner tracker the best possible resolution of the muon transverse-momentum (p_T) can be achieved. Fig. 3.5 shows the relative resolution of the different components and their combined performance as a function of the p_T in different pseudorapidity regions.

3.1.6. Trigger System

For protons, beam crossings happen at a frequency of up to 40 MHz. To cope with the enormous amount of data resulting from the high number of events, a trigger system marks the start of the physics event selection process. In the first step the Level-1 Trigger consisting of custom-designed electronics reduces the output rate to around 30 kHz with a maximal output rate of 100 kHz[26]. For the second step the High-Level Trigger (HLT), a farm of about a thousand processors running a special filter software, has access to the whole readout data and can already

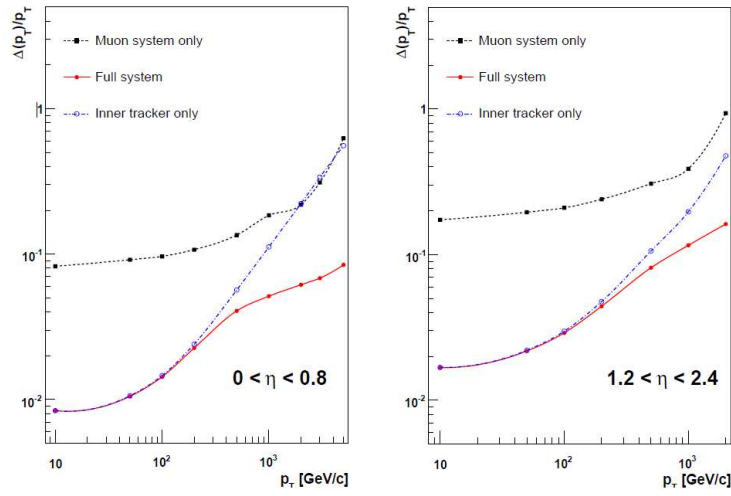


Figure 3.5.: Expected relative resolution of the muon transverse-momentum as presented in Ref. [25] in 2008.

perform calculations similar to the ones used for offline reconstruction. Together both triggers are designed to reduce the rate of events down to about 1000/s.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

4. Event Reconstruction and Simulation

In this section the reconstruction and identification algorithm of objects in the CMS detector as well as the prediction of the specific backgrounds made with Monte Carlo simulations are discussed.

4.1. Particle Reconstruction

Particles resulting from the proton-proton collisions in the CMS are reconstructed with the particle-flow (PF) algorithm described in Ref. [27]. It combines the information of all CMS subsystems to describe one event per bunch crossing containing all recorded data. As a first step the tracks of charged particles in the inner tracker and clusters of deposited energy in the calorimeters are reconstructed. This information is then linked together through the different detector components to derive fully reconstructed particles. This happens iteratively: muons, which are identifiable with high precision, are processed at first and all associated cluster and track information is removed. Next, identified electrons are removed and the remaining clusters and tracks are linked together and analyzed to identify the rest of the collision objects, an overview is given in the following sections. The reconstructed particle information can further be used for jet reconstruction, reconstruction of the missing transverse momentum and energy (see Sec. 4.1.5) or for identification of hadronic τ decays.

4.1.1. Muons

To combine information about muons from the inner tracker and the muon system several different approaches are used. "Standalone-muon tracks" are built by starting from a group of drift tube (DT) or CSC segments registering a hit and exploiting all information along its way using a Kalman-filter technique [28]. "Tracker muon tracks" are reconstructed by using a track of a muon with a p_T above 0.5 GeV and a total momentum of more than 2.5 GeV from the inner tracker system and extrapolating it to the muon system, looking for muon segments matching the trajectory while taking into account possible interactions of the muon with

the detector material along its flight path. A match is valid if the respective segment is no further than 3 cm away from the track if the ratio of the distance to the track uncertainty is smaller than 4. "Global muons" are built with an "outside-in" approach, matching tracks in the muon system to tracks of the inner tracker. If both a global muon and a tracker muon get reconstructed from the same tracker track the result is merged to one particle. About 99% of produced muons with a trajectory in physical reach of the muon system are detected in at least one of those two categories.

4.1.2. Electrons

Electrons are reconstructed using a combination of the information of the charged track in the Inner Tracker system and the deposited energy in the ECAL. The detector is optimized for electrons in the p_T range from 5 to 50 GeV [29]. Difficulties in the reconstruction of electrons arise from the strong magnetic field aligned with the beam axis as well as from the large amount of matter located in front of the inner tracking system in which electrons radiate bremsstrahlung while decelerating. This results in a large ϕ spread of the energy arriving at the ECAL. To actually register the electron energy at the primary vertex the bremsstrahlung photons have to be collected. It has been shown that about 97% of the incident energy of electrons and photons is contained in a 5×5 crystal array in the ECAL [30]. To account for spread in ϕ due to bremsstrahlung photons a "supercluster" is built by clustering several clusters in ϕ direction and summing up the registered energy deposits. The main backgrounds in the detection of electrons stem from misidentified hadron overlaps in jets and prompt electrons produced by semileptonic decay of mainly charm and bottom quarks. Boosted decision trees, trained on simulated events, are used to estimate whether an electron is fake or real. Additionally there is a background originating from photons traversing the tracker and being converted into e^+e^- pairs. These converted photons can be identified by analyzing the associated R_9 shower shape variable defined as the fraction of energy of a 3×3 crystal array centered on the most energetic crystal divided by the total energy in the supercluster [31].

4.1.3. Hadronic Tau Decays

The hadron-plus-strips (HPS) algorithm reconstructs the different hadronic decay modes of the τ lepton (2.4) by starting from the constituents of reconstructed jets. Neutral pions associated with different τ_h decay modes promptly decay into a photon pair which has a high likelihood of converting into an electron-positron pair getting separated in the $\phi \times \eta$ plane by the strong magnetic field on its path through the detector material. The photon and electron candidates are clustered

together in rectangular $\Delta\phi \times \Delta\eta$ regions called 'strips' to reconstruct the original neutral pions. While earlier versions of the HPS algorithm used fixed strip sizes [32], the current version supports variable sizing, based on the p_T of highest- p_T photon or electron not yet included in a strip. This accounts for low- p_T electrons and photons resulting from charged pions undergoing nuclear interactions in the tracker material and photons and electrons resulting from π^0 decays, which experience scattering and bremsstrahlung effects multiple times, which could all end up outside of a fixed sized strip and affect the isolation of a τ_h candidate despite directly resulting from the τ_h decay.

Reduction of $\text{jet} \rightarrow \tau_h$ misidentification can be achieved by imposing certain isolation requirements on the τ_h candidates. Two types of discriminants are used, the first being isolation sum discriminants. The isolation sum is defined as the sum of the scalar p_T of photons and of charged particles within the isolation cone of $\Delta R=0.5$ of the τ_h candidate while excluding charged hadrons and photon-constituents of the τ_h candidate. To exclude pileup contributions, charged particles originating in a distance $d_z > 0.2\text{cm}$ from the production vertex of the τ_h candidate are not included in the p_T sum. A **loose**, **medium** and **tight** working point is defined for the isolation sum being less than 2.5, 1.5 or 0.8 GeV, respectively [33]. The second type of discriminant is MVA-based and use the isolation complemented by other variables such as $p_T^{\text{strip,outer}}$, which is the p_T sum of electrons and photons outside of the signal cone but still included in the strip and p_T weighted ΔR , $\Delta\phi$ and $\Delta\eta$ of photons and electrons in involved strips as inputs on a classifier based on a boosted decision tree. Further included input parameters are the reconstructed τ_h decay mode, τ lifetime sensitive variables and the multiplicity of photons and electron candidates with $p_T > 0.5\text{GeV}$ in the signal and isolation cones, as well as the transverse impact parameter together with its significance [33, 32]. For the MVA-based discriminants the six working points **very loose**, **loose**, **medium**, **tight**, **very tight** and **very-very tight** are defined by BDT discriminant requirements [33].

Bremsstrahlung photons emitted by electrons in the detector material can mimic neutral pions. Together with the original electrons a misreconstruction of τ_h candidates in either the h^\pm or $h^\pm\pi^0$ decay mode can occur. The main handle to minimize misidentification of those electrons as τ_h candidates is the application of another boosted decision tree using variables associated with the energy deposition of electron and photons in the ECAL and HCAL, the (η, ϕ) distance between the photons in any strip and the leading charged particle, the mass of the τ_h candidate, the number of hits in the silicon pixel and strip detector using different reconstruction algorithms and the χ^2 per degree-of-freedom of the track reconstructed using a Gaussian sum filter [32]. The number of photons in the associated strips and the fraction of τ_h energy carried by photons complements the input variables [33].

To discriminate muons against τ_h objects in the h^\pm decay mode a veto is imposed when a signal near the τ_h candidate direction is found in the muon detector.

4.1.4. Jets

The collision of protons at high energies results in the scattering of the contained quarks, anti-quarks and gluons. In order to fulfill confinement the separated fragments hadronize and build a cone of material called jet flying in the direction of the initial free particle. Information about the properties of the original partons is carried in the resulting jets which is why it is important to study them extensively.

The reconstruction of jets for this analysis was done with the anti- k_T jet clustering algorithm with a jet distance parameter of $R=0.4$ [34]. A key characteristic of the anti- k_T algorithm is its prioritization of hard particles when reconstructing jets. The boundaries of the jet are only affected by other hard particles closer than a distance $2R$ in the $\phi - \eta$ plane. If no other hard particles are present in that radius, all soft particles in the radius R are said to be accumulated by the perfectly conical hard jet. Hard neighbors closer than $1R$ are merged into one jet with the final cone being the union of each separate cone and a centered cone of radius R .

To account for systematic errors in the reconstruction of jets, like detector noise or non-linear calorimeter response, jet energy corrections are derived [35]. For the identification of jets originating from the decay of a bottom quark as shown in Fig. 4.1 the combined secondary vertex v2 b tagging algorithm is used [36] for the data observed in the 2016 run. Those jets are identifiable due to the specific lifetime of hadrons containing b quarks of about 1.5 ps which results in a typical displacement for the corresponding secondary vertex of a few mm to one cm. For 2017 data, the DeepCSV algorithm is used for the identification of b-tagged jets. Both algorithms classify a jet with a transverse momentum $p_T > 20$ and a pseudorapidity $|\eta| < 2.4$ as b-tagged if their b-tag discriminator surpasses the medium working point of the respective algorithm.

4.1.5. Missing Transverse Energy

All SM particles created in the pp-collision at the LHC can be detected by one of the several subsystems of the CMS detector except for Neutrinos. But even without direct measurements one can derive information about those particles via the laws of conservation of energy and momentum. The exact momentum prior to the collision can not be known exactly due to indeterminable momentum sharing of the constituents of a proton but the momentum in the plane perpendicular to the beam axis must equal zero. By determining the transverse momentum of all collision products information about the not detectable particles can be collected.

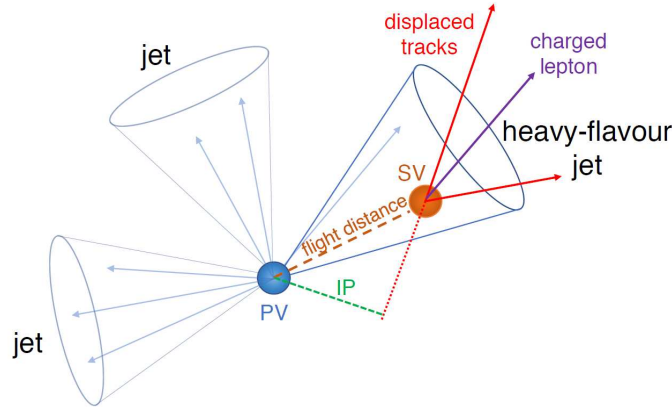


Figure 4.1.: Schematics of a heavy-flavor jet (resulting from a bottom or charm quark decay) with a typical displacement of the secondary vertex defined by the decay time of the quark [36].

An overview of the kinematic relations is shown in Fig. 4.2. The value of the missing transverse momentum ($\vec{\cancel{E}}_T$) and its scalar value, called missing transverse energy ($\cancel{E}_T = |\vec{\cancel{E}}_T|$), is important for many different analyses, for example for the search for BSM or SUSY particles as well as for SM analyses and the initial discovery of the Higgs boson [3]. Due to the very process of quantifying \cancel{E}_T it is very sensitive to the quality of detector signals or inaccuracies in the reconstruction of particles. Resolution and scale of the \cancel{E}_T measurement is determined by comparing the well-measured vector boson momentum of events with an identified Z boson or direct photon where no genuine \cancel{E}_T is expected, which is balanced with the momentum of the hadronic system, to event samples with a hadronic system similar to SM processes, for example $t\bar{t}$ +jets or W+jets.

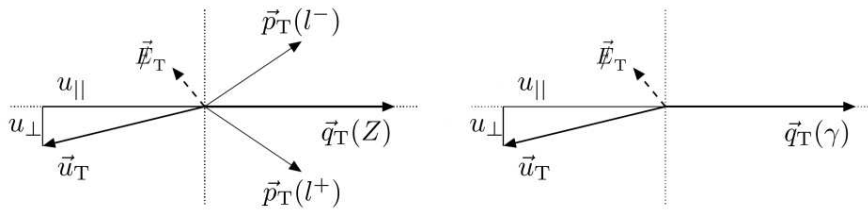


Figure 4.2.: Transverse plane kinematics of $Z \rightarrow l^+l^-$ and direct-photon events [37].

4.2. Simulation

Events generated with Monte Carlo methods are used to simulate the different processes happening during pp-collisions for a better understanding of the observed data. Different software packages are used for different types of processes.

The underlying hard subprocess can be analyzed with lowest order precision perturbative QCD. For the W +jets and Z/γ^* +jets samples the event generator **MadGraph** [38] is used. Single top and $t\bar{t}$ samples are generated with **POWHEG** [39] at next to leading order precision. Hadronization and parton showering, as well as τ lepton decays are simulated using the **PYTHIA8**[40] generator with the **CUETP8M1** [41] tune. To extract simulated event datasets as if they were measured by the real CMS instruments, all events are passed through a **GEANT-4** [42] simulation of the detector system. The final events are obtained by passing the data through the same reconstruction process as collision data. Several corrections have to be applied to account for selected trigger path efficiencies and efficiencies in the tracking, identification and determination of isolation for electrons and muons as well as for the τ_h candidates [43].

5. Summary of the Measurement of Higgs boson production and decay to the $\tau\tau$ final state

The basis of the analysis is the data collected at the CMS experiment in 2016 and 2017 combining all events in the $e\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels at a center-of-mass energy of 13 TeV, amounting to an integrated luminosity of 77.4 fb^{-1} . The goal was to measure the inclusive cross section $\sigma_{incl}\mathcal{B}(H \rightarrow \tau\tau)$ for Higgs boson production with a subsequent decay into a pair of two τ leptons as well as the respective cross sections for individual production modes. Additionally kinematic production properties were investigated in terms of simplified template cross sections. To estimate the expected SM background processes a variety of methods was used as shown in Tab. 5.1. The data driven fake factor method discussed in this thesis and the τ -embedding technique are able to cover about 90% of the background events in the channels with at least one hadronically decaying τ lepton.

5.1. Event categorization

Classification into signal and background categories for events passing the preselection is done via fully connected feed-forward neural networks using two hidden layers with 200 nodes each using a hyperbolic tangent activation function and resulting in five to eight output nodes, depending on the final state. Each output node stands for a distinctive event class, as predicted by the neural network, using two signal categories ($gg\rightarrow H$ and VBF) and three to six background categories. The value of the output nodes can be interpreted as the probability of the event being in the respective class, it is then assigned to the category with the highest output value. A detailed list of event classes per final state is given in Tab. 5.2. For each channel and year a separate neural network is trained. Datasets created with simulation samples are used as training samples, except for QCD multijet events since there is no reasonably well modelled simulation for this category. Instead, data events from sideband regions are used as QCD training samples. This means events with same sign tau pair candidates for the $\mu\tau_h$ and $e\tau_h$ channels and

Process	Misidentification	$e\mu$	$e\tau_h$	$\mu\tau_h$	$\tau_h\tau_h$
$Z \rightarrow \tau\tau$		EMB	EMB	EMB	EMB
W+jets		MC	F_F	F_F	F_F
QCD multijet		CR	F_F	F_F	F_F
$Z \rightarrow \ell\ell$	Jet $\rightarrow \tau_h$	MC	F_F	F_F	F_F
	$\ell \rightarrow \tau_h$		MC	MC	MC
Diboson+single t	Jet $\rightarrow \tau_h$	MC	F_F	F_F	F_F
	$\ell \rightarrow \tau_h$		MC	MC	MC
$t\bar{t}$	Jet $\rightarrow \tau_h$	MC	F_F	F_F	F_F
	$\ell \rightarrow \tau_h$		MC	MC	MC

Table 5.1.: SM background processes contributing to the event selection[43]. Processes marked with "MC" are estimated from simulation, " F_F " stands for estimation via the fake factor method presented in this thesis, "EMB" indicates estimation using the τ embedding technique discussed in Section 7.1 and "CR" stands for estimation of shape and normalization of QCD multijet events from control regions in data. Genuine decays into τ leptons from the diboson+single t and $t\bar{t}$ backgrounds are also taken from embedded samples.

tau pairs with an anti-isolated sub-leading τ_h in the $\tau_h\tau_h$ channel.

5.2. Results

The measured Higgs boson production cross section multiplied with the branching fraction into a pair of τ leptons, inclusively and split by production modes, is listed in Tab. 5.3. Additionally the obtained results were presented as simplified template cross sections as defined in Ref. [15]. For the signal strength a value of $\mu_{incl} = 0.75 \pm 0.18$ has been found which is consistent with the expected value from the SM within the experiment's statistical precision with a p-value of 0.11[43].

Process	$e\mu$	$e\tau_h$	$\mu\tau_h$	$\tau_h\tau_h$
$gg \rightarrow H$	ggH	ggH	ggH	ggH
VBF	qqH	qqH	qqH	qqH
$Z \rightarrow \tau\tau$	ztt	ztt	ztt	ztt
QCD	qcd	qcd	qcd	qcd
$t\bar{t}$	tt	tt	tt	
$Z \rightarrow \ell\ell$	misc	zll	zll	misc
W+jets		wj	wj	
Diboson	db			
Single t	st	misc	misc	

Table 5.2.: Assignment table between signal or background processes and their respective classification category used in the analysis depending on the $\tau\tau$ final state.

Higgs boson production cross section and subsequent decay	Measured result
$\sigma_{incl} \mathcal{B}(H \rightarrow \tau\tau)$	$2.56 \pm 0.48(\text{stat}) \pm 0.34(\text{syst}) \text{pb}$
$\sigma(gg \rightarrow H, bbH) \mathcal{B}(H \rightarrow \tau\tau)$	$1.11 \pm 0.81(\text{stat}) \pm 0.78(\text{syst}) \text{pb}$
$\sigma(VBF + V(qq)H) \mathcal{B}(H \rightarrow \tau\tau)$	$0.34 \pm 0.08(\text{stat}) \pm 0.09(\text{syst}) \text{pb}$

Table 5.3.: Summary of the measurement results for the product of the Higgs boson production cross section with the branching ratio into a pair of τ leptons inclusively and split by the Higgs production modes[43].



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

6. The Fake Factor Method

Misidentification of jets as hadronic τ lepton decays amounts to a large fraction of the backgrounds for the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ final states. The *fake factor method* provides an estimation of the so-called background of *fake taus* which is entirely data-driven except for corrections partially based on simulated events to remove biases in the prediction. Previous analyses using this method include the search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state and the measurement of the $Z/\gamma^* \rightarrow \tau\tau$ cross section described in Ref. [44, 45].

6.1. Calculation

The idea of the fake factor method is to measure the number of events in an Application Region (AR) which is distinct from the Signal Region (SR) by the τ_h isolation requirement. In this implementation the `tight` isolation must fail while the `vloose` criterion is fulfilled, a state called anti-isolation. The anti-isolation leads to an AR which is orthogonal to the SR and very pure in fake taus. To estimate the number of background events from a background in the SR, the number of events in the AR is weighted with the FF, which is defined by the ratio of events that fulfill the `tight` working point over events that fulfill the `vloose` but fail the `tight` working point of the τ_h ID discriminant.

$$\text{FF}_i = \frac{n_{\text{Tight}}}{n_{\text{VeryLoose}}}, i \in \{\text{QCD}, \text{W+jets}, \text{t}\bar{\text{t}}\} \quad (6.1)$$

For each background i the fake factor is determined in a dedicated Determination Region (DR_i). Other backgrounds are estimated from simulation and subtracted from the numerator and denominator of the FF ratio. Using the results from DR_i in the SR assumes the ratio of fake taus from the specific background i in the SR to be equal to the determined FF ratio in the DR_i . The fake factor is determined differentially as a function of different variables, taking into account the most important kinematic dependencies, to ensure the validity of the assumption. Corrections are determined to account for biases when applying the FF.

The applied combined FF is the weighted average of all different fake tau contributions, weighted with the expected fraction of contribution of each background in the AR. The predicted number of events with jets misidentified as τ_h in a given

range of a kinematic variable is calculated via Eq. (6.2) by multiplying the number of events in the AR with the fake factor.

$$N_{jet \rightarrow \tau_h} = N_{AR} \cdot \sum_i w_i \cdot FF_i \quad (6.2)$$

The weights of the fake factor in Eq. (6.2) are determined in the AR by estimating the ratio of the number of events of a specific background over the total amount of events in the AR.

$$w_i = \frac{N_{AR}^i}{\sum_j N_{AR}^j}, i, j \in \{\text{QCD}, \text{W+jets}, t\bar{t}\} \quad (6.3)$$

For each background category the fake factor is estimated distinctively for events with a jet multiplicity of zero (0-jet) and for events with a jet multiplicity of one or more (1-jet). A function consisting of a Landau function and a polynomial of degree two is fit in dependence of the τ_h candidate's transverse momentum, $p_{T,\tau}$.

6.1.1. QCD Fake Factors

To obtain a region very pure in QCD multijet events several cuts have to be modified with respect to the event selection of the SR. In Tab. 6.1 the requirements for events in the $\mu\tau_h$ and $e\tau$ channels are listed.

	DR	SR
Electric charges of light lepton and τ_h candidate	Same sign (SS): $q_l \cdot q_{\tau_h} > 0$	Opposite sign (OS): $q_l \cdot q_{\tau_h} < 0$
Relative light lepton isolation ($\mu\tau_h$)	In range 0.05 - 0.15	<0.15
Relative light lepton isolation ($e\tau_h$)	In range 0.02 - 0.15	<0.15

Table 6.1.: Definition of the QCD DR in the $\mu\tau_h$ and $e\tau_h$ channels

In the $\tau_h\tau_h$ channel only the same sign requirement is used to define the DR. Since in this channel two τ_h candidates are present, the fake factor calculation is first done with one of them being selected as the leading τ_h required to pass the **tight** isolation working point, and then a second time with the other one being the leading τ_h candidate to avoid biases when selecting the leading τ_h , and finally weighted by a factor of 0.5. The QCD multijet background is the dominating background in the $\tau_h\tau_h$ final state by a large margin, so the QCD fake factors can be used to estimate the other backgrounds as a reasonably good approximation. Results of the QCD fake factor calculation are shown in Fig. 6.1.

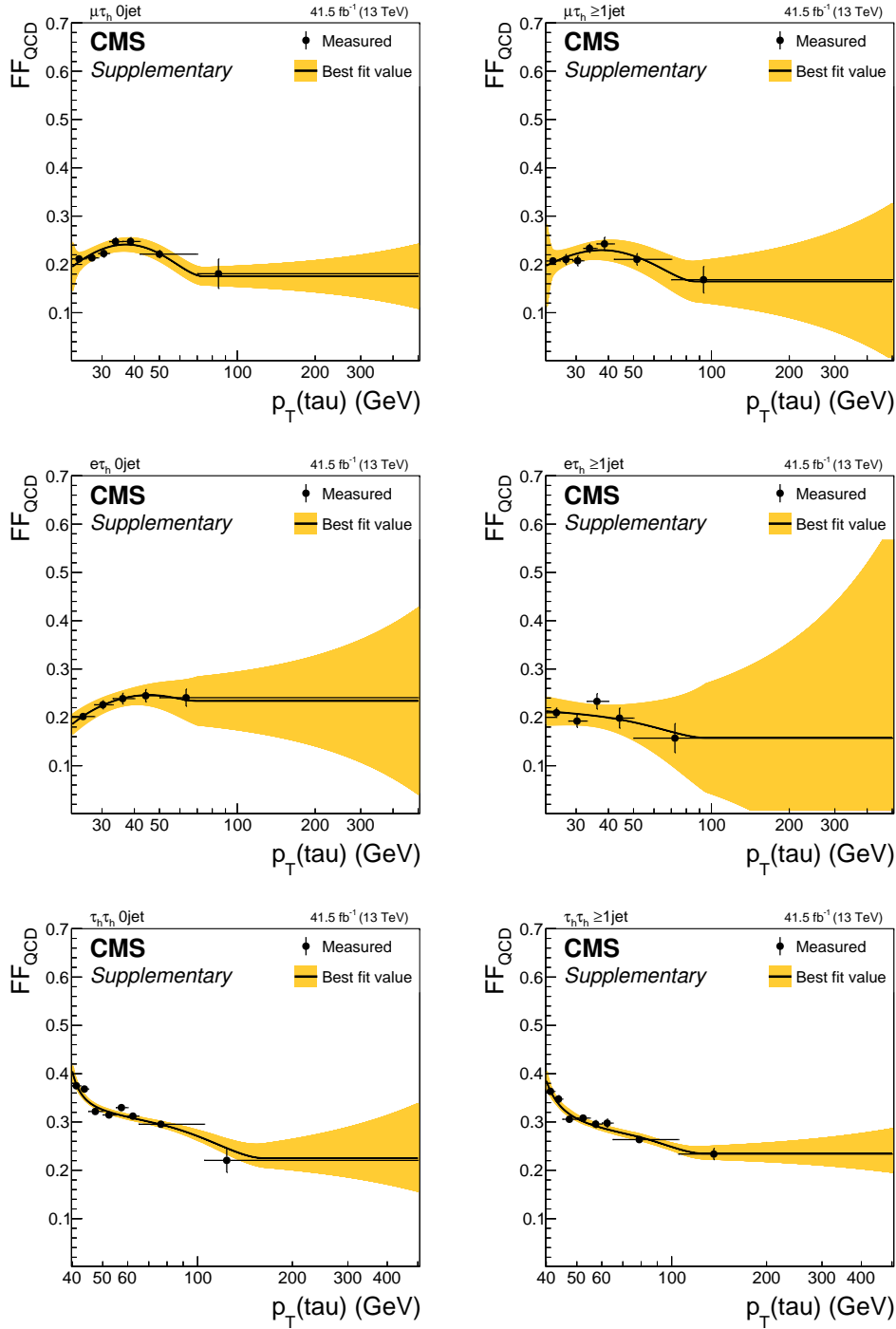


Figure 6.1.: The uncorrected fit in p_T dependency of the QCD fake factors calculated in the QCD DR for the $\mu\tau_h$ (first row), $e\tau_h$ (second row) and $\tau_h\tau_h$ (third row) channels. The fake factors in the left column are for events with jet multiplicity zero, the ones in the right column are for events with a jet multiplicity of at least one.

6.1.2. W+jets Fake Factors

For the $\mu\tau_h$ and $e\tau_h$ channels the DR of the W+jets fake factors is defined by the same requirements as for the SR except for two additional ones regarding the transverse mass

$$m_T^l = \sqrt{2p_T^l \cancel{E}_T (1 - \cos\Delta\phi)} \quad (6.4)$$

on the one hand, where $\Delta\phi$ defines the angle between the light lepton momentum p_T^l and the missing energy vector $\vec{\cancel{E}}_T$, since W+jets events usually have a larger m_T value. The second requirement targets the number of b-tagged jets in the event to reduce contamination by $t\bar{t}$ events. Both requirements are listed in Tab. 6.2.

	DR	SR
m_T	>70 GeV	<50 GeV
Number of b-tagged jets	= 0	≤ 1

Table 6.2.: Definition for the W+jets DR

In Fig. 6.2 the p_T dependency of the uncorrected fake factors is presented.

6.1.3. $t\bar{t}$ Fake Factors

Fake factors for $t\bar{t}$ processes are calculated using simulated events after the preselection since there is no region pure in $t\bar{t}$ events with enough events to provide sufficient statistical significance to be suitable as DR. The obtained $t\bar{t}$ fake factors are shown in Fig. 6.3. As can be seen in Fig. 6.4, the number of fake taus resulting from $t\bar{t}$ processes is notably lower than the number of fakes from other processes, this holds true for every channel.

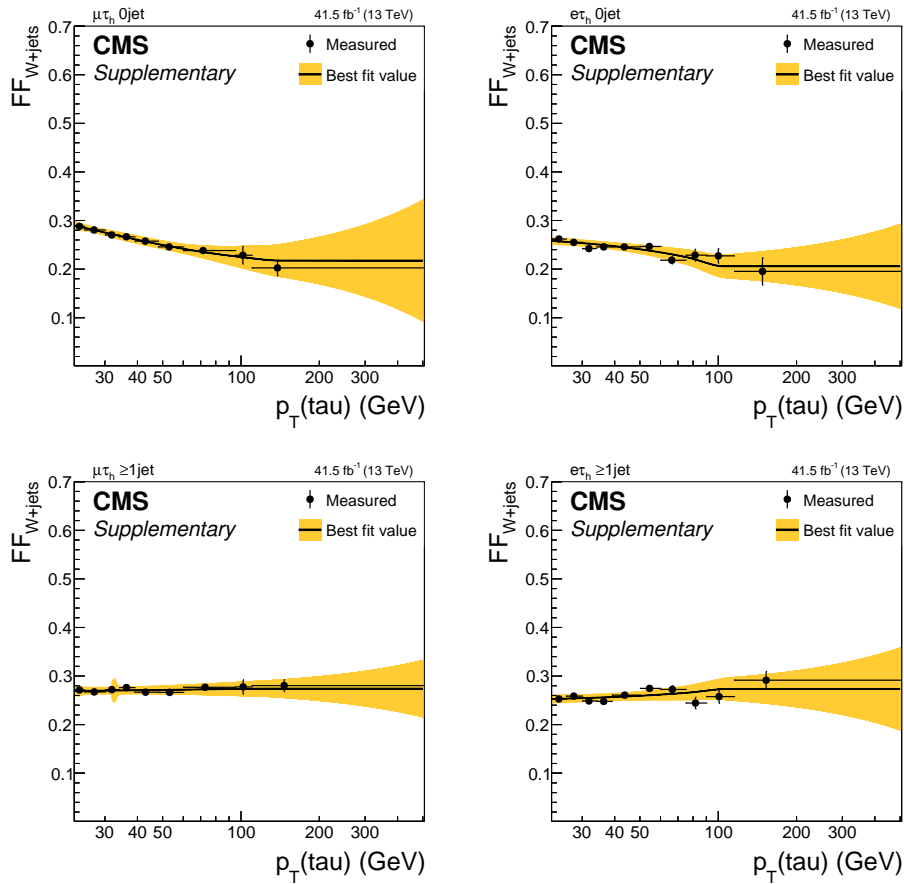


Figure 6.2.: The uncorrected fit in p_T dependency of the W +jets fake factors calculated in the W DR for the $\mu\tau_h$ (left column) and $e\tau_h$ (right column) channels. The fake factors in the first row are for events with jet multiplicity zero, the ones in the second row are for events with a jet multiplicity of at least one.

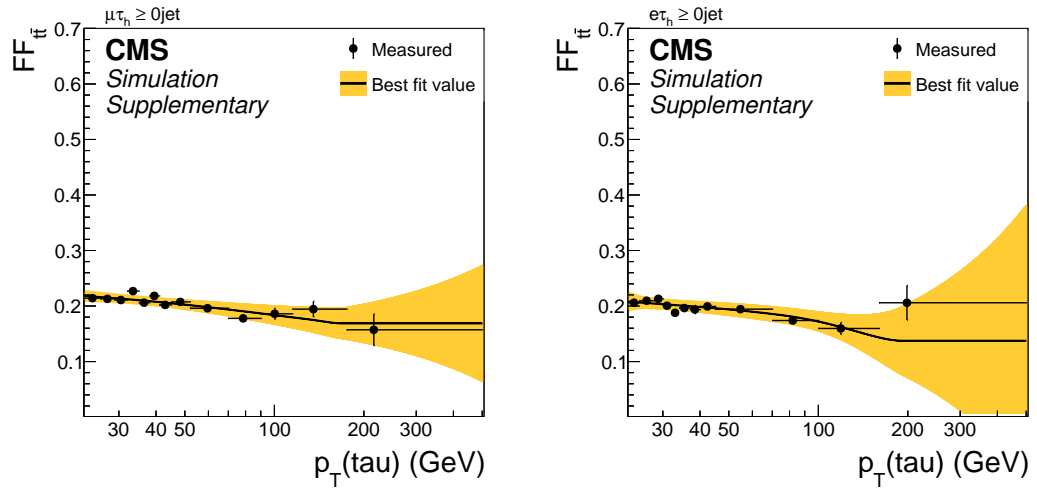


Figure 6.3.: The uncorrected fit in p_T dependency of the $t\bar{t}$ fake factors calculated from simulated events in the $\mu\tau_h$ (left) and $e\tau_h$ (right) channels.

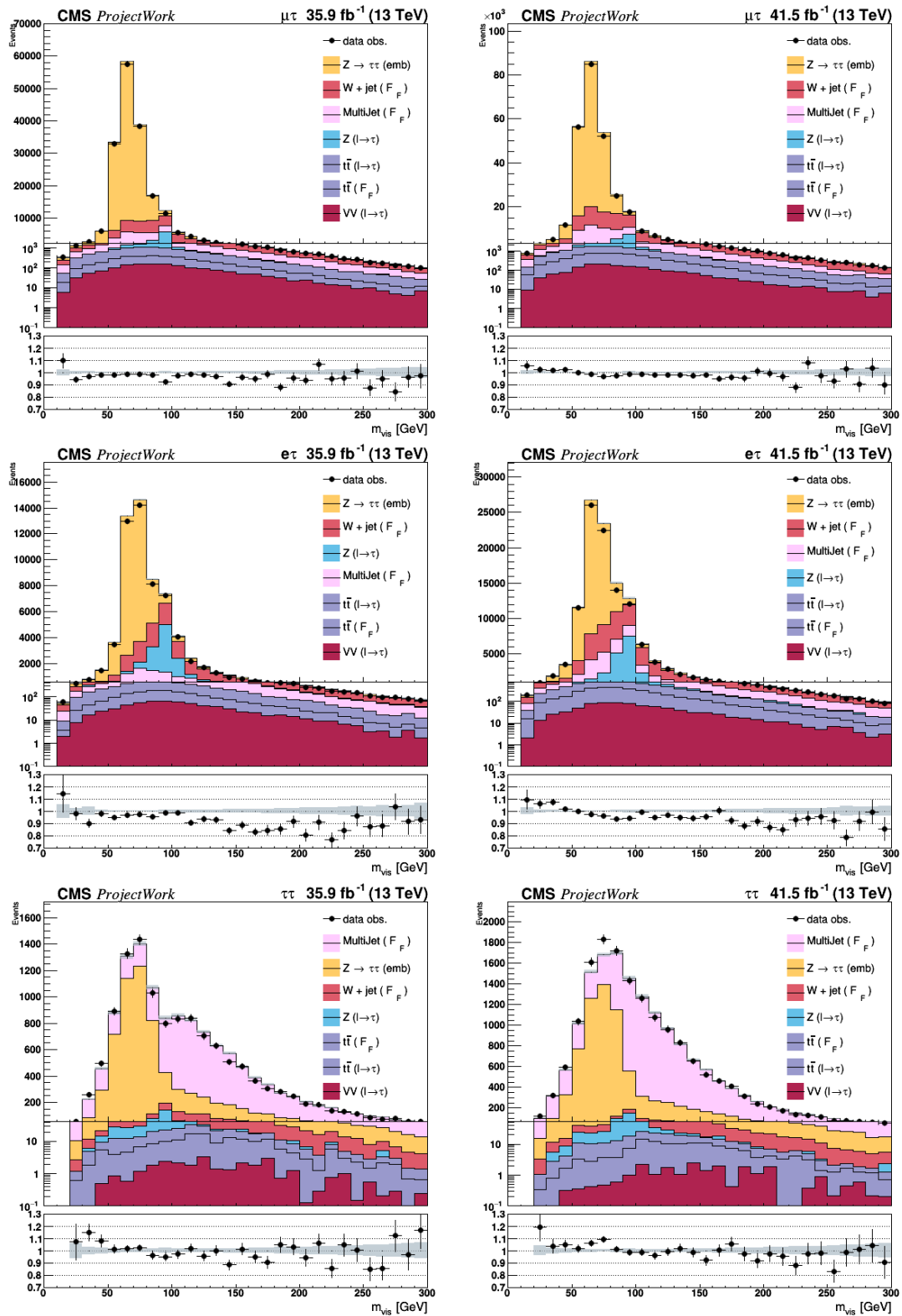


Figure 6.4.: Comparison of data versus prediction in the $\mu\tau_h$, $e\tau_h$ and $\tau\tau_h$ signal regions with the different fake τ_h contributions plotted individually, for 2016 (left) and 2017 (right).

6.1.4. Fake Factor Weights

The combined fake factor used on an event in the AR is a composition of the three different fake factor categories, weighted by the expected fraction of events being in the respective categories in the AR.

The QCD background fraction is estimated from data in the application region, but with the additional requirement that the charges of the two detected particles are of the same sign(SS). From the total number of selected SS events an estimation of SS events originating from events of one of the other categories is subtracted based on MC simulation. The other background categories (TT, W+ZJ, $\tau_h \rightarrow \tau_h + \text{lepton} \rightarrow \tau_h$) are estimated via MC simulation.

In practice, the background fractions are not calculated during the fake factor calculation but have to be provided from an external source to the fake factors for an analysis. This way maximal flexibility is ensured, for the assigned variable and binning of the fractions can be chosen to fulfill the specific requirements of the respective analysis.

6.2. Corrections

After deriving the raw fake factors in the DRs, several corrections are derived to account for eventual mismodelling of the estimated fake background. Imperfect modelling can occur due to differences in the composition of the DR and SR arising from the respective selection criteria, or due to an eventual dependency on variables not used to model the fake factor, which is handled by non-closure corrections.

6.2.1. QCD fake factor corrections

Three different corrections are derived for the QCD multi-jet fake factor: the non-closure correction depending on m_{vis} , a correction depending on the lepton isolation and a correction for the extrapolation of the fake factor, which is derived in the same-sign region, to the opposite-sign region. This is necessary because some properties of the second lepton in the event can affect the isolation of the first τ lepton and thus have to be considered when applying the raw fake factors. The non-closure correction is calculated in the same region as the initial raw fake factors by estimating the fake background as a function of m_{vis} and comparing it with the observed background. The ratio of these two is then used as multiplicative factor on the raw fake factor as non-closure correction after smoothing it with a Gaussian kernel.

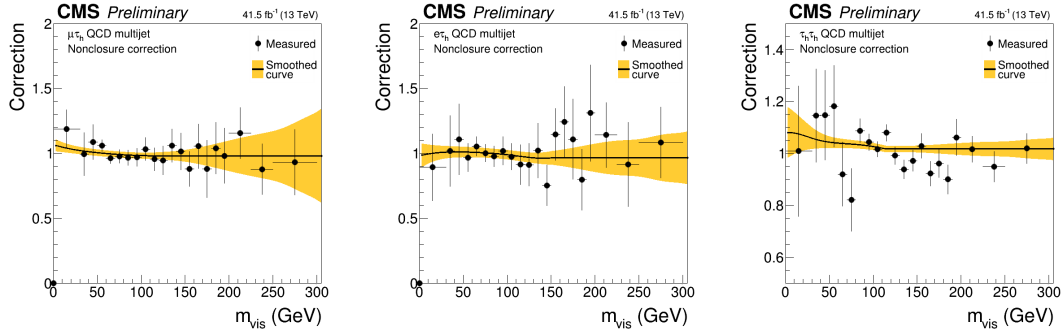


Figure 6.5.: The smoothed QCD multi-jet non-closure corrections for the $\mu\tau_h$ (left), $e\tau_h$ (middle) and $\tau_h\tau_h$ (right) channels for the 2017 FF..

The non-closure corrected fake factors are then used to derive the lepton-isolation correction in the SS region by estimating the fake background as a function of the lepton isolation and calculating the ratio of the estimation and the observed background, smoothed with a Gaussian kernel. The reason for this correction is the correlation of the τ_h and the second lepton, for example in QCD events through their flavor or their p_T due to momentum balance which affect the relative isolation. For low values of the lepton isolation there is a non-negligible contamination by mostly W +jets events which is subtracted based on simulation with an associated uncertainty propagated to the correction.

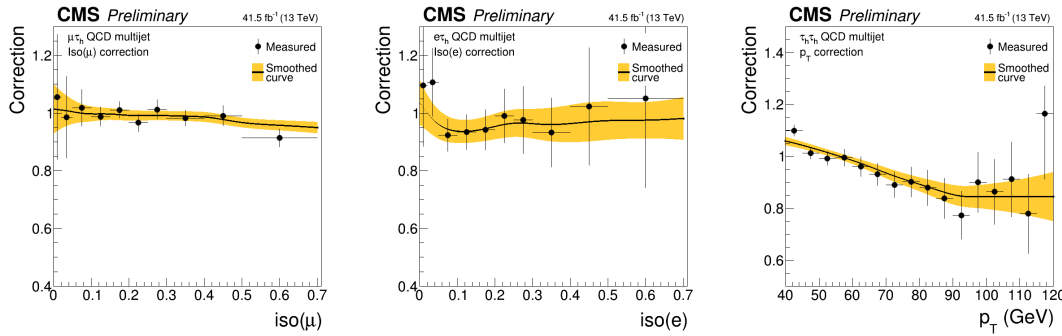


Figure 6.6.: The smoothed QCD multi-jet lepton corrections for the $\mu\tau_h$ (left), $e\tau_h$ (middle) and $\tau_h\tau_h$ (right) channels for the 2017 FF..

To account for the extrapolation of the fake factors to the OS region, a correction is derived in a region where the lepton isolation is in an interval between the upper DR isolation limit and that very limit plus 0.1 for the $\mu\tau_h$ and $e\tau$ channels, or where the non-leading tau is anti-isolated, fulfilling the VLOOSE isolation criteria but failing the TIGHT requirement in the $\tau_h\tau_h$ channel. For this region raw fake

factors and their non-closure correction are calculated and are then applied to a region only differing by a charge product requirement. The ratio of estimated and observed background is stored as OS→SS correction as a function of m_{vis} .

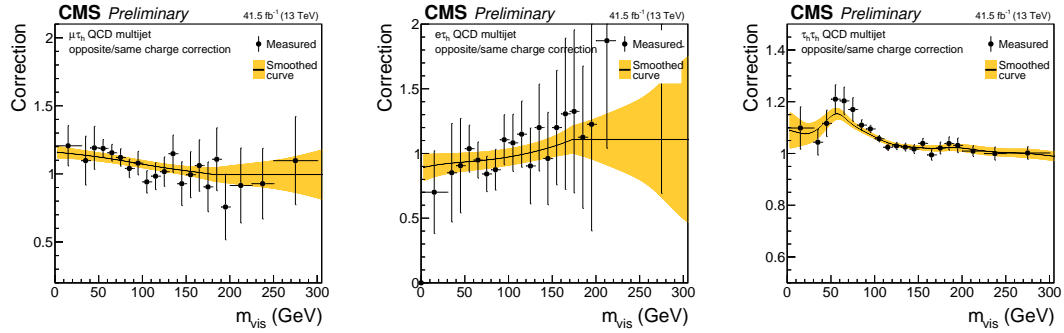


Figure 6.7.: The smoothed QCD multi-jet OS→SS corrections for the $\mu\tau_h$ (left), $e\tau_h$ (middle) and $\tau_h\tau_h$ (right) channels for the 2017 FF.

6.2.2. W+jets fake factor corrections

For the W+jets fake factors two different corrections are derived. A non-closure correction as a function of the visible mass m_{vis} is calculated by estimating the fake background in the region the raw fake factors were derived and comparing the result with the observed fake background. The ratio of these two is used as correction depending on m_{vis} .

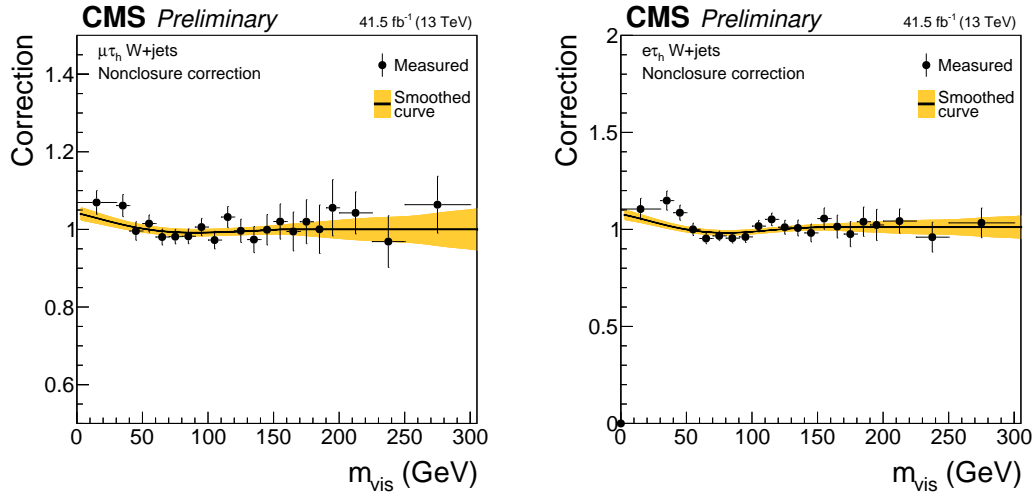


Figure 6.8.: The smoothed W non-closure corrections for the $\mu\tau_h$ (left) and $e\tau_h$ (right) channels for the 2017 FF.

By requiring a minimum p_T on the light lepton for the fake factor calculation, events with no hard W boson recoil and subsequently no hard jets tend to be underrepresented in the low m_T^l region. This affects the τ_h isolation since it is strongly correlated to the p_T difference of the τ_h candidate and the associated jet. To mitigate this effect a second, m_T -dependent, correction is derived by recalculating raw fake factors and the non-closure correction from simulation and estimating the W fake background in simulated events. The ratio of the result and the original simulated W background is further used as m_T -correction.

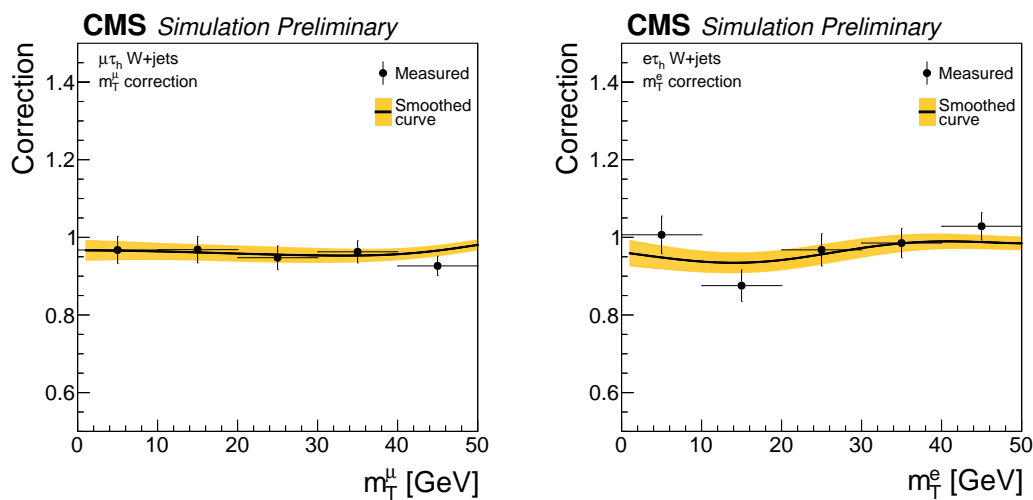


Figure 6.9.: The smoothed W m_T corrections for the $\mu\tau_h$ (left) and $e\tau_h$ (right) channels for the 2017 FF.

6.2.3. $t\bar{t}$ fake factor corrections

To select events for the control region for the calculation of the $t\bar{t}$ fake factor corrections two requirements have to be imposed additionally to the SR selection.

- The number of b-tagged jets in the event must be greater or equal to one (SR: 0 or 1 b-tagged jet)
- There is at least one isolated muon and one isolated electron in the event (SR: veto on a third lepton)

As visible in Fig. 6.10 the selected region is relatively sparsely populated, large statistical uncertainties are the consequence. The requirement of an additional lepton also typically selects events where the τ_h candidate is a jet from a B meson decay and both W bosons decay leptonically, whereas in the SR the τ_h candidate often originates from hadronic W decays, which is why the control region is only

used to derive a constant correction factor as a ratio of the $t\bar{t}$ fake factors in data and simulation, with the difference between those two as uncertainty.

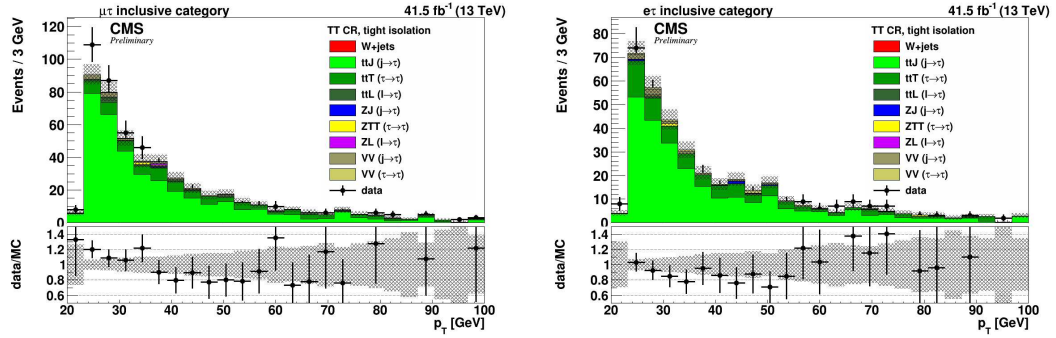


Figure 6.10.: Simulated population of the $t\bar{t}$ +jets control region as a function of the p_T of the τ_h candidate for the $\mu\tau_h$ (left) and $e\tau_h$ channels (right). The expected QCD multijet contribution is not included in the graphic.

The derived correction factors for 2016 and 2017 are as follows:

- $\mu\tau_h$ 2016: 0.88
- $\mu\tau_h$ 2017: 0.98
- $e\tau_h$ 2016: 0.80
- $e\tau_h$ 2017: 0.97

The non-closure correction in dependence of m_{vis} for $t\bar{t}$ fake factors is derived from simulation and is shown in Fig. 6.11.

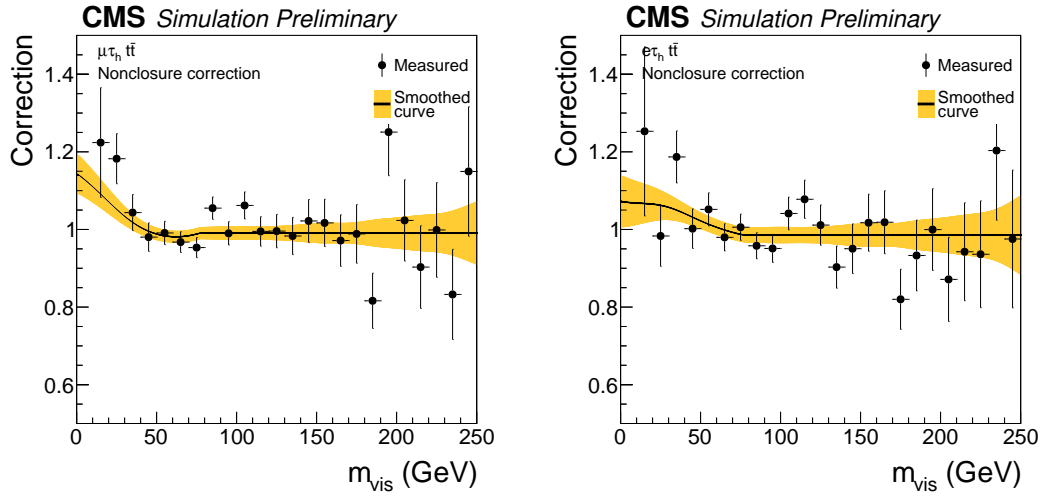


Figure 6.11.: The smoothed $t\bar{t}$ non-closure corrections for the $\mu\tau_h$ (left) and $e\tau_h$ (right) channels for the 2017 FF.

6.3. Uncertainties

The list of fake factor related uncertainties consists of:

- Statistical uncertainty of the fake factor due to uncertainties when fitting the p_T dependence. These are uncorrelated and constitute six nuisance parameters in the $\mu\tau_h$ and $e\tau_h$ channels and two in the $\tau_h\tau_h$ channel.
- Systematic uncertainties related to the QCD multi-jet fake factor corrections. The squared uncertainty of each correction is added for each channel individually resulting in one nuisance parameter per channel.
- Systematic uncertainties related to the W fake factor corrections. The squared uncertainty of each correction is added for each channel individually resulting in one nuisance parameter for the $\mu\tau_h$ and $e\tau_h$ channels.
- Systematic uncertainties related to the $t\bar{t}$ fake factor corrections. The squared uncertainty of each correction is added for each channel individually resulting in one nuisance parameter for the $\mu\tau_h$ and $e\tau_h$ channels.
- Systematic uncertainties on the fraction of W/Z+jets and $t\bar{t}$ events with one misidentified τ_h in the AR, which adds two nuisance parameters for the $\tau_h\tau_h$ channel.

- Systematic uncertainties related to the subtraction of the true- τ background in the anti-isolated region which are correlated across all categories.

All uncertainties of the fake factors are shown in the plots in the Sections 6.1 and 6.2. Not all of the uncertainties have a significant impact on the results. Except for the uncertainty on the subtraction of the true- τ -background in the anti-isolated signal region all uncertainties enter as shape uncertainty and are normalized to the same area as the nominal shape. The squared normalization factors are added and act as separate nuisance parameters, separately for the statistical uncertainties on the raw fake factors and the systematic uncertainties on the corrections. Since all uncertainties are dominantly statistical in nature, they are treated as uncorrelated across channels, except for the uncertainties on $t\bar{t}$ corrections for $\mu\tau_h$ and $e\tau_h$ as they are determined in the same DR.

7. Extension of the fake factor method for the $H \rightarrow \tau\tau$ analysis

7.1. Determination of genuine $\tau\tau$ Backgrounds via Embedding

With this technique the $Z \rightarrow \tau\tau$ background is estimated by hybrid events, merging observed data with simulated events. At first genuine $Z \rightarrow \mu\mu$ events are selected by the application of several selection cuts to obtain a set of data with a purity of 99%[46] for the $Z \rightarrow \mu\mu$ process. The largest remaining background originates from $t\bar{t}$ events. In a second step the energy deposits detected in the calorimeters that are associated with the tracks of the two selected muons in the inner tracker and the muon system and in clusters of the calorimeter are removed. The energy and momentum of the removed muons are then used as input to simulate two τ leptons with the same kinematic properties as the selected muons in an otherwise empty detector via PYTHIA[47]. A correction in the form of a boost of the muon four-momenta is applied to account for the mass difference of muons and τ leptons. From each seeded event a distinct sample for each of the final states $e\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ is created by enforcing the respective decays of the embedded τ particle. The simulated energy deposits are merged to the event from which the selected muons were removed so subsequently all lepton identification is based on the full hybrid event. The entire process of replacing muons with simulated tau leptons is pictured schematically in Fig. 7.1. For validation purposes the same procedure can also be done with the injection of two simulated electrons or muons instead of τ leptons. As visible in Fig. 7.2 the $t\bar{t}$ background represents the biggest contamination in the selected samples. Most of the contribution stems from $t\bar{t}(\mu\mu)$ events, with other contributions being below the required accuracy of this method[46]. The described substitution of muons by τ leptons provides an additional estimate of the final state $t\bar{t}(\tau\tau)$ based on data.

The fake factor method has been adapted to use embedded $Z \rightarrow \tau\tau$ samples instead of samples generated with MC simulation. For both the QCD and W+jet FF the use of embedded samples replaces the MC events for $Z \rightarrow \tau\tau$, $t\bar{t}(\tau\tau)$ and diboson($\tau\tau$) processes from MC background estimations. Both in the estimation of the raw fake factors as well as for the calculation of the respective corrections em-

bedded samples were used for these categories. Only $t\bar{t}$ fake factors are calculated using MC samples entirely and later merged to the other fake factors calculated with embedded samples. This is done due to the $t\bar{t}$ control region requirement of having at least one isolated muon and one isolated electron in each event which is never fulfilled for events generated with the embedding method. In Fig. 7.4 a comparison of the performance when using embedded samples and MC samples is shown. The embedded samples model the background more accurately and reduce outlier estimations, this is notably especially in the lower m_{vis} region.

The lower m_{vis} region is dominated by $Z \rightarrow \tau\tau$ events. Better modelling of this process by the embedded samples leads to general better performance and reduced outliers. The affected fake factors also show a clear difference as visible in Fig. 7.3. As shown in Fig. 7.4 regions where the background of jets misidentified as taus is dominating also show better agreement and less outliers, especially in the $\mu\tau_h$ and $\tau_h\tau_h$ channels.

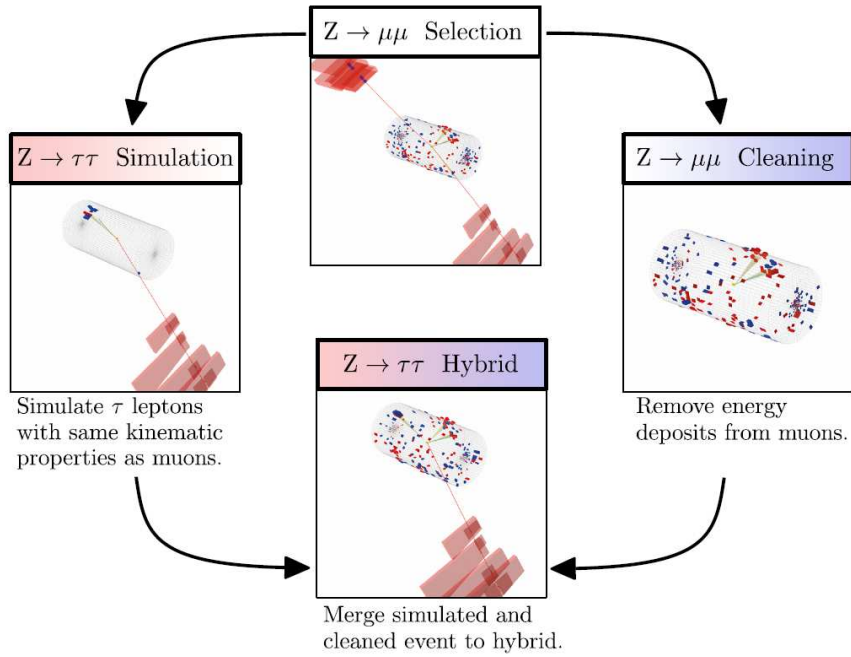


Figure 7.1.: Schematic overview of the τ -embedding process [46]

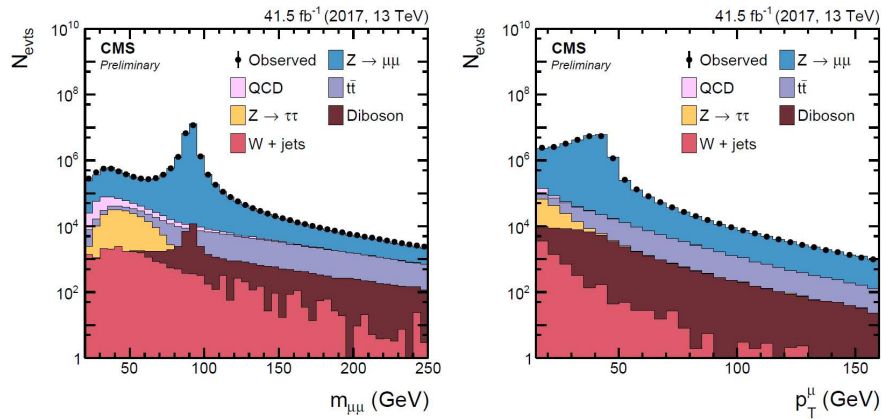


Figure 7.2.: Event distribution with respect to the invariant mass of the selected dimuon pair $m_{\mu\mu}$ (left), and the p_T of the trailing muon after the event selection (right) [46]

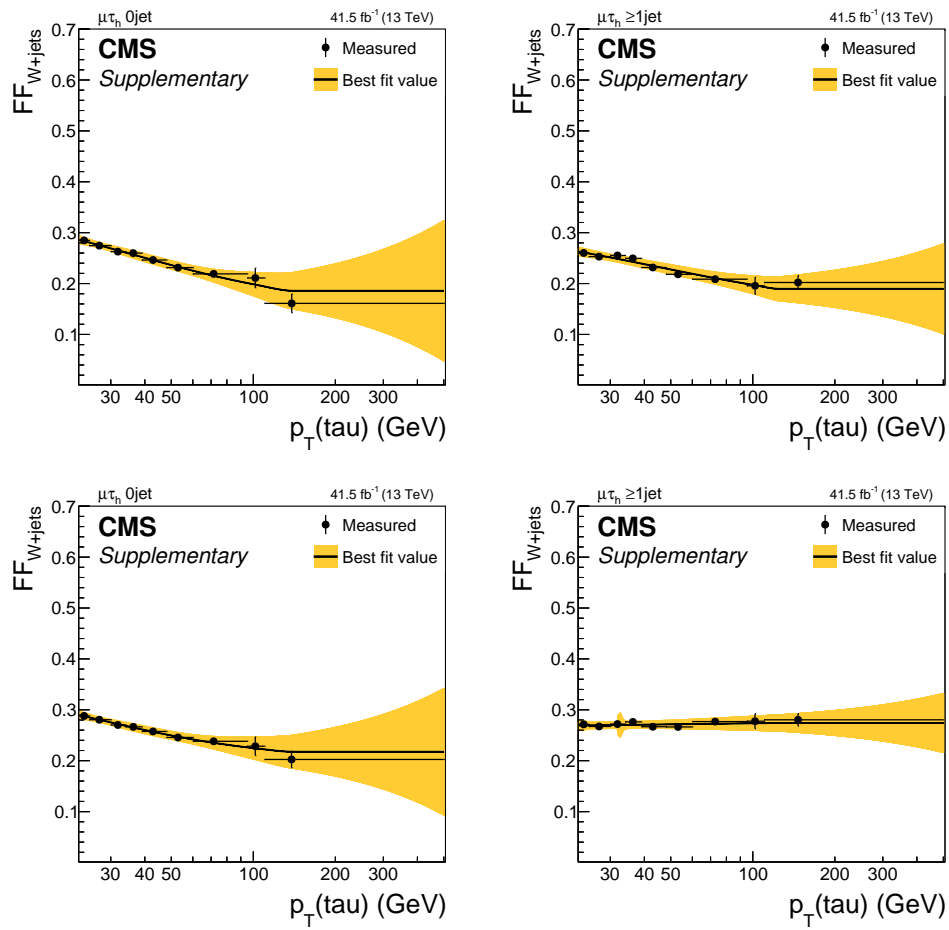


Figure 7.3.: Raw fake factors derived in the $W+jets + \text{Drell-Yan}$ region exemplarily for 2017 data in the $\mu\tau_h$ channel. The top row shows the raw FF for 0-jet (left) and n-jet (right) events calculated without using embedded samples, the FF in the bottom row were calculated with embedded samples.

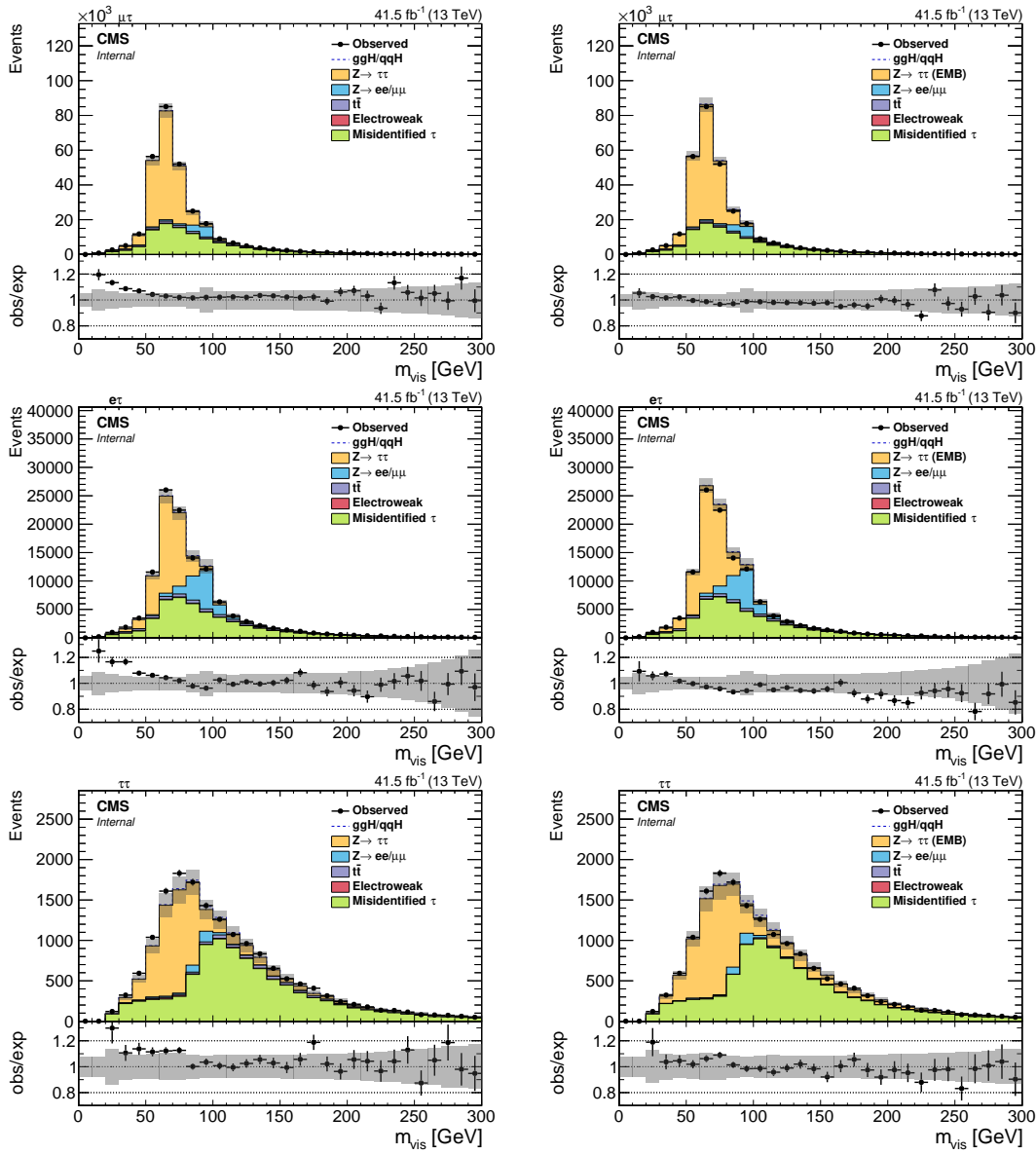


Figure 7.4.: Performance plots for $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels binned in m_{vis} , with fake factors using simulation for $Z \rightarrow \tau\tau$ events (left) and fake factors using embedded samples (right).

7.2. Decay Mode Binning

The past versions of the FF method used to calculate distinct fake factors for events with one charged particle in the τ decay final state (1-prong) and events with three charged particles in the τ decay final state (3-prong). This splitting, however,

naturally leads to higher uncertainty and less accurate fits due to fewer events in the respective categories. Especially the 3-prong category is less populated. It was investigated whether merging the two bins together could help getting a more robust fit result with minimal effect on the general predictive performance of the FF. A comparison of the $p_{T,\tau}$ fits of the fake factors is included in Fig. 7.5. It can be seen that the general shape of the fake factor is dominated by the 1-prong region, as the merged FF produces a graph more similar to the 1-prong FF. In the higher $p_{T,\tau}$ range the merged FF represents the average estimation of the individual 1-prong and 3-prong estimates, which in general compensates for not distinguishing between the decay modes anymore. The plot also shows lower uncertainties for each individual bin as well as the fitted fake factor curve, this is attributable to the higher number of events in the merged version leading to a more robust fit. Applying the fake factors without distinction of decay modes to data does not change the quality of fake tau background estimations as visible in Fig. 7.6. Due to the not noticeable change in predictive performance the merged decay mode version is used from now on, due to the better and more robust p_T fit.

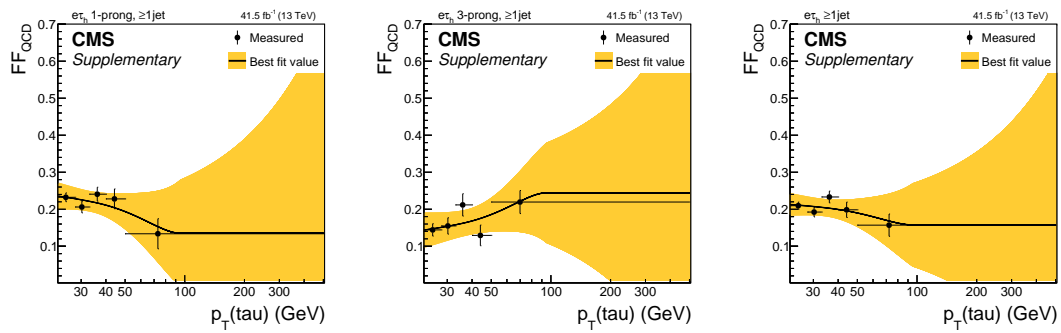


Figure 7.5.: Comparison of QCD fake factors for events with one or more jets with decay mode distinction ((left) 1-prong and (middle) 3-prong) and (right) with merged decay mode bins.

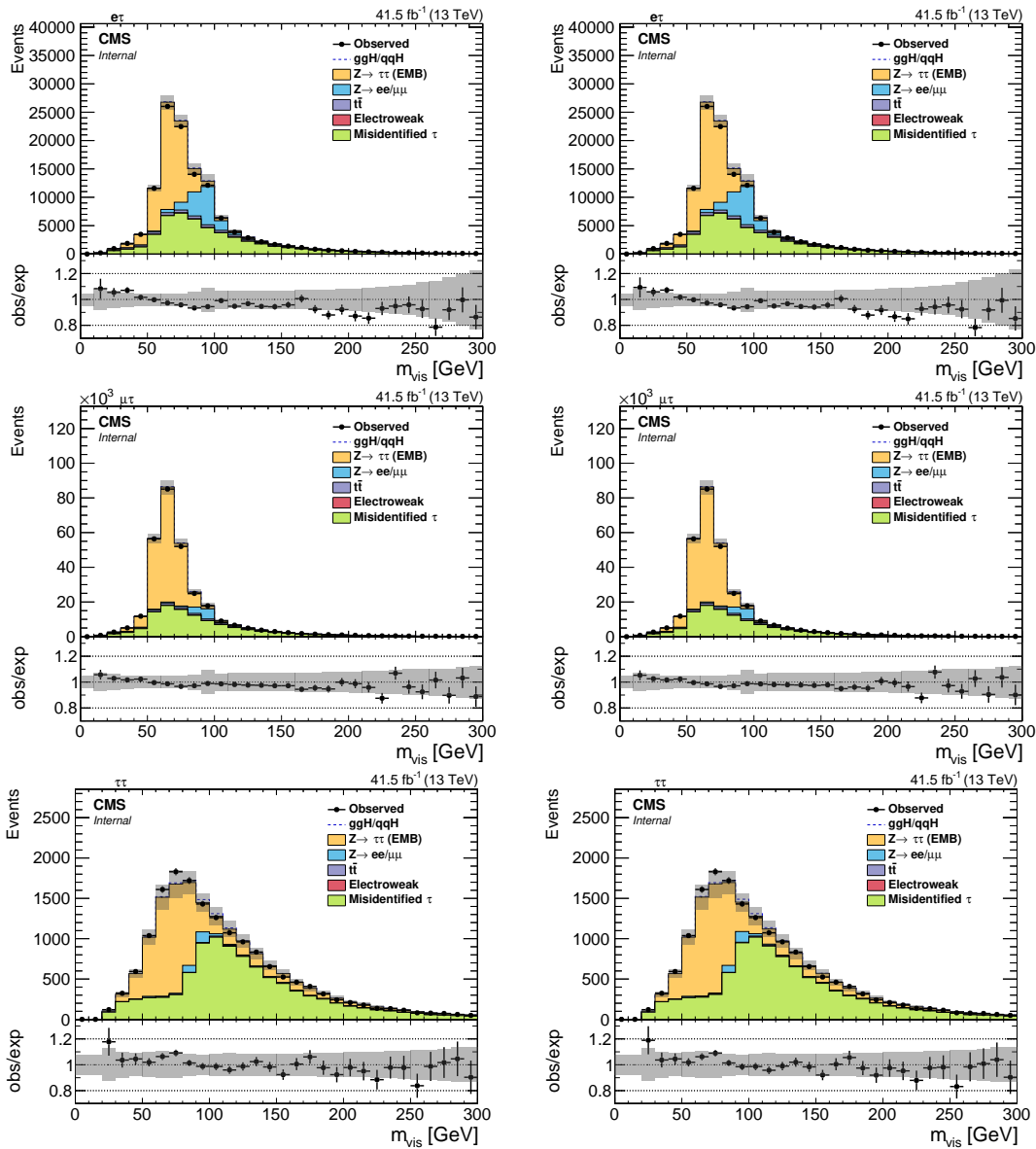


Figure 7.6.: On the left side performance plots for all three channels using fake factors with 1-prong/3-prong distinction are shown, on the right side the corresponding plots created with fake factors using merged bins are visible.

7.3. Separation of Corrections into 0-jet/1-jet Categories

The fake factors are calculated separately for events containing zero jets (0-jet) and events containing one or more jets (1-jet). Corrections to the derived fake factors however are calculated without such a distinction and applied equally on both categories with no further subdivision. This leaves room to the question if splitting up the corrections could enhance performance due to better adaptation to the underlying data. To test this, a feature was built into the FF creation tool to enable a 0-jet/1-jet binning for the corrections. In Fig. 7.7 and Fig. 7.8 the QCD and W+jet nonclosure corrections are shown exemplarily for a comparison. Due to the division of events in 0-jet/1-jet there are higher uncertainties per bin in the split corrections. However the general distributions are found to be very similar within the respective uncertainties which also applies to the smoothed function which is later directly applied on the FF. Performance plots of both FF versions are shown in Fig. 7.9 and Fig. 7.10 for the m_{vis} and n_{jets} variables. As expected from the earlier direct comparison of the corrections there appear no effective differences in the performance of the FF estimation with only very marginal differences in sparsely populated bins. This is interpreted as further proof that the FF method already is very robust. The option to enable the binning of corrections in n_{jets} is included in the FF creation tool, however the current version of the FF does not use it to ensure robust fits with low uncertainties in the calculation of the smoothed corrections.

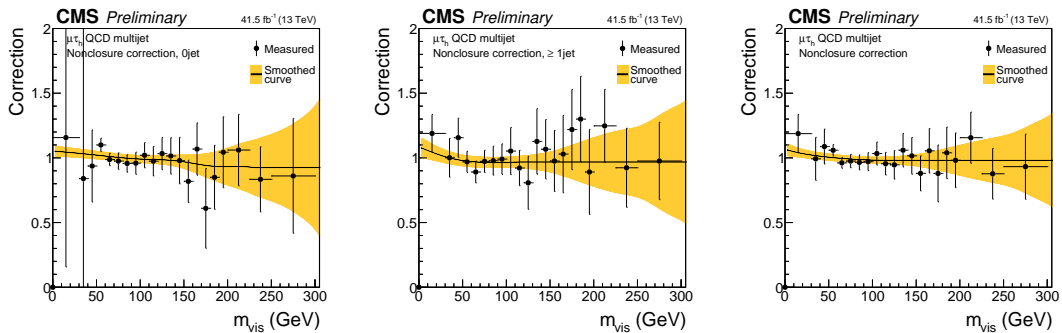


Figure 7.7.: QCD nonclosure corrections with separately derived corrections for events containing zero jets (left) and events containing one or more jets (middle), compared to the correction calculated from the entire dataset (right).

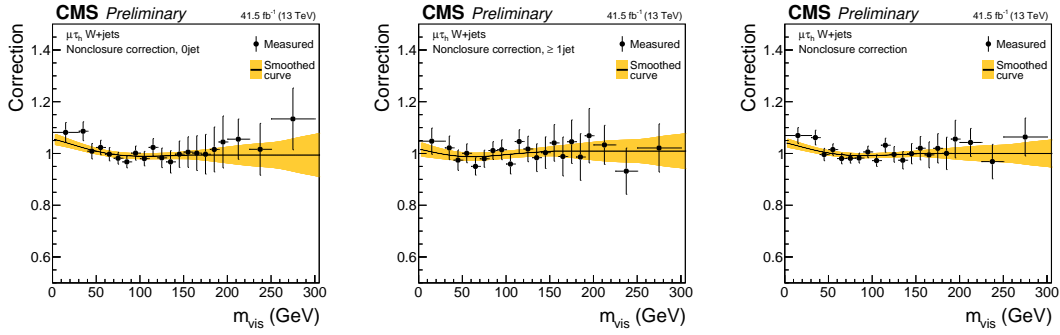


Figure 7.8.: W+jets nonclosure corrections with separately derived corrections for events containing zero jets (left) and events containing one or more jets (middle), compared to the correction calculated from the entire dataset (right).

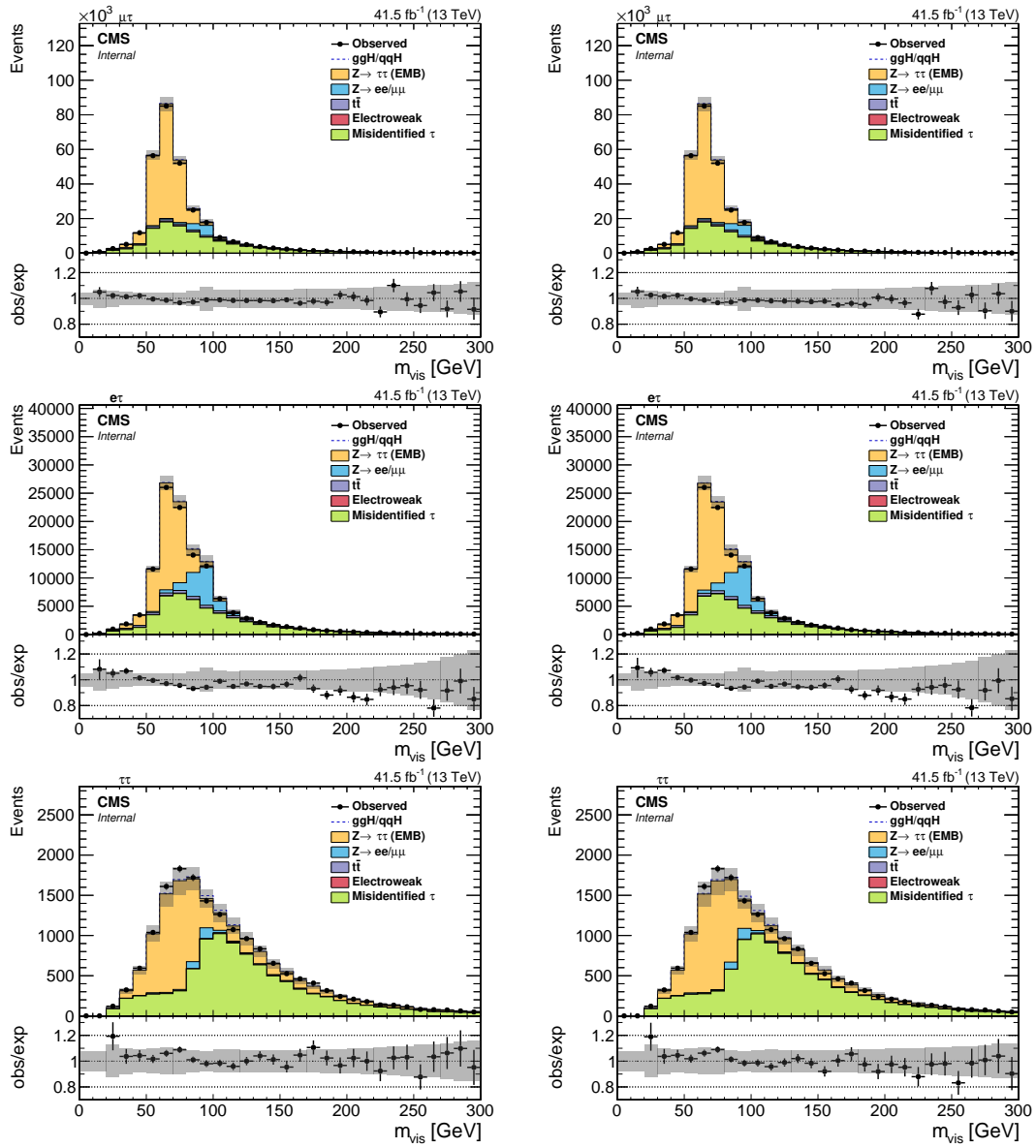


Figure 7.9.: Performance plots in the inclusive SR for the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels binned in m_{vis} , with corrections distinctive in 0-jet/1-jet (left) and corrections without distinction (right).

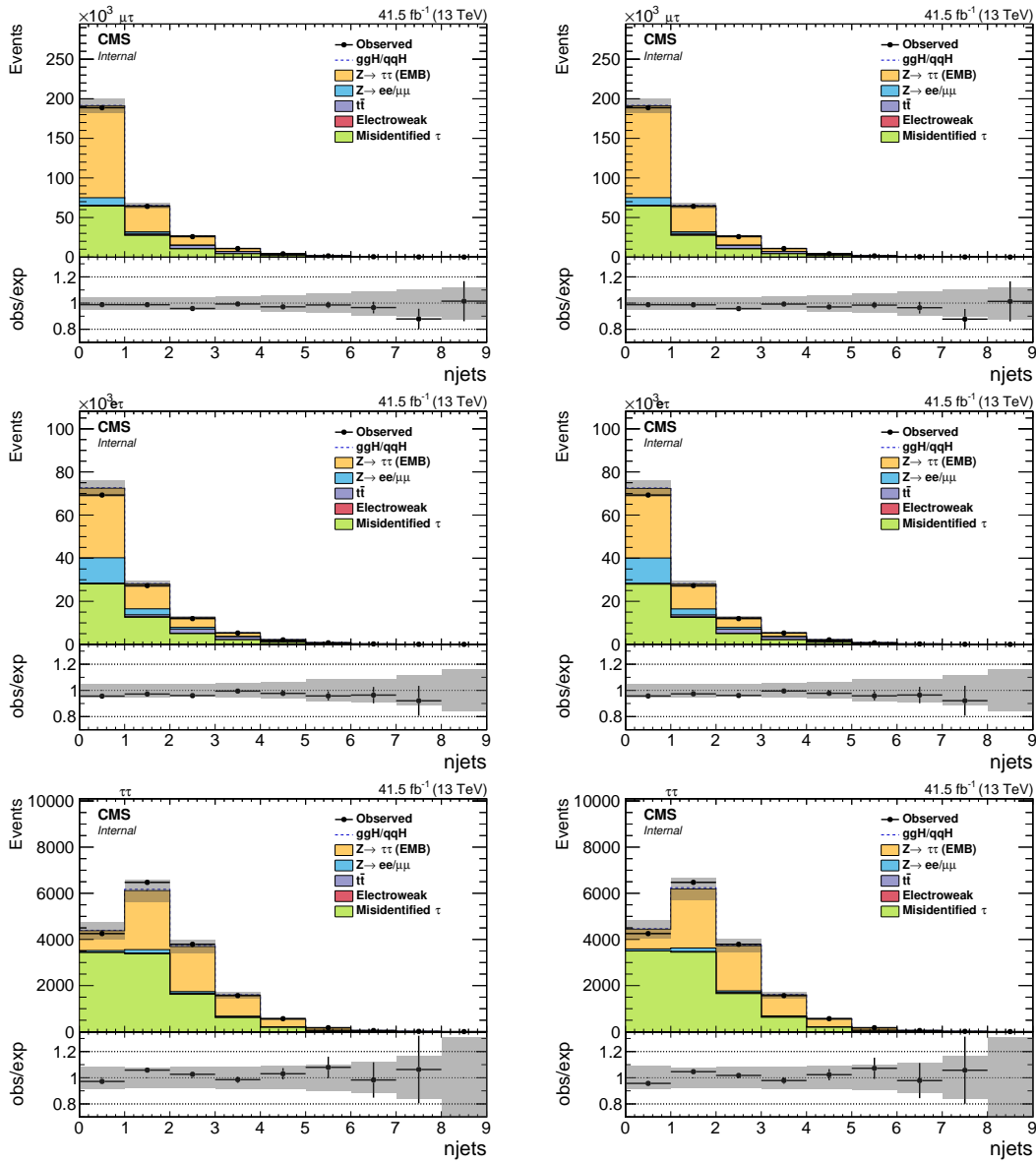


Figure 7.10.: Performance plots in the inclusive SR for the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels binned in n_{jets} , with corrections distinctive in 0-jet/1-jet (left) and corrections without distinction (right).

7.4. Application of the Fake Factor Method to 2016 and 2017 Data

All discussed methods and changes were applied separately to 2016 and 2017 data to derive fake factors for further use in the analysis. In Fig. 7.11 and Fig. 7.12 performance plots for the respective datasets are shown. The plots compare the prediction of backgrounds to the observed data by stacking the misidentified τ background estimated with the FF method, $Z \rightarrow \tau\tau$ events generated with the τ -embedding technique and the backgrounds from $Z \rightarrow \ell\ell$, $t\bar{t}$ and $EWKZ$ estimated via MC simulation. The background fractions used for the performance plots were calculated in dependence of the visible mass m_{vis} . For further analysis of the event distribution of a specific variable the fractions can be calculated in dependence of the respective variable. All plots show the distribution for the respective categories depicted in dependence of different kinematic variables prior to the fit used for the signal extraction (pre-fit). A full list of resulting performance plots and FF plots is included in Appendix C, instructions on how to set up the fake factor tool are listed in Appendix B.

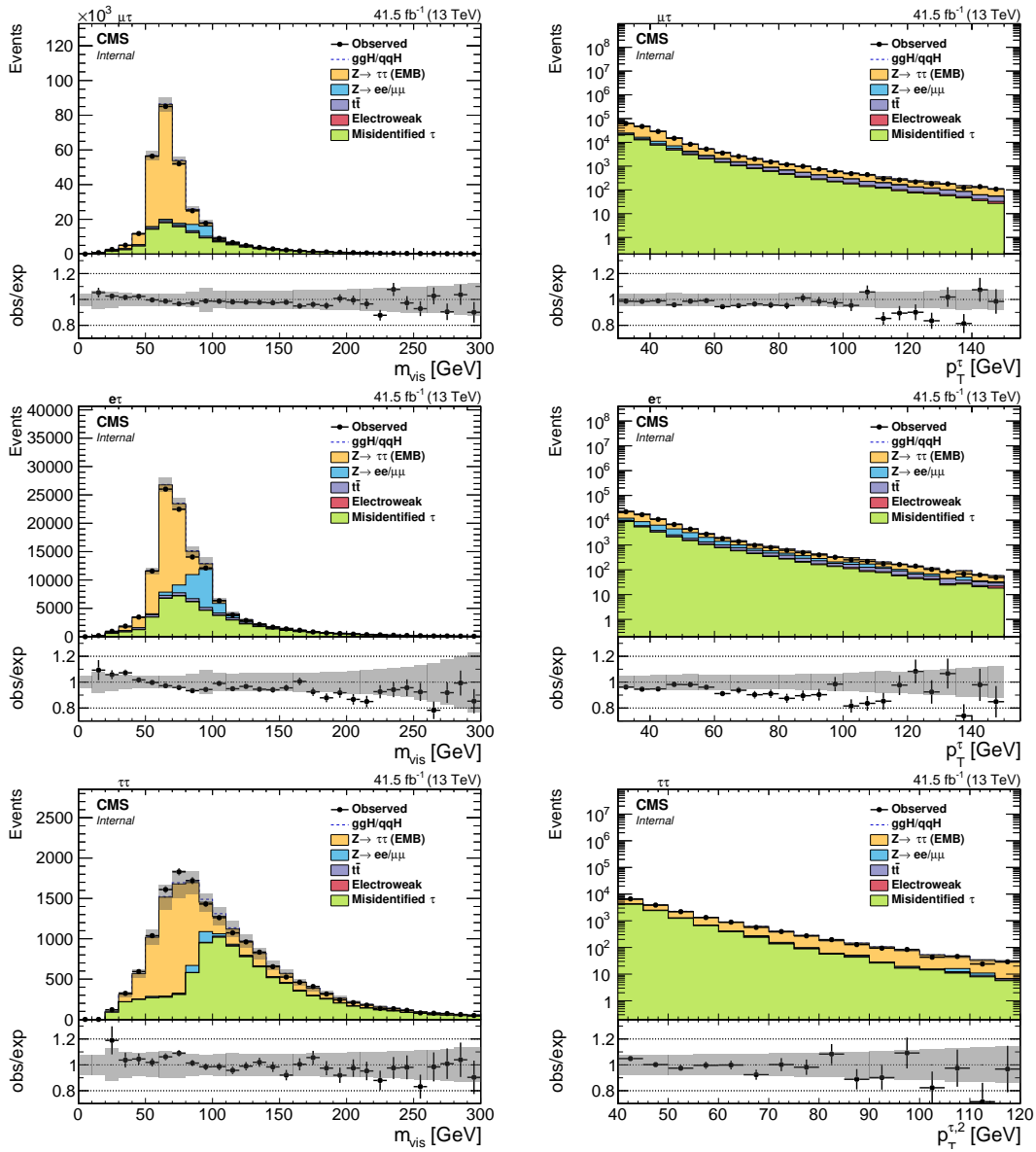


Figure 7.11.: Performance plots for 2017 data in the inclusive SR of the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels, binned in m_{vis} and $p_{T,\tau}$.

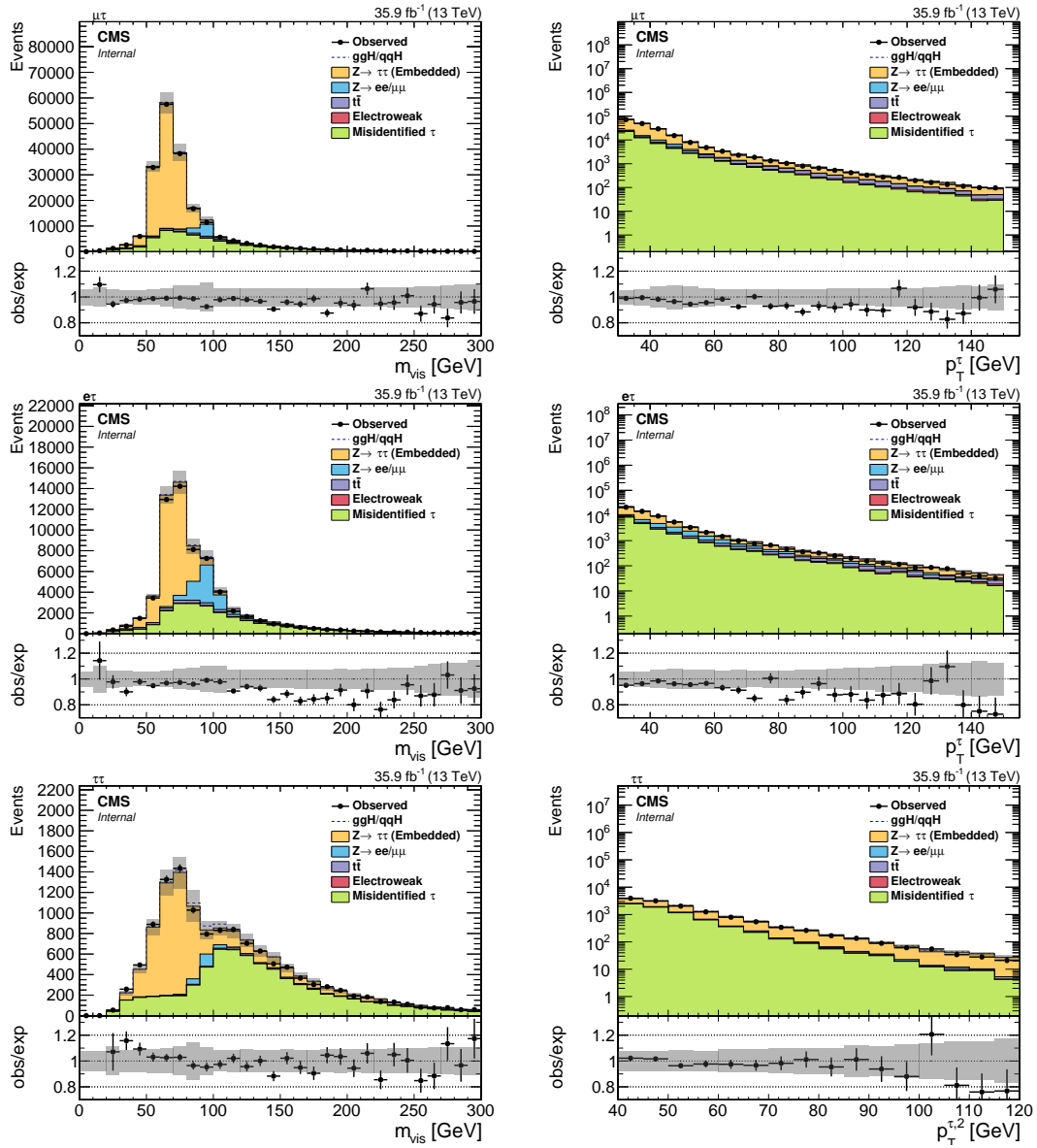


Figure 7.12.: Performance plots for 2016 data in the inclusive SR of the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels, binned in m_{vis} and $p_{T,\tau}$.

8. Conclusions

The data driven fake factor method as a means to estimate the background of misidentified hadronically decaying tau leptons was extended and applied to data collected by the CMS experiment in 2016 and 2017 at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 35.9 and 41.5 fb⁻¹ (for 2016 and 2017, respectively). A set of fake factors for each year has been created for the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels. Samples generated by the τ -embedding method estimate the $Z \rightarrow \tau\tau$ background process more precisely than samples generated with MC simulation. It was shown that this also improves performance of the FF method. Only $t\bar{t}$ fake factors and the respective corrections have to be derived using MC generated samples since the control region constraint of requiring events with at least one isolated muon and electron is never fulfilled for events in the embedded samples due to the creation process of those samples. Another focus of this thesis was to explore possibilities to generally make the FF method more robust and reliable. The approach to calculate fake factors separately for 1-prong and 3-prong events led to high uncertainties in the p_T fits of the respective fake factors due to sparsely populated regions in the phase space. To increase robustness in those fits the method was adapted to merge both regions under the requirement that the overall performance does not decrease, which was shown to be true. Similar investigations were done for the corrections derived for the different fake factor categories, which were calculated using no 0-jet/1-jet distinction, unlike the derived fake factors. A comparison was created to determine whether the introduction of such a distinction could improve fake factor performance while keeping uncertainties low. Marginal improvements with significantly lowered confidence for the calculation of the corrections led to the decision to not use the n_{jet} splitting. The results of the extended FF method are used in the recent preliminary measurements of Higgs boson properties of the CMS $H \rightarrow \tau\tau$ working group for the data recorded in 2016 and 2017. It will also be applied in the upcoming legacy publications for the 2016-2018 dataset.

For future improvement of the FF method a focus can be laid on the generation of the background fractions. Categorization of background events into the respective classes via machine learning instead of the cut-based approach and the generation of fractions for each different variable could improve the overall precision of the FF method.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

A. Acronyms

ALICE	A Large Ion Collider Experiment
AR	Application Region
ATLAS	A Toroidal LHC Apparatus
BSM	Beyond the Standard Model
CERN	European Organization for Nuclear Research
CMS	Compact Muon Solenoid
CP	Charge-Parity
CR	Control Region
CSC	Cathode Strip Chamber
DR	Determination Region
DT	Drift Tube
ECAL	Electromagnetic Calorimeter
EMB	Embedded Samples
FF	Fake Factor
HB	Hadron Barrel
HCAL	Hadron Calorimeter
HE	Hadronic Endcaps
HF	Forward Hadron Calorimeters
HLT	High-Level Trigger
HO	Outer Hadron Calorimeter
HPS	Hadron-Plus-Strips
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty
LHCf	Large Hadron Collider forward
MC	Monte Carlo
MSSM	Minimal Supersymmetric Extension of the Standard Model
MVA	Multivariate Analysis
MoEDAL	Monopole and Exotics Detector at the LHC
OS	Opposite Sign
P	Parity
PF	Particle-Flow
QCD	Quantum Chromodynamics
RPC	Resistive Plate Chamber

SM	Standard Model of Particle Physics
SR	Signal Region
SS	Same Sign
SUSY	Supersymmetry
TOTEM	Total Elastic and Diffractive Cross Section Measurement
VBF	Vector Boson Fusion

B. Software Instructions

Setting up the CMSSW environment

For running the code available at

- https://github.com/fspreitzer/FakeFactor/tree/FF2017ML_Embedding

a CMSSW environment has to be set up with the following commands.

- `cmsrel CMSSW_8_0_25`
- `cd CMSSW_8_0_25/src/; cmsenv`

This assumes the user to be working on the CERN cluster as a CMS user.

Setting up the FF tool

The fake factor tool and all necessary dependencies are installed with the following commands.

- `git clone https://github.com/fspreitzer/FakeFactor.git .`
 `-b FF2017ML_Embedding`
- `cd HTTutilities/Jet2TauFakes`
- `git checkout v0.2.1`
- `cd ../../`
- `scram b -j 8`
- `git init`

To configure the tool the user-specific paths in `ViennaTool/Settings.h` have to be customized for the specific user, as well as all configuration options related to the fake-factor creation. Further instructions can be found in `README.md` and `instructions.txt`. The tool is started by running the script `ViennaTool/steerAll.sh`



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

C. Performance Plots

In the following section the modeling for several variables with the fake factor method for the inclusive signal region is presented as pre-fit plots. The weight fractions used for the fake factors were calculated as a function of m_{vis} .

Re-evaluating the fractions for each different variable would increase the accuracy for the respective modeling, however at the cost of significantly more complexity when applying the FF method. The general good agreement shows that the modeling is also satisfactory when only using one set of fractions. The plots C.1-C.3 show the performance plots distributions for the visible mass m_{vis} , $p_{T,\tau}$, \cancel{E}_T , the number of jets and the jet- p_T for the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels for 2017, in plots C.4-C.6 the respective distributions for 2016 are shown.

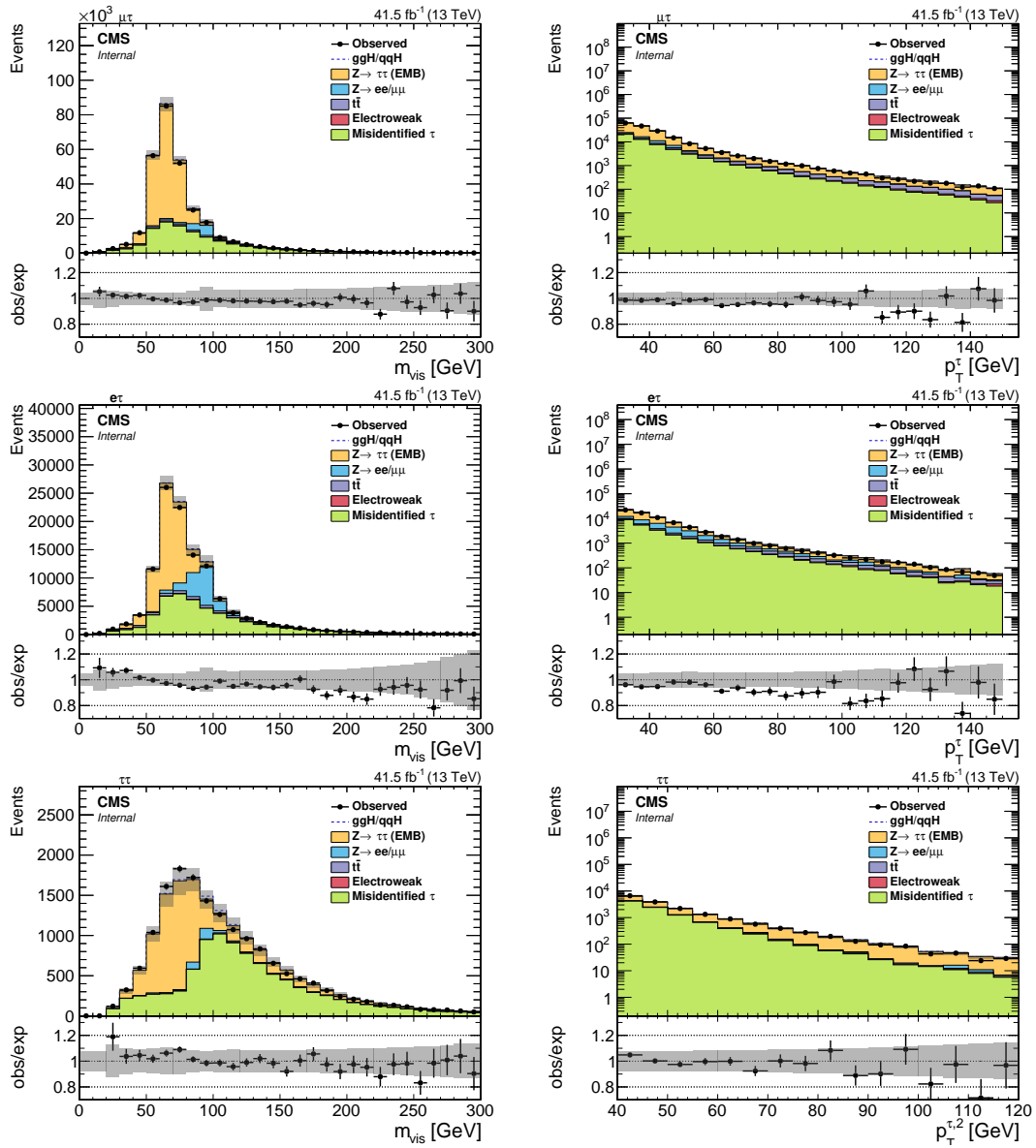


Figure C.1.: Performance plots for 2017 data in the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels, binned in m_{vis} and $p_{T,\tau}$.

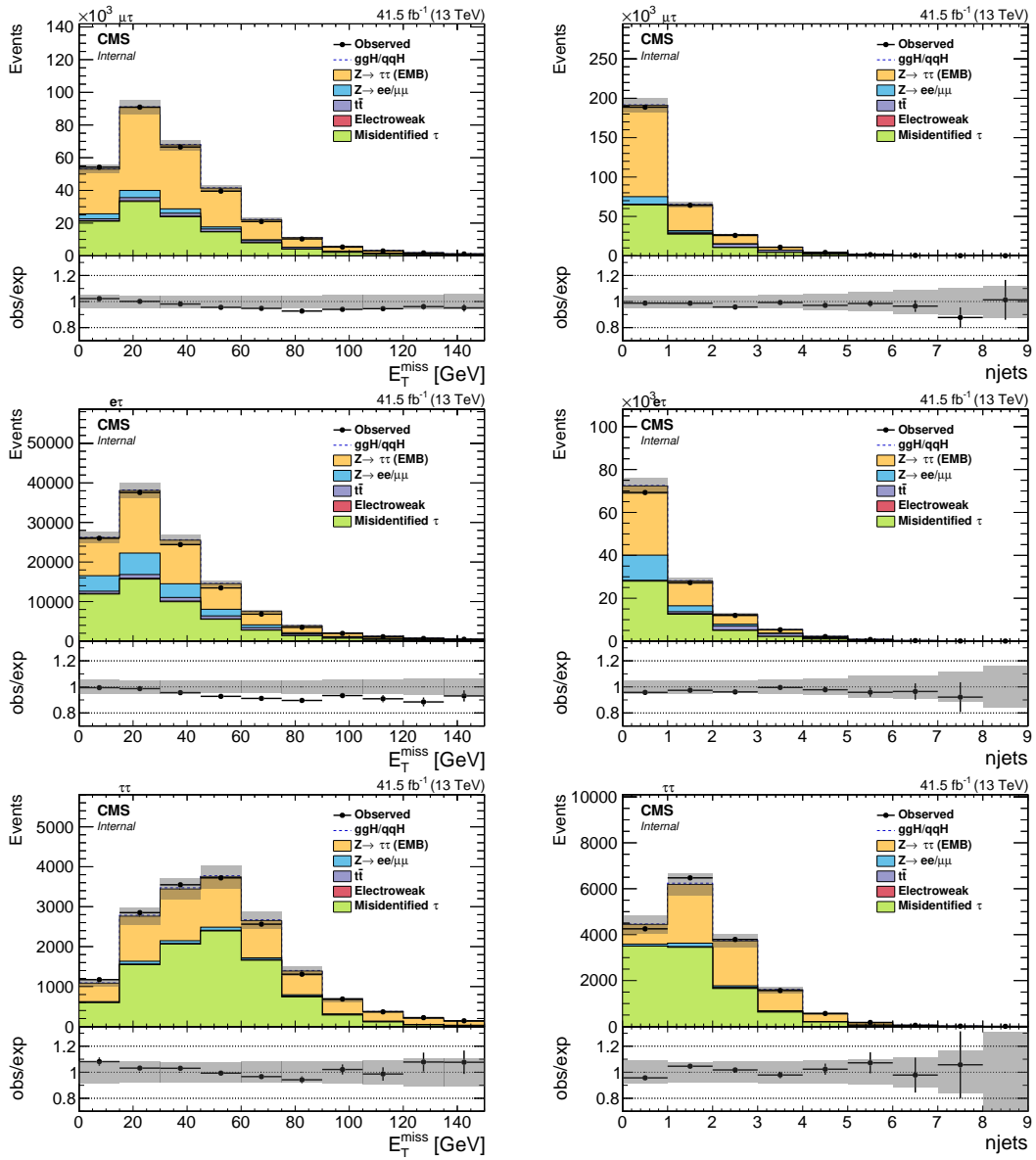


Figure C.2.: Performance plots for 2017 data in the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels, binned in E_T^{miss} and n_{jets} .

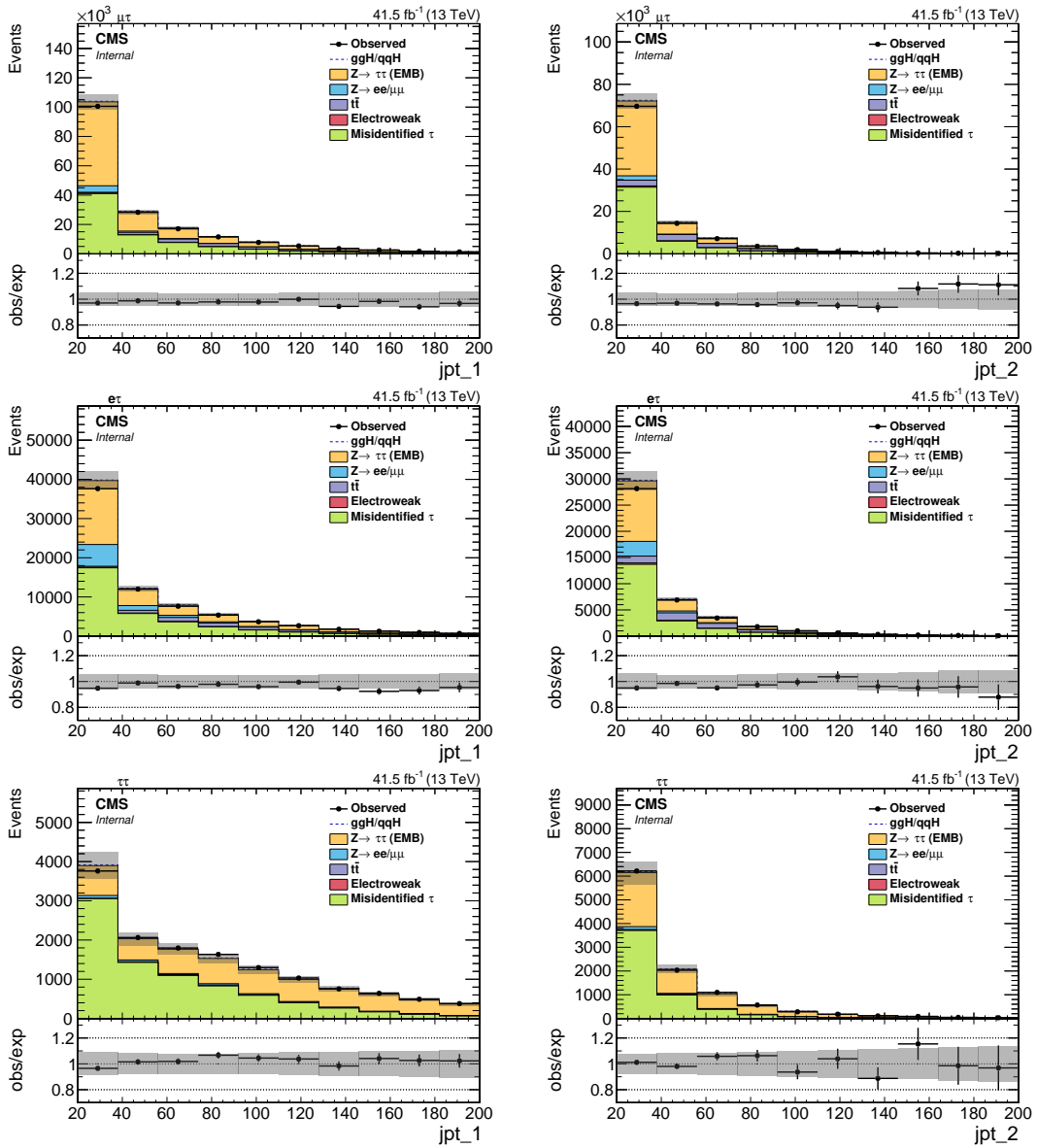


Figure C.3.: Performance plots for 2017 data in the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels, binned in the p_T of the leading jet (j_{pT_1}) and the p_T of the trailing jet (j_{pT_2}).

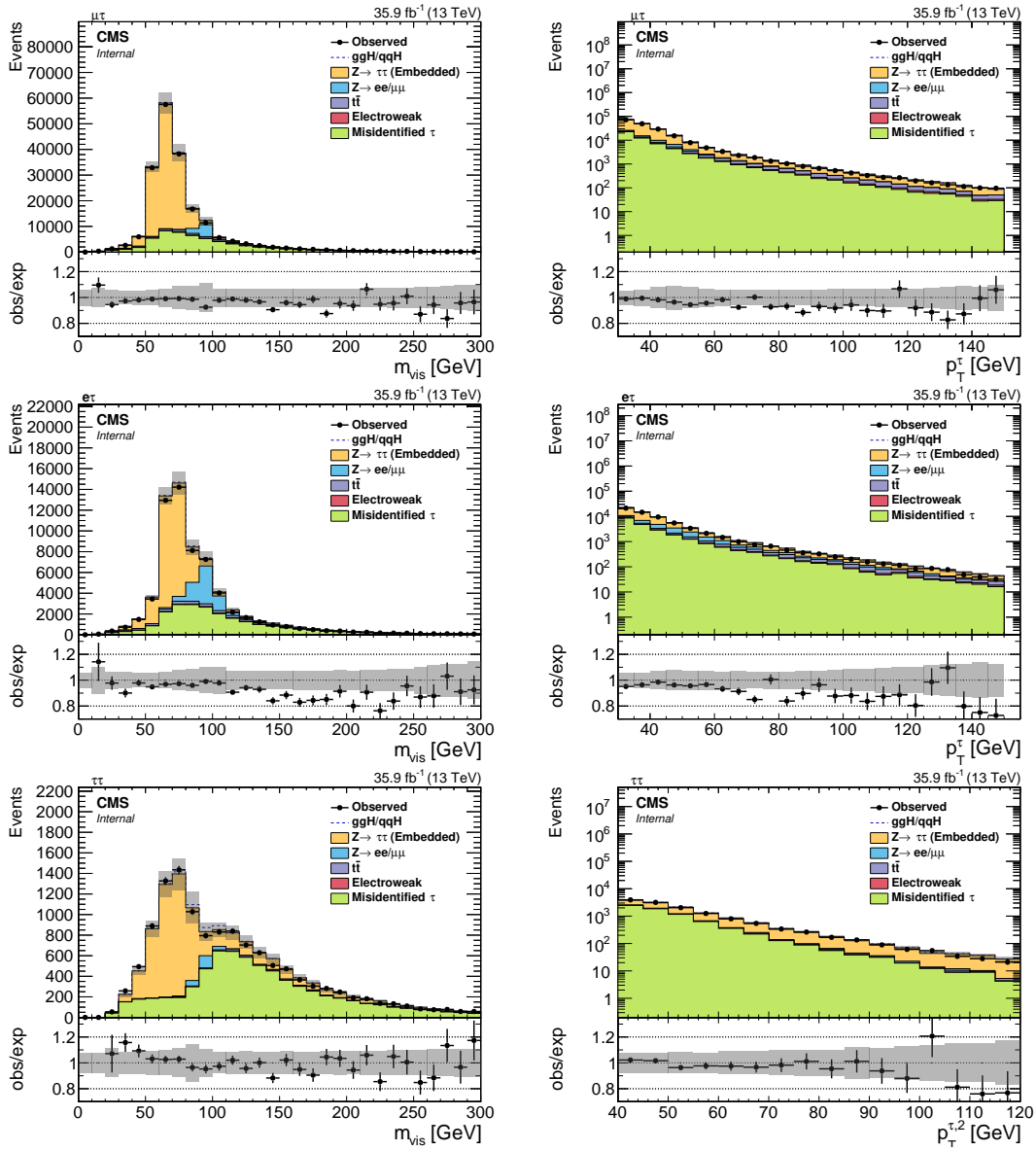


Figure C.4.: Performance plots for 2016 data in the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels, binned in m_{vis} and $p_{T,\tau}$.

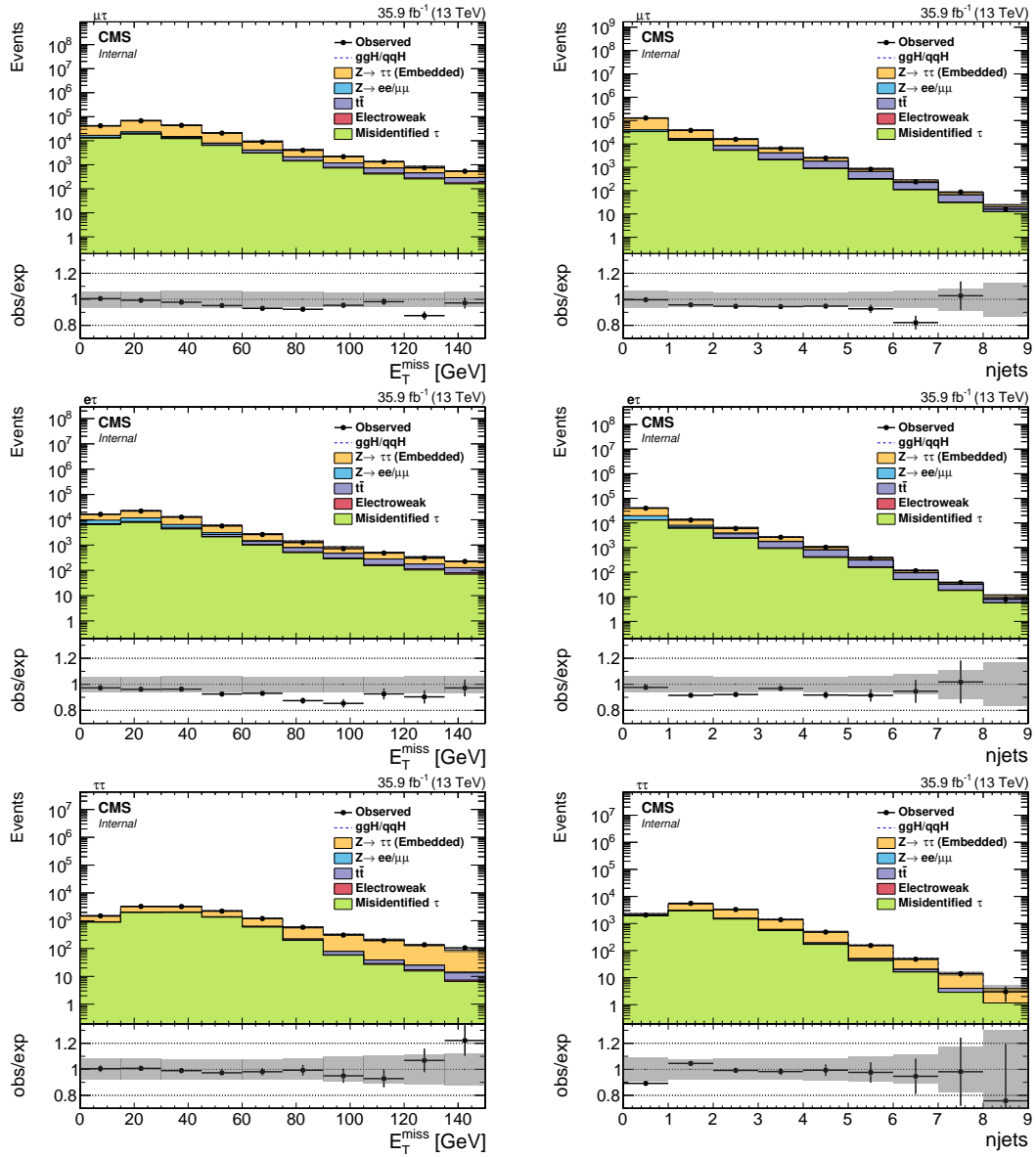


Figure C.5.: Performance plots for 2016 data in the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels, binned in E_T^{miss} and n_{jets} .

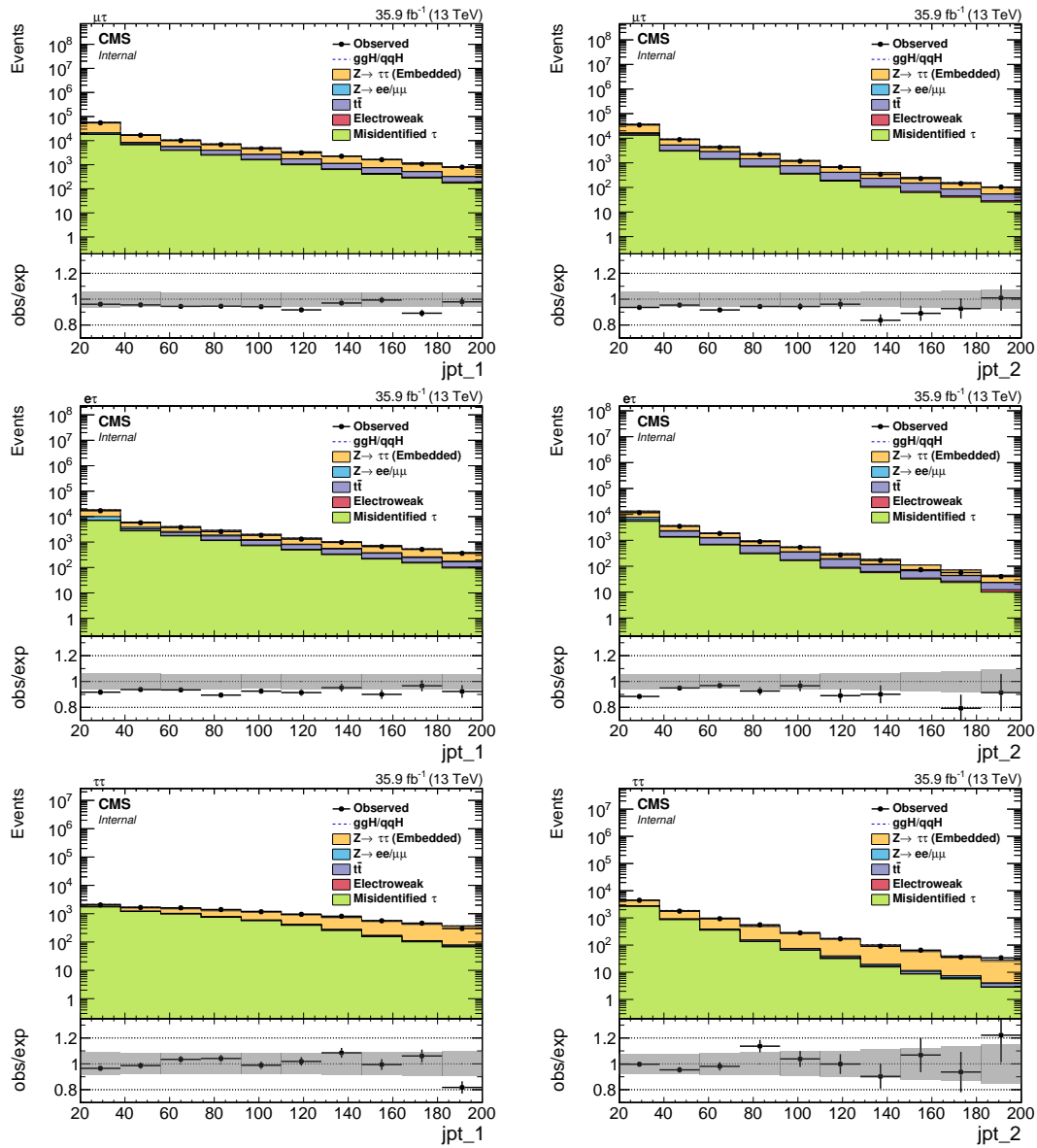


Figure C.6.: Performance plots for 2016 data in the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ channels, binned in the p_T of the leading jet (j_{pT_1}) and the p_T of the trailing jet (j_{pT_2}).



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Bibliography

- [1] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [2] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693.
- [3] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693.
- [4] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98.3 (2018), 030001. 1898 p. DOI: 10.1103/PhysRevD.98.030001. URL: <http://cds.cern.ch/record/2636832>.
- [5] W. Cottingham and D. Greenwood. “An introduction to the standard model of particle physics”. Cambridge: Cambridge Univ. Press, 1998. URL: <https://cds.cern.ch/record/396082>.
- [6] S. L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (Feb. 1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2. URL: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [7] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [8] C. N. Yang and R. L. Mills. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. In: *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191. URL: <https://link.aps.org/doi/10.1103/PhysRev.96.191>.
- [9] CMS Collaboration. “Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying t leptons at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* 2018.8 (Aug. 2018). DOI: 10.1007/jhep08(2018)066. URL: [https://doi.org/10.1007/jhep08\(2018\)066](https://doi.org/10.1007/jhep08(2018)066).

- [10] ATLAS Collaboration. “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector”. In: (2018). DOI: 10.1016/j.physletb.2018.07.035. eprint: arXiv:1806.00425.
- [11] A. Djouadi. “The anatomy of electroweak symmetry breaking”. In: *Physics Reports* 457.1-4 (Feb. 2008), pp. 1–216. DOI: 10.1016/j.physrep.2007.10.004. URL: <https://doi.org/10.1016/j.physrep.2007.10.004>.
- [12] A. Denner et al. “Standard Model Higgs-Boson Branching Ratios with Uncertainties”. In: *The European Physical Journal C* 71.9 (Sept. 2011). DOI: 10.1140/epjc/s10052-011-1753-8. URL: <https://doi.org/10.1140/epjc/s10052-011-1753-8>.
- [13] ATLAS Collaboration. “Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector”. In: (2018). DOI: 10.1016/j.physletb.2018.09.013. eprint: arXiv:1808.08238.
- [14] CMS Collaboration. “Observation of Higgs boson decay to bottom quarks”. In: *Physical Review Letters* 121.12 (Sept. 2018). DOI: 10.1103/physrevlett.121.121801. URL: <https://doi.org/10.1103/physrevlett.121.121801>.
- [15] LHC Higgs Cross Section Working Group. “CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector”. en. In: (2017). DOI: 10.23731/cyrm-2017-002. URL: <https://e-publishing.cern.ch/index.php/CYRM/issue/view/32>.
- [16] ATLAS and CMS Collaborations. “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV”. In: *Journal of High Energy Physics* 2016.8 (Aug. 2016). DOI: 10.1007/jhep08(2016)045. URL: [https://doi.org/10.1007/jhep08\(2016\)045](https://doi.org/10.1007/jhep08(2016)045).
- [17] M. L. Perl et al. “Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation”. In: *Phys. Rev. Lett.* 35 (1975). [,193(1975)], pp. 1489–1492. DOI: 10.1103/PhysRevLett.35.1489.
- [18] Planck Collaboration. “Planck 2018 results. VI. Cosmological parameters”. 2018. eprint: arXiv:1807.06209.
- [19] Y. Fukuda et al., Super-Kamiokande Collaboration. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.

- [20] S. Dimopoulos and H. Georgi. “Softly broken supersymmetry and SU(5)”. In: *Nuclear Physics B* 193.1 (Dec. 1981), pp. 150–162. DOI: 10.1016/0550-3213(81)90522-8. URL: [https://doi.org/10.1016/0550-3213\(81\)90522-8](https://doi.org/10.1016/0550-3213(81)90522-8).
- [21] C. Csaki. “The Minimal Supersymmetric Standard Model (MSSM)”. In: (1996). DOI: 10.1142/S021773239600062X. eprint: [arXiv:hep-ph/9606414](https://arxiv.org/abs/hep-ph/9606414).
- [22] A. Djouadi et al. The Minimal Supersymmetric Standard Model: Group Summary Report. Tech. rep. hep-ph/9901246. 1998. eprint: [arXiv:hep-ph/9901246](https://arxiv.org/abs/hep-ph/9901246). URL: <http://cds.cern.ch/record/376049>.
- [23] B. Bajc et al. “Threshold corrections to dimension-six proton decay operators in non-minimal SUSY SU(5) GUTs”. In: *Nuclear Physics B* 910 (Sept. 2016), pp. 1–22. DOI: 10.1016/j.nuclphysb.2016.06.017. URL: <https://doi.org/10.1016/j.nuclphysb.2016.06.017>.
- [24] CERN. “LHC Guide”. Mar. 2017. URL: <http://cds.cern.ch/record/2255762>.
- [25] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: <https://doi.org/10.1088/1748-0221/3/08/s08004>.
- [26] CMS Collaboration. “The CMS trigger system”. In: *Journal of Instrumentation* 12.01 (Jan. 2017), P01020–P01020. DOI: 10.1088/1748-0221/12/01/p01020. URL: <https://doi.org/10.1088/1748-0221/12/01/p01020>.
- [27] CMS Collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003. DOI: 10.1088/1748-0221/12/10/p10003. URL: <https://doi.org/10.1088/1748-0221/12/10/p10003>.
- [28] CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 13.06 (June 2018), P06015–P06015. DOI: 10.1088/1748-0221/13/06/p06015. URL: <https://doi.org/10.1088/1748-0221/13/06/p06015>.
- [29] S. Baffioni et al. “Electron reconstruction in CMS”. In: *The European Physical Journal C* 49.4 (Jan. 2007), pp. 1099–1116. DOI: 10.1140/epjc/s10052-006-0175-5. URL: <https://doi.org/10.1140/epjc/s10052-006-0175-5>.
- [30] CMS Collaboration. “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. Tech. rep. CERN-LHCC-2006-001, CMS-TDR-8-1. 2006. URL: <http://cds.cern.ch/record/922757>.

- [31] CMS Collaboration. “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”. In: *Journal of Instrumentation* 10.06 (June 2015), P06005–P06005. DOI: 10.1088/1748-0221/10/06/p06005. URL: <https://doi.org/10.1088/1748-0221/10/06/p06005>.
- [32] CMS Collaboration. “Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS”. In: *Journal of Instrumentation* 11.01 (Jan. 2016), P01019–P01019. DOI: 10.1088/1748-0221/11/01/p01019. URL: <https://doi.org/10.1088/1748-0221/11/01/p01019>.
- [33] CMS Collaboration. “Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 13.10 (Oct. 2018), P10005–P10005. DOI: 10.1088/1748-0221/13/10/p10005. URL: <https://doi.org/10.1088/1748-0221/13/10/p10005>.
- [34] M. Cacciari, G. Salam, and G. Soyez. “The anti- k_t jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063. URL: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [35] CMS Collaboration. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. DOI: 10.1088/1748-0221/12/02/p02014. URL: <https://doi.org/10.1088/1748-0221/12/02/p02014>.
- [36] CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. DOI: 10.1088/1748-0221/13/05/p05011. URL: <https://doi.org/10.1088/1748-0221/13/05/p05011>.
- [37] CMS Collaboration. “Performance of the CMS missing transverse energy reconstruction in pp data at $\sqrt{s} = 8$ TeV”. In: *Journal of Instrumentation* 10.02 (Feb. 2015), P02006–P02006. DOI: 10.1088/1748-0221/10/02/p02006. URL: <https://doi.org/10.1088/1748-0221/10/02/p02006>.
- [38] F. Maltoni and T. Stelzer. “MadEvent: Automatic Event Generation with MadGraph”. In: *Journal of High Energy Physics* 2003.02 (Feb. 2003), pp. 027–027. DOI: 10.1088/1126-6708/2003/02/027. URL: <https://doi.org/10.1088/1126-6708/2003/02/027>.
- [39] S. Frixione, P. Nason, and C. Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), pp. 070–070. DOI: 10.1088/1126-

6708/2007/11/070. URL: <https://doi.org/10.1088/1126-6708/2007/11/070>.

- [40] T. Sjöstrand, S. Mrenna, and P. Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Computer Physics Communications* 178.11 (June 2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. URL: <https://doi.org/10.1016/j.cpc.2008.01.036>.
- [41] CMS Collaboration. “Underlying Event Tunes and Double Parton Scattering”. Tech. rep. CMS-PAS-GEN-14-001. Geneva: CERN, 2014. URL: <http://cds.cern.ch/record/1697700>.
- [42] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 2003), pp. 250–303. DOI: 10.1016/s0168-9002(03)01368-8. URL: [https://doi.org/10.1016/s0168-9002\(03\)01368-8](https://doi.org/10.1016/s0168-9002(03)01368-8).
- [43] CMS Collaboration. “Measurement of Higgs boson production and decay to the $\tau\tau$ final state”. Tech. rep. CMS-PAS-HIG-18-032. Geneva: CERN, 2019. URL: <https://cds.cern.ch/record/2668685>.
- [44] CMS Collaboration. “Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: (2018). DOI: 10.1007/JHEP09(2018)007. eprint: [arXiv:1803.06553](https://arxiv.org/abs/1803.06553).
- [45] CMS Collaboration. “Measurement of the $Zg^* \rightarrow tt$ cross section in pp collisions at $\sqrt{s} = 13$ TeV and validation of t lepton analysis techniques”. In: *The European Physical Journal C* 78.9 (Sept. 2018). DOI: 10.1140/epjc/s10052-018-6146-9. URL: <https://doi.org/10.1140/epjc/s10052-018-6146-9>.
- [46] CMS Collaboration. “An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data”. In: *Journal of Instrumentation* 14.06 (June 2019), P06032–P06032. DOI: 10.1088/1748-0221/14/06/p06032. URL: <https://doi.org/10.1088/1748-0221/14/06/p06032>.
- [47] T. Sjöstrand et al. “An Introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (June 2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. URL: <https://doi.org/10.1016/j.cpc.2015.01.024>.